

# BAB I

## PENDAHULUAN

### A. Latar Belakang

Secara formal, pendidikan diselenggarakan disekolah. Hal itu sering dikenal dengan pengajaran dimana proses belajar mengajar yang melibatkan banyak faktor baik pengajar, pelajar, bahan/materi, fasilitas maupun lingkungan. Pengajaran dilaksanakan tidak hanya untuk kesenangan atau bersifat mekanis saja tetapi mempunyai misi atau tujuan bersama. Dalam usaha untuk mencapai misi dan tujuan itu perlu diketahui apakah usaha yang dilakukan sudah sesuai dengan tujuan? Untuk mengetahui apakah tujuan pendidikan sudah tercapai perlu diadakan tes. Sebuah tes yang dapat dikatakan baik sebagai alat pengukur harus memenuhi persyaratan tes, yaitu : validitas, reliabilitas, objektivitas, praktibilitas dan ekonomis.

Ketika mengkonstruksi sebuah tes, maka pertanyaan yang muncul adalah apakah interpretasi skor yang dihasilkan dari penggunaan instrument sebagai alat ukur tepat, bermanfaat dan dapat digunakan pada keadaan sekolah yang ada? Ada banyak macam tes dan bervariasi pula kegunaannya tergantung dari tujuan yang hendak dicapai. Tes dan instrument evaluasi dalam kategori *paper* dan *pencil* dapat digunakan untuk tes pencapaian hasil belajar, yang didalamnya tercakup focus-fokus penting tujuan pilihan, penempatan, diagnosis, dan sertifikasi profesi. Tes kecakapan digunakan untuk memprediksi keberhasilan siswa dalam prospek belajar, atau dalam prospek bekerja. Tes penghargaan digunakan untuk mengestimasi keberhasilan siswa, tes perkembangan social, dan juga mengetahui problem siswa dan beberapa tes lainnya. Agar supaya instrument tersebut memiliki kemampuan mengevaluasi secara baik maka perlu memenuhi beberapa karakteristik penting dan perlu. Karakteristik penting tersebut adalah; validitas, reliabilitas dan dapat digunakan (*Usability*).

### B. Rumusan Masalah

- a. Apakah yang dimaksud dengan analisis kualitatif dan kuantitatif?
- b. Apakah yang dimaksud dengan daya beda?
- c. Apakah fungsi dari pengecoh?
- d. Apakah yang dimaksud dengan validitas?
- e. Apakah yang dimaksud dengan realibilitas?

### **C. Tujuan dan Manfaat Pembuatan Makalah**

Setelah membaca dan mempelajari makalah ini mahasiswa diharapkan mampu :

- a. Mendeskripsikan analisis kualitatif dan kuantitatif serta mampu membedakannya
- b. Menjelaskan pengertian daya beda
- c. Menjelaskan fungsi dari pengecoh
- d. Menjelaskan validitas serta mampu menggunakan metode pengujiannya
- e. Menjelaskan realibilitas serta mampu menggunakan metode pengujiannya
- f. Sebagai calon pengajar dan pendidik, materi ini bisa menjadi bekal mahasiswa saat sudah menjadi guru.

### **D. Metode Penulisan**

Makalah ini disusun berdasarkan referensi (tinjauan pustaka) dari buku, jurnal maupun dari internet.

## **BAB II**

### **PEMBAHASAN**

#### **A. ANALISIS KUALITATIF (TEORITIK)**

Telaah soal atau penilaian butir soal secara kualitatif merupakan analisis teoritis, dilakukan sendiri atau bantuan teman sejawat, dilakukan sebelum soal diujikan dengan kriteria berdasarkan: (1) isi atau materi, (2) konstruksi, dan (3) bahasa.<sup>1</sup>

Pada teknik analisis kualitatif ini terdapat 2 cara yang dapat dilakukan yaitu:

##### **1. Teknik Moderator**

Merupakan teknik berdiskusi yang didalamnya terdapat satu orang sebagai penengah. Berdasarkan teknik ini , setiap butir soal didiskusikan secara bersama-sama dengan beberapa ahli dan dimoderatori oleh satu orang. Kelebihan teknik ini adalah setiap butir soal dapat dituntaskan secara bersama-sama, sedangkan kelemahannya adalah teknik ini memerlukan waktu lama untuk mendiskusikan setiap satu butir soal.

##### **2. Teknik Panel**

---

<sup>1</sup> Ahmad Sofyan, *Evaluasi Pembelajaran IPA Berbasis Kompetensi*, (Jakarta : UIN Jakarta Press, 2006), h. 101

Merupakan suatu teknik menelaah butir soal yang setiap butir soalnya ditelaah berdasarkan kaidah penulisan butir soal. Para penelaah biasanya dipersilahkan memperbaiki langsung pada teks soal dan memberikan komentarnya serta memberikan nilai pada setiap butir soalnya yang kriterianya berdasarkan: isi materi, konstruksi, dan bahasa.

Aspek Analisis	Kriteria Penilaian	1	2	3	4
Isi dan Materi	1. Soal sesuai dengan indicator				
	2. Pilihan jawaban homogen dan logis				
	3. Hanya ada satu kunci jawaban yang paling tepat.				
Kontruksi	4. Pokok soal dirumuskan dengan singkat, jelas dab tegas.				
	5. Rumusan pokok soal dan pilihan jawaban merupakan pernyataan yang diperlukan.				
	6. Pokok soal tidak memberi petunjuk kearah kunci jawaban.				
	7. Pokok soal bebas dari pernyataan yang bersifat negative ganda.				
	8. Gambar/grafik/tabel/diagram dan sejenisnya jelas dan berfungsi.				
	9. Panjang pilihan jawaban relatif sama.				
	10. Pilihan jawaban tidak menggunakan pernyataan yang berbunyi “semua jawaban di atas salah”, “a,b, dan c salah” dan sejenisnya.				
	11. Pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan urutan besar kecilnya angka atau kronologis.				
	12. Butir soal tidak bergantung pada jawaban soal sebelumnya.				
Bahasa	13. Menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia.				
	14. Menggunakan bahasa yang komunikatif.				
	15. Tidak menggunakan bahasa yang berlaku setempat.				

	16. Pilihan jawaban tidak mengulang kata/kelompok kata yang sama.				
--	---	--	--	--	--

## B. ANALISIS KUANTITATIF

Analisis kuantitatif, merupakan analisis empiris, dilakukan setelah soal diujikan dikoreksi dan diberikan skor. Hasil analisis kuantitatif berupa angka-angka numerik, yang dapat digunakan untuk mengetahui daya serap (mastery learning) siswa secara individu dan kelas, mengetahui tingkat kesukaran butir, daya beda, validitas, dan reabilitas.<sup>2</sup>

### 1. Tingkat Kesukaran ( Difficulty Indeks)

Dasar pertimbangan untuk menentukan proporsi jumlah soal yang termasuk mudah, sedang, dan sukar adalah berdasarkan pada kurva normal, artinya sebagian besar soal berada dalam kategori sedang, sebagian lagi termasuk ke dalam kategori mudah dan sukar, dapat dibuat  $15\% = 70\% = 15\%$ , artinya 15% soal mudah, 70% soal sedang dan 15% soal sukar, misalnya : jumlah seluruh soal 60, maka yang mudah 10 soal, yang sedang 40 soal dan yang sukar 10 soal<sup>3</sup>.

Tingkat kesukaran juga merupakan salah satu analisis kuantitatif konvensional paling sederhana dan mudah. Hasil hitungannya merupakan proporsi atau perbandingan antara siswa yang menjawab benar dengan keseluruhan siswa yang mengikuti tes. Indeks kesukaran rentangannya dari 0,0 – 1,0. Semakin besar indeks menunjukkan semakin mudah butir soal, karena dapat dijawab dengan benar oleh sebagian besar atau seluruh siswa. Sebaliknya, jika sebagian kecil atau tidak ada sama sekali siswa yang menjawab benar menunjukkan butir sukar. Indeks 0,0 menunjukkan butir sangat sukar, sedangkan indeks 1,0 menunjukkan butir sangat mudah. Perhitungan analisa tingkat kesukaran soal yaitu dengan melakukan judgement oleh guru, kemudian soal diuji cobakan dan dianalisa, apakah judgement guru sesuai atau tidak, misalnya : soal no. 1 di judgement ke dalam kategori mudah, soal no.2 sedang dan soal no.3 sukar. Setelah dilakukan uji coba, kemudian dianalisa apakah nomor-nomor soal tadi sesuai dengan judgement atau tidak. Cara melakukan analisa untuk menentukan tingkat kesukaran soal tersebut dengan menggunakan rumus dibawah:

Rumus:  $P = B/N$

<sup>2</sup> Ahmad Sofyan, *Evaluasi Pembelajaran IPA Berbasis Kompetensi*, (Jakarta : UIN Jakarta Press, 2006), h. 102

<sup>3</sup> Wahyudin Uyu, dkk, *Evaluasi Pembelajaran SD* (Bandung: UPI Press, 2006), h. 98

P = Proporsi (Indeks Kesukaran)

B = Jumlah siswa yang menjawab benar

N = Jumlah peserta tes

Tingkat kesukaran yang baik adalah : P = 0,5 atau 0,15 d'' Pd'' 0,75

Ketentuan : P = 0-0,25 → sukar

P = 0,26-0,75 → sedang;

P = 0,76 - 1— mudah.

Contoh :

Tes formatif IPA, 10 soal bentuk pilihan ganda, option 4, dengan proporsi 2 soal mudah, 6 soal

sedang dan 2 soal sukar, jumlah siswa = 20 orang.

No Soal	Kemampuan yang diukur	Judgement P soal	Jumlah siswa yang menjawab benar	Indeks kesukaran	Indeks kesukaran
1	Pengetahuan	Mudah	18	0,90	Mudah
2	Pengetahuan	Mudah	12	0,60	Sedang
3	Pemahaman	Sedang	10	0,50	Sedang
4	Aplikasi	Sedang	12	0,60	Sedang
5	Aplikasi	Sedang	9	0,45	Sedang
6	Pemahaman	Sednag	20	1,00	Mudah
7	Analisa	Sedang	6	0,30	Sukar
8	Pemahaman	Sedang	10	0,50	Sedang
9	Sintesa	Sukar	4	0,20	Sukar
10	Sintesa	Sukar	9	0,45	Sedang

Dalam mencari indeks kesukaran menggunakan rumus yang telah di tulis diatas:

$$P = B/N = 18/20$$

$$P = 0,90$$

Dari contoh diatas, diperoleh hasil, yaitu : soal nomor 1, 3, 4, 5, 8 dan 9, terdapat kesesuaian antara judgement dengan hasil analisa, soal nomor 2 yang di judgement mudah ternyata termasuk soal sedang, soal nomor 6 yang di judgement sedang ternyata termasuk

soal mudah, soal nomor 7 yang dijudgement sedang, ternyata termasuk sukar dan soal nomor 10 yang dijudgement sukar, ternyata termasuk soal sedang.

Atas dasar hasil diatas , soal yang harus diperbaiki adalah:

Soal nomor 2, diturunkan ke dalam kategori mudah

Soal nomor 6, dinaikan ke dalam kategori sedang

Soal nomor 7 diturunkan ke dalam kategori sedang

Soal nomor 10, dinaikan ke dalam kategori sukar.

## **2. Daya beda/ daya pembeda (Discriminating power)**

Analisa daya pembeda mengkaji butir-butir soal dengan tujuan untuk mengetahui kesanggupan soal dalam membedakan siswa yang tergolong mampu/tinggi prestasinya dengan siswa yang terolong kurang/rendah prestasinya<sup>4</sup>. artinya soal yang bersangkutan bila diujikan untuk anak yang prestasinya bagus maka hasilnya akan bagus pula sedangkan untuk anak yang prestasinya kurang maka hasilnya akan kurang. Dalam hal ini suatu butir soal dikatakan tidak memiliki pembeda atau daya beda jika soal yang diberikan pada dua kelompok siswa yang berbeda kemampuannya dan hasilnyaapun sama artinya soal itu tidak menunjukkan gambaran hasil kemampuan siswa yang sesungguhnya, maka dikatakan tidak memiliki daya pembeda.

Dapat disimpulkan bahwa soal yang baik adalah soal yang dapat membedakan peserta didik yang berprestasi baik dan peserta didik yang berprestasi kurang. Soal digunakan oleh seorang evaluator untuk menguji kelompok yang diuji. Soal akan berfungsi dengan baik jika dapat membedakan kemampuan orang-orang dalam kelompok tersebut.

Manfaat daya pembeda yaitu ;

- 1) Untuk meningkatkan mutu setiap butir soal melalui data empiriknya. Berdasarkan indeks daya pembeda, setiap butir soal dapat diketahui apakah butir soal itu baik, direvisi, atau ditolak.
- 2) Untuk mengetahui seberapa jauh setiap butir soal dapat mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan guru. Apabila suatu butir soal tidak dapat membedakan kedua

---

<sup>4</sup> Wahyudin uyu,dkk. Evaluasi pembelajaran SD.(Bandung:UPI Press,2006).hal 101

kemampuan siswa itu, maka butir soal itu dapat dicurigai "kemungkinannya" seperti berikut ini.

- Kunci jawaban butir soal itu tidak tepat.
- Butir soal itu memiliki 2 atau lebih kunci jawaban yang benar
- Kompetensi yang diukur tidak jelas
- Pengecoh tidak berfungsi
- Materi yang ditanyakan terlalu sulit, sehingga banyak siswa yang menebak
- Sebagian besar siswa yang memahami materi yang ditanyakan berpikir ada

Mengetahui daya pembeda sangat penting. Butir-butir item tes hasil belajar haruslah mampu memberikan hasil tes yang menunjukkan adanya perbedaan-perbedaan kemampuan yang terdapat dikalangan siswa. Daya pembeda ini dapat diketahui dengan melihat besar kecilnya angka indeks diskriminasi. Angka indeks diskriminasi adalah angka atau bilangan yang menunjukkan besar kecilnya daya pembeda yang dimiliki oleh sebutir item. Indeks diskriminasi dapat diberikan lambang D. D ini angkanya berkisar antara 0-1 tetapi memungkinkan juga betanda minus, jika angka indeksnya positif maka dapat dikatakan memiliki daya pembeda yang bagus bagi butir soal. Untuk angka positif berarti kelompok yang berprestasi baik lebih banyak menjawab benar dan yang kurang berprestasi lebih banyak menjawab salah. Jika nilainya nol maka dikatakan tidak memiliki daya pembeda<sup>5</sup>, hal ini dapat dikatakan soal tidak memiliki kemampuan untuk membedakan kemampuan siswa yang kurang dan mampu. patokan untuk Angka indeks diskriminasi dapat dilihat sebagai berikut;

Besarnya angka indeks diskriminasi item (D)	Klasifikasi	Interpretasi
Kurang dari 0,2	Poor	Butir item yang bersangkutan daya pembedanya lemah sekali (jelek) dianggap tidak memiliki daya pembeda yang baik.
0,2 – 0,4	satisfactory	Butir item yang bersangkutan telah memiliki daya pembeda yang cukup(sedang)
0,4 – 0,7	Good	Butir item yang bersangkutan

<sup>5</sup> Anas Sudjiono, Pengantar Evaluasi Pendidikan (Jakarta: PT Raja Grafindo Persada, 2011) hal 389.

		memiliki daya pembeda yang baik
0,7 – 1,0	Excellent	Butir item yang bersangkutan telah memiliki daya pembeda yang baik sekali
Bertanda negatif	-	Butir item yang bersangkutan daya pembedanya negatif (jelek sekali)

Menurut Ebel soal dapat diklasifikasikan dalam empat katagori yaitu: sangat baik, baik, cukup dan kurang/ jelek. Klasifikasi soal berdasarkan indeks daya beda soal selanjutnya ditampilkan dalam tabel 2.1.

Tabel 2.1 Klasifikasi Indeks Daya Beda Soal<sup>6</sup>

Indeks Diskriminasi	Kategori Soal
Kurang dari 0,19	Jelek – Soal tidak dipakai/dibuang
0,20 – 0,29	Kurang – soal diperbaiki
0,30 – 0,39	Baik – soal diterima tetapi perlu diperbaiki
Lebih dari 0,40	Sangat Baik – soal diterima

Untuk mengetahui daya pembeda soal bentuk pilihan ganda adalah dengan menggunakan rumus berikut ini<sup>7</sup>.

$$D = (Ba - Bb) / 0,5N$$

Keterangan ;

DP = daya pembeda soal, daya beda yang baik adalah  $D > 0,3$

BA = jumlah yang menjawab benar pada kelompok atas.

BB = jumlah yang menjawab benar pada kelompok bawah.

N = jumlah peserta tes

<sup>6</sup>Ebel Robert L, Essential of Educational Measurement Fifth Edition ( America: Prentice Hall Inc, 1991) hal 232.

<sup>7</sup> Ahmad Sofyan, Dkk, Evaluasi Pembelajaran IPA Berbasis Kompetensi (Jakarta: Lemlit UIN Jakarta, 2006) hal 104.

Akhirnya sebagai tindak lanjut atas hasil penganalisisan mengenai daya pembeda item tes hasil belajar tersebut adalah;

- a. Butir-butir yang sudah memiliki daya pembeda item yang baik hendaknya dicatat pada bank soal tes hasil belajar. dan dapat dikeluarkan lagi pada tes hasil belajar yang akan datang karena dinilai kualitasnya cukup memadai.
- b. Butir-butir soal yang masih rendah mendapat 2 kemungkinan perlakuan;
  1. Ditelusuri untuk diperbaiki ,di ajukan dalam tes lagi, dianalisis kembali
  2. Dibuang untuk tes yang akan datang
- c. Khusus yang memiliki hasil minus atau negatif sebaiknya tidak perlu dikeluarkan lagi<sup>8</sup>.

### 3. Fungsi pengecoh(distracter function)

Pada saat membicarakan tes objektif bentuk multiple choice *item* tersebut untuk setiap butir *item* yang dikeluarkan dalam tes hasil belajar telah dilengkapi dengan beberapa kemungkinan jawab, atau yang sering dikenal dengan istilah option atau alternatif. Option atau alternatif itu jumlahnya berkisar antara 3 sampai dengan 5 buah, dan dari kemungkinan-kemungkinan jawaban yang terpasang pada setiap butir *item* itu, salah satu diantaranya adalah merupakan jawaban betul (kunci jawaban), sedangkan sisanya adalah merupakan jawaban salah. Jawaban-jawaban salah itulah yang biasa dikenal dengan istilah *distractor* (pengecoh).<sup>9</sup>

Fungsi pengecoh dimaksudkan untuk mengetahui seberapa besar peserta yang tidak memiliki kunci jawaban ( option) pada bentuk soal pilihan ganda. Untuk soal pilihan ganda, alternatif jawaban menurut kaidah harus homogen dan logis sehingga setiap pilihan jawaban (option) dapat berfungsi atau ada yang memilih. Setiap pengecoh dapat dikatakan berfungsi apabila ada yang memilih. Setiap pengecoh dapat dikatakan berfungsi apabila terpilih minimal sebanyak 5% dari jumlah peserta.untuk menghitungnya dapat digunakan rumus sebagai berikut<sup>10</sup>

$$(\text{Jumlah siswa yang memilih option yang salah}) / (\text{jumlah seluruh peserta tes}) \times 100\%$$

---

<sup>8</sup> Anas Sudjiono, Pengantar Evaluasi Pendidikan (Jakarta: PT Raja Grafindo Persada, 2011) hal 408.

<sup>9</sup> Anas Sudjiono, Op. cit. hal 409.

<sup>10</sup> Ahmad Sofyan, Dkk, Lok. cit. hal 104.

Menganalisis fungsi distraktor sering dikenal dengan istilah lain, yaitu : menganalisis pola penyebaran jawaban *item*. Adapun yang dimaksud dengan pola penyebaran jawaban *item* adalah suatu pola yang dapat menggambarkan bagaimana *testee* menentukan pilihan jawabnya terhadap kemungkinan-kemungkinan jawab yang telah dipasangkan pada setiap butir *item*.

Suatu kemungkinan dapat terjadi, yaitu bahwa dari keseluruhan alternatif yang dipasang pada butir *item* tertentu, sama sekali tidak dipilih oleh *testee*. Dengan kata lain, *testee* menyatakan “blangko”. Pernyataan blangko ini sering dikenal dengan istilah omiet dan biasa diberi lambang dengan huruf O.

Sebagai tindak lanjut atas hasil penganalisaan terhadap fungsi distraktor tersebut maka distraktor yang sudah dapat menjalankan fungsinya dengan baik dapat dipakai lagi pada tes-tes yang akan datang, sedangkan distraktor yang belum dapat berfungsi dengan baik sebaiknya diperbaiki atau diganti dengan distraktor yang lain.

#### **4. VALIDITAS**

Validitas merupakan syarat yang penting dalam suatu alat evaluasi. Validitas berasal dari kata *validity*, dapat diartikan tepat atau shahih, yakni sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya.<sup>11</sup> Beberapa kriteria dipilih untuk memperlihatkan keefektifan terhadap peramalan performance yang akan datang (yang akan terjadi), kriteria yang lain untuk menunjukkan status yang muncul, kriteria yang lain lagi untuk menimbulkan sifat-sifat yang representatif dari luasnya isi atau tingkah laku, dan kriteria yang lain lagi untuk (melengkapi) penyediaan data atau untuk menunjang atau menolak beberapa teori psikologis. Sebagaimana dikemukakan oleh Scarvia B. Anderson dalam bukunya “Encyclopedia of Educational Evaluation” disebutkan bahwa “ A test is valid it measure what it purpose to measure” (sebuah tes dikatakan valid apabila tes tersebut mengukur apa yang hendak diukur).

Validitas suatu instrumen evaluasi, tidak lain adalah derajat yang menunjukkan di mana suatu tes mengukur apa yang hendak diukur. Validitas suatu instrumen evaluasi mempunyai beberapa makna penting di antaranya seperti berikut.

---

<sup>11</sup> Ahmad Sofyan, Dkk, Evaluasi Pembelajaran IPA Berbasis Kompetensi (Jakarta: Lemlit UIN Jakarta, 2006) hal 105.

1. Validitas berhubungan dengan ketepatan interpretasi hasil tes atau instrumen evaluasi untuk grup individual dan bukan instrumen itu sendiri.
2. Validitas diartikan sebagai derajat yang menunjukkan kategori yang bisa mencakup kategori rendah, menengah, dan tinggi.
3. Prinsip suatu tes valid, tidak universal. Validitas suatu tes yang perlu diperhatikan oleh para peneliti adalah bahwa ia hanya valid untuk suatu tujuan tertentu saja. Tes valid untuk bidang studi metrologi industri belum tentu valid untuk bidang yang lain misalnya bidang mekanika teknik.<sup>12</sup>

Validitas suatu alat evaluasi, bukanlah merupakan ciri yang absolut atau mutlak. Suatu tes dapat memiliki validitas yang tinggi, sedang, rendah, tergantung kepada tujuannya. 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*).

#### **a) Validitas Isi**

Validitas isi artinya ketepatan daripada suatu tes dilihat dari segi isi tersebut. Suatu tes hasil belajar dikatakan valid, apabila materi tes tersebut benar-benar merupakan bahan-bahan yang representatif terhadap bahan-bahan pelajaran yang diberikan. Untuk mendapatkan validitas isi memerlukan dua aspek penting, yaitu valid isi dan valid teknik sampling. Valid isi mencakup khususnya, hal-hal yang berkaitan dengan apakah item-item evaluasi menggambarkan pengukuran dalam cakupan yang ingin diukur. Sedangkan valid teknik sampling pada umumnya berkaitan dengan bagaimanakah baiknya suatu sample item tes mempresentasikan total cakupan isi.

Oleh karena materi yang diajarkan tertera dalam kurikulum maka validitas ini sering disebut validitas kurikuler. Validitas isi dapat diusahakan tercapainya sejak penyusunan dengan memerinci materi kurikulum atau materi buku pelajaran. Misalnya untuk siswa kelas I SMU akan diberikan tes Matematika, maka item-itemnya harus diambil dari materi pelajaran kelas I, apabila kita sisipkan dan item-item yang diambil dari materi pelajaran kelas III maka tes tersebut sudah tidak valid lagi.

Validitas isi pada umumnya ditentukan melalui pertimbangan para ahli. Tidak ada formula matematis untuk menghitung dan tidak ada cara untuk menunjukkan secara pasti. Akan tetapi, untuk memberikan gambaran bagaimana suatu tes divalidasi dengan

---

<sup>12</sup> Sukardi, Evaluasi Pendidikan Prinsip dan Operasionalnya (Jakarta: Bumi Aksara, 2011) hal 31.

menggunakan validitas isi, pertimbangan ahli tersebut dilakukan dengan cara seperti berikut. Pertama, para ahli diminta untuk mengamati secara cermat semua item dalam tes yang hendak divalidasi. Kemudian mereka diminta untuk mengoreksi interpretasi item-item yang telah dibuat. Pada akhir perbaikan, mereka juga diminta untuk memberikan pertimbangan-pertimbangan tentang bagaimana baik interpretasi tes evaluasi tersebut menggambarkan cakupan isi yang hendak diukur. Pertimbangan ahli tersebut biasanya juga menyangkut, apakah semua aspek yang hendak diukur telah dicakup melalui interpretasi item pertanyaan dalam tes. Atau dengan kata lain perbandingan dibuat antara apa yang harus dimasukkan dengan apa yang ingin diukur yang telah direfleksikan menjadi tujuan tes.<sup>13</sup>

#### **b) Validitas Konstruk**

Validasi konstruk merupakan derajat yang menunjukkan suatu tes mengukur sebuah konstruk sementara. Untuk menentukan adanya validitas konstruk suatu tes dikorelasikan dengan suatu konsepsi atau teori, item-item dalam tes itu harus sesuai dengan ciri-ciri yang disebutkan dalam konsepsi tadi, yaitu konsepsi tentang obyek yang akan dites.<sup>14</sup>

Untuk mengetahui apakah suatu tes memenuhi syarat-syarat validitas konstruksi atau tidak maka kita harus membandingkan susunan tersebut telah memenuhi syarat-syarat penyusunan tes maka berarti tes tersebut memenuhi syarat validitas konstruksi, apabila tidak memenuhi syarat-syarat penyusunan tes berarti tidak memenuhi validitas konstruksi.

Proses melakukan validasi konstruk dapat dilakukan dengan cara melibatkan hipotesis testing yang dideduksi dari teori yang menyangkut dengan konstruk yang relevan. Misalnya jika suatu teori kecemasan menyatakan bahwa seseorang yang memiliki kecemasan yang lebih tinggi akan bekerja lebih lama dalam menyelesaikan suatu problem, dibanding dengan orang yang memiliki tingkat kecemasan rendah. Jika terjadi orang yang cemasnya tinggi ternyata kemudian bekerja sebaliknya yaitu lebih cepat, ini bukan berarti bahwa tes yang sudah baku tadi berarti tidak mengukur kecemasan orang. Atau dengan kata lain hipotesis yang berhubungan dengan tingkah laku seseorang dengan kecemasan tinggi tidak benar. Dari kasus tersebut mengindikasikan bahwa konstruksi yang berhubungan dengan orang yang memiliki kecemasan tinggi memerlukan kajian ulang, guna mengadakan koreksi dan penyesuaian kembali.

#### **c) Validitas Konkuren**

---

<sup>13</sup> Sukardi, Op.cit.hal33.

<sup>14</sup> Ngalim Purwanto, Prinsip-Prinsip dan Teknik Evaluasi Pengajaran (Bandung: Remaja Karya, 1986) hal178.

Jika hasil suatu tes mempunyai korelasi yang tinggi dengan hasil dari suatu alat pengukur lain terhadap bidang yang sama pada waktu yang sama pula, maka tes itu dikatakan memiliki konkuren validity.

Validitas ini lebih umum dikenal dengan validitas empiris. Sebuah tes dikatakan memiliki validitas empiris jika hasilnya sesuai dengan pengalaman. Jika istilah “sesuai” tentu ada dua hal yang dipasangkan. dalam hal ini hasil tes dipasangkan dengan hasil pengalaman. Pengalaman selalu mengenai hal yang telah lampau sehingga data pengalaman tersebut sekarang sudah ada.

Cara-cara membuat tes dengan validitas konkruen dapat dilakukan dengan beberapa langkah seperti berikut.

- a. Administrasi tes yang baru yang dilakukan terhadap grup atau anggota kelompok
- b. Catat tes baku yang ada termasuk berapa koefisien validitasnya jika ada
- c. Hubungkan atau korelasikan dua tes skor tersebut

Hasil yang dicapai atau koefisien validitas yang muncul menunjukkan derajat hubungan validitas tes yang baru. jika koefisien tinggi, berarti tes yang baru tersebut mempunyai validitas konkruen yang baik. Sebaliknya, tes yang baru dikatakan mempunyai validitas konkruen yang jelek, apabila koefisien yang dihasilkan rendah.

Tes mental merupakan contoh nyata terapan suatu tes pembeda (validitas konkruen yang melibatkan penentuan tes ) yang sering ditemui dalam kasus psikologi. Jika hasil skor suatu tes dapat digunakan dengan cara benar untu mengklarifikasi orang yang satu dengan orang lainnya, maka validitas konkruen tes tersebut memiliki daya pembeda yang baik.

#### **d) Validitas Prediksi**

Memprediksi artinya meramal, dan meramal selalui mengenai hal yang akan datang jadi sekarang belum terjadi. Sebuah tes dikatakan memiliki validitas prediksi apabila mempunyai kemampuan untuk meramalkan apa yang akan terjadi pada masa yang akan mendatang. Jenis validitas ini menunjukkan kenyataan jika ujian yang dimaksud dihubungkan dengan kriteria-kriteria tentang hasil karya atau kesuksesan dimasa depan. Demikianlah jika suatu tes bakat skolastik diberikan pada siswa-siswa SMU dikorelasikan dengan prestasi mereka di perguruan tinggi, maka kenyataan yang diperoleh itu akan menunjukkan validitas prediksi.

Instrumen validitas prediksi mungkin bervariasi bentuknya tergantung beberapa faktor misalnya kurikulum yang digunakan, buku pegangan yang dipakai, intensitas mengajar dan letak geografis atau daerah sekolah. Yang perlu diperhatikan saat melakukan tes validasi ini

yaitu perlu memperhatikan proses dan cara membandingkan instrumen yang divalidasi dengan tes yang dibakukan. Perlu disadari bahwa skor tes yang dihasilkan juga memiliki sifat ketidak sempurnaan.

Ketika kriteria telah diidentifikasi dan ditentukan, prosedur selanjutnya adalah menentukan validitas prediksi suatu tes dengan cara seperti berikut.

- a. Buat item tes sesuai dengan tujuan yang hendak dicapai
- b. Tentukan kelompok yang dijadikan subyek dalam *pilot study*
- c. Identifikasi kriteria prediksi yang hendak dicapai
- d. Tunggu sampai tingkah laku yang diprediksi atau variabel kriteria muncul dan terpenuhi dalam kelompok yang telah ditentukan
- e. Capai ukuran-ukuran kriteria tertentu
- f. Korelasikan dua set skor yang dihasilkan<sup>15</sup>

Sebagai contoh, kita akan menyelenggarakan tes untuk menentukan validitas prediksi tes pada mahasiswa yang mengikuti mata kuliah matematika teknik. Langkah pertama yang harus dilakukan adalah membuat tes item, kemudian memberikannya kepada kelompok mahasiswa potensi yang mengambil mata kuliah tersebut. Kemudian kita menunggu selama satu semester penuh pada kelompok mahasiswa yang hendak diprediksi pada mata kuliah yang sama dengan mengukur melalui nilai ujian akhir. Hasil korelasi antara dua set nilai akan menentukan validitas prediksi tes. Jika hasilnya menunjukkan koefisien korelasi tinggi, berarti tes tersebut mempunyai validitas prediksi tinggi.

## 5. Pengujian Validitas Instrumen

### a) Rumus “Product Moment Correlation”

Menurut Husein Umar (1998: 195) untuk menguji tingkat validitas instrumen dalam penelitian digunakan teknik analisis Koefisien Korelasi Produk-Moment Pearson (*Pearson Product-Moment Correlation Coefficient*) dengan rumus sebagai berikut:

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

Dimana:

$r_{xy}$  : Koefisien korelasi Pearson antara item instrumen yang akan digunakan dengan variabel yang bersangkutan

---

<sup>15</sup> Sukardi, Op.cit.hal36.

- X : Skor item instrumen yang akan digunakan
- Y : Skor semua item instrumen dalam variabel tersebut
- n : Jumlah responden

Untuk menguji keberartian koefisien  $r_{xy}$  valid atau tidak valid akan digunakan uji t, yang dilakukan dengan membandingkan antara  $t_{hitung}$  dengan  $t_{tabel}$ . Dimana  $t_{hitung}$  dicari dengan menggunakan rumus dari Husein Umar (1998: 197) sebagai berikut:

$$t_{hitung} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Dimana r adalah koefisien korelasi Pearson dan db adalah derajat bebas.

Keputusan pengujian validitas instrumen dengan menggunakan taraf signifikansi 5% adalah sebagai berikut:

- a. Item instrumen dikatakan valid jika  $t_{hitung}$  lebih besar atau sama dengan  $t_{0,05}$ ; maka item instrumen tersebut dapat digunakan.
- b. Item instrumen dikatakan tidak valid jika  $t_{hitung}$  lebih kecil dari  $t_{0,05}$ ; maka item instrumen tersebut tidak dapat digunakan.

#### Interpretasi Koefisien Korelasi Nilai r

Interval Koefisien	Tingkat Hubungan
0,80 – 1,000	Sangat Kuat
0,60 – 0,799	Kuat
0,40 – 0,599	Cukup Kuat
0,20 – 0,399	Rendah
0,00 – 0,199	Sangat Rendah

#### **Contoh : "Hubungan Motivasi dengan Kinerja Dosen STAI Daruttaqwa Gresik"**

Motivasi (X) : 60; 70; 75; 65; 70; 60; 80; 75; 85; 90; 70; dan 85

Kinerja (Y) : 450; 475; 450; 470; 475; 455; 475; 470; 485; 480; 475; dan 480.

No	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
1.	60	450	3600	202500	27000
2.	70	475	4900	225625	33250
3.	75	450	5625	202500	33750
4.	65	470	4225	220900	30550
5.	70	475	4900	225625	33250
6.	60	455	3600	207025	27300
7.	80	475	6400	225625	38000
8.	75	470	5625	220900	35250
9.	85	485	7225	235225	41225
10.	90	480	8100	230400	43200
11.	70	475	4900	225625	33250
12.	85	480	7225	230400	40800
Statistik	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
Jumlah	885	5640	66325	2652350	416825

Mencari  $r_{hitung}$  dengan cara masukkan angka statistik dari tabel penolong dengan rumus ;

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

$$r_{xy} = \frac{12(416.825) - (885).(5.460)}{\sqrt{\{12.(66.325) - (885)^2\} . \{12.(2.652.350) - (5.640)^2\}}}$$

$$r_{xy} = \frac{169.900}{\sqrt{133.463.835.000}} = \frac{169.00}{365.327,02} = 0,465$$

Dengan melihat hasil koefisien korelasi ( $r$ ) = 0,465 ini berarti bahwa antara kedua kelompok tersebut terdapat korelasi yang cukup kuat.

**b) Rumus “Rank Method of Correlation” (Metode Spearman)**

Berikut ini adalah data hubungan antara Nilai Ujian Tengah Semester (UTS) (X) dengan nilai Ujian Akhir Semester (UAS) (Y) dari 12 responden :

X	Y
12	19
13	19
14	11
15	12
16	17

16	15
14	16
20	18
17	14
11	13
12	18
18	13

Sumber : Data fiktif

Buktikanlah hipotesis yang berbunyi “terdapat hubungan yang positif antara Nilai UTS dan UAS” dengan menggunakan data tersebut!

1. Menentukan taraf signifikansi

Nilai  $\alpha = 0,05$

2. Menghitung hitung  $\rho$  :

Untuk menghitung nilai  $\rho$  , maka harus disusun tabel sebagai berikut :

X	Y	RX	RY	RX-RY	(RX-RY) <sup>2</sup>

12	19	2,5	11,5	-9	81
13	19	4	11,5	-7,5	56,25
14	11	5,5	1	4,5	20,25
15	12	7	2	5	25
16	17	8,5	8	0,5	0,25
16	15	8,5	6	2,5	6,25
14	16	5,5	7	-1,5	2,25
20	18	12	9,5	2,5	6,25
17	14	10	5	5	25
11	13	1	3,5	-2,5	6,25
12	18	2,5	9,5	-7	49
18	13	11	3,5	7,5	56,25
					$\Sigma b = 334$

$$\rho = 1 - \frac{(6\sum b)}{(n(n^2-1))} \text{----- Rumus 1.1}^{16}$$

$$\rho = 1 - \frac{(6.334)}{(12(12^2-1))}$$

$$\rho = 1 - 1,168$$

$$\rho = -0,168$$

Jadi,

Nilai hitung  $\rho = -0,168$  (tanda “-“ hanya menunjukkan arah hubungan dua variabel negatif).

Nilai tabel ( $n=12$ ) = 0,591 (=0,05)

### 3. Kesimpulan

Untuk itu, karena nilai hitung  $\rho <$  tabel, maka  $H_0$  diterima. Kesimpulannya adalah hipotesis yang berbunyi : “tidak terdapat hubungan antara nilai UTS dan UAS” diterima.

## 6. RELIABILITAS

Reliabilitas adalah tingkat atau derajat konsistensi dari suatu instrumen, reliabilitas tes berkenaan dengan dengan pertanyaan, apakah suatu tes teliti dan dapat dipercaya sesuai dengan kriteria yang telah ditetapkan. Suatu tes dapat dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu yang sama pada waktu atau kesempatan yang berbeda.

Menurut Gronlund, ada empat faktor yang dapat mempengaruhi reliabilitas, yaitu :

---

<sup>16</sup> Ngalim Purwanto, Op.cit.hal179.

1. Panjang tes, yaitu banyaknya soal tes. Ada kecenderungan, semakin panjang suatu tes akan lebih tinggi tingkat reliabilitas suatu tes, karena semakin banyak soal, maka akan semakin banyak sampel yang diukur dan proporsi jawaban yang benar semakin semakin banyak, sehingga faktor tebakan akan semakin rendah.
2. Sebaran skor, besarnya sebaran skor akan membuat tingkat reliabilitas menjadi lebih tinggi, Karena koefisien reliabilitas yang lebih besar diperoleh ketika peserta didik tetap pada posisi yang relative sama dalam satu kelompok pengujian ke pengujian berikutnya. Dengan kata lain, peluang selisih dari perubahan posisi dalam kelompok dapat memperbesar koefisien reliabilitas.
3. Tingkat kesukaran, dalam penilaian yang menggunakan pendekatan penilaian acuan norma, baik untuk soal yang mudah maupun sukar, cenderung menghasilkan tingkat reliabilitas yang rendah. Hal ini disebabkan antara hasil tes yang mudah dengan hasil tes yang sukar keduanya dalam satu sebaran skor yang terbatas. Untuk tes yang mudah, skor akan berada dibagian atas dan akhir dari skala penilaian. Bagi kedua tes (mudah dan sukar), perbedaan antarpeserta didik kecil sekali dan cenderung tidak dapat dipercaya. Tingkat kesukaran soal yang ideal untuk meningkatkan koefisien reliabilitas adalah soal yang menghasilkan sebaran skor berbentuk genta atau kurva normal.
4. Objektivitas, menunjukkan skor tes kemampuan yang sama antara peserta didik yang satu dengan peserta didik lainnya. Peserta didik memperoleh hasil yang sama dalam mengerjakan suatu tes. Jika peserta didik memiliki tingkat kemampuan yang sama, maka akan memperoleh hasil tes yang sama pada saat mengerjakan tes yang sama. Objektivitas prosedur tes yang tinggi akan memperoleh reliabilitas hasil tes yang tidak dipengaruhi oleh prosedur penskoran.

Konsep reliabilitas mendasari kesalahan pengukuran yang mungkin terjadi pada suatu proses pengukuran atau pada nilai tunggal tertentu, sehingga menimbulkan perubahan pada susunan kelompoknya. Misalnya, guru mengetes peserta didik dengan instrumen tertentu dan mendapat nilai 70. Kemudian pada kesempatan yang berbeda dengan instrumen yang sama, guru melakukan tes kembali, ternyata peserta didik tersebut mendapat nilai 75. Artinya, tes tersebut tidak reliabel, karena terjadi kesalahan pengukuran. Tes yang reliabel adalah apabila koefisien reliabilitasnya tinggi dan kesalahan baku pengukurannya rendah.

Menurut perhitungan *product-moment* dari Person, ada tiga macam reliabilitas, yaitu koefisien stabilitas, koefisien ekuivalen dan koefisien konsistensi internal.

### **a) Koefisien Stabilitas**

Jenis reliabilitas yang menggunakan teknik *test and retest*, yaitu memberikan tes kepada sekelompok individu, kemudian diadakan pengulangan tes pada kelompok yang sama dengan waktu yang berbeda. Cara memperoleh koefisien stabilitas adalah dengan mengorelasikan hasil tes pertama dengan hasil tes kedua dari kelompok yang sama, tes yang sama, pada waktu yang berbeda. Jika antara waktu tes pertama dengan tes yang kedua cukup lama, kemudian diadakan latihan-latihan tambahan, maka bisa jadi nilai tes yang kedua akan lebih besar daripada tes yang pertama. Sebaliknya, jika antara waktu tes pertama dengan tes kedua relatif pendek, maka nilai tes kedua bisa jadi sama atau lebih besar daripada tes pertama karena soal dan jawaban masih dapat diingat.

Kesalahan teknis ini dapat bersumber dari berbagai faktor, sehingga menyebabkan peserta didik mempunyai skor yang berbeda pada saat dua kali mengerjakan tes yang sama. Bisa saja perubahan skor yang terjadi bukan disebabkan perubahan hal yang diukur, tetapi memang karena situasi yang berbeda atau pengalaman dari peserta didik pada saat mengikuti tes yang pertama, sehingga ketika mengerjakan tes yang kedua, peserta didik lebih berhati-hati dan lebih baik hasilnya. Keunggulan teknik ini adalah dapat memperkecil kemungkinan masuknya sumber kesalahan yang lain. Namun, patut juga dipertimbangkan bahwa penggunaan kelompok yang sama dan tes yang sama dalam dua kali tes akan mempengaruhi hasil tes yang kedua, karena responden sudah memiliki pengalaman mengerjakan tes yang pertama. Hal ini sekaligus menunjukkan kelemahan teknik *test and retest*.

### **b) Koefisien ekuivalen**

Jika mengorelasikan dua buah tes yang parallel pada kelompok dan waktu yang sama. Metode yang digunakan untuk memperoleh koefisien ekuivalen adalah metode dengan menggunakan dua buah bentuk tes *parallel or alternate-forms method*. Syarat-syarat yang harus dipenuhi kedua tes parallel adalah criteria yang dipakai pada kedua tes sama., masing-masing tes dikonstruksikan tersendiri, jumlah item, isi, dan corak sama, tingkat kesukaran sama, petunjuk waktu yang disediakan untuk mengerjakan tes dan contoh-contoh juga sama. Kemungkinan kesalahan pada teknik ini bersumber dari derajat keseimbangan antara dua tes

tersebut, serta kondisi tempat yang mungkin berbeda pada kelompok tes pertama dengan kelompok tes kedua, meskipun dilakukan pada waktu yang sama.

### c) Koefisien konsistensi internal

Reliabilitas yang didapat dengan jalan mengorelasikan dua buah tes dari kelompok yang sama, tetapi diambil dari butir-butir yang bernomor genap untuk tes yang pertama dan butir-butir bernomor ganjil untuk tes yang kedua. Teknik ini sering disebut *split-half method*. Split berarti membelah dan half berarti setengah atau separuh. Jadi, split-half adalah tes yang dibagi menjadi dua bagian yang sama, kemudian mengorelasikan butir soal yang bernomor ganjil dalam belahan pertama (X) dan yang bernomor genap dalam belahan kedua (Y). Untuk membagi tes menjadi dua bagian dapat juga dilakukan dengan jalan mengambil nomor soal secara acak, tetapi jumlahnya tetap harus sama untuk masing-masing kelompok. Disamping itu, pembagian tes dapat juga dilakukan dengan cara setengah bagian pertama untuk kelompok pertama dan setengah lagi untuk kelompok kedua.

Untuk menghitung koefisien stabilitas, koefisien ekuivalen dan koefisien konsistensi internal dapat digunakan analisis korelasi seperti pada pengujian validitas. Khusus bagi perhitungan koefisien konsistensi internal, korelasi tersebut baru sebagian dari seluruh tes. Untuk memperoleh angka koefisien korelasi secara menyeluruh dari tes tersebut harus dihitung dari nomor-nomor kedua tes itu dengan rumus Spearman Brown.

$$r_{nn} = \frac{2r_{1.2}}{1+(n-1)r_{1.2}}$$

Ket : r : korelasi

n : panjang tes yang selalu sama dengan 2 karena seluruh tes = 2 x ½

Contoh :

10 orang peserta didik dites dalam mata pelajaran IPA dan IPS, jumlah soal masing-masing lima buah. Dua buah nomor genap diambil dari hasil tes IPA dan tiga buah nomor ganjil diambil dari hasil tes IPS. Data diperoleh sebagai berikut.

### **Nilai 10 Orang Peserta Didik Dalam Mata Pelajaran IPA dan IPS**

Nama	Skor IPA		Skor IPS		
	No.Genap (2 dan 4)		No.Ganjil (1,3 dan 5)		
A	8	6	8	7	10
B	7	7	6	7	5
C	5	6	6	6	6
D	8	6	7	6	9
E	5	6	5	5	5
F	4	7	4	6	6
G	5	9	7	5	5
H	7	5	8	5	4
I	7	8	4	9	7
J	9	5	9	9	4

**Perhitungan Koefesien Konsistensi Internal**

X	Y	$x_i - \bar{x}$	$y_i - \bar{y}$	$x^2$	$y^2$	xy
14	25	1	6	1	36	6
14	18	1	-1	1	1	-1
11	18	-2	-1	4	1	2

14	22	1	3	1	9	3
11	15	-2	-4	4	16	8
11	16	-2	-3	4	9	6
14	17	1	-2	1	4	-2
12	17	-1	-2	1	4	2
15	20	2	2	4	1	2
14	22	1	3	1	9	3
$\sum X = 130$	$\sum Y = 190$			$\sum x^2 = 22$	$\sum y^2 = 90$	$\sum xy = 29$
$\bar{x} = 13$	$\bar{y} = 19$					

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{29}{\sqrt{(22)(90)}} = \frac{29}{\sqrt{1980}} = \frac{29}{44.50} = 0.65$$

Untuk menghitung seluruh tes itu, dapat digunakan rumus Spearman brown sebagai berikut :

$$r_{nn} = \frac{2r_{1.2}}{1+(n-1)r_{1.2}} = \frac{(2)(0.65)}{1+(2-1)(0.65)} = \frac{1.30}{1.65} = 0,787$$

### BAB III

#### PENUTUP

##### A. Kesimpulan

Hubungan antara validitas dengan realibilitas adalah jika suatu tes dinyatakan mempunyai instrumen valid maka akan mempunyai reabilitas yang baik juga, sedangkan jika suatu instrumen yang reliable belum tentu valid. Scarvia B. Anderson menyatakan bahwa persyaratan bagi tes yaitu validitas dan reabilitas, dalam hal ini validitas lebih penting dan reabilitas ini perlu karena menyokong terbentuknya validitas. Sebuah tes mungkin reliabel tetapi tidak valid. Sedangkan sebuah tes yang valid biasanya reliabel. Tuntutan bahwa instrumen evaluasi harus valid menyangkut harapan diperolehnya data yang valid, sesuai dengan kenyataan. Dalam hal reabilitas ini tuntutannya tidak jauh berbeda. Jika validitas terkait dengan ketepatan objek yang tidak lain adalah tidak menyimpangnya data dari kenyataan, artinya bahwa data tersebut benar, maka konsepreabilitasnya terkait dengan

pemotretan berkali-kali. Instrument yang baik adalah instrument yang dapat dengan ajeg memberikan data yang sesuai dengan kenyataan.

## **B. Saran**

Ketika kita menjadi pengajar dan pendidik, sebaiknya dalam penyusunan instrument tes, seperti soal tes hendaknya disesuaikan dengan criteria penyusunan soal yang baik dan benar. Sehingga ketika diuji dengan validitas maupun realibilitas sesuai dengan kualitas dan metode pembelajaran yang menjunjung tinggi cita-cita guru Indonesia untuk mencerdaskan kehidupan bangsa.

## **DAFTAR PUSTAKA**

Arifin, Zaenal. *Evaluasi Pembelajaran*. Bandung : Remaja Rosdakarya . 2010

Ebel,Robert L. *Essential of Educational Measurement Fifth Edition*. America : Prentice Hall Inc.1991

Purwanto, Ngalim. *Prinsip-Prinsip Dan Teknik Evaluasi Pengajaran*. Bandung : Remadja Karya.1986

Sofyan, Ahmad,dkk. *Evaluasi Pembelajaran IPA Berbasis Kompetensi*. Jakarta : Lemlit UIN Jakarta.2006

Sudjiono, Anas. *Pengantar Evaluasi Pendidikkan*. Jakarta: PT Raja Grafindo Persada. 2011

Sukardi. *Evaluasi Pendidikan Prinsip dan Operasionalnya*. Jakarta:Bumi Aksara. 2011

Wahyudin uyu,dkk. *Evaluasi pembelajaran SD*. Bandung : UPI Press. 2006

[http://file.upi.edu/Direktori/FIP/JUR.\\_ADMINISTRASI\\_PENDIDIKAN/196807291998021\\_SURYADI/VALIDITAS\\_tes.pdf](http://file.upi.edu/Direktori/FIP/JUR._ADMINISTRASI_PENDIDIKAN/196807291998021_SURYADI/VALIDITAS_tes.pdf) diakses 5 Desember 2012