

## BAB II

### LANDASAN TEORI

#### A. Tes

##### 1. Pengertian Tes

Tes berasal dari bahasa Latin *testum* yang berarti alat untuk mengukur tanah. Dalam bahasa Prancis kuno, kata tes berarti ukuran yang dipergunakan untuk membedakan antara emas dengan perak serta logam lainnya<sup>11</sup>. *Testing* adalah saat pengambilan tes, *testee* adalah responden yang sedang mengerjakan tes sedangkan *tester* adalah subjek evaluasi<sup>12</sup>.

Sedangkan dilihat dari segi istilah, ada berbagai macam pendapat, diantaranya:

- a. Anne Anastasi (1976) dalam bukunya *Psychological Testing* mengatakan bahwa tes pada dasarnya merupakan suatu pengukuran yang obyektif dan standart terhadap sampel perilaku<sup>13</sup>.
- b. Frederick G Brown (1976) mengatakan bahwa tes adalah prosedur yang sistematis guna mengukur sampel perilaku seseorang. Sistematis juga memiliki pengertian obyektif, standart dan syarat-syarat kualitas lainnya.<sup>14</sup>

---

<sup>11</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 43

<sup>12</sup> Suharsimi Arikunto, *Dasar-Dasar* .....hal. 53

<sup>13</sup> Syaifuddin Azwar, *Tes Prestasi*..... hal. 3

<sup>14</sup> Syaifuddin Azwar, *Tes Prestasi*.....hal. 3

- c. Menurut Lee J. Cronbach dalam bukunya yang berjudul *Essential of Psychological Testing*, menyatakan bahwa tes adalah suatu prosedur yang sistematis untuk membandingkan tingkah laku dua orang atau lebih<sup>15</sup>.
- d. Menurut Sumardi Suryabrata (1984) tes adalah pertanyaan-pertanyaan yang harus dijawab dan atau perintah-perintah yang harus dijalankan, yang mendasarkan harus bagaimana *testee* menjawab pertanyaan-pertanyaan atau melakukan perintah-perintah itu, penyelidik mengambil kesimpulan dengan cara membandingkan dengan standart atau *testee* lainnya<sup>16</sup>.
- e. *Test is a systematic procedure for comparing the behavior of two or more individuals*. Tes merupakan prosedur sistematis yang direncanakan oleh evaluator guna membandingkan perilaku dua orang atau siswa atau lebih. Dalam kenyataannya tes pada umumnya terdiri atas sekumpulan pertanyaan atau tugas yang harus dijawab oleh para peserta didik atau *test is a grup of questions or tasks to which a student is to respond*<sup>17</sup>.
- f. Tes menurut Muchtar Buchori, yang dikutip oleh Suharsimi Arikunto, adalah suatu percobaan yang diadakan untuk mengetahui ada atau tidaknya hasil-hasil pelajaran tertentu pada seorang murid atau kelompok murid<sup>18</sup>.

---

<sup>15</sup> Anas Sudjiono, *Pengantar Evaluasi Pendidikan*, (Jakarta: PT. Raja Grafindo Persada, 1996), hal. 66

<sup>16</sup> Chabib Toha, *Teknik Evaluasi*.....hal. 43

<sup>17</sup> Sukardi, *Evaluasi Pendidikan*.....hal. 20

<sup>18</sup> Suharsimi Arikunto, *Dasar-Dasar*.....hal.32

- g. Dalam bukunya yang berjudul *Evaluasi Pendidikan* Drs. Amir Daien Indrakusuma yang dikutip oleh Suharsimi Arikunto mengatakan bahwa tes adalah suatu alat atau prosedur yang sistematis dan obyektif untuk memperoleh data-data atau keterangan-keterangan yang diinginkan tentang seseorang, dengan cara yang boleh dikatakan tepat dan cepat<sup>19</sup>.
- h. Webster's Collegiate mendefinisikan tes sebagai *any series of questions or exercises or other means of measuring the skill, knowledge, intelligence, capacities of aptitudes or an individual or group*. Yang lebih kurang artinya demikian: tes adalah serentetan pertanyaan atau alat lain yang digunakan untuk mengukur ketrampilan, pengetahuan, intelegensi, kemampuan atau bakat yang dimiliki oleh individu atau kelompok<sup>20</sup>.

Dari beberapa uraian dan kutipan di atas jika dikaitkan dengan evaluasi pendidikan dapat ditarik kesimpulan bahwa tes adalah prosedur yang sistematis, obyektif dan standart yang berupa serentetan pertanyaan atau latihan yang harus dijawab oleh *testee* untuk menghasilkan suatu nilai yang mencerminkan tingkah laku atau prestasi *testee*.

## 2. Fungsi Tes

Secara umum, ada dua macam fungsi yang dimiliki oleh tes, yaitu:

---

<sup>19</sup> Suharsimi Arikunto, *Dasar-Dasar*.....hal.32

<sup>20</sup> Suharsimi Arikunto, *Dasar-Dasar*.....hal.32

1. Sebagai alat pengukur terhadap peserta didik. Dalam hubungan ini tes berfungsi mengukur tingkat perkembangan atau kemajuan yang telah dicapai oleh peserta didik setelah mereka menempuh proses belajar-mengajar dalam jangka waktu tertentu.
2. Sebagai alat pengukur keberhasilan program pengajaran, sebab melalui tes tersebut akan dapat diketahui sudah seberapa jauh program pengajaran yang telah ditentukan, telah dapat dicapai.

Sedangkan menurut Suharsimi Arikunto dalam bukunya *Dasar-Dasar Evaluasi Pendidikan*, fungsi tes dapat ditinjau dari tiga hal:

- a. Fungsi untuk kelas.
- b. Fungsi untuk bimbingan.
- c. Fungsi untuk administrasi.

Adapun perbandingan dari ketiga fungsi tersebut adalah<sup>21</sup>:

**Tabel 2.1.**  
**Fungsi Tes**

Fungsi Untuk Kelas	Fungsi Untuk Bimbingan	Fungsi Untuk Administrasi
a. Mengadakan diagnosis terhadap kesulitan belajar siswa.	a. Menentukan arah pembicaraan dengan orang tua tentang anak-anak mereka.	a. Memberi petunjuk dalam mengelompokkan siswa.
b. Mengevaluasi celah antara bakat dengan pencapaian.	b. Membantu siswa dalam menentukan pilihan.	b. Penempatan siswa baru.
c. Menaikkan tingkat prestasi.	c. Membantu siswa mencapai tujuan pendidikan dan jurusan.	c. Membantu siswa memiliki kelompok.
d. Mengelompokkan		d. Menilai kurikulum.
		e. Memperluas hubungan masyarakat (public relation).
		f. Menyediakan informasi

---

<sup>21</sup> Suharsimi Arikunto, *Dasar-Dasar*.....hal.152

siswa dalam kelas pada waktu metode kelompok. e. Merencanakan kegiatan proses belajar mengajar untuk siswa secara perseorangan. f. Menentukan siswa mana yang memerlukan bimbingan khusus. g. Menentukan tingkat pencapaian untuk setiap anak.	d. Memberikan kesempatan kepada pembimbing, guru, dan orang tua dalam memahami kesulitan anak.	untuk badan lain di luar sekolah.
---	--	-----------------------------------

Fungsi tes yang lain di kemukakan oleh Saifuddin Azwar dalam bukunya *Tes Prestasi*, yaitu sebagai motivator dalam belajar. Walaupun nilai yang diperoleh dalam tes hendaknya tidak dijadikan tujuan utama bagi siswa dalam belajar akan tetapi tes dapat digunakan sebagai sarana peningkatan motivasi untuk belajar siswa. Pengalaman menunjukkan bahwa siswa akan belajar lebih giat dan berusaha lebih keras apabila mereka mengetahui bahwa di akhir program yang sedang ditempuh akan diadakan tes untuk mengetahui nilai dan prestasi mereka<sup>22</sup>.

### 3. Klasifikasi Tes

Secara umum tes dibedakan berdasarkan obyek pengukurannya dapat dibagi menjadi dua, yaitu tes kepribadian (*personality test*) dan tes hasil belajar (*Achievement test*)<sup>23</sup>.

---

<sup>22</sup> Syaifuddin Azwar, *Tes Prestasi*,.....hal. 15

<sup>23</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 44

a. Tes Kepribadian (Personality Test)

Adalah tes yang dilaksanakan dengan tujuan mengungkap ciri-ciri khas dari seseorang yang banyak sedikitnya bersifat lahiriyah, seperti gaya bicara, cara berpakaian, nada suara, hobi atau kesenangan dan lain-lain<sup>24</sup>. Yang termasuk dalam jenis tes ini dan banyak digunakan dalam kependidikan adalah:

- 1) Pengukuran sikap.
- 2) Pengukuran minat.
- 3) Pengukuran bakat.
- 4) Tes intelegensi.

b. Tes Hasil Belajar (Achievement Test)

Adalah tes yang dipergunakan untuk menilai hasil-hasil pelajaran yang telah diberikan oleh guru kepada murid-muridnya, atau oleh dosen kepada mahasiswanya, dalam jangka waktu tertentu<sup>25</sup>. Menurut fungsinya tes hasil belajar dapat dibedakan menjadi empat macam yaitu:

1). Tes Penempatan (*Plecement test*)

Tes penempatan adalah tes untuk mengukur kemampuan dasar yang dimiliki oleh anak didik; kemampuan tersebut dapat dipakai untuk meramalkan kemampuan peserta didik pada masa mendatang, sehingga kepadanya dapat

---

<sup>24</sup> Anas Sudijono, *Pengantar Evaluasi*.....hal. 73

<sup>25</sup> Ngalim Purwanto, *Prinsip-Prinsip*.....hal. 33

dibimbing, diarahkan atau ditempatkan pada jurusan yang sesuai dengan kemampuan dasarnya<sup>26</sup>.

## 2) Tes Diagnostic

Adalah tes yang dilaksanakan untuk menentukan secara tepat, jenis kesukaran yang dihadapi oleh para peserta didik dalam suatu mata pelajaran tertentu. Dengan diketahuinya jenis-jenis kesukaran yang dihadapi oleh peserta didik itu maka lebih lanjut akan dapat dicarikan upaya berupa pengobatan (*therapy*) yang tepat. Tes diagnostic juga bertujuan untuk menemukan jawaban atas pertanyaan “Apakah peserta didik sudah dapat menguasai pengetahuan yang merupakan dasar atau landasan untuk dapat menerima pengetahuan selanjutnya?”<sup>27</sup>

## 3) Tes Formatif

Adalah tes hasil belajar yang bertujuan untuk mengetahui, sudah sejauh manakah peserta didik “telah terbentuk” (sesuai dengan tujuan pengajaran yang telah ditentukan) setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu<sup>28</sup>. Tes formatif juga bertujuan untuk mencari umpan balik (*feedback*), yang selanjutnya hasil penilaian tersebut dapat digunakan untuk memperbaiki proses belajar mengajar yang sedang atau yang sudah dilaksanakan<sup>29</sup>.

---

<sup>26</sup> Chabib Toha, *Teknik* .....hal. 46

<sup>27</sup> Anas Sudijono, *Pengantar Evaluasi*.....hal. 70

<sup>28</sup> Anas Sudijono, *Pengantar Evaluasi*.....hal. 71

<sup>29</sup> Ngalim Purwanto, *Prinsip-Prinsip*.....hal. 26

#### 4) Tes Sumatif

Adalah tes yang dilakukan untuk memperoleh data atau informasi sampai dimana penguasaan atau pencapaian belajar siswa terhadap bahan pelajaran yang telah dipelajarinya<sup>30</sup>. Tes ini mengukur keberhasilan belajar peserta didik secara menyeluruh, materi yang diujikan seluruh pokok bahasan dan tujuan pengajaran dalam satu program tahunan atau semesteran, masing-masing pokok bahasan terwakili dalam butir-butir soal yang diujikan<sup>31</sup>.

Klasifikasi tes hasil belajar menurut tingkatannya dapat dibedakan menjadi:

##### 1) Tes Standart

Pengertian tes standart secara sempit adalah tes yang disusun oleh satu tim ahli, atau disusun oleh lembaga yang khusus menyelenggarakan secara professional. Yang dituntut dalam tes standart bukan standart prestasi peserta didik dari penguasaan materi yang diajarkan pada suatu tingkat, lembaga pendidikan tertentu, melainkan adanya persamaan performance pada kelompok peserta didik atau lembaga pendidikan disebabkan adanya kesamaan tolak ukur<sup>32</sup>. Tes standar ini merupakan tes yang mengalami proses standardisasi,

---

<sup>30</sup> Ngalm Purwanto, *Prinsip-Prinsip*.....hal. 26

<sup>31</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 48.

<sup>32</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 51.

yaitu proses validasi dan keandalan (reliability) sehingga tes tersebut benar-benar valid dan andal untuk suatu tujuan dan bagi suatu kelompok tertentu.<sup>33</sup>

## 2) Tes Nonstandart

Adalah tes yang disusun oleh seorang pendidik yang belum memiliki keahlian professional dalam menyusun tes secara baik<sup>34</sup>.

Sedangkan menurut bentuknya, tes dapat dibedakan menjadi 3 macam, yakni:

### 1) Tes Tindakan

Adalah tes dimana respon atau jawaban yang dituntut dari peserta didik berupa tindakan, tingkah laku konkrit. Alat yang dapat digunakan untuk melakukan tes ini adalah observasi atau pengamatan terhadap tingkah laku tersebut<sup>35</sup>.

### 2) Tes Lisan

Tes lisan merupakan sekumpulan item pertanyaan dan atau pernyataan yang disusun secara terencana, diberikan oleh seorang guru kepada para siswanya tanpa media tulis<sup>36</sup>.

Dari segi persiapan dan cara bertanya tes lisan dapat dibedakan menjadi dua, yakni:

---

<sup>33</sup> Ngalim Purwanto, *Prinsip-Prinsip*.....hal. 33

<sup>34</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 52

<sup>35</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 63

<sup>36</sup> Sukardi, *Evaluasi Pendidikan*.....hal. 93

- a) Tes lisan bebas: artinya, pendidik dalam memberikan soal kepada peserta didik tanpa menggunakan pedoman yang dipersiapkan secara tertulis.
- b) Tes lisan berpedoman: pendidik menggunakan pedoman tertulis tentang apa yang akan ditanyakan kepada peserta didik<sup>37</sup>.

### 3) Tes Tertulis

Yaitu tes yang terdiri dari serangkaian soal, pertanyaan (item) atau tugas secara tertulis dan jawaban yang diberikan secara tertulis juga. Tes tertulis secara umum dapat dibedakan menjadi 2 macam, yakni:

#### a) Tes subyektif

Tes subyektif, yang pada umumnya berbentuk esai (uraian). Tes bentuk esai adalah sejenis tes kemajuan belajar yang memerlukan jawaban yang bersifat pembahasan atau uraian kata-kata<sup>38</sup>.

#### b) Tes obyektif

Yaitu tes yang terdiri dari butir-butir soal (item) yang dapat dijawab, oleh *testee* dengan jalan memilih salah satu (atau lebih) diantara beberapa kemungkinan jawaban yang telah dipasangkan dengan masing-masing item dengan jalan menuliskan (mengisi) jawabannya berupa kata-kata atau simbol-simbol tertentu pada tempat atau ruang yang telah disediakan untuk masing-masing butir item yang bersangkutan<sup>39</sup>.

---

<sup>37</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 61

<sup>38</sup> Suharsimi Arikunto, *Dasar-Dasar* .....hal.162

<sup>39</sup> Anas Sudijono, *Pengantar Evaluasi*.....hal. 106

Adapun macam-macam tes obyektif adalah sebagai berikut:

(1) Tes Melengkapi (*completion test*)

Adalah salah satu bentuk tes jawaban bebas, dimana butir-butir soalnya berupa satu kalimat dimana bagian-bagian tertentu yang dianggap penting dikosongkan, kepada *testee* diminta untuk mengisi bagian-bagian yang ditiadakan tersebut<sup>40</sup>.

(2) Tes benar-salah (*true-false test*)

Soal-soalnya berupa pernyataan-pernyataan (*statement*). *Statement* tersebut ada yang benar dan ada yang salah. Orang yang ditanya bertugas untuk menandai masing-masing pernyataan itu dengan melingkari huruf B jika pernyataan itu betul menurut pendapatnya dan melingkari huruf S jika pernyataan itu salah.

Bentuk benar-salah ada dua macam (dilihat dari segi mengerjakan/menjawab soal) yakni, dengan pembetulan yaitu siswa diminta membetulkan bila ia memilih jawaban yang salah atau siswa hanya diminta untuk melingkari huruf B atau S tanpa memberikan jawaban yang betul (tanpa pembetulan)<sup>41</sup>.

(3) Tes pilihan ganda (*multiple choice test*)

Tes pilihan ganda terdiri atas suatu keterangan atau pemberitahuan tentang suatu pengetahuan yang belum lengkap. Dan untuk melengkapinya

---

<sup>40</sup> Chabib Toha, *Teknik Evaluasi*.....hal. 67

<sup>41</sup> Suharsimi Arikunto, *Dasar-Dasar* .....hal.166

harus memilih satu dari beberapa kemungkinan jawaban yang telah disediakan<sup>42</sup>.

(4) Menjodohkan (*matching test*)

Tes bentuk menjodohkan merupakan bentuk khusus dari pilihan jamak. Bentuk ini terdiri atas dua macam kolom paralel, tiap kolom berisi statement yang satu menempati posisi sebagai soal dan satunya sebagai jawaban, kemudian peserta didik diminta untuk menjodohkan kesesuaian antar dua statement tersebut. Tes ini sering digunakan untuk mengukur informasi tentang fakta; pengertian; hubungan dan pengertian simbol tertentu<sup>43</sup>.

(5) *Rearrangement exercises*

Yang dimaksud dengan *Rearrangement exercises* adalah bentuk tes yang berupa rangkaian kalimat utuh dan benar, kemudian diceraikan secara tidak beraturan, sehingga bentuk aslinya sulit dikenali, peserta didik diminta menyusun kembali sesuai dengan urutan yang benar. Bentuk tes ini banyak digunakan untuk mata pelajaran bahasa Inggris<sup>44</sup>.

c. Penggolongan lain

Dilihat dari segi banyaknya orang yang mengikuti tes, tes dapat dibedakan menjadi dua, yaitu:

---

<sup>42</sup> Suharsimi Arikunto, *Dasar-Dasar*..... hal.168.

<sup>43</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 84

<sup>44</sup> Chabib Toha, *Teknik Evaluasi*.....hal. 84

- 1) Tes individual (*individual test*), yakni tes dimana *tester* hanya berhadapan dengan satu orang *testee* saja.
- 2) Tes kelompok (*group test*), yakni tes dimana *tester* berhadapan dengan lebih dari satu orang *testee*<sup>45</sup>.

Dilihat dari segi waktu yang disediakan bagi *testee* untuk menyelesaikan tes, tes dapat dibedakan menjadi dua golongan yaitu:

- 1) *Power test*, yaitu tes dimana waktu yang disediakan buat *testee* untuk menyelesaikan tes tersebut tidak dibatasi.
- 2) *Speed test*, yaitu tes dimana waktu yang disediakan buat *testee* untuk menyelesaikan tes tersebut dibatasi<sup>46</sup>.

Dilihat dari segi bentuk responnya tes dibedakan menjadi dua golongan, yaitu:

- 1) *Verbal test*, tes yang menghendaki respon (jawaban) tertuang dalam bentuk ungkapan kata atau kalimat.
- 2) *Non verbal test*, tes yang menghendaki respon (jawaban) tertuang dalam bentuk tindakan atau tingkah laku<sup>47</sup>.

#### 4. Ciri-Ciri Tes Yang Baik

Suatu tes dapat dikatakan baik sebagai alat pengukur jika dapat memenuhi syarat-syarat tes yang baik, di antara syarat-syarat tes tersebut adalah:

---

<sup>45</sup> Anas Sudijono, *Pengantar Evaluasi* .....hal. 74

<sup>46</sup> Anas Sudijono, *Pengantar Evaluasi* .....hal. 75

<sup>47</sup> Anas Sudijono, *Pengantar Evaluasi*.....hal. 75

a. Validitas.

Menurut Anastasi dalam Sumarna Surapranata, validitas adalah suatu tingkatan yang menyatakan bahwa suatu alat ukur telah sesuai dengan apa yang diukur<sup>48</sup>.

b. Reliabilitas.

Reliabilitas juga dapat diartikan sama dengan konsistensi atau keajegan. Suatu instrument evaluasi, dikatakan memenuhi nilai reliabilitas tinggi, apabila tes yang dibuat mempunyai hasil yang konsisten dalam mengukur yang hendak diukur<sup>49</sup>.

c. Obyektivitas.

Adalah kualitas yang menunjukkan identitas atau kesamaan dari skor-skor atau diagnosis-diagnosis yang diperoleh dari data yang sama dari penskor-penskor kompeten yang sama<sup>50</sup>.

d. Praktibilitas (Practibility).

Sebuah tes dikatakan memiliki praktibilitas yang tinggi apabila tes tersebut bersifat praktis, mudah mengadministrasikannya, praktis disini juga termasuk dalam pelaksanaan, pemeriksaan dan juga pemberian petunjuk-petunjuk yang jelas sehingga dapat diberikan/diwakili oleh orang lain<sup>51</sup>.

---

<sup>48</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 50.

<sup>49</sup> Sukardi, *Evaluasi Pendidikan* .....hal. 43

<sup>50</sup> Ngalim Purwanto, *Prinsip-Prinsip*.....hal. 137

<sup>51</sup> Suharsimi Arikunto, *Dasar-Dasar*.....hal. 62

e. Ekonomis.

Yang dimaksud ekonomis disini ialah bahwa pelaksanaan tersebut tidak membutuhkan ongkos atau biaya yang mahal, tenaga yang banyak, dan waktu yang lama<sup>52</sup>.

**B. Validitas.**

1. Pengertian Validitas

Validitas berasal dari kata *validity* yang mempunyai arti sejauh mana ketepatan dan kecermatan suatu instrument pengukuran (tes) dalam melakukan fungsi ukurnya. Suatu tes dapat dikatakan mempunyai validitas yang tinggi apabila tes tersebut menjalankan fungsi ukurnya<sup>53</sup>.

Valid, menurut Gronlund (1985), dalam Sukardi dapat diartikan sebagai ketepatan interpretasi yang dihasilkan dari skor tes atau instrument evaluasi. Gay (1983) dan Johnson dan Johnson (2002) mengatakan suatu instrument dikatakan valid apabila instrument yang digunakan dapat mengukur apa yang hendak diukur<sup>54</sup>.

Nunnaly (1972) menyatakan bahwa validitas tes perlu dilakukan untuk mengetahui kualitas tes dalam kaitannya mengukur hal yang seharusnya diukur.

---

<sup>52</sup> Suharsimi Arikunto, *Dasar-Dasar*.....hal. 62

<sup>53</sup> Syaifuddin Azwar, *Tes Prestasi*, ..... hal. 173

<sup>54</sup> Sukardi, *Evaluasi Pendidikan* .....hal. 30

Menurut Anastasi (1988) validitas adalah suatu tingkatan yang menyatakan bahwa suatu alat ukur telah sesuai dengan apa yang diukur<sup>55</sup>.

Scarvia B. Anderson, dkk, dalam bukunya *Encyclopedia Of Educational Evaluation* yang dikutip oleh Suharsimi Arikunto, menyatakan, “*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<sup>56</sup>.

Lindgren, H. C. (1967: 445) mengatakan bahwa: “*...validity the extent to wich they measure what they are expected to measure*”. Jadi validitas itu merupakan tingkat ketepatan tes tersebut dalam mengukur materi dan perilaku yang harus diukur<sup>57</sup>.

Dari beberapa pengertian diatas dapat diketahui bahwa tidak ada validitas yang berlaku secara umum untuk semua tujuan pengukuran, dengan kata lain suatu tes valid untuk mengukur suatu grup atau karakteristik belum tentu valid untuk grup atau karakteristik lainnya. Seperti yang diungkapkan oleh Saifuddin Azwar, bahwa tidak ada validitas yang berlaku secara umum untuk semua tujuan pengukuran. Suatu tes hanya menghasilkan ukuran yang valid untuk satu tujuan pengukuran yang spesifik saja<sup>58</sup>.

---

<sup>55</sup> Sumarna Surapranata, *Analisis, Validitas*, .....hal. 50

<sup>56</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi* .....hal. 65

<sup>57</sup> Ign. Masidjo, *Penilaian Pencapaian*.....hal.40

<sup>58</sup> Syaifuddin Azwar, *Tes Prestasi, Fungsi*.....hal. 174

## 2. Bentuk-Bentuk Validitas.

Secara metodologis, validitas suatu tes dapat dibedakan menjadi empat macam, yaitu validitas isi (*content validity*), validitas konstruk (*construct validity*), validitas konkuren (*Concurrent validity*) dan validitas prediksi (*Predictive validity*). Keempat macam validitas tersebut sering pula dikelompokkan menjadi dua macam menurut rentetan berfikir kedua macam validitas itu, yaitu validitas logis dan validitas empiris.

### a. Validitas logis

Istilah logis berasal dari kata “logika”, yang berarti penalaran. Dengan makna demikian maka validitas logis untuk sebuah instrument evaluasi menunjukkan pada sebuah kondisi bagi sebuah instrument yang memenuhi persyaratan valid berdasarkan hasil penalaran. Kondisi valid tersebut dipandang terpenuhi karena instrument yang bersangkutan sudah dirancang secara baik, mengikuti teori dan ketentuan yang ada<sup>59</sup>.

Ada dua macam validitas logis yang dapat dicapai oleh sebuah instrument, yaitu:

#### 1) Validitas isi (*Content validity*)

Yang dimaksud validitas isi ialah derajat dimana sebuah tes evaluasi mengukur cakupan substansi yang ingin diukur<sup>60</sup>. Validitas isi sering pula

---

<sup>59</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi* ..... hal. 63

<sup>60</sup> Sukardi, *Evaluasi Pendidikan Teknik* ..... hal. 32

dinamakan validitas kurikulum yang mengandung arti bahwa suatu alat ukur dipandang valid apabila sesuai dengan isi kurikulum yang hendak diukur<sup>61</sup>.

Sebuah tes dikatakan memiliki validitas isi apabila mengukur tujuan khusus tertentu yang sejajar dengan materi atau isi pelajaran yang diberikan<sup>62</sup>.

Validitas isi menunjukkan sejauh mana aitem-aitem dalam tes mencakup keseluruhan kawasan isi yang hendak diukur oleh tes tersebut. Pengertian mencakup keseluruhan kawasan isi tidak saja berarti tes itu harus komprehensif akan tetapi isinya harus pula tetap relevan dan tidak keluar dari batasan tujuan pengukuran<sup>63</sup>.

Menurut Gwon (1977), validitas isi dapat ditentukan dengan prosedur sebagai berikut<sup>64</sup>:

- a) Mendefinisikan domain yang hendak diukur.
- b) Menentukan domain yang akan diukur oleh masing-masing soal.
- c) Membandingkan masing-masing soal dengan domain yang sudah ditetapkan.

Selain dapat dianalisis secara teoritis validitas isi juga dapat dianalisis secara empiris. Menganalisis validitas isi (khusus untuk tes pilihan ganda) secara empiris dapat digunakan *Item and Test Analysis* (Iteman). Iteman merupakan perangkat lunak yang diciptakan khusus untuk analisis statistik butir soal dan tes. Program ini dibuat dengan pendekatan analisis statistik butir soal

---

<sup>61</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 51

<sup>62</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi* .....hal. 67

<sup>63</sup> Syaifuddin Azwar, *Tes Prestasi, Fungsi* ..... hal. 175

<sup>64</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 53

secara klasikal yang berguna untuk menentukan kualitas butir soal dan tes berdasarkan data empiris hasil uji coba. Alat analisis ini dimaksudkan untuk mendapatkan informasi tentang: validitas butir soal yang meliputi indeks kesukaran butir tes, indeks daya beda butir, dan keberfungsian pengecoh. Disamping itu, juga untuk menentukan: korelasi biserial titik (*point biserial correlation*), dan keseimbangan isi atau keterwakilan materi yang hendak diukur. Secara empiris kelima informasi tersebut tersebut dibutuhkan karena saling berhubungan antara satu dengan yang lainnya, dimana keberfungsian pilihan daya beda butir soal, dan indeks kesukaran dan daya beda butir dapat mempengaruhi interkorelasi butir, dan secara keseluruhan kelima informasi tersebut merupakan penentu tingkat reliabilitas tes.

## 2) Validitas konstruk (*Construct validity*)

Validitas konstruk merupakan derajat yang menunjukkan suatu tes mengukur sebuah konstruk sementara atau *hypotetical construct*. Secara definitive, konstruk merupakan suatu sifat yang tidak dapat diobservasi, tetapi kita dapat merasakan pengaruhnya melalui salah satu atau dua indera kita<sup>65</sup>.

Validitas konstruk mengandung arti bahwa suatu alat ukur dikatakan valid apabila telah cocok dengan konstruksi teoritik dimana tes itu dibuat. Dengan kata lain sebuah tes memiliki validitas konstruksi apabila soal-soalnya mengukur setiap aspek berfikir seperti yang diuraikan dalam standart

---

<sup>65</sup> Sukardi, *Evaluasi Pendidikan* .....hal. 33

kompetensi, kompetensi dasar maupun indikator yang terdapat dalam kurikulum<sup>66</sup>.

Jadi konstruksi disini merupakan rekaan psikologi atau teoritik terhadap suatu konsep tertentu. Seperti halnya aspek-aspek ingatan, pemahaman, aplikasi, analisis, sintesis dan evaluasi. Untuk pengujian validitas konstruk, diperlukan analisis yang kompleks seperti analisis faktor<sup>67</sup>.

Analisis faktor konfirmasi (*Confirmatory Factor Analisis*), yang digunakan untuk menganalisis instrument nontes secara empiris, dengan menggunakan metode ekstraksi komponen utama (*Principle Component Extraction*). Analisis tersebut bertujuan untuk menguji kebenaran konstruk teori yang dijadikan acuan dalam pengembangan instrumen, dengan cara menentukan struktur atau model faktor dari sejumlah butir instrument berdasarkan muatan faktor (*factor loading*), jumlah varians (*eigenvalue*) dan proporsi varian (*communality*).

Selain itu ada juga pengujian validitas konstruksi ini yang lebih sederhana, yaitu dengan cara<sup>68</sup>:

a) Pengujian validitas konvergen.

Yaitu mengkorelasikan skor total dengan skor faktor, dengan asumsi antara skor total dengan skor faktor terdapat korelasi yang signifikan.

---

<sup>66</sup> Sumarna Surapranata, *Analisis, Validitas*,..... hal. 53

<sup>67</sup> Syaifuddin Azwar, *Tes Prestasi, Fungsi*..... hal. 176

<sup>68</sup> Chabib Toha, *Teknik Evaluasi*..... hal. 111

b) Pengujian validitas diskriminan.

Adalah dengan mengkorelasikan skor faktor yang satu dengan skor faktor yang lain, dengan asumsi bahwa masing-masing faktor tidak berkorelasi secara signifikan, sehingga tiap faktor secara khusus mengukur faktor tertentu.

c) Pengujian stabilitas dan keajegan.

Yaitu dengan cara tes-retes, uji konsistensi dengan uji belah dua, sedangkan untuk stabilitasnya, dapat dilakukan dengan cara membandingkan hasil antar kelompok.

b. Validitas Empiris.

Istilah “validitas empiris” memuat kata “empiris” yang artinya “pengalaman”<sup>69</sup>. Dinamakan demikian karena validitas tersebut ditentukan dengan menghubungkan performansi sebuah tes terhadap kriteria penampilan tes lainnya dengan menggunakan formulasi statistic<sup>70</sup>. Hal ini berarti validitas empiris tidak dapat diperoleh hanya dengan menyusun instrument berdasarkan ketentuan seperti halnya validitas logis.

Ada dua macam validitas empiris, yaitu:

1) Validitas konkuren (*Concurrent validity*)

Validitas konkuren atau validitas *ada sekarang* menunjukkan pada hubungan antara tes skor yang dicapai dengan keadaan sekarang. Sebuah tes

---

<sup>69</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi*.....hal. 66

<sup>70</sup> Sukardi, *Evaluasi Pendidikan* .....hal. 32

dikatakan memiliki validitas konkuren apabila hasilnya sesuai dengan pengalaman<sup>71</sup>.

Validitas ini juga disebut sebagai validitas pengukuran serentak, hal ini dikarenakan tes tersebut memiliki kesesuaian dengan hasil pengukuran lain yang dilaksanakan pada saat itu dengan menggunakan alat ukur yang berbeda<sup>72</sup>. Tes dengan validitas konkuren biasanya diadministrasikan dalam waktu yang sama atau dengan kriteria valid yang sudah ada<sup>73</sup>.

Jadi tes dapat dikatakan memiliki validitas konkuren yang tinggi adalah tes yang mempunyai korelasi yang tinggi dengan hasil suatu alat ukur lain terhadap bidang yang sama dalam waktu yang sama.

## 2) Validitas prediksi (*Predictive validity*)

*Predictive validity* menunjukkan hubungan antara tes skor yang diperoleh peserta tes dengan keadaan yang akan terjadi diwaktu yang akan datang. Sebuah tes dikatakan memiliki validitas prediksi apabila mempunyai kemampuan untuk memprediksikan apa yang akan terjadi di masa yang akan datang<sup>74</sup>. Tepat-tidaknya ramalan tersebut dapat dilihat dari korelasi koefisien antara hasil tes itu dengan hasil alat ukur lain pada masa mendatang<sup>75</sup>.

Validitas prediksi pada suatu tes pada umumnya ditentukan dengan membangun hubungan antara skor tes dan beberapa ukuran keberhasilan

---

<sup>71</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 55

<sup>72</sup> Chabib Toha, *Teknik Evaluasi*.....hal. 116

<sup>73</sup> Sukardi, *Evaluasi Pendidikan* .....hal. 34

<sup>74</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 54

<sup>75</sup> Ngalim Purwanto, *Prinsip-Prinsip Dan*.....hal. 138

dalam situasi tertentu yang digunakan untuk memprediksi keberhasilan, yang selanjutnya sebagai prediktor. Sedangkan tingkah laku yang hendak diprediksi pada umumnya disebut sebagai kriteria. Dalam membuat validitas prediksi, suatu tes biasanya mempunyai sekuensi seperti berikut. Mengidentifikasi dan mendefinisikan secara teliti kriteria yang hendak diinginkan. Criteria yang terpilih harus mengukur validitas terhadap tingkah laku yang diprediksi<sup>76</sup>.

### 3. Faktor-Faktor Yang Mempengaruhi Validitas.

Banyak faktor yang dapat mempengaruhi hasil tes evaluasi tidak valid. Beberapa faktor tersebut secara garis besar dapat dibedakan menurut sumbernya, yaitu:

#### a. Faktor yang berasal dari dalam tes.

Beberapa sumber yang pada umumnya berasal dari faktor internal tes evaluasi di antaranya sebagai berikut.

- 1) Arahan tes yang disusun dengan makna tidak jelas sehingga dapat mengurangi validitas tes.
- 2) Kata-kata yang digunakan dalam struktur instrument evaluasi, terlalu sulit.
- 3) Item-item tes dikonstruksi dengan jelek.
- 4) Tingkat kesulitan item tes tidak tepat dengan materi pembelajaran yang diterima siswa.

---

<sup>76</sup> Sukardi, *Evaluasi Pendidikan Teknik*.....hal. 36

- 5) Waktu yang dialokasikan tidak tepat, hal ini termasuk kemungkinan terlalu kurang atau terlalu longgar.
- 6) Jumlah item tes terlalu sedikit sehingga tidak mewakili sample materi pembelajaran.
- 7) Jawaban masing-masing item evaluasi bisa diprediksi siswa.

b. Faktor yang berasal dari administrasi dan skor.

- 1) Waktu pengerjaan tidak cukup sehingga siswa dalam memberikan jawaban dalam situasi yang tergesa-gesa.
- 2) Adanya kecurangan dalam tes sehingga tidak bisa membedakan antara siswa yang belajar dengan yang melakukan kecurangan.
- 3) Pemberian petunjuk dari pengawas yang tidak dapat dilakukan pada semua siswa.
- 4) Teknik pemberian skor yang tidak konsisten, misalnya pada tes esai, juga dapat mengurangi validitas tes evaluasi.
- 5) Siswa tidak dapat mengikuti arahan yang diberikan dalam tes esai, juga dapat mengurangi validitas tes evaluasi.
- 6) Siswa tidak dapat mengikuti arahan yang diberikan dalam tes baku.

c. Faktor-faktor yang berasal dari jawaban siswa.

Seringkali terjadi bahwa interpretasi terhadap item-item tes evaluasi tidak valid, karena dipengaruhi oleh jawaban siswa dari pada interpretasi item-item pada tes evaluasi.

## C. Reliabilitas

### 1. Pengertian Reliabilitas.

Syarat lain yang juga penting bagi suatu instrument evaluasi adalah reliabilitas. Kata *reliabilitas* dalam bahasa Indonesia diambil dari kata *reliability* dalam bahasa Inggris, berasal dari kata *reliable* yang artinya dapat dipercaya<sup>77</sup>. Reliabilitas juga dapat diartikan sama dengan konsistensi atau keajegan<sup>78</sup>, ketetapan atau ketelitian<sup>79</sup>.

Selain itu reliabilitas juga diartikan dengan keajegan dan stabilitas. Diartikan keajegan bilamana tes tersebut diujikan berkali-kali hasilnya relative sama, artinya setelah hasil tes pertama dengan tes berikutnya dikorelasikan terdapat hasil korelasi yang signifikan. Reliabilitas diartikan dengan stabilitas bilamana tes itu diujikan dan hasilnya diadakan analisis reliabilitas dengan menggunakan kriteria internal dalam tes tersebut<sup>80</sup>.

Nunally (1970), Allen dan Yen (1979) dan Anastasi (1986) menyatakan bahwa reliabilitas adalah kestabilan skor yang diperoleh orang yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda atau dari satu pengukuran ke pengukuran lainnya<sup>81</sup>.

---

<sup>77</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi* .....hal. 59

<sup>78</sup> Sukardi, *Evaluasi Pendidikan Teknik dan* .....hal. 43

<sup>79</sup> Ngalm Purwanto, *Prinsip-Prinsip Dan*..... hal. 139

<sup>80</sup> Chabib Toha, *Teknik Evaluasi*.....hal. 118-119

<sup>81</sup> Sumarna Surapranata, *Analisis, Validitas*,..... hal. 69

Walaupun reliabilitas memiliki berbagai nama lain seperti keterpercayaan, keterandalan, keajegan, konsistensi, kestabilan dan sebagainya, namun ide pokok dalam konsep reliabilitas adalah sejauh mana hasil suatu pengukuran dapat dipercaya. Hasil ukur adalah dapat dipercaya apabila dalam beberapa kali pengukuran terhadap kelompok subjek yang sama diperoleh hasil yang relative sama<sup>82</sup>.

Secara empirik, tinggi-rendahnya reliabilitas ditunjukkan oleh suatu angka yang disebut koefisien reliabilitas. Koefisien korelasi antara dua variable dilambangkan oleh huruf  $r$ . apabila skor pada tes pertama diberi lambing  $X$  dan skor pada tes yang ke dua yang pararel diberi lambing  $X'$ , maka koefisien korelasi anantara kedua tes tersebut adalah  $r_{XX'}$ . Symbol inilah yang digunakan sebagai symbol koefisien reliabilitas.

Besarnya koefisien reliabilitas berkisar antara 0,0 sampai dengan 1,0. Meskipun koefisien korelasi juga dapat bertanda negative (-), namun koefisien reliabilitas selalu mengacu pada angka positif (+) dikarenakan angka yang negative tidak ada artinya bagi interpretasi reliabilitas hasil ukur.<sup>83</sup>

## 2. Estimasi Reliabilitas.

Tujuan utama mengestimasi reliabilitas adalah untuk menentukan seberapa besar reliabilitas yang terjadi akibat adanya kesalahan pengukuran dan seberapa

---

<sup>82</sup> Syaifuddin Azwar, *Tes Prestasi, Fungsi*..... hal. 180

<sup>83</sup> Syaifuddin Azwar, *Reliabilitas dan Validitas*, (Yogyakarta: Pustaka Pelajar Offset, 1997), hal. 9

besar variabilitas skor sebenarnya<sup>84</sup>. Estimasi reliabilitas dapat dilakukan melalui salah satu pendekatan umum, yaitu pendekatan tes-ulang (tes-retest), pendekatan tes-sejajar (*alternate-forms*), dan pendekatan konsistensi internal.

a. Pendekatan tes ulang (*test-retest method*).

Metode ini sering pula dinamakan metode stabilitas, merupakan pendekatan yang paling tua yang digunakan untuk mengestimasi reliabilitas, dan juga dinamakan *single-test-double-trial method*<sup>85</sup>.

Asumsi yang menjadi dasar dalam cara ini adalah bahwa suatu tes yang reliable tentu akan menghasilkan skor tampak yang relative sama apabila dikenakan dua kali pada waktu yang berbeda. Semakin besar variasi perbedaan skor subyek antara kedua pengenaan itu berarti semakin sulit untuk mempercayai bahwa tes itu memberikan hasil ukur yang konsisten.<sup>86</sup>

Prinsip estimasinya adalah mengenakan suatu instrument pengukuran dua kali dengan tenggang waktu tertentu, terhadap sekelompok subjek yang sama. Kelemahan pendekatan tes ulang adalah kurang praktisnya pengenaan tes dua kali dan besarnya kemungkinan tes terjadi efek bawaan (*carry-over effects*) dari satu pengenaan tes ke pengenaan yang kedua<sup>87</sup>.

---

<sup>84</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 89

<sup>85</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 93

<sup>86</sup> Syaifuddin Azwar, *Reliabilitas dan*, .....hal. 36

<sup>87</sup> Syaifuddin Azwar, *Tes Prestasi*, .....hal. 182

Adapun langkah-langkah yang dapat ditempuh pada uji reliabilitas ini adalah sebagai berikut<sup>88</sup>:

- 1) Menyusun sebuah tes yang akan diukur reliabilitasnya.
- 2) Menguji tes yang tersusun tersebut (tahap I).
- 3) Menghitung skor hasil tes tahap I.
- 4) Menguji ulang tes yang tersusun tersebut (tahap II).
- 5) Menghitung skor hasil tes ulang (tahap II).
- 6) Menghitung reliabilitas tes tersebut dengan jalan mengkorelasikan skor tes I dengan skor tes II dengan rumus korelasi *Product Moment Pearson* sebagai berikut:

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

b. Pendekatan tes sejajar.

Pendekatan tes-sejajar atau ekuivalen sering pula dinamakan *alternate-forms methods* atau *double test-double trial method*<sup>89</sup>. Metode ini berkenaan dengan penggunaan dua buah tes yang sama atau relative sama kepada peserta didik yang sama.<sup>90</sup>

---

<sup>88</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 120

<sup>89</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 97

<sup>90</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 97

Pendekatan ini hanya dapat dilakukan apabila tersedia dua bentuk instrument pengukuran yang dapat dianggap memenuhi asumsi paralel, yakni mengukur obyek psikologis yang sama, berdasarkan *bleu-print* yang sama serta spesifikasi yang sama pula<sup>91</sup>. Spesifikasi ini meliputi antara lain tujuan ukur, batasan objek ukur dan operasionalisasinya, indikator-indikator perilakunya banyaknya item, format item, juga jika perlu meliputi taraf kesukaran item dan lain sebagainya.<sup>92</sup>

Cara ini dapat digunakan untuk mengetahui koefisien stabilitas tes dengan asumsi bahwa sistem yang diukur dengan tes tersebut tidak akan berubah dengan hanya digunakan dua bentuk tes. Adapun langkah yang ditempuh adalah<sup>93</sup>:

- 1) Menyusun dua buah tes yang ekuivalen.
- 2) Menguji kedua tes tersebut (dalam waktu yang bersamaan atau beriringan).
- 3) Memberikan skor hasil tes yang telah diujikan, disusun dengan memisahkan antara tes A dan tes B.
- 4) Mencari koefisien stabilitas kedua tes (A dan B) dengan jalan mencari korelasinya dengan rumus korelasi *Product Moment*.

---

<sup>91</sup> Syaifuddin Azwar, *Tes Prestasi*, .....hal. 182

<sup>92</sup> Syaifuddin Azwar, *Reliabilitas dan*, .....hal. 39

<sup>93</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 123

Bilamana dua buah tes sudah disusun berdasarkan petunjuk tersebut dan sudah diujikan kepada satu sample, maka hasil kedua buah tes tersebut dikorelasikan dengan rumus korelasi *Product Moment* seperti digunakan di atas.

Walaupun pendekatan reliabilitas bentuk paralel dapat menghilangkan masalah penentuan tenggang waktu yang tepat, akan tetapi pendekatan ini tidak menghilangkan sama sekali kemungkinan terjadinya efek bawaan. Adapun kelemahan utama dalam pendekatan ini terletak pada sulitnya menyusun dua tes yang paralel itu sendiri.<sup>94</sup>

c. Pendekatan konsistensi internal

Estimasi reliabilitas dengan pendekatan konsistensi internal didasarkan pada data dari sekali penguasaan satu bentuk alat ukur pada sekelompok subjek (*Single trial administration*)<sup>95</sup>. Pendekatan reliabilitas konsistensi internal bertujuan melihat konsistensi antar item atau antar bagian dalam tes itu sendiri.<sup>96</sup>

Diantara metode-metode yang digunakan dalam pendekatan ini adalah sebagai berikut:

1) Metode belah dua (*Split half methods*)

Metode ini memungkinkan mengestimasi reliabilitas tanpa harus menyelenggarakan tes dua kali. Dengan demikian beberapa kelemahan seperti

---

<sup>94</sup> Syaifuddin Azwar, *Reliabilitas dan*.....hal. 41

<sup>95</sup> Syaifuddin Azwar, *Tes Prestasi, Fungsi*.....hal. 182

<sup>96</sup> Syaifuddin Azwar, *Reliabilitas dan*.....hal. 42

*carry-over effect* dan khususnya terhadap perolehan skor sebenarnya dapat di minimalisasi<sup>97</sup>.

Ada dua cara membelah butir soal ini yaitu<sup>98</sup>:

- a) Membelah item-item genap dan item-item ganjil yang selanjutnya disebut ganjil-genap dan,
- b) Membelah atas item-item awal dan item-item akhir yaitu separuh jumlah pada nomer-nomer awal dan separuh pada nomer-nomer akhir yang selanjutnya disebut belahan awal-akhir.

Beberapa rumus untuk mencari teknik reliabilitas yang menggunakan teknik belah dua adalah:

- a) Rumus Spearman-Brown

Adapun cara-cara yang dapat ditempuh adalah<sup>99</sup>:

- (1) Membelah skor tes ke dalam skor ganjil dan genap.
- (2) Skor ganjil menjadi variable X, dan skor genap menjadi variable Y.
- (3) Menghitung koefisien korelasi  $\frac{1}{2}$  tes menggunakan korelasi *Product moment*.
- (4) Menghitung koefisien korelasi satu tes penuh dengan rumus Spearman

Brown sebagai berikut:

---

<sup>97</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 100

<sup>98</sup> Suharsimi Arikunto, *Dasar-Dasar*.....hal. 93

<sup>99</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 125

$$\text{Rumus: } r_{XX'} = \frac{2 \times r_{\frac{11}{22}}}{\left(1 + r_{\frac{11}{22}}\right)}$$

Keterangan:

$r_{XX'}$  = Reliabilitas Instrumen (satu tes penuh)

$r_{\frac{11}{22}}$  = Reliabilitas ½ tes/setengah tes

- (5) Setelah diketahui koefisien korelasi satu tes penuh, dilanjutkan dengan tes signifikansi table r *Product Moment*.

Formula Spearman-brown hanya akan menghasilkan estimasi reliabilitas yang cermat apabila koefisien korelasi diantara kedua belahan tes itu tinggi, karena tingginya korelasi antara kedua belahan merupakan pula indikasi terpenuhinya asumsi paralelisme.<sup>100</sup>

b) Rumus Rulon

Cara mencari reliabilitas dengan menggunakan rumus Rulon ini tidak lagi menggunakan korelasi *Product Moment* namun didasarkan pada selisih skor subjek pada kedua belah tersebut.

Adapun langkah-langkahnya adalah sebagai berikut<sup>101</sup>:

- (1) Membelah hasil tes menjadi dua, yaitu belahan genap, serta skor total.
- (2) Mencari deviasi antara skor ganjil dan skor genap.

---

<sup>100</sup> Syaifuddin Azwar, *Reliabilitas dan*.....hal. 71

<sup>101</sup> Chabib Toha, *Teknik Evaluasi* .....hal. 131

(3) Mencari standar deviasi kuadrat dari deviasi nilai tersebut dan standar deviasi kuadrat dari skor total.

(4) Menghitung besarnya reliabilitas dengan formula Rulon:

$$r_{XX'} = 1 - \frac{SD_D^2}{SD_t^2}$$

Keterangan:

$r_{XX'}$  = Reliabilitas tes

$SD_D^2$  = Standar deviasi kuadrat (varian) dari selisih skor ganjil dan genap.

c) Rumus Flanagan

Persamaan lain yang juga dapat digunakan untuk menentukan reliabilitas belah dua adalah persamaan Flanagan, yaitu<sup>102</sup>:

$$r_{XX'} = 2 \left( 1 - \frac{SD_1^2 + SD_2^2}{SD_t^2} \right)$$

Keterangan:

$r_{XX'}$  = Reliabilitas tes

$SD_1$  = Standart deviasi pada belahan pertama

$SD_2$  = Standart deviasi pada belahan kedua

$SD_t$  = standart deviasi skor total

2) Metode untuk mengukur homogenitas tes

---

<sup>102</sup> Syaifuddin Azwar, *Tes Prestasi, Fungsi.....*hal. 184

Cara mengukur konsistensi internal yang lain, tidak harus membagi tes menjadi dua bagian. Prosedur ini menilai konsistensi antar item atau homogenitas item-item dalam satu tes evaluasi yang direncanakan<sup>103</sup>. Terdapat beberapa teknik dan persamaan yang digunakan dalam hal ini, yakni:

a) Koefisien Alpha

Koefisienn alpha dihitung dengan menggunakan persamaan sebagai berikut<sup>104</sup>:

$$r_{xx'} = \frac{k}{k-1} \left( 1 - \frac{\sum S_i^2}{S_t^2} \right)$$

Keterangan:

$r_{xx'}$  = Reliabilitas tes

k = Jumlah soal

$S_i^2$  = Jumlah varian (Standart deviasi kuadrat) dari skor soal

$S_t^2$  = Jumlah varian (Standart deviasi kuadrat) dari skor total

Alpha Cronbach pada prinsipnya termasuk mengukur homogenitas yang di dalamnya memfokuskan pada dua aspek penting, yaitu aspek isi (*content*) dan aspek heterogenitas dari tes tersebut<sup>105</sup>. Selain itu dengan perkembangan teknologi informasi dan semakin intensifnya penggunaan computer untuk mendukung suatu penelitian, perhitungan formula Alpha

---

<sup>103</sup> Sukardi, *Evaluasi Pendidikan Teknik*..... hal. 49

<sup>104</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 114

<sup>105</sup> Sukardi, *Evaluasi Pendidikan Teknik*..... hal. 50

sudah banyak dilakukan dengan program computer dengan Alpha sebagai indeks reliabilitas. Salah satu program computer tersebut adalah *Item and Test Analysis* (iteman) versi 3.00.

b) Rumus Kuder-Richardson

Salah satu indeks homogenitas yang paling banyak digunakan dan sering ditemukan dalam proses penelitian evaluasi adalah formula Kuder Richardson (KR). Ada dua macam Kuder Richardson, yaitu KR-20 dan KR-21<sup>106</sup>.

Adapun formula KR-20 adalah:

$$r_{xx'} = \left( \frac{k}{k-1} \right) \left( \frac{S^2 - \sum pq}{S^2} \right)$$

Keterangan:

$r_{xx'}$  = Reliabilitas menggunakan KR-20

p = Proporsi peserta tes menjawab benar

q = Proporsi peserta tes menjawab salah (q=1-p)

$\sum pq$  = Jumlah perkalian antara p dan q

k = Banyaknya soal

S = Standart deviasi

Sedangkan rumus KR-21 adalah sebagai berikut:

---

<sup>106</sup> Sukardi, *Evaluasi Pendidikan Teknik*.....hal. 49

$$r_{xx'} = \left( \frac{k}{k-1} \right) \left( 1 - \frac{M(k-M)}{k \cdot SD_t^2} \right)$$

Keterangan:

M = Mean skor (skor total dibagi N)

$r_{xx'}$  = Reliabilitas menggunakan KR-21

k = Banyak soal

$SD_t$  = Standart deviasi

c) Persamaan Hoyt

Persamaan Hoyt adalah sebagai berikut<sup>107</sup>:

$$r_{xx'} = \frac{S_t^2 - S_2^2}{S_t^2} \text{ atau } r_{xx'} = 1 - \frac{S_2^2}{S_t^2}$$

Keterangan:

$r_{xx'}$  = Reliabilitas tes

$S_t^2$  = Varian peserta

$S_2^2$  = Varian sisa

### 3. Faktor-Faktor Yang Mempengaruhi Reliabilitas

Koefisien reliabilitas dapat dipengaruhi di antaranya oleh waktu penyelenggaraan tes-retes. Interval penyelenggaraan yang terlalu dekat atau terlalu jauh akan mempengaruhi koefisien reliabilitas.

---

<sup>107</sup> Sumarna Surapranata, *Analisis, Validitas*,.....hal. 117

Faktor-faktor lain yang juga mempengaruhi reliabilitas instrument evaluasi di antaranya sebagai berikut:

- a. Panjang tes, semakin panjang suatu tes evaluasi, semakin banyak jumlah item materi pembelajaran diukur.
- b. Penyebaran skor, koefisien reliabilitas secara langsung dipengaruhi oleh bentuk sebaran skor dalam kelompok siswa yang diukur.
- c. Kesulitan tes, tes normative yang terlalu mudah atau terlalu sulit untuk siswa cenderung menghasilkan skor reliabilitas rendah.
- d. Obyektivitas, yang dimaksud dengan obyektif yaitu derajat di mana siswa dengan kompetensi sama, mencapai hasil sama<sup>108</sup>.

Sedangkan menurut Ngalim Purwanto<sup>109</sup>, faktor-faktor yang dapat mempengaruhi keandalan suatu tes adalah:

- a. Luas-tidaknya sampling yang di ambil. Makin luas suatu sampling, berarti tes semakin andal.
- b. Perbedaan bakat dan kemampuan murid yang di tes. Makin variabel kemampuan peserta tes, berarti makin tinggi keadaan koefisien tes.
- c. Suasana dan kondisi testing. Suasana ketika berlangsung testing, seperti tenang, gaduh, banyak gangguan, pengetes yang marah-marrah dapat mengganggu mengerjakan tes sehingga dengan demikian mempengaruhi pula hasil dan keandalan tes.

---

<sup>108</sup> Sukardi, *Evaluasi Pendidikan Teknik*.....hal. 52

<sup>109</sup> Ngalim Purwanto, *Prinsip-Prinsip Dan*.....hal. 141

#### **D. Hubungan Antara Taraf Validitas dan Reliabilitas Suatu Tes**

Suatu tes yang reliable atau handal adalah suatu tes yang hasil pengukurannya dalam satu atau berbagai pengukuran menunjukkan hasil yang konsisten atau hasil yang tepat dan teliti. Akan tetapi hasil pengukuran yang konsisten atau tepat dan teliti dari suatu tes belum menjamin bahwa hasil pengukuran yang demikian itu merupakan hasil yang dikehendaki oleh tes tersebut. Dengan kata lain hasil pengukuran dari suatu tes yang konsisten belum tentu valid. Reliabilitas pengukuran instrument evaluasi diperlukan untuk mencapai hasil pengukuran yang valid. Dalam kaitannya dengan posisi konsistensi, para penilai bisa memiliki instrumen evaluasi yang reliable tanpa valid, sebaliknya kita mempunyai instrument valid dengan reliabilitas yang baik.

Apabila tes yang valid ini dicapai dalam satu atau berbagai pengukuran, maka akan tetap atau konsisten mengukur apa yang seharusnya diukur. Hal ini dikarenakan suatu tes yang valid adalah suatu tes yang mampu mengukur apa yang seharusnya diukur. Karena tes yang valid tersebut telah disusun berdasarkan perencanaan yang baik dan petunjuk-petunjuk konstruksi.

Misalnya suatu tes evaluasi pengajaran yang direncanakan berdasarkan langkah-langkah perencanaan secara tepat (berdasarkan kompetensi dasar, rincian bahan pelajaran dan visualisasi kisi-kisi yang sesuai) dan disusun berdasarkan petunjuk-petunjuk konstruksi serta sempat diujicobakan dalam suatu pengukuran (bersama tes pengajaran lainnya yang valid), maka dari tes evaluasi pengajaran

tersebut selain dapat dicari validitas isi dan konstruksinya dapat juga di cari validitas kriterianya serta reliabilitasnya<sup>110</sup>.

## **E. Hipotesis**

Hipotesis dapat diartikan sebagai suatu jawaban yang bersifat sementara terhadap permasalahan penelitian sampai terbukti melalui data yang terkumpul. Adapun rumusan hipotesis pada penelitian ini adalah:

1. Hipotesis nol atau *null hypotheses* (Ho), yang menyatakan:

Tes mata pelajaran PAI kelas VII SLTPN 13 Surabaya tahun ajaran 2008/2009 tidak valid.

Tes mata pelajaran PAI kelas VII SLTPN 13 Surabaya tahun ajaran 2008/2009 tidak reliabel.

2. Hipotesis kerja atau hipotesis alternative (Ha), yang menyatakan:

Tes mata pelajaran PAI kelas VII SLTPN 13 Surabaya tahun ajaran 2008/2009 valid.

Tes mata pelajaran PAI kelas VII SLTPN 13 Surabaya tahun ajaran 2008/2009 reliabel.

---

<sup>110</sup> Ign. Masidjo, *Penilaian Pencapaian* ..... hal. 257