

# **PENILAIAN HASIL BELAJAR**

**Sudji Munadi**

## **BAGIAN 1**

### **A. Pendahuluan**

Teknologi merupakan penerapan ilmu dasar sehingga tujuan pembelajaran juga menekankan peningkatan kemampuan membangun dan menerapkan pengetahuan, informasi secara logis, kritis, kreatif, dan inovatif. Sebagai ilmu terapan, teknologi mengkaji berbagai persoalan yang berkait dengan perancangan/rekayasa untuk menemukan produk baru yang dapat memenuhi kebutuhan manusia dalam segala aspek kehidupan, baik yang berkait dengan aspek ideologi, politik, sosial budaya, maupun pertahanan dan keamanan. Dalam perkembangannya produk teknologi bukan hanya berupa produk kebendaan, tetapi juga pengembangan suatu sistem yang mendukung layanan/jasa.

Tujuan pembelajaran teknologi lebih banyak pada kegiatan yang bersifat praktik dengan mengintegrasikan pengetahuan, keterampilan dan sikap sebagai proses pembentukan kompetensi. Dengan demikian, kompetensi dalam pembelajaran ini adalah integrasi dari pengetahuan, keterampilan, dan sikap yang diperlukan oleh seseorang untuk melaksanakan suatu tugas di dunia kerja. Pada level pendidikan dasar dan menengah kajian teknologi lebih berfokus pada aspek keterampilan untuk melakukan tindakan yang berbasis teknologi, yang meliputi keterampilan gerak/psikomotor dalam ragam teknologi, bisnis, dan seni.

Peserta didik dinyatakan berkompeten dalam pekerjaan tertentu manakala ia memiliki pengetahuan, keterampilan dan sikap kerja minimum yang digunakan untuk menyelesaikan pekerjaan tersebut dalam bentuk unjuk kerja/kinerja. Unjuk kerja adalah tingkah laku yang membuahkan suatu hasil, khususnya tingkah laku yang dapat mengubah lingkungan dengan cara-cara tertentu. Dalam pembelajaran, unjuk kerja merupakan penampilan peserta didik dalam mengerjakan sesuatu tugas yang terkait dengan pembelajaran yang dilakukan.

Hasil belajar peserta didik dilihat dari 3 (tiga) aspek, yaitu:

1. Hasil belajar yang berkaitan dengan perkembangan kognitif.
2. Hasil belajar yang berkaitan dengan perkembangan afektif,
3. Hasil belajar yang berkaitan dengan perkembangan keterampilan (psikomotor).

3H, Head, Hand, Heart

## **B. Batasan-Batasan**

### **1. Pengukuran**

Pengukuran diartikan sebagai pemberian angka kepada suatu atribut atau karakteristik tertentu yang dimiliki oleh orang, hal, atau objek tertentu menurut aturan atau formulasi yang jelas. Dalam pengukuran terdapat dua karakteristik utama, yaitu penggunaan angka atau skala tertentu dan menurut aturan atau formula tertentu. Skala atau angka dalam pengukuran dapat diklasifikasikan kedalam 4 (empat) kategori, yaitu: skala nominal, skala ordinal, skala interval, dan skala rasio.

<b>Skala</b>	<b>Contoh nilai ujian</b>
Rasio	0, 1,2, 3, .....100
Interval	0, 1,2, 3, .....100
Ordinal	A = sangat memuaskan B = memuaskan C = kurang memuaskan
Nominal	Lulus, Tidak lulus

### **2. Penilaian**

Penilaian adalah proses memberikan atau menentukan nilai kepada objek tertentu berdasarkan suatu kriteria tertentu. Penilaian merupakan kegiatan menafsirkan atau mendeskripsikan hasil pengukuran. Penilaian adalah proses

untuk mengambil keputusan dengan menggunakan informasi yang diperoleh melalui pengukuran hasil belajar, baik yang menggunakan instrumen tes maupun non tes.

Secara garis besar, penilaian dapat dibagi menjadi dua, yaitu:

### **Penilaian formatif**

Penilaian formatif dilakukan dengan maksud memantau sejauhmanakah suatu proses pendidikan telah berjalan sebagaimana yang direncanakan

### **Penilaian sumatif.**

Penilaian sumatif dilakukan untuk mengetahui sejauhmanakah peserta didik telah dapat berpindah dari suatu unit pembelajaran ke unit berikutnya.

Untuk melakukan penilaian hasil belajar perlu memperhatikan prinsip-prinsip dan teknik penilaian.

### **Prinsip-prinsip penilaian**

Penilaian hasil belajar peserta didik didasarkan pada prinsip-prinsip sebagai berikut.

1. *Mendidik*, yakni mampu memberikan sumbangan positif terhadap peningkatan pencapaian belajar peserta didik. Hasil penilaian harus dapat memberikan umpan balik dan memotivasi peserta didik untuk lebih giat belajar.
2. *Terbuka/transparan*, yakni prosedur penilaian, kriteria penilaian, dan dasar pengambilan keputusan diketahui oleh pihak yang terkait.
3. *Menyeluruh*, yakni meliputi berbagai aspek kompetensi yang akan dinilai. Penilaian yang menyeluruh meliputi ranah pengetahuan (kognitif), keterampilan (psikomotor), sikap dan nilai (afektif) yang direfleksikan dalam kebiasaan berfikir dan bertindak.
4. *Terpadu dengan pembelajaran*, yakni menilai apapun yang dikerjakan peserta didik dalam kegiatan belajar mengajar itu dinilai, baik kognitif, psikomotorik dan afektifnya. Dengan demikian, penilaian tidak hanya

- dilakukan setelah peserta didik menyelesaikan pokok bahasan tertentu melainkan saat mereka sedang melakukan proses pembelajaran.
5. *Objektif*, yakni tidak terpengaruh oleh pertimbangan subjektif penilai.
  6. *Sistematis*, yakni penilaian dilakukan secara berencana dan bertahap untuk memperoleh gambaran tentang perkembangan belajar peserta didik sebagai hasil kegiatan belajarnya.
  7. *Berkesinambungan*, yakni dilakukan secara terus menerus sepanjang berlangsungnya kegiatan pembelajaran.
  8. *Adil*, yakni tidak ada peserta didik yang diuntungkan atau dirugikan berdasarkan latar belakang sosial-ekonomi, budaya, agama, bahasa, suku bangsa, warna kulit, dan jender.
  9. *Menggunakan acuan kriteria*, yakni menggunakan kriteria tertentu dalam menentukan kelulusan peserta didik.

#### **a. Teknik Penilaian**

Dalam memperoleh data, pendidik dapat menggunakan berbagai teknik penilaian secara komplementer (saling melengkapi) sesuai dengan kompetensi yang dinilai, sebagaimana diuraikan dalam panduan penilaian masing-masing kelompok mata pelajaran. Teknik-teknik tersebut antara lain terdiri atas:

##### **1) Tes kinerja**

Tes kinerja dapat berbentuk tes keterampilan tertulis, tes identifikasi, tes simulasi, dan uji petik kerja. Melalui tes kinerja peserta didik diminta mendemonstrasikan kinerjanya.

##### **2) Observasi**

Observasi dilakukan untuk mengumpulkan data kualitatif dan kuantitatif sesuai dengan kompetensi yang dinilai, baik dilakukan secara formal maupun informal. Observasi formal dilakukan dengan cara menggunakan instrumen yang sudah dirancang sebelumnya, sedangkan observasi informal dilakukan tanpa menggunakan instrumen yang dirancang terlebih dahulu.

### **3). Penugasan**

Penugasan dapat dilaksanakan dalam bentuk proyek atau tugas rumah. Proyek adalah sejumlah kegiatan yang dirancang, dilakukan, dan diselesaikan oleh peserta didik di luar kegiatan kelas dan harus dilaporkan secara tertulis maupun lisan dalam waktu tertentu. Tugas rumah adalah tugas yang harus diselesaikan peserta didik di luar kegiatan kelas, misalnya menyelesaikan soal-soal dan melakukan latihan.

### **4). Portofolio**

Portofolio adalah kumpulan dokumen dan karya-karya peserta didik dalam bidang tertentu yang diorganisasikan untuk mengetahui minat, perkembangan, prestasi, dan kreativitas peserta didik.

### **5). Tes tertulis**

Tes tertulis dilakukan dalam bentuk tes yang jawabannya berupa pilihan dan isian. Tes yang jawabannya berupa pilihan meliputi pilihan ganda, benar-salah, menjodohkan, dll. Adapun tes yang jawabannya berupa isian berbentuk isian singkat dan uraian.

### **6). Tes lisan**

Tes lisan dilaksanakan melalui komunikasi langsung tatap muka antara peserta didik dengan seorang atau beberapa penguji. Pertanyaan dan jawaban diberikan secara lisan dan spontan. Tes jenis ini memerlukan daftar pertanyaan dan pedoman pensekoran.

### **7). Jurnal**

Jurnal merupakan catatan pendidik selama proses pembelajaran yang berisi informasi kekuatan dan kelemahan peserta didik yang berkaitan dengan kinerja ataupun sikap peserta didik yang dipaparkan secara deskriptif.

### **8). Wawancara**

Wawancara dilakukan untuk memperoleh informasi secara mendalam tentang wawasan, pandangan, atau aspek kepribadian peserta didik yang jawabannya diberikan secara lisan dan spontan.

### **9). Inventori**

Inventori merupakan skala psikologis yang dipakai untuk mengungkapkan sikap, minat, dan persepsi peserta didik terhadap sesuatu objek psikologis. Inventori antara lain berupa skala Thurstone, skala Likert, atau skala berdiferensiasi semantik.

### **10). Penilaian Diri**

Penilaian diri merupakan teknik penilaian dengan cara meminta peserta didik untuk mengemukakan kelebihan dan kekurangan dirinya dalam berbagai hal.

### **11). Penilaian antarteman**

Penilaian antarteman merupakan teknik penilaian dengan cara meminta peserta didik untuk mengemukakan kelebihan dan kekurangan temannya dalam berbagai hal.

Kombinasi penggunaan berbagai teknik penilaian akan memberikan informasi yang lebih akurat tentang kemajuan belajar peserta didik. Rangkuman bentuk penilaian beserta bentuk instrumennya disajikan dalam tabel berikut.

### **b. Acuan Penilaian**

Dilihat dari perencanaan dan penafsiran hasil tes, pengukuran dalam bidang pendidikan dapat berdasarkan acuan norma/relatif atau acuan kriteria/patokan. Kedua acuan tersebut menggunakan asumsi yang berbeda tentang kemampuan seseorang. Penafsiran hasil tes antara kedua acuan itu juga berbeda, sehingga menghasilkan informasi yang berbeda maknanya. Pemilihan

acuan ditentukan oleh karakteristik mata pelajaran yang akan diukur dan tujuan yang akan dicapai.

Penilaian acuan norma berasumsi bahwa kemampuan orang berbeda dan dapat digambarkan menurut distribusi normal. Perbedaan itu harus ditunjukkan oleh hasil pengukuran, misalnya setelah mengikuti pembelajaran selama satu semester, peserta didik dites. Hasil tes seorang peserta didik dibandingkan dengan kelompoknya, sehingga dapat diketahui posisi peserta didik tersebut di kelas itu.

Penilaian acuan kriteria berasumsi bahwa hampir semua orang dapat belajar apa saja, meskipun dengan waktu yang berbeda. Dalam acuan kriteria, penafsiran skor hasil tes selalu dibandingkan dengan kriteria yang telah ditetapkan. Bagi peserta didik yang telah mencapai kriteria yang telah ditetapkan (standar) diberi pelajaran tambahan yang biasa disebut pengayaan, sedangkan bagi peserta didik yang belum mencapai standar diberi remedi.

Ketuntasan belajar setiap indikator yang telah ditetapkan dalam suatu kompetensi dasar atau kriteria ketuntasan minimal (KKM) adalah 75%. Satuan pendidikan harus menentukan kriteria ketuntasan minimal dengan mempertimbangkan tingkat kemampuan rata-rata peserta didik dan pendukung penyelenggaraan pembelajaran. Satuan pendidikan diharapkan meningkatkan kriteria ketuntasan belajar ini terus menerus untuk mencapai kriteria ketuntasan 75% atau lebih. Khusus untuk SMK, nilai mata pelajaran Kompetensi Keahlian Kejuruan kriteria ketuntasannya 7,00.

### **3. Tes**

Tes adalah sehimpunan pertanyaan yang harus dijawab, atau pernyataan-pernyataan yang harus dipilih, ditanggapi, atau tugas-tugas yang harus dilakukan oleh orang yang diuji dengan tujuan untuk mengukur suatu aspek tertentu dari orang yang diuji tersebut.

Tes merupakan sejumlah pertanyaan yang memiliki jawaban benar atau salah, pertanyaan yang membutuhkan jawaban, pertanyaan yang harus diberikan tanggapan dengan tujuan mengukur tingkat kemampuan seseorang

atau mengungkap aspek tertentu dari orang yang dikenai tes. Dengan demikian, setiap tes menuntut keharusan adanya respons dari orang yang dites yang dapat disimpulkan sebagai suatu atribut yang dimiliki oleh orang tersebut yang sedang dicari informasinya.

Tes dapat dipilah-pilahkan berdasarkan bentuk, tipe dan ragamnya, seperti berikut:

- a. Menurut bentuknya:** tes bentuk uraian/esei dan tes bentuk objektif.
- b. Menurut tipenya:** tes uraian dapat dipilah menjadi tes uraian terbatas dan tes uraian bebas, sedangkan tes bentuk objektif dapat berbentuk Benar-Salah, Menjodohkan, Pilihan Ganda biasa, Sebab-Akibat, Kompleks.
- c. Menurut ragamnya:**
  - a. Tes uraian terbatas: tes jawaban singkat, tes melengkapi dan tes uraian terbatas sederhana.
  - b. Tes uraian bebas: tes uraian bebas sederhana dan tes uraian ekspresif.
  - c. Tes objektif benar-salah: tes benar -salah sederhana dan tes benar salah dengan koreksi.
  - d. Tes objektif menjodohkan: tes menjodohkan sederhana dan tes menjodohkan sebab akibat.
  - e. Tes objektif pilihan ganda: tes pilihan ganda biasa, tes pilihan ganda hubungan antar hal, tes pilihan ganda analisis kasus, tes pilihan ganda kompleks, dan tes pilihan ganda membaca diagram.

#### **4. Manfaat pengukuran, penilaian dan tes hasil belajar:**

- a. Umpan balik bagi:
  - 1) peserta didik
  - 2) Guru
  - 3) Guru lain (misal guru BP)
  - 4) Orang tua



- 5) Institusi
- 6) Pejabat
- b. Seleksi
- c. Diagnosis ..... remedial
- d. Pengembangan ilmu pengetahuan ttg pengukuran hasil belajar
- e. Menumbuhkan motivasi
- f. Pengembangan kurikulum
- g. Penempatan (*placement test*)

## **5. Keterbatasan Pengukuran, Penilaian, dan Tes Hasil Belajar**

Keterbatasan dapat ditinjau dari aspek-aspek:

- a. Si pembuat tes
- b. Peserta tes
- c. Instrumen/alat
- d. Lingkungan

## **BAB II**

### **PERENCANAAN DAN PENGEMBANGAN TES**

#### **A. Perencanaan tes**

Dalam merencanakan tes perlu memperhatikan dan mempertimbangkan hal-hal sebagai berikut:

1. Pengambilan sampel dan pemilihan butir soal.
2. Aspek yang akan diuji.
3. Tipe tes yang akan digunakan.
4. Format butir soal.
5. Jumlah butir soal.
6. distribusi tingkat kesukaran butir soal (25% mudah, 50% sedang, 25% sukar)

#### **2. Pengembangan tes**

Untuk mengembangkan tes dapat diikuti langkah-langkah sebagai berikut:

1. Menyusun kisi-kisi tes.
2. Menulis soal.
3. Menelaah soal.
4. Mengujicoba soal
5. Menganalisis butir soal.
6. Memperbaiki tes.
7. Merakit tes.
8. Melaksanakan tes.
9. Menafsirkan hasil tes.

#### **Penyusunan Kisi-kisi**

Kisi-kisi digunakan untuk acuan pengembangan instrumen, baik bentuk maupun item instrumen. Dosen dalam penyusunan kisi-kisi perlu menelusuri dan mengacu pada pengembangan kurikulum, silabus, dan pengalaman belajar mahasiswa.

Kisi-kisi asesmen berbasis kompetensi digunakan untuk menunjukkan keterkaitan antara kompetensi, kompetensi dasar, indikator pencapaian, dan strategi asesmen yang direncanakan (yang meliputi metode asesmen, bentuk asesmen, dan item instrumen). Kisi-kisi asesmen berbasis kompetensi tersebut dapat ditampilkan dalam bentuk Tabel 1 berikut.

Tabel 1. Kisi-kisi Asesmen Berbasis Kompetensi untuk Tingkat Mata Pelajaran

Program Studi: .....

Mata Pelajaran: .....

Standar kompetensi mata kuliah	Kompetensi dasar	Indikator pencapaian	Strategi Asesmen		
			Metode	Bentuk instrumen	Nomor item instrumen

## BAGIAN III

### KONSTRUKSI BUTIR SOAL

#### A. Tes Bentuk Uraian

Tes uraian Tes uraian adalah perangkat tes yang butir soalnya mengandung pertanyaan atau tugas yang jawaban atau pengerjaan soal tersebut harus dilakukan dengan cara **mengekspresikan pikiran** peserta tes. Ciri khas tes uraian adalah bahwa jawaban soal tidak disediakan oleh orang yang mengkonstruksi tes, tetapi harus dipasok oleh peserta tes. Peserta tes bebas menjawab pertanyaan yang diajukan. Setiap peserta tes dapat memilih, menghubungkan dan menyampaikan gagasannya dengan menggunakan kata-katanya sendiri. Dengan demikian, pemberian skor terhadap jawaban soal tidak mungkin dilakukan secara objektif.

Tes bentuk uraian sangat tepat untuk mengukur hasil belajar:

1. Mengaplikasikan prinsip.
2. Menginterpretasi hubungan.
3. Mengenal dan menyatakan inferensi.
4. Mengenal relevansi dari suatu informasi
5. Merumuskan dan mengenal hipotesis.
6. Merumuskan dan mengenal kesimpulan yang sah.
7. Mengidentifikasi asumsi yang mendasarkan suatu kesimpulan.
8. Mengenal keterbatasan data.
9. Mengenal dan menyatakan masalah.
10. Mendesain prosedur eksperimen.

Tes uraian dapat dikelompokkan menjadi dua

1. Uraian bebas/terbuka
2. Uraian terbatas/tertutup/terstruktur

## **Tes Uraian Bebas**

### Contoh

1. Bersifat ingatan yang terpilih.

Misal: Sebutkan tiga cara mencegah terjadinya korosi pada baja..

2. Bersifat ingatan evaluatif.

Misal: Sebutkan nama dua tokoh yang paling besar peranannya dalam reformasi di Indonesia dalam abad 21.

3. Membandingkan dua hal terbatas.

Misal: Bandingkanlah taktik dan strategi perjuangan mencapai kemerdekaan antara Ir. Soekarno dan Drs. Moh. Hatta.

4. Membandingkan dua hal secara umum.

Misal: Bandingkanlah logam ferro dan non ferro.

5. Mengambil keputusan, baik dalam arti menentang atau mendukung sesuatu.

Misal: Apakah sebaiknya hukuman mati diterapkan dalam negara yang berdasarkan Pancasila? Berikan alasan pendapatmu.

6. Menguraikan sebab akibat.

Misal: Apakah sebabnya tumbuh-tumbuhan yang selalu terlindung dari sinar matahari kelihatan kurus dan kemudian mati?

7. Menjelaskan penggunaan atau pengertian suatu frasa atau pernyataan dalam suatu karangan.

Misal: Definisikan arti frasa "makan hati" dalam kalimat berikut ini.

"Ibu tua itu selalu makan hati melihat kelakuan anaknya".

8. Meringkas suatu karangan yang telah dibaca.

Misal: Uraikanlah secara singkat siklus air (tidak lebih dari 100 kata)

9. Menganalisis.

Misal: Dalam setiap perundingan antara Republik Indonesia dan Belanda pada masa perjuangan mempertahankan kemerdekaan Indonesia, Belanda selalu mengusulkan agar Indonesia berbentuk negara serikat. Alasan-alasan politis apakah yang mendasari usul belanda tersebut?

10. Menyatakan hubungan.

Misal: Apakah sebabnya rumah harus mempunyai ventilasi yang cukup?

11. Memberi ilustrasi atau contoh.

Misal: Berilah dua contoh tindakan manusia yang menyebabkan terganggunya keseimbangan alam?

12. Mengklasifikasi.

Misal: Masuk golongan apakah binatang berikut ini? Sapi, kerbau, kambing, kijang, rusa, jerapah. Beri alasan.

13. Menerapkan prinsip atau aturan kedalam suatu situasi baru.

Misal: Andaikan ada sebuah balon diisi dengan gas ringan, kemudian dilepaskan dalam sebuah kamar. Balon tersebut mengambang di antara lantai dan langit-langit. Bila kemudian gas dalam balon tersebut dipanaskan apakah yang akan terjadi?

14. Membahas sesuatu.

Misal: Bahaslah hubungan antara panjang tangkai suatu pendulum dengan jangka waktu berayunnya.

15. Menyatakan maksud atau tujuan.

Misal: Tulislah interpretasi anda secara singkat apa maksud pengarang sajak "Aku" menyatakan bahwa "Aku ingin hidup seribu tahun lagi"

16. Mengeritik secara tepat, terpercaya, dan relevan.

Misal: Coba tulis kritik atau pertahankan pendapat yang menyatakan bahwa semua bakteri berbahaya bagi kesehatan manusia.

7. Membuat garis besar.

Misal: Tulislah secara garis besar cara untuk menghitung harga satuan suhu dari skala Celcius ke skala Fahrenheit.

18. Mengorganisasi ulang (reorganisasi) fakta.

Misal: Telusurilah kembali perkembangan bahasa Indonesia dari bahasa Melayu sehingga menjadi bahasa negara dan bahasa pengantar di Nusantara.

19. Merumuskan permasalahan atau pertanyaan dari beberapa kenyataan atau asumsi yang ditegakkan terlebih dahulu.

Misal: Kenyataan menunjukkan bahwa laju peningkatan penduduk di Indonesia masih berkisar antara 1,5% sampai dengan 2,0% untuk masa 25 tahun mendatang, dan laju pertumbuhan ekonomi kita akan berkisar antara 2% sampai dengan 5%. Rumuskanlah tiga masalah pokok yang akan timbul pada awal abad ke 21 yang akan datang di Indonesia.

20. Menyatakan metode atau prosedur baru.

Misal: Dalam keadaan biasa (normal) tumbuh-tumbuhan yang baru ditanam akan tumbuh dengan pucuk mengarah ke atas dan akar mengarah ke bawah. Dapatkah anda jelaskan kapan keadaan tersebut tidak berlaku? Tuliskan persyaratan yang harus dipenuhi.

## 2. Tes Uraian Tipe Jawaban Singkat

Ada dua model yaitu bentuk pertanyaan dan bentuk asosiasi.

### a. Bentuk pertanyaan

Siapakah pembaca teks  
Proklamasi pada 17-8-1945? \_\_\_\_\_

Apakah nama tarian khas  
Sumatera Barat? \_\_\_\_\_

$p + 12 = 25$  berapakah  $p$ ? \_\_\_\_\_

Bagaimanakah simbol  
kimia asam belerang \_\_\_\_\_

### b. Bentuk asosiasi

Apakah nama kesenian daerah dari daerah-daerah berikut?

Bali \_\_\_\_\_

Yogyakarta \_\_\_\_\_

Surabaya \_\_\_\_\_

Ponorogo \_\_\_\_\_

## 2. Tes Bentuk Objektif

Tes bentuk objektif adalah perangkat tes yang butir-butir soalnya mengandung alternatif jawaban yang harus dipilih atau dikerjakan oleh peserta tes. Alternatif jawaban telah dipasok oleh pengkonstruksi butir soal. Peserta tes hanya memilih jawaban dari alternatif jawaban yang telah disediakan. Dengan demikian, pemberian skor terhadap jawaban soal dapat dilakukan secara objektif oleh pemeriksa. Karena sifatnya yang objektif ini maka penskorannya tidak saja bisa dilakukan oleh manusia, melainkan juga oleh mesin. Seperti mesin scanner.



Secara umum ada tiga tipe tes bentuk objektif, yaitu: benar-salah, menjodohkan, dan pilihan ganda.

### **a. Contoh Tes Objektif Pilihan Ganda**

1. Saripati permasalahan harus ditempatkan pada pokok soal (stem).

Contoh :

(Kurang baik)

Pulau Jawa adalah pulau yang .....

- a. menghasilkan banyak minyak
- b. penduduknya terpadat
- c. dijadikan objek wisata
- d. mendapat julukan pulau Perca

(Sebaiknya)

Pulau yang terpadat penduduknya di Indonesia adalah pulau.....

- a. Sumatera
- b. Jawa
- c. Kalimantan
- d. Sulawesi

2. Kalau pokok soal merupakan pernyataan yang belum lengkap, maka kata atau kata-kata yang melengkapinya harus diletakkan pada ujung pernyataan.

Contoh :

Kurang baik

Menurut De Bakey, ..... adalah penyebab penyakit penyempitan pembuluh darah.

- a. kolesterol
- b. kelebihan berat
- c. merokok
- d. tekanan batin

Sebaiknya

Menurut De Bakey, penyakit penyempitan pembuluh darah disebabkan oleh.....

- a. kolesterol
- b. kelebihan berat
- c. merokok
- d. tekanan batin

3. Hindari penggunaan kata-kata teknis atau ilmiah atau istilah yang aneh atau mentereng.

Contoh :

Kurang baik

Apakah kritik utama ahli psikologi terhadap tes?

- a. Tes menimbulkan anxiety.
- b. Tes selalu disertai cultural bias
- c. Tes hanya mengukur hal-hal yang trivial
- d. Tes tergantung hanya pada kemampuan kognitif guru

Sebaiknya:

Apakah kritik utama ahli psikologi terhadap tes?

- a. Tes menimbulkan cemas
- b. Tes selalu disertai bias budaya
- c. Tes hanya mengukur hal-hal yang tampak
- d. Tes tergantung hanya pada kemampuan guru

4. Semua pilihan jawaban harus **homogen** dan dimungkinkan sebagai jawaban yang benar.

Contoh :

Kurang baik

Siapakah di antara nama-nama di bawah ini yang menemukan telepon?

- a. Bell
- b. Marconi
- c. Morse
- d. Pasteur

Sebaiknya

Siapakah di antara nama-nama di bawah ini yang menemukan telepon?

- a. Bell
- b. Marconi
- c. Morse
- d. Edison

5. Hindari adanya petunjuk/indikator pada jawaban yang benar.

Contoh :

Kurang baik

Agar air panas dalam teko tidak cepat dingin maka teko tersebut dibungkus dengan.....

- a. kain
- b. seng
- c. tembaga
- d. timah

Sebaiknya

Air panas akan bertahan panas jika disimpan dalam bejana yang terbuat dari.....

- a. aluminium
- b. keramik
- c. seng
- d. plastik

6. Pokok soal diusahakan tidak menggunakan ungkapan atau kata-kata yang bermakna tidak tentu, misalnya: kebanyakan, seringkali, kadang-kadang dan sebagainya.

Contoh :

Kurang baik

Kebanyakan hewan yang hidupnya dalam air bernafas dengan .....

- a. insang
- b. kulit
- c. paru-paru
- d. insang dan paru-paru

Sebaiknya

Berudu bernafas dengan .....

- a. insang
- b. kulit
- c. paru-paru
- d. insang dan paru-paru

Semua contoh di atas merupakan contoh tes pilihan ganda biasa. Selain itu, pilihan ganda juga dapat dikelompokkan pada pilihan ganda analisis hubungan antar hal, pilihan ganda analisis kasus, pilihan ganda kompleks, dan pilihan ganda yang menggunakan diagram, gambar, grafik atau tabel.

## **b. Pilihan Ganda Analisis Hubungan Antar Faktor**

Contoh:

Untuk soal berikut ini pilihlah:

- A. Jika kedua pernyataan benar dan keduanya menunjukkan sebab akibat.
- B. Jika pernyataan pertama dan kedua benar tapi tidak menunjukkan hubungan sebab akibat.
- C. Jika salah satu dari pernyataan tersebut salah
- D. Jika kedua pernyataan salah

Contoh

Frekuensi detak nadi seseorang yang baru berlari cepat akan naik.

SEBAB

Pada waktu lari cepat denyut jantung bertambah cepat

## **c. Pilihan Ganda Analisis Kasus**

Kasus

Kadit Lantas Polda DIY menjelaskan jumlah kecelakaann lalu lintas di DIY bulan Januari-Juni 2008 sebanyak 5000 kasus atau meningkat 5,25% dibanding tahun 2007. Meningkatnya kecelakaan itu antara lain dikarenakan terhentinya Operasi Zebra menjadi operasi rutin lalu lintas. Di samping itu pengguna jalan hanya berdisiplin jika petugas.

Pertanyaan

Meningkatnya kecelakaan lalu lintas di DIY bukan hanya disebabkan oleh terhentinya Operasi Zebra tetapi juga disebabkan:

- a. pengawas lalu lintas yang tidak pernah kendor
- b. volume kendaraan di jalan makin bertambah
- c. angkutan yang terlibat dalam pengaturan lalu lintas dikurangi jumlahnya
- d. potensi polisi lalu lintas belum dikerahkan secara maksimal

#### **d. Pilihan Ganda Kompleks**

Untuk soal berikut pilihlah:

- A. Jika (1) dan (2) benar
- B. Jika (1) dan (3) benar
- C. Jika (2) dan (3) benar
- D. Jika semuanya benar

Salah satu vitamin yang larut dalam lemak adalah vitamin A yang terdapat dalam.....

- (1) minyak ikan dan telur
- (2) bayam, ikan dan telur
- (3). air susu dan wortel

Pilihan Ganda menggunakan Diagram, Gambar, Grafik atau Tabel pada prinsipnya sama dengan ganda biasa. Yang harus diperhatikan adalah gambar atau grafik atau tabel atau bentuk lain yang sejenis harus dibuat sejelas dan sebaik mungkin.

## **BAGIAN IV**

### **INSTRUMEN NON TES**

Informasi tentang hasil belajar peserta didik tidak saja hanya dapat diperoleh melalui tes, tetapi juga dapat melalui non tes, misalnya pedoman observasi, skala sikap, daftar cek, catatan anekdotal, dan jaringan sosiometrik.

Alat ukur yang berupa non tes digunakan untuk mengukur perubahan tingkah laku yang berkenaan dengan ranah kognitif, afektif, maupun psikomotorik terutama yang berhubungan dengan apa yang dapat dikerjakan dari pada apa yang diketahui atau dipahami. Alat ukur non tes yang banyak digunakan untuk mengukur hasil belajar antara lain:

- a. bagan partisipasi (*participation charts*)
- b. daftar cek (*check list*)
- c. skala lajuan (*rating scale*)
- d. skala sikap (*attitude scale*)

Skala lajuan dapat dibagi dalam empat tipe yaitu:

- a. *numerical rating scale*
- b. *descriptive graphic rating scale*
- c. *ranking methods rating scale*
- d. *paired comparison rating scale*

Skala sikap dapat dibedakan:

- a. Skala Likert
- b. Skala Thurstone
- c. Skala Guttman

## 1. Bagan Partisipasi (*Participations Charts*)

Mata Pelajaran :

Topik :

Tanggal :

Waktu :

Tujuan :

No	Nama	Kualitas kontribusi			
		Sangat Berarti	Penting	Meragukan	Tidak Relevan
1	A	III	III	-	-
2	B	II	II	I	I
3	C	III	III	-	-
4	D	I	II	-	II
5	E	III	-	I	II
6	F	I	-	III	II
7	G	-	II	-	-
8	H	I	II	-	-
9	I	IIII	I	-	I
10	J	II	III	-	-

- \*) Sangat berarti: mengemukakan gagasan baru yang penting dalam diskusi  
Penting : mengemukakan alasan-alasan penting dalam pendapatnya  
Meragukan : pendapat yang tak didukung oleh data atau fakta  
Tidak relevan : gagasan yang diajukan tidak relevan dengan masalah

## 2. Cek List

Petunjuk: Berilah tanda cek (V) di tempat yang telah disediakan dalam tabel berikut untuk setiap pernyataan yang disajikan

No	Aspek yang diamati	Cek
1	Menyatakan rasa gembira secara lisan	.....
2	Menyatakan rasa sedih secara lisan	.....
3	Memperlihatkan sikap gembira	.....
4	Memperlihatkan sikap sedih	.....
5	Mengucapkan selamat kepada orang lain yang berbahagia	.....
6	Meniru tingkah laku orang dewasa	.....
7	Dst.	

#### FORMAT OBSERVASI

1. Apakah semua siswa membaca petunjuk dengan baik sebelum praktikum
2. Apakah mereka melakukan:
  - a. pemeriksaan terhadap semua alat yang tersedia
  - b. pemeriksaan terhadap larutan yang disediakan
  - c. pemeriksaan terhadap sambungan baterai (sumber arus)
3. Apakah semua siswa ikut serta dalam diskusi

#### 3. Skala Nilai (Rating Scale)

##### a. Numerical rating scale

Contoh: Nyatakanlah tingkatan setiap pernyataan atau berikut ini dengan cara melingkari salah satu angka yang ada di depan pernyataan tersebut. Angka-angka itu mengandung makna:

- 1 = tidak memuaskan
- 2 = di bawah rata-rata
- 3 = rata-rata
- 4 = di atas rata-rata
- 5 = sempurna



1. Keaktifan peserta didik berpartisipasi dalam kegiatan kelas 1 2 3 4 5
2. Interaksi peserta didik dengan kelompoknya 1 2 3 4 5

**a. Descriptive Graphic Rating Scale**

1. Bagaimanakah aktivitas peserta didik dalam diskusi kelas? sangat aktif, \_\_, \_\_, \_\_, \_\_, \_\_, tidak aktif

## **BAB IV**

### **ANALISIS KUALITAS SOAL**

Analisis kualitas perangkat soal tes hasil belajar dapat dilakukan dengan dua cara yaitu: analisis secara teoritik (kualitatif) dan analisis secara empiris (kuantitatif). Analisis secara teoritis adalah telaah soal yang difokuskan pada aspek materi, konstruksi, dan bahasa. Aspek materi berkaitan dengan substansi keilmuan yang ditanyakan serta tingkat berpikir yang terlibat, aspek konstruksi berkaitan dengan teknik penulisan soal, dan aspek bahasa berkaitan dengan kekomunikatifan/kejelasan hal yang ditanyakan. Analisis empiris adalah telaah soal berdasarkan data lapangan (uji coba) yang mencakup tingkat kesukaran, daya beda, keberfungsian pengecoh, kebermaknaan butir, dan reliabilitas.

#### **A. Analisis Kualitas Soal Secara Teoritis**

Secara teoritis, kualitas soal dapat ditelaah dengan memperhatikan hal-hal sebagai berikut:

##### **1. Tes Bentuk Obyektif**

###### **a. Materi:**

- 1) Soal harus sesuai dengan indikator.
- 2) Pengecoh berfungsi.
- 3) Setiap soal harus mempunyai satu jawaban yang benar atau yang paling benar.

###### **b. Konstruksi**

- 4) Pokok soal harus dirumuskan secara jelas.
- 5) Rumusan pokok soal dan pilihan jawaban harus merupakan pernyataan yang diperlukan saja.
- 6) Pokok soal jangan memberi petunjuk ke arah jawaban benar.
- 7) Pokok soal jangan mengandung pernyataan yang bersifat negatif ganda.
- 8) Pilihan jawaban harus homogen dan logis ditinjau dari segi materi.
- 9) Panjang rumusan pilihan jawaban relatif sama.
- 10) Pilihan jawaban tidak menggunakan pernyataan "Semua jawaban di atas salah" atau "{Semua jawaban di atas benar" dan sejenisnya.
- 11) Pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan urutan besar kecilnya angka atau kronologis waktunya.
- 12) Gambar/grafik/tabel/diagram dan sejenisnya harus jelas dan berfungsi.

13) Butir tes jangan tergantung pada jawaban sebelumnya.

### **c. Bahasa**

14) Setiap tes harus menggunakan bahasa yang sesuai dengan kaidah bahasa Indoensia.

15) Menggunakan bahasa yang komunikatif sehingga mudah dimengerti.

16) Jangan menggunakan bahasa yang berlaku setempat.

17) Pilihan jawaban jangan mengulang kata atau frase yang bukan merupakan satu kesatuan pengertian.

## **2. Tes Bentuk Uraian**

### **a. Materi:**

1) Soal harus sesuai dengan indikator.

2) Batasan pertanyaan dan jawaban yang diharapkan (ruang lingkup) harus jelas..

3) Isi materi sesuai dengan petunjuk pengukuran.

4) Isi materi yang ditanyakan sudah sesuai dengan jenjang, jenis sekolah, atau tingkat kelas.

### **b. Konstruksi**

5) Rumusan kalimat soal atau pertanyaan harus menggunakan kata-kata tanya atau perintah yang menuntut jawaban terurai, seperti: mengapa, uraikan, jelaskan, bandingkan, hubungkan, tafsirkan, buktikan, dan hitunglah.

6) Buatlah petunjuk yang jelas tentang cara mengerjakan soal.

7) Buatlah pedoman penskoran segera setelah soalnya ditulis dengan cara menguraikan komponen yang akan dinilai atau kriteria peskorannya.

8) Hal-hal lain yang menyertai tes seperti tabel, gambar, grafik, peta, atau yang sejenisnya harus disajikan dengan jelas dan terbaca.

### **c. Bahasa**

9) Rumusan kalimat tes harus komunikatif.

10) Butir tes menggunakan bahasa Indoensia yang baik dan benar.

11) Rumusan tes tidak menggunakan kata-kata/kalimat yang menimbulkan penafsiran ganda atau salah penafsiran..

12) Jangan menggunakan bahasa yang berlaku setempat, jika tes akan digunakan untuk daerah lainj atau nasional.

13) Rumusan tes tidak mengandung kata-kata yang dapat menyinggun g perasaan peserta ujianj.

## **2. Analisis Secara Empiris**

Analisis karakteristik butir soal mencakup analisis parameter kuantitatif dan kualitatif butir soal. Parameter kuantitatif berkaitan dengan analisis butir soal berdasarkan atas tingkat kesukaran, daya beda, dan keberfungsian alternatif pilihan jawaban. Parameter kualitatif berkaitan dengan analisis butir soal berdasarkan atas pertimbangan ahli (*expert judgement*).

#### a. Tingkat Kesukaran

Tingkat kesukaran adalah angka yang menunjukkan besarnya proporsi peserta tes yang menjawab betul pada suatu butir. Rentang angka ini adalah 0,00 sampai 1,00. Jika suatu butir soal memiliki tingkat kesukaran 0,00 berarti tidak ada peserta tes yang menjawab butir soal tersebut dengan benar. Dengan kata lain butir soal terlalu sukar. Sebaliknya, jika butir soal memiliki tingkat kesukaran 1,00 berarti semua peserta tes dapat menjawab butir soal dengan benar. Dengan kata lain, butir soal terlalu mudah.

Untuk menghitung besarnya tingkat kesukaran suatu butir dapat digunakan rumus sebagai berikut:

$$\text{Tingkat Kesukaran (TK)} = \frac{\text{Jumlah peserta tes yang menjawab benar}}{\text{Jumlah seluruh peserta tes}}$$

Contoh: Pd butir 1, peserta tes yang menjawab benar ada 15 orang dari 35 orang siswa. Berapakah Indeks Kesukaran butir tsb.

$$IK (1) = 15/35 \times 100 = 0,45$$

Berdasarkan rumus tersebut maka dapat dikatakan bahwa tingkat kesukaran suatu butir soal dipengaruhi oleh tingkat kemampuan dari anggota kelompok peserta tes. Hal ini berarti bahwa tingkat kesukaran butir soal tidak semata-mata menunjukkan ukuran kesukaran butir soal, tetapi juga menunjukkan kemampuan rata-rata peserta tes. Misalnya, pada suatu kelompok tes, tingkat kesukaran butir soal nomor 2 adalah 0,65, angka ini dapat

diinterpretasikan bahwa butir soal nomor 2 ini memiliki tingkat kesukaran 0,65 untuk kelompok tersebut.

Berapakah tingkat kesukaran yang paling baik untuk suatu butir soal? Tidak ada referensi yang secara pasti menyebutkan besarnya tingkat kesukaran tertentu untuk menyatakan bahwa suatu butir soal memiliki tingkat kesukaran yang baik. Namun, dalam beberapa literatur disebutkan bahwa rentang tingkat kesukaran yang dapat digunakan sebagai kriteria adalah: lebih kecil dari 3,00 masuk kategori sukar, antara 0,30 – 0,80 termasuk cukup/sedang, dan lebih besar dari 0,80 termasuk mudah.

Untuk membantu bagaimana cara menentukan tingkat kesukaran butir soal dan sekaligus tingkat kesukaran perangkat soal dapat diikuti contoh berikut. Misal, hasil tes mata pelajaran Matematika untuk 10 orang siswa kelas 11 SMK seperti pada Tabel 2. Berdasarkan data Tabel 2. tersebut dapat dihitung tingkat kesukaran untuk masing-masing butir. Misal, butir nomor 1, pada butir ini peserta tes yang menjawab betul ada 10 orang dan jumlah peserta tes seluruhnya ada 10 orang, jadi tingkat kesukaran butir nomor 1 tersebut adalah:  $10/10 = 1$ . Pada butir nomor 2, peserta tes yang menjawab betul pada butir ini ada 7 orang dan jumlah peserta tes seluruhnya ada 10 orang, jadi tingkat kesukaran butir nomor 2 tersebut adalah:  $= 7/10 = 0,70$ . Dengan cara yang sama dapat dihitung tingkat kesukaran untuk butir-butir lain yang hasilnya adalah sebagai berikut: tingkat kesukaran butir nomor 3 =  $7/10 = 0,70$ ; butir nomor 4 =  $0,70$ ; butir nomor 5 =  $0,80$ ; butir nomor 6 =  $0,90$ , butir nomor 7 =  $0,70$  dan butir nomor 8 =  $0/10 = 0,00$ . Butir nomor 1 adalah butir yang terlalu mudah, sedangkan butir nomor 8 adalah butir yang terlalu sukar.

Untuk menghitung tingkat kesukaran perangkat soal tersebut (terdiri dari 8 butir soal) dapat digunakan cara sebagai berikut:

$$\text{Tingkat Kesukaran Naskah Soal} = \frac{\text{Jumlah tingkat kesukaran semua butir}}{\text{Jumlah seluruh butir soal}}$$

Tabel 2. Hasil Tes Mata Pelajaran Matematika  
(Siswa Kelas 11 SMK = 10 orang)

Butir Siswa	1	2	3	4	5	6	7	8	9	10	Total
A	1	1	1	1	1	1	0	0	0	1	7
B	1	1	1	1	1	1	0	0	1	0	7
C	1	0	1	1	1	0	1	0	0	0	5
D	1	1	0	1	0	1	0	0	0	1	5
E	1	0	1	0	0	1	1	0	0	1	5
F	1	0	1	0	0	0	1	0	0	0	3
G	1	1	1	0	0	1	0	0	0	0	4
H	1	1	1	1	1	1	1	0	1	1	9
I	1	0	0	0	1	1	0	0	0	0	3
J	1	1	0	1	1	1	1	0	1	1	8
K	1	1	1	1	1	1	0	0	0	1	7
L	1	1	1	1	1	1	0	0	1	0	7
M	1	0	1	1	1	0	1	0	1	0	6
N	1	1	0	1	0	1	0	0	0	0	5
O	1	0	1	0	0	1	1	0	0	0	4
P	1	0	0	0	1	1	1	0	1	0	5
Q	1	1	1	1	1	1	0	0	0	1	7
R	1	0	1	1	1	1	1	0	1	0	7
S	1	1	0	0	1	1	0	0	0	1	6
T	1	1	0	1	0	1	0	0	1	0	5
Jumlah	20	12	13	13	13	17	9	0	8	9	
IK	1	0,60	0,65	0,65	0,65	0,85	0,45	0	0,40	0,45	

Keterangan: IK = Indeks Kesukaran

$$\text{Tingkat Kesukaran} = \frac{1,00+0,60+0,65+0,65+ 0,65+0,45+0,70+0,00+0,40+0,45}{10}$$

$$\text{Tingkat Kesukaran Naskah Soal} = 0,55,5 \text{ (sedang)}$$

Dalam satu perangkat soal perlu dipertimbangkan juga proporsi butir-butir soal yang mudah, sedang/cukup, dan sukar. Agar tingkat kesukaran antar butir soal dalam perangkat soal berimbang dapat disusun butir soal dengan proporsi: 25% mudah, 50% cukup, dan 25% sukar. Kriteria tingkat kesukaran butir soal dapat dilihat pada Tabel 3.

Tabel 3. Kriteria Tingkat Kesukaran

Indeks Kesukaran	Kriteria
0,00 – 0,29	Sukar
0,30 – 0,80	Sedang
0,81 – 1,00	Mudah

### b. Daya Beda

Di samping tingkat kesukaran, butir soal juga harus dianalisis daya bedanya. Daya beda butir soal adalah indeks yang menggambarkan tingkat kemampuan suatu butir soal untuk membedakan kelompok yang pandai dari kelompok yang kurang pandai. Interpretasi daya beda selalu dikaitkan dengan kelompok peserta tes. Artinya, suatu daya beda butir soal yang dianalisis berdasarkan data kelompok tertentu belum tentu dapat berlaku pada kelompok yang lain. Dengan kata lain, interpretasi daya beda butir soal untuk peserta tes kelas A tidak mungkin sama dengan interpretasi daya beda kelas B untuk mata

pelajaran yang sama. Hal ini sangat tergantung pada kemampuan masing-masing kelompok.

Untuk menghitung indeks daya beda dapat digunakan rumus sebagai berikut:

$$IB = PA/JA - PB/JB$$

IB = Indeks daya beda

PA = Jumlah kelompok atas yang menjawab benar pada butir

JA = Jumlah seluruh kelompok atas

PB = Jumlah kelompok bawah yang menjawab benar pada butir

JB = Jumlah seluruh kelompok bawah

Berdasarkan rumus tersebut, rentang indeks daya beda suatu butir adalah dari -1 sampai +1. Indeks daya beda = -1 memberikan arti bahwa semua anggota kelompok atas menjawab salah pada butir, sedangkan kelompok bawah semua anggotanya menjawab benar. Butir seperti ini termasuk kategori tidak baik. Sebaliknya, indeks daya beda = +1 memberikan arti bahwa semua anggota kelompok atas menjawab benar pada butir, sedangkan semua anggota kelompok bawah menjawab salah.

Contoh:

Pada butir 1 dari 15 orang kelompok atas yang menjawab benar ada 12 orang, sedang kelompok bawah yang menjawab benar ada 6 orang, berapa Indeks Beda butir 1

$$IB (1) = 12/15 - 6/15 = 6/15 = 0,40$$

Pada butir 1 dari 15 orang kelompok atas yang menjawab benar ada 5 orang, sedang kelompok bawah yang menjawab benar ada 10 orang, berapa Indeks Beda btr 1

$$IB (1) = 5/15 - 10/15 = -5/15 = - 0,33$$



Indeks Beda	Kriteria
0,71 – 1,00	Baik Sekali
0,41 – 0,70	Baik
0,21 – 0,40	Cukup
0,00 – 0,20	Jelek
Negatif	Semua butir soal yang indeks bedanya bernilai negatif adalah butir soal yang kurang baik

### c. Keberfungsian Alternatif Pilihan Jawaban

Dalam tes hasil belajar berbentuk objektif dengan model pilihan ganda, umumnya memiliki (4) empat atau (5) lima alternatif pilihan jawaban dimana salah satu alternatif jawabannya adalah jawaban yang benar (kunci jawaban). Alternatif pilihan jawaban yang salah sering disebut dengan istilah pengecoh (*distractor*). Alternatif pilihan jawaban dalam suatu butir soal dikatakan berfungsi jika semua pilihan jawaban tersebut dipilih oleh peserta tes dengan kondisi di mana jawaban yang benar lebih dipilih dari pada alternatif pilihan jawaban yang lain. Pengecoh berfungsi jika paling sedikit 5% dari peserta tes memilih jawaban tersebut.

### d. Omit

Omit adalah proporsi peserta tes yang tidak menjawab pada semua alternatif jawaban). Butir soal yang baik jika omit paling banyak 10% dari peserta tes.

Misal, pada butir 1 setelah dianalisis diperoleh data sbb:

Pilihan jawaban	a	b	c*	d	omit	Jumlah
Kelompok Atas	5	7	15	3	0	30
Kelompok Bawah	8	8	6	5	3	30
Jumlah	13	15	21	8	3	60

$$IK (1) = 21/60 = 0,35$$

$$IB (1) = 15/30 - 6/30 = 0,30$$

Keberfungsian Pengecoh: berfungsi (>5% peserta tes memilih pengecoh)

Omit: butir soal baik karena <10% dari peserta tes yang tidak memilih semua alternatif jawaban

#### **e. Validitas**

Suatu instrumen dapat dikatakan memiliki validitas yang tinggi apabila instrumen tersebut menjalankan fungsi ukurnya, atau memberikan hasil ukur yang tepat dan akurat sesuai dengan maksud dikenakannya pengukuran tersebut. Hal lain yang juga sangat penting dalam konsep validitas adalah kecermatan pengukuran, yaitu kemampuan untuk mendeteksi perbedaan-perbedaan kecil sekalipun yang ada dalam atribut yang diukurnya.

Pengertian validitas berkaitan erat dengan tujuan pengukuran. Tidak ada validitas yang dapat berlaku secara umum untuk semua tujuan pengukuran. Suatu tes hanya valid untuk satu tujuan pengukuran yang spesifik. Sebagai contoh, tes matematika mungkin valid sebagai prediktor kemampuan di bidang ilmu pengetahuan (*science*) tapi mungkin kurang valid untuk dijadikan prediktor kemampuan di bidang sejarah.

Secara empiris, suatu instrumen dapat dikatakan valid apabila memenuhi dua criteria, yaitu: a. instrumen tersebut harus mengukur konsep atau variabel yang diharapkan hendak diukur dan harus tidak mengukur konsep atau variabel lain yang tidak diharapkan untuk diukur, dan b. instrumen tersebut dapat memprediksi perilaku yang lain yang berhubungan dengan variabel yang diukur. Analisis validitas dapat dilakukan pada dua kawasan yaitu analisis untuk keseluruhan isi instrumen dan analisis untuk masing-masing butir soal atau tes.

#### **Macam Validitas**

Kajian validitas dapat dilakukan melalui telaah mendalam tentang teori atau konsep dan melalui hasil pengalaman empiris di lapangan. Fokus analisis validitas yang dilakukan melalui kajian mendalam tentang suatu teori atau

konsep adalah berkaitan dengan validitas logis (*logical validity*). Fokus analisis validitas yang dilakukan melalui hasil pengalaman empiris adalah berkaitan dengan validitas empiris (*empirical validity*). Secara umum, validitas dapat dikelompokkan menjadi tiga, yaitu: validitas isi, validitas berdasarkan kriteria, dan validitas konstruk. Ada juga yang mengelompokkan validitas kedalam kelompok seperti berikut, validitas logis dikelompokkan menjadi validitas isi (*content validity*) dan validitas konstruk (*construct validity*), sedangkan validitas empiris dapat dikelompokkan menjadi: validitas yang ada sekarang (*concurrent validity*) dan validitas prediksi (*predictive validity*). Tipe validitas dapat dikelompokkan berdasarkan estimasi yang disesuaikan dengan sifat dan fungsi tes, yaitu validitas isi, validitas konstruk, dan validitas kriteria.

### **Validitas Isi.**

Validitas isi menunjuk seberapa jauh instrumen mengukur keseluruhan kawasan pokok bahasan dan perilaku yang hendak diukur. Batasan keseluruhan kawasan pokok bahasan tidak saja mengindikasikan bahwa instrumen tersebut harus komprehensif, melainkan juga isinya harus tetap relevan dan tidak keluar dari batasan tujuan pengukuran. Analisis validitas isi banyak dilakukan terhadap perangkat instrumen untuk pengukuran prestasi belajar (*achievement test*). Analisis validitas isi juga dapat digunakan untuk menganalisis efektifitas pelaksanaan program yaitu dengan jalan membandingkan antara isi instrumen dengan isi rancangan

Validitas isi ditentukan melalui analisis teoritik dan empirik, yang meliputi: (1) kejelasan pokok kompetensi dan sub kompetensi, (2) penetapan pokok kompetensi dan sub kompetensi yang diukur oleh setiap butir tes, (3) kecocokan butir-butir tes dengan kompetensi dan sub kompetensi yang terukur. Validitas isi dapat dianalisis dari validitas tampak (*face validity*) dan validitas logik (*logical validity*). Analisis validitas tampak berkaitan dengan analisis terhadap format tampilan perangkat tes. Analisis validitas logik berkaitan dengan analisis yang menunjuk pada sejauhmana isi tes merupakan representasi dari ciri-ciri atribut yang hendak diukur.

## **Validitas Konstrak**

Validitas konstrak merujuk pada seberapa jauh suatu instrumen mengukur konstrak teori yang hendak diukurnya. Analisis validitas konstrak memiliki asumsi bahwa instrumen yang digunakan mengandung definisi operasional yang tepat. Dengan demikian, analisis validitas konstrak pada dasarnya sama dengan merumuskan suatu konsep yang bersifat abstraksi dan generalisasi yang perlu dirunuskan definisinya sedemikian rupa sehingga dapat diukur. Langkah untuk menganalisis validitas konstrak dimulai dari menganalisis unsur-unsur suatu konstrak, kemudian menilai apakah unsur-unsur tersebut logis mengukur suatu konstrak. Langkah terakhir adalah menghubungkan konstrak yang sedang diukur dengan konstrak lainnya dan menelaah kaitan antara konstrak pertama dengan unsur-unsur tertentu pada konstrak yang tadi.

## **Validitas Kriteria**

Validitas kriteria merupakan validitas yang selalu dikaitkan dengan kriteria eksternal yang dijadikan dasar pegujian skor tes. Dengan kata lain, validitas berdasar kriteria dapat dilakukan dengan komputasi korelasi skor tes dengan skor kriteria. Berdasarkan analisis ini maka validitas kriteria dapat dilihat dari validitas prediktif dan validitas kongkuren.

Validitas prediktif merujuk pada daya prediksi suatu instrumen. Prediksi menunjukkan bahwa kriteria penilaian diperoleh pada masa yang akan datang. Sebagai contoh, misalnya: tes psikologis bagi calon-calon siswa SMK. Tes ini dikatakan memiliki validitas prediktif yang tinggi apabila calon siswa yang mendapat nilai tinggi ternyata juga memiliki prestasi akademik yang tinggi selama belajar di SMK. Dengan kata lain, instrumen yang memiliki validitas prediksi dapat digunakan untuk memprediksi keberhasilan akademik siswa. Koefisien korelasi antara skor tes dan skor kriteria merupakan indikator yang menunjuk pada saling hubung antara skor tes dengan skor kriteria. Dengan demikian, jika tes telah teruji validitasnya maka tes tersebut akan memiliki fungsi prediktif yang sangat berguna bagi peserta tes.

Validitas kongkuren atau validitas yang ada sekarang (*concurrent validity*) merupakan validitas yang selalu dikaitkan dengan kriteria yang ada dan dapat diperoleh dalam waktu yang bersamaan. Validitas kongkuren menjadi sangat penting artinya untuk keperluan diagnostik. Koefisien korelasi antara skor tes dan skor kriteria menunjukkan sejauhmana kesesuaian antara hasil ukur tes dengan hasil ukur tes lain yang sudah teruji kualitasnya. Perbedaannya dengan validitas prediktif adalah tersedianya kriteria, dimana pada validitas prediktif kriteria eksternal diperoleh setelah tes sedangkan pada validitas kongkuren kriteria eksternal diperoleh dalam waktu yang bersamaan.

## **f. Reliabilitas**

### **Pengertian Reliabilitas**

Istilah reliabel dapat diartikan tetap atau konstan. Instrumen yang reliabel adalah kemampuan instrumen untuk menghasilkan data yang mendekati sama bila instrumen tersebut digunakan berulang-ulang untuk mengukur objek yang sama dan dengan cara yang sama. Dengan kata lain, reliabilitas adalah indeks yang menggambarkan sejauhmana suatu instrumen dapat diandalkan.

Analisis reliabilitas selalu dikaitkan dengan konsistensi pengukuran, yaitu bagaimana hasil pengukuran tetap (konstan) dari satu pengukuran ke pengukuran yang lain. Untuk lebih memahami makna reliabilitas dapat didekati dengan memperhatikan tiga aspek yang terkait dengan alat ukur, yaitu: kemantapan, ketepatan, dan homogenitas. Kemantapan merujuk pada hasil pengukuran yang sama pada pengukuran berulang-ulang dalam kondisi yang sama. Ketepatan merujuk pada istilah tepat dan benar dalam mengukur dari sesuatu yang diukur. Artinya, instrumen tersebut memiliki pernyataan-pernyataan yang jelas, mudah dimengerti, dan detail. Homogenitas merujuk pada tingkat keterkaitan yang erat antar unsur-unsurnya.

Hasil-hasil pengukuran umumnya menampilkan sekor yang terbatas yang diperoleh pada waktu tertentu. Hasil pengukuran yang sempurna konsisten tidak mungkin untuk dicapai karena adanya faktor-faktor yang terlibat di dalamnya. Sebagai contoh, misalnya: kelompok yang sama dikenakan pengukuran dua kali

secara berurutan, proses pengukuran ini akan memberikan berbagai variasi skor karena adanya fluktuasi pada memori sesaat, perhatian, usaha, kelelahan, ketegangan emosional, tebak tepat dan sebagainya. Sebaliknya, jika pengukuran dikenakan pada kelompok yang sama tetapi dalam selang waktu yang cukup lama maka variasi skor akan terjadi dikarenakan adanya pengaruh pengalaman belajar, perubahan kesehatan, lupa dan sebagainya.

Validitas dan reliabilitas merupakan persyaratan utama dalam kualitas instrumen. Instrumen yang teruji validitasnya biasanya memiliki tingkat reliabilitas yang tinggi. Namun, instrumen yang reliabel belum menjamin bahwa instrumen tersebut valid. Berkaitan dengan hal ini, Linn dan Gronlund (1995) mengemukakan bahwa *reliability is a necessary but not sufficient condition for validity*. Oleh karena itu, meskipun suatu instrumen berdasarkan analisis dikatakan valid, tapi perlu dilakukan analisis tingkat reliabilitasnya.

### **Pengujian Reliabilitas**

Analisis interpretasi reliabilitas suatu instrumen dapat dilakukan melalui berbagai cara. Sebagai contoh, suatu instrumen dikatakan reliabel jika skor amatan memiliki korelasi tinggi dengan skor sebenarnya. Kuadrat koefisien korelasi antara skor amatan dan skor sebenarnya dinyatakan sebagai koefisien reliabilitas. Reliabilitas dapat juga dinyatakan sebagai koefisien korelasi antara skor amatan dua instrumen yang parallel. Jika dua buah instrumen yang parallel diberikan kepada kelompok yang sama, kemudian skor amatan kedua instrumen tersebut dikorelasikan, koefisien korelasi tersebut dinamakan sebagai koefisien reliabilitas. Dalam banyak kasus adalah sangat sulit mendapatkan skor amatan yang sebenarnya dan juga sangat sulit membuat dua instrumen yang betul-betul parallel. Oleh karena itu, reliabilitas perlu ditaksir dengan metode-metode tertentu.

Pengujian reliabilitas suatu instrumen dapat dikerjakan secara internal dan eksternal. Pengujian reliabilitas secara internal (*internal consistency*) berkaitan dengan analisis konsistensi butir-butir yang ada dalam instrumen. Pengujian reliabilitas secara eksternal dapat dilakukan melalui analisis tes ulang (*test-*

*retest*) berkaitan dengan stabilitas tes (*stability*), tes paralel (*parallel test*). Dengan ketiga metode tersebut, yaitu metode tes ulang, tes paralel dan konsistensi internal, akan menghasilkan taksiran koefisien reliabilitas yang berbeda. Koefisien reliabilitas yang sebenarnya adalah sulit untuk dapat diamati sehingga yang diperoleh hanyalah koefisien reliabilitas taksiran.

### **Metode Tes Ulang (Test-Retest)**

Pengujian reliabilitas dengan metode tes ulang adalah menganalisis tingkat reliabilitas sebuah instrumen yang digunakan berulang (2 kali) dalam waktu yang berbeda untuk responden yang sama dan dalam kondisi yang relatif sama. Perhitungan koefisien reliabilitas dilakukan dengan cara mengkorelasikan skor hasil tes yang pertama dengan skor hasil tes yang kedua. Jika hasilnya positif dan signifikan maka dikatakan instrumen tersebut reliabel. Metode ini tampaknya sederhana, namun memiliki beberapa kelemahan di antaranya:

1. Sangat besar kemungkinannya para responden masih ingat dengan materi soal tes yang pertama (*carry-over effect*) sehingga akan mengulang kembali jawaban yang pernah diberikan pada tes kedua. Untuk itu selang waktu tes pertama dan kedua perlu diperhatikan.
2. Kemungkinan terjadinya perbedaan kesiapan responden pada saat pengukuran pertama dibandingkan dengan pengukuran kedua.

### **Tes Paralel (*Parallel Test*)**

Pengujian reliabilitas dengan cara tes paralel adalah menganalisis reliabilitas dua buah instrumen yang secara teoritis diasumsikan paralel diujikan pada responden yang sama dengan waktu yang sama pula. Perhitungan koefisien reliabilitasnya adalah dengan cara mengkorelasikan skor hasil tes kedua instrumen tersebut. Jika hasilnya positif dan signifikan maka dikatakan bahwa instrumen tersebut reliabel. Namun demikian, untuk membuat dua buah instrumen yang betul-betul paralel memang tidak mungkin.

## Metode Konsistensi Internal

Selain dengan cara-cara di atas, pengujian reliabilitas dapat dilakukan dengan cara yang lebih efisien yaitu dengan cara: instrumen diujikan sekali saja untuk kemudian dilakukan analisis reliabilitas dengan menggunakan teknik-teknik tertentu. Cara seperti ini bisa disebut dengan istilah *internal consistency*. Metode ini tidak cocok jika tes tidak dapat dibagi dua menjadi butir-butir yang parallel atau jika tes tidak memiliki butir-butir independent yang dapat dipisahkan.

Menurut Allen dan Yen (1979) ada tiga cara yang bisa digunakan untuk membagi suatu tes menjadi dua bagian, yaitu: metode ganjil genap, maksudnya, mengelompokkan butir-butir bernomor ganjil menjadi satu kelompok pertama dan butir-butir bernomor genap menjadi kelompok yang kedua; metode belah dua dengan nomor urut, maksudnya semua butir dibagi menjadi dua kelompok, misalnya ada 40 butir, kelompok pertama terdiri dari butir nomor 1 sampai dengan 20 dan kelompok kedua butir nomor 21 sampai dengan 40; dan metode *matched random subset*.

Beberapa teknik yang banyak digunakan untuk menghitung reliabilitasnya adalah dengan teknik belah dua (*split half*) dari Spearman Brown, KR-20, KR-21, dan Alpha Cronbach.

## Teknik Belah Dua (*Split-Half*)

Analisis reliabilitas dengan teknik belah dua adalah membagi banyaknya butir dalam instrumen menjadi dua bagian. Pembagian dapat dilakukan dengan dua cara, yaitu butir ganjil-genap dan separoh atas separoh bawah. Selanjutnya, dilakukan korelasi antar sekor pada masing-masing bagian. Dengan menggunakan rumus Spearman Brown dapat dihitung besarnya koefisien reliabilitas instrumen.

Rumus Spearman Brown:

$$r_{11} = \frac{2 r_{gg}}{1 + r_{gg}}$$



Keterangan:  $r_{11}$  = koefisien reliabilitas

$r_{gg}$  = koefisien korelasi skor antar bagian/belahan

### Kuder Richardson – 20 (KR-20)

$$r_{11} = \frac{k}{(k-1)} \left[ \frac{\sigma_t^2 - \sum pq}{\sigma_t^2} \right]$$

Keterangan:  $r_{11}$  = koefisien reliabilitas

$k$  = jumlah butir

$p$  = proporsi jumlah subjek yang menjawab betul pada butir

$q = 1 - p$

$\sigma_t^2$  = varian sekor total

### Kuder Richardson – 21 (KR-21)

$$r_{11} = \frac{k}{(k-1)} \left[ 1 - \frac{M_t(k - M_t)}{k \sigma_t^2} \right]$$

Keterangan:  $r_{11}$  = koefisien reliabilitas

$k$  = jumlah butir

$M_t$  = rerata sekor total

$\sigma_t^2$  = varian sekor total

Perhitungan koefisien reliabilitas dengan menggunakan rumus Kuder-Richardson lebih cocok untuk instrumen dengan butir-butir soal yang betul-betul homogen. Apabila butir-butirnya tidak homogen akan diperoleh hasil taksiran reliabilitas yang cenderung lebih kecil dari pada reliabilitas sesungguhnya.

## Alpha Cronbach (Koefisien Alpha)

$$r_{11} = \frac{k}{(k-1)} \left[ 1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right]$$

Keterangan:  $r_{11}$  = koefisien reliabilitas

$k$  = jumlah butir

$\sum \sigma_i^2$  = rerata sekor total

$\sigma_t^2$  = varian sekor total

Di samping menyangkut masalah validitas dan reliabilitas, analisis kualitas perangkat tes juga perlu memperhatikan masalah objektivitas, dan efisien. Suatu instrumen dikatakan memiliki objektivitas jika dalam memberikan bentuk kuantitatif hasil pengukuran tidak melibatkan unsur-unsur subjektivitas penilai. Suatu instrumen dikatakan efisien jika dalam menyusunnya mudah dan murah serta mudah menggunakannya (praktis). Di samping itu, waktu yang digunakan untuk mengoreksi hasil pengukurannya tidak membutuhkan waktu yang terlalu lama.

## Contoh Cara Menghitung Koefisien Reliabilitas

### Teknik Belah Dua (*Split-Half*)

Tabel 3 adalah hasil penyekoran terhadap 8 butir soal Mekanika Teknik dari 10 orang siswa. Tabel ini merupakan tabel persiapan untuk menghitung koefisien reliabilitas dengan teknik belah dua. Berdasarkan Tabel 11. dapat disusun tabel baru yang berisikan skor-skor untuk butir-butir soal bernomor ganjil dan genap seperti tampak pada tabel 12. Butir soal bernomor ganjil adalah butir-butir nomor: 1, 3, 5, dan 7, dan butir-butir bernomor genap 2, 4, 6, dan 8.

