

BAB I

PENDAHULUAN

A. Evaluasi

1) Pengertian Evaluasi

Selain dari istilah evaluasi (*evaluation*) dan asesmen (*assessment*) dikenal pula beberapa istilah lainnya yaitu pengukuran (*measurement*), tes (*test*) dan testing. Diantara ketiga istilah tersebut, tes merupakan istilah yang paling akrab dengan pendidik. Hal tersebut disebabkan karena Tes prestasi belajar (*Achievement test*) seringkali dijadikan sebagai satu-satunya alat untuk menilai hasil belajar siswa. Padahal tes sebenarnya hanya merupakan salah satu alat ukur hasil belajar. Tes prestasi belajar (*Achievement test*) sering kali dipertukarkan pemakaiannya oleh pendidik dengan konsep pengukuran hasil belajar (*measurement*). Dengan demikian, perlu adanya upaya untuk memperkenalkan kepada pendidik tentang pengertian dan esensi tentang konsep evaluasi, asesmen, tes dan pengukuran yang sesungguhnya. Diantara peristilahan tersebut, Asesmen merupakan istilah yang belum dikenal secara umum. Para pendidik seringkali salah dalam menafsirkan makna asesmen yang sesungguhnya. Istilah asesmen perlu diperkenalkan kepada pendidik. Hal ini disebabkan karena asesmen telah menjadi khazanah peristilahan dalam dunia pendidikan kita. Selain dari itu, pemahaman tentang asesmen juga dapat mendukung keberhasilan pendidik dalam melaksanakan praktek penilaian pembelajaran di kelas.

Evaluasi (penilaian) dan assessment (penilaian) memiliki makna yang berbeda walaupun memiliki arti yang sama. Secara umum bisa dikatakan evaluasi ruang lingkupnya lebih luas dibandingkan penilaian. Menurut Guba dan Lincoln (1985), mendefinisikan evaluasi sebagai “suatu proses untuk menggambarkan evaluan (orang yang dievaluasi) dan menimbang makna dan nilainya”. Six (1980) juga berpendapat “evaluasi adalah suatu proses dimana pertimbangan atau keputusan suatu nilai dibuat dari berbagai pengamatan, latar belakang serta pelatihan dari evaluator”.

Dari dua pendapat di atas tentang evaluasi ini, dapat kita peroleh gambaran bahwa evaluasi adalah suatu proses yang sistematis dan berkelanjutan untuk menentukan kualitas (nilai dan arti) dari pada sesuatu,

berdasarkan pertimbangan dan kriteria tertentu untuk membuat suatu keputusan (Arifin, 2012). Kriteria yang digunakan dapat saja berasal dari apa yang dievaluasi itu sendiri (internal), tetapi bisa juga berasal dari luar apa yang dievaluasi (eksternal), baik yang bersifat kuantitatif maupun kualitatif.

Pada dasarnya evaluasi merupakan suatu bagian penting dalam pembelajaran yang memberikan keuntungan bagi pendidik dan siswa, serta sebagai alat untuk memonitor apakah pembelajaran telah dipahami oleh siswa, mengetahui kemajuan siswa sebagai individu maupun sebagai kelompok, untuk merekam apa yang telah siswa capai, dan untuk membantu siswa dalam belajar.

Evaluasi menurut Kumano (2001) merupakan penilaian terhadap data yang dikumpulkan melalui kegiatan asesmen. Sementara itu menurut Calongesi (1995) evaluasi adalah suatu keputusan tentang nilai berdasarkan hasil pengukuran. Sejalan dengan pengertian tersebut, Zainul dan Nasution (2001) menyatakan bahwa evaluasi dapat dinyatakan sebagai suatu proses pengambilan keputusan dengan menggunakan informasi yang diperoleh melalui pengukuran hasil belajar, baik yang menggunakan instrumen tes maupun non tes.

Secara garis besar dapat dikatakan bahwa evaluasi adalah pemberian nilai terhadap kualitas sesuatu. Selain dari itu, evaluasi juga dapat dipandang sebagai proses merencanakan, memperoleh, dan menyediakan informasi yang sangat diperlukan untuk membuat alternatif-alternatif keputusan. Dengan demikian, Evaluasi merupakan suatu proses yang sistematis untuk menentukan atau membuat keputusan sampai sejauhmana tujuan-tujuan pengajaran telah dicapai oleh siswa (Purwanto, 2002).

Cronbach (Harris, 1985) menyatakan bahwa evaluasi merupakan pemeriksaan yang sistematis terhadap segala peristiwa yang terjadi sebagai akibat dilaksanakannya suatu program. Sementara itu Arikunto (2003) mengungkapkan bahwa evaluasi adalah serangkaian kegiatan yang ditujukan untuk mengukur keberhasilan program pendidikan. Tayibnapis (2000) dalam hal ini lebih meninjau pengertian evaluasi program dalam konteks tujuan yaitu sebagai proses menilai sampai sejauhmana tujuan pendidikan dapat dicapai. Berdasarkan tujuannya, terdapat pengertian evaluasi sumatif dan evaluasi formatif. Evaluasi formatif dinyatakan sebagai upaya untuk memperoleh feedback perbaikan program, sementara itu evaluasi sumatif merupakan upaya

menilai manfaat program dan mengambil keputusan (Lehman, 1990).

Jadi evaluasi adalah proses penentuan ukuran atau nilai dari data yang terkumpul (Kirkendall, Grubel dan Johnson. 1980) . Evaluasi meliputi kedua tes dan pengukuran. Hubungan timbal balik antara tes, pengukuran dan evaluasi dapat diperhatikan. Evaluasi mencakup semuanya

B. Assessment

1) Pengertian Assessment

Assessment atau disebut juga dengan penilaian adalah suatu penerapan dan penggunaan berbagai cara dan alat untuk mendapatkan serangkaian informasi tentang hasil belajar dan pencapaian kompetensi dari peserta didik. Yang pada dasarnya, assessment yaitu istilah lain dari penilaian. Istilah Assessment sangat berkaitan dengan istilah evaluasi yaitu metode untuk mendapatkan hasil belajar siswa. Sehingga proses assessment ini dilaksanakan bertujuan untuk mengetahui sejauh apa penguasaan belajar dari para peserta didik. Pengertian lain dari assesment yaitu proses untuk memperoleh data atau informasi dari proses pembelajaran dan juga memberikan umpan balik terhadap pendidik ataupun kepada peserta didik.

Istilah asesmen (*assessment*) diartikan oleh Stiggins (1994) sebagai penilaian proses, kemajuan, dan hasil belajar siswa (*outcomes*). Sementara itu asesmen diartikan oleh Kumano (2001) sebagai “ *The process of Collecting data which shows the development of learning*”. Dengan demikian dapat disimpulkan bahwa asesmen merupakan istilah yang tepat untuk penilaian proses belajar siswa. Namun meskipun proses belajar siswa merupakan hal penting yang dinilai dalam asesmen, faktor hasil belajar juga tetap tidak dikesampingkan.

Gabel (1993) mengkategorikan asesmen ke dalam kedua kelompok besar yaitu asesmen tradisional dan asesmen alternatif. Asesmen yang tergolong tradisional adalah tes benar-salah, tes pilihan ganda, tes melengkapi, dan tes jawaban terbatas. Sementara itu yang tergolong ke dalam asesmen alternatif (non-tes) adalah essay/uraian, penilaian praktek, penilaian proyek, kuesioner, inventori, daftar Cek, penilaian oleh teman sebaya/sejawat, penilaian diri (*self assessment*), portofolio, observasi, diskusi dan interviu (wawancara).

Wiggins (1984) menyatakan bahwa asesmen merupakan sarana yang

secara kronologis membantu pendidik dalam memonitor siswa. Oleh karena itu, maka Popham (1995) menyatakan bahwa asesmen sudah seharusnya merupakan bagian dari pembelajaran, bukan merupakan hal yang terpisahkan. Resnick (1985) menyatakan bahwa pada hakikatnya asesmen menitikberatkan penilaian pada proses belajar siswa. Berkaitan dengan hal tersebut, Marzano *et al.* (1994) menyatakan bahwa dalam mengungkap penguasaan konsep siswa, asesmen tidak hanya mengungkap konsep yang telah dicapai, akan tetapi juga tentang proses perkembangan bagaimana suatu konsep tersebut diperoleh. Dalam hal ini asesmen tidak hanya dapat menilai hasil dan proses belajar siswa, akan tetapi juga kemajuan belajarnya.

2) Fungsi Assessment

Dalam kegiatan belajar mengajar, assessment atau penilaian mempunyai peranan yang penting. Karena assessment mempunyai dua fungsi yakni fungsi formatif dan fungsi sumatif.

- **Fungsi Formatif**

Fungsi formatif yaitu dimana assessment dipakai untuk memberikan umpan balik atau feedback terhadap para pendidik untuk dijadikan dasar ketika memperbaiki dan membenarkan proses pembelajaran dan juga mengadakan remedial untuk para peserta didik.

- **Fungsi Sumatif**

Adalah fungsi sebagai penentu nilai belajar siswa dalam satu mata pelajaran tertentu, sehingga selanjutnya bisa dijadikan bahan memberikan laporan, menentukan kenaikan kelas serta menentukan lulus atau tidaknya siswa.

3) Tujuan Assessment

Menurut Chittenden (1994) menyatakan bahwa tujuan penilaian “assessment purpose” adalah “keeping track”, checking up, finding out and summing up

- **Keeping Track**

Keeping track yaitu untuk menelusuri dan melacak proses belajar peserta didik yang sesuai dengan rencana pelaksanaan pembelajaran yang sudah diterapkan. Maka dari itu pendidik wajib mengumpulkan data dan informasi dalam kurun waktu tertentu dari berbagai jenis dan teknik penilaian untuk mendapatkan gambaran suatu pencapaian kemajuan belajar peserta didik.

- **Checking Up**

Checking Up adalah untuk mengecek pencapaian kemampuan peserta didik didalam proses belajar dan kekurangan-kekurangan peserta didik ketika mengikuti proses pembelajaran. Dengan kata lain, pendidik penting melaksanakan penilaian untuk tahu bagian mana dari materi yang telah dikuasai peserta didik dan bagian dari materi yang belum dikuasai.

- **Finding Out**

Finding Out adalah mencari, menemukan dan mendeteksi kekurangan kesalahan atau kelemahan peserta didik didalam proses belajar, sehingga pendidik bisa dengan tanggap mencari alternatif penyelesaiannya.

- **Summing Up**

Summing Up adalah cara untuk menyimpulkan tingkat penguasaan siswa terhadap kompetensi yang sudah ditetapkan. Hasil dari penyimpulan ini bisa digunakan pendidik dalam menyusun laporan kemajuan belajar ke berbagai pihak yang saling membutuhkan.

4) Jenis-Jenis Assessment

Adapun jenis-jenis assessment yang sering dipakai, antara lain tes tertulis yang disajikan kepada siswa untuk menjawabnya yaitu:

- **Performance Assessment**

Performance assessment yaitu jenis assessment yang menyuruh para peserta didik untuk melakukan demonstrasi bersamaan mengaplikasikan pengetahuan dalam berbagai situasi yang dikehendaki.

- **Penilaian Portofolio Dan Penilaian Proyek**

Penilaian proyek ini adalah tugas dalam bentuk suatu investigasi diawali dari pengumpulan selanjutnya pengorganisasian dan evaluasi hingga dengan penyajian data.

- **Product Assessment Dan Self Assessment**

Product Assessment adalah penilaian keterampilan dengan cara membuat suatu produk tertentu. Sedangkan Self Assessment dilaksanakan sendiri oleh peserta didik atau pendidik yang bersangkutan untuk kepentingan pengelolaan Kegiatan Belajar Mengajar di tingkat kelas, terakhir, jenis

assessment juga dapat berbentuk penilaian sikap dan penilaian dengan basis kelas.

5) Contoh Assessment

Contoh dari assessment adalah pemberian tugas ketika belajar atau adanya UAS. Penilaian dilakukan oleh pendidik berdasarkan assessment berupa lembar jawaban tugas atau ujian. Pendidik memberikan nilai, bisa berupa angka atau huruf terhadap hasil pekerjaan peserta didik. Setelah semua hasil assessment dinilai/diukur maka memasuki tahap evaluasi. Semua hasil peserta didik diklasifikasikan, ada yang lulus atau tidak lulus

C. Tes

1. Pengertian Tes

Tes (*test*) merupakan suatu alat penilaian dalam bentuk tulisan untuk mencatat atau mengamati prestasi siswa yang sejalan dengan target penilaian (Jacobs & Chase, 1992; Alwasilah, 1996). Jawaban yang diharapkan dalam tes menurut Sudjana dan Ibrahim (2001) dapat secara tertulis, lisan, atau perbuatan. Menurut Zainul dan Nasution (2001) tes didefinisikan sebagai pertanyaan atau tugas atau seperangkat tugas yang direncanakan untuk memperoleh informasi tentang suatu atribut pendidikan atau suatu atribut psikologis tertentu. Setiap butir pertanyaan atau tugas tersebut mempunyai jawaban atau ketentuan yang dianggap benar. Dengan demikian apabila suatu tugas atau pertanyaan menuntut harus dikerjakan oleh seseorang, tetapi tidak ada jawaban atau cara pengerjaan yang benar dan salah maka tugas atau pertanyaan tersebut bukanlah tes.

Tes merupakan salah satu upaya pengukuran terencana yang digunakan oleh pendidik untuk mencoba menciptakan kesempatan bagi siswa dalam memperlihatkan prestasi mereka yang berkaitan dengan tujuan yang telah ditentukan (Calongesi, 1995). Tes terdiri atas sejumlah soal yang harus dikerjakan siswa. Setiap soal dalam tes menghadapkan siswa pada suatu tugas dan menyediakan kondisi bagi siswa untuk menanggapi tugas atau soal tersebut.

Tes menurut Arikunto dan Jabar (2004) merupakan alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dengan menggunakan cara atau aturan yang telah ditentukan. Dalam hal ini harus dibedakan pengertian antara tes, testing, testee, tester. Testing adalah saat pada waktu tes tersebut dilaksanakan (saat pengambilan tes). Sementara itu Gabel (1993) menyatakan bahwa testing menunjukkan proses pelaksanaan tes. Testee adalah responden yang mengerjakan tes. Mereka inilah yang akan dinilai atau diukur kemampuannya. Sedangkan Tester adalah seseorang yang disertai tugas untuk melaksanakan pengambilan tes kepada responden.

Dewasa ini tes masih merupakan alat evaluasi yang umum digunakan untuk mengukur keberhasilan siswa dalam mencapai tujuan pendidikan dan pengajaran (Subekti & Firman, 1989). Menurut Faisal (1982), seringkali skor tes ini dipergunakan sebagai satu-satunya indikator

dalam menilai penguasaan konsep, efektivitas metode belajar, pendidik serta aspek lainnya terhadap siswa di dalam praktek pendidikan. Padahal dengan mempergunakan tes, aspek kemampuan afektif siswa kurang terukur, sehingga sangatlah penting untuk tidak membuat generalisasi kemampuan siswa hanya melalui tes saja.

Wayan Nurkencana (1990)! tes adalah suatu cara untuk mengadakan penilaian yang berbentuk suatu tugas yang harus dikerjakan anak atau sekelompok anak sehingga menghasilkan suatu nilai tentang tingkah laku atau prestasi anak tersebut yang kemudian dapat dibandingkan dengan nilai yang dicapai oleh anak-anak lain atau standar yang telah ditetapkan.

Menurut Riduwan (2006) tes sebagai instrumen pengumpulan data adalah serangkaian pertanyaan/latihan yang digunakan untuk mengukur ketrampilan pengetahuan/intelektual. Menurut Allen Philips (1979) , *test is commonly defined as a tool or instrument of measurement that is used to obtain data about a specific trait or characteristic of an individual or group.* (Test biasanya diartikan sebagai alat atau instrumen dari pengukuran yang digunakan untuk memperoleh data tentang suatu karakteristik atau ciri yang spesifik dari individu atau kelompok.) Menurut Rusli Lutan (2000) tes adalah sebuah instrument yang dipakai untuk memperoleh informasi tentang seseorang atau obyek.

Overton Terry (2008) *test is a method to determine a student's ability to complete certain tasks or demonstrate mastery of a skill or knowledge of content. Some types would be multiple choice tests or a weekly spelling test. While it commonly used interchangeably with assesment, or even evaluation, it can be distinguished by the fact that a test is one form of an assessment.* Tes adalah suatu metode untuk menentukan kemampuan siswa menyelesaikan sejumlah tugas tertentu atau mendemonstrasikan penguasaan suatu keterampilan atau pengetahuan pada suatu materi pelajaran. beberapa tipe tes misalnya tes pilihan ganda atau tes mengeja mingguan. seringkali penggunaannya tertukar dengan asesmen! atau bahkan evaluasi (penilaian)! yang mana sebenarnya tes dapat dengan mudah dibedakan berdasarkan kenyataan bahwa tes adalah salah satu bentuk asesmen.

Istilah tes berasal dari bahasa latin "*testum*" yang berarti sebuah piring atau jambangan dari tanah liat. Istilah tes ini kemudian dipergunakan dalam lapangan psikologi dan selanjutnya hanya dibatasi sampai pada metode psikologi, yaitu suatu cara untuk menyelidiki seseorang. Penyelidikan tersebut dilakukan mulai dari pemberian suatu tugas kepada seseorang atau untuk menyelesaikan suatu masalah tertentu (Arifin, 2012: 6).

2. Jenis-Jenis Tes

Berbagai macam bentuk tes, antara lain ;

- 1) Tes Betul-Salah (*True False*)
- 2) Tes Pilihan Ganda (*Multiple Choice*)
- 3) Tes Menjodohkan (*Matching*)
- 4) Tes Analisa Hubungan (*Relationship Analysis*)

a) Dari segi fungsi tes di sekolah

1) Tes Formatif

Tes Formatif, yaitu tes yang diberikan untuk memonitor kemajuan belajar selama proses pembelajaran berlangsung. Tes ini diberikankan dalam tiap satuan unit pembelajaran. Manfaat tes formatif bagi peserta didik adalah :

- a) Untuk mengetahui apakah peserta didik sudah menguasai materi dalam tiap unit pembelajaran.
- b) Merupakan penguatan bagi peserta didik.

- c) Merupakan usaha perbaikan bagi siswa, karena dengan tes formatif peserta didik mengetahui kelemahan-kelemahan yang dimilikinya.
- d) Peserta didik dapat mengetahui bagian dari bahan yang mana yang belum dikuasainya.

2) Tes Summatif

Tes sumatif diberikan dengan maksud untuk mengetahui penguasaan atau pencapaian peserta didik dalam bidang tertentu. Tes sumatif dilaksanakan pada tengah atau akhir semester.

3) Tes Penempatan

Tes penempatan adalah tes yang diberikan dalam rangka menentukan jurusan yang akan dimasuki peserta didik atau kelompok mana yang paling baik ditempati atau dimasuki peserta didik dalam belajar.

4) Tes Diagnostik

Tes diagnostik adalah tes yang digunakan untuk mendiagnosis penyebab kesulitan yang dihadapi seseorang baik dari segi intelektual, emosi, fisik dan lain-lain yang mengganggu kegiatan belajarnya.

3. Ciri-ciri Tes Yang Baik

Sebuah tes dikatakan baik jika memenuhi persyaratan:

- 1) Bersifat valid atau memiliki validitas yang cukup tinggi. Suatu tes dikatakan valid bila tes itu isinya dapat mengukur apa yang seharusnya diukur, artinya alat ukur yang digunakan tepat
- 2) Bersifat reliable, atau memiliki reliabilitas yang baik. Reliabilitas sering diartikan dengan keterandalan. Suatu tes dikatakan reliabel jika tes itu diberikan berulang-ulang memberikan hasil yang sama.
- 3) Bersifat praktis atau memiliki kepraktisan. Tes memiliki sifat kepraktisan artinya praktis dari segi perencanaan, pelaksanaan tes dan memiliki nilai ekonomi tetapi harus tetap mempertimbangkan kerahasiaan tes.

Namun syarat minimum yang harus dimiliki oleh sebuah tes yang baik adalah valid dan reliable.

4. Langkah-langkah Pengembangan Tes

Ada enam tahap dalam merencanakan dan menyusun tes agar diperoleh tes yang baik, yaitu:

1) Pengembangan spesifikasi tes

Spesifikasi tes adalah suatu ukuran yang menunjukkan keseluruhan kualitas tes dan ciri-ciri yang harus dimiliki oleh tes yang akan dikembangkan. Hal yang perlu diperhatikan adalah :

- a) Menentukan tujuan, tujuan pembelajaran yang baik hendaklah berorientasi kepada peserta didik, bersifat menguraikan hasil belajar, harus jelas dan dapat dimengerti, mengandung kata kerja yang jelas (kata kerja operasional), serta dapat diamati dan dapat di ukur.
- b) Menyusun kisi-kisi soal, penyusunan kisi-kisi soal bertujuan untuk merumuskan setepat mungkin ruang lingkup, tekanan dan bagian-bagian tes sehingga perumusan tersebut dapat menjadi petunjuk yang efektif bagi penyusun tes.
- c) Memilih tipe soal, dalam memilih tipe soal perlu diperhatikan kesesuaian antara tipe soal dengan materi, tujuan evaluasi, scoring, pengelolaan hasil evaluasi, penyelenggaraan tes, serta ketersediaan dana dan kepraktisan.
- d) Merencanakan tingkat kesukaran soal, untuk soal objektif dapat diketahui melalui uji coba atau dapat juga diperkirakan berdasarkan berat ringannya beban penyelesaian soal tersebut
- e) Merencanakan banyak soal
- f) Merencanakan jadwal penerbitan soal

2) Penulisan soal

- 3) Penelaahan soal, yaitu menguji validitas soal yang bertujuan untuk mencermati apakah butir-butir soal yang disusun sudah tepat untuk mengukur tujuan pembelajaran yang sudah dirumuskan, ditinjau dari segi isi/materi, kriteria dan psikologis.
- 4) Pengujian butir-butir soal secara empiris, kegiatan ini sangat penting jika soal yang dibuat akan dibakukan.
- 5) Penganalisisan hasil uji coba.
- 6) Pengadministrasian soal

5. Menganalisis Tes

Menganalisis instrument (alat evaluasi) bertujuan untuk mengetahui apakah alat ukur yang digunakan atau yang akan digunakan sudah memenuhi syarat-syarat sebagai alat ukur yang baik, tepat mengukur sesuatu sesuai tujuan

yang telah dirumuskan. Sebuah instrument dikatakan baik jika memenuhi syarat validitas, reliabilitas dan bersifat praktis.

1. **Validitas Tes**

Suatu tes dikatakan valid jika tes itu dapat mengukur apa yang seharusnya diukur. Valid disebut juga sahih, terandalkan atau tepat. Tes hasil belajar yang valid, harus dapat menggambarkan hasil belajar yang diukur

Macam-macam validitas

a) Validitas isi (content validity)

Validitas isi sering juga disebut validitas logis atau validitas rasional. Validitas isi dapat dianalisis dengan bantuan kisi-kisi tes dan pedoman penelaahan butir soal.

Penelaahan butir soal secara umum ditinjau dari tiga aspek yaitu:

- 1) Aspek materi
- 2) Aspek bahasa
- 3) Aspek konstruksi

b) Validitas ramalan (predictive validity)

Suatu tes dikatakan memiliki validitas ramalan, apabila hasil pengukuran yang dilakukan dengan tes itu dapat digunakan untuk meramalkan, atau tes itu mempunyai daya prediksi yang cukup kuat. Untuk mengetahui apakah suatu tes hasil belajar dapat dinyatakan sebagai tes yang memiliki validitas ramalan dapat dilakukan dengan mengkorelasikan tes hasil belajar yang sedang diuji dengan kriteria yang ada.

c) Validitas bandingan (concurrent validity)

Suatu tes dikatakan memiliki validitas concurrent, apabila tes tersebut mempunyai kesesuaian dengan hasil pengukuran lain yang dilaksanakan saat itu. Misalnya, membandingkan hasil tes dari soal yang sedang dicari validitasnya dengan hasil tes dari soal standar. Jika terdapat korelasi yang positif antara kedua tes tersebut, berarti soal tes yang dibuat mempunyai validitas concurrent.

d) Construct validity (validitas konstruk)

Validitas konstruk artinya butir-butir soal dalam tes tersebut membangun setiap aspek berpikir seperti yang tercantum dalam tujuan pembelajaran yang telah dirumuskan. Penganalisisan validitas ini dapat dilakukan dengan jalan melakukan

pencocokan antara aspek berpikir yang dikehendaki diungkapkan oleh tujuan pembelajaran, yaitu melalui penelaahan butir-butir soal.

Meski terdapat beberapa jenis validitas, dalam periode terakhir validitas dianggap sebagai suatu konsep utuh, tidak dipilah-pilah sebagai jenis validitas.

2. Cara menentukan validitas instrumen

Validitas instrument dapat diketahui dengan mencari korelasi hasil instrument dengan dengan kriterium atau melakukan analisis butir. Apabila data yang digunakan adalah data interval maka dapat digunakan rumus Product Moment Korelasi, sebagai berikut :

$$r_{xy} = \frac{N \sum xy(x)(\sum y)}{\sqrt{\{N \sum x^2 - (\sum x)^2\} \{N \sum y^2 - (\sum y)^2\}}}$$

r_{xy} = koefisien korelasi

n = Jumlah sampel

X = Tes

Y = Re tes (Sudjana , 2009).

3. Cara menentukan validitas tiap butir soal

Tinggi rendahnya validitas soal secara keseluruhan berhubungan dengan validitas tiap butir soal. Validitas butir soal dapat dicari dalam hubungannya dengan skor total tiap individu yang ikut serta dalam evaluasi. Langkah-langkah yang ditempuh sebagai berikut :

- a) Skor suatu instrument dengan baik dan teliti Untuk individu yang benar diberi angka 1, sedangkan yang salah diberi angka nol.
- b) Jumlahkan skor total untuk tiap individu.
- c) Gunakan rumus product moment correlation atau korelasi biserial.

4. Reliabilitas

Suatu alat ukur dikatakan reliabel, apabila alat ukur itu dicobakan kepada objek yang sama secara berulang-ulang maka hasilnya akan tetap sama, konsisten, stabil atau relatif sama.

- Faktor-faktor yang mempengaruhi reliabilitas
 - a) Konstruksi item yang tidak tepat, sehingga tidak dapat mempunyai daya pembeda yang kuat.

- b) Panjang/pendeknya suatu instrumen
- c) Evaluasi yang subjektif akan menurunkan reliabilitas
- d) Ketidaktepatan waktu yang diberikan
- e) Kemampuan yang ada dalam kelompok
- f) Luas/tidaknya sampel yang diambil.
- Teknik pengujian reliabilitas tes hasil belajar
 - a. Bentuk objektif

1) Metode Belah dua

Dalam pelaksanaannya, seorang penilai hanya melakukan ujian satu kali terhadap sejumlah peserta, sehingga tidak ada pengaruh dari instrumen yang terdahulu. Jumlah butir soal yang diberikan harus genap sehingga dapat dibagi dua dan tiap kelompok mempunyai jumlah butir yang sama. Koefisien reliabilitas akan menunjukkan internal konsistensi dari pada butir soal dalam keseluruhan instrumen. Cara membelah dua instrumen tersebut dapat dilakukan dengan cara nomor genap dan ganjil, awal dan akhir. Untuk menentukan reliabilitas kedua bagian instrumen tersebut dapat digunakan Product Moment Coorelation, sedangkan untuk mencari reliabilitas keseluruhan instrumen dapat digunakan rumus Spearman Brown, sebagai berikut :

2) Metode Ulangan

Pelaksanaannya dilakukan dua kali kepada sejumlah subjek yang sama, dalam waktu yang berbeda. Reliabilitas metode ulangan ini untuk melihat bagaimana stabilnya skor setiap individu apabila dilakukan pengujian dalam waktu yang berbeda, dengan kondisi dan perlengkapan yang sama/ hampir bersamaan. Rumus yang digunakan untuk menentukan metode ulangan ini adalah Product Moment Correlation.

3) Metode Bentuk Paralel

Bentuk ini dapat digunakan untuk memperkirakan reliabilitas dari semua tipe, tetapi koefisien yang dihasilkan hanya menggambarkan ekivalensi antara kedua instrumen. Tidak akan menunjukkan ekivalensi dalam kesukaran butir dan isi. Kedua bentuk instrumen yang diberikan mengukur hal yang sama, dengan memiliki tingkat kesukaran yang sama, pengetahuan dan keterampilan yang sama dengan sistematika yang tidak berbeda antara kedua bentuk instrumen tersebut, tetapi dalam bentuk pertanyaan yang berbeda. Rumus yang dapat digunakan

untuk menentukan reliabilitas instrumen dalam bentuk paralel ini adalah product moment correlation dan Rank order correlation.

4) Bentuk essay

Rumus yang digunakan untuk mengukur reliabilitas tes berbentuk uraian dinamakan rumus Alpha, yaitu :

Interpretasi terhadap koefisien reliabilitas tes pada umumnya digunakan patokan sebagai berikut:

$0,80 < r_{11}$	1,00	reliabilitas sangat tinggi
$0,60 < r_{11}$	0,80	reliabilitas tinggi
$0,40 < r_{11}$	0,60	reliabilitas sedang
$0,20 < r_{11}$	0,40	reliabilitas rendah
$0,00 < r_{11}$	0,20	reliabilitas sangat rendah

Nilai r yang diperoleh dibandingkan dengan r_{tabel} . Jika $r_{\text{hitung}} > r_{\text{tabel}}$ maka dapat disimpulkan bahwa soal tes reliabel

D. Pengukuran

1) Pengertian Pengukuran

Seorang pendidik dalam melakukan evaluasi dalam pembelajaran, salah satu langkah yang harus dilakukannya adalah melakukan pengukuran. Dalam penilaian pendidikan, pendidik harus mengetahui standar penilaian yang telah ditetapkan oleh pemerintah sebagai acuan dasar, sehingga pendidik mampu melakukan pengukuran sesuai dengan apa yang hendak diukur dalam bidang pendidikan. Pengukuran akan dapat dilakukan dengan baik apabila pendidik mengetahui objek apa yang diukur, sehingga pendidik dapat menentukan instrumen dengan tepat dalam pengukuran, yaitu memiliki derajat validitas dan reliabilitas yang tinggi.

Menurut Cangelosi (1995) yang dimaksud dengan pengukuran (*Measurement*) adalah suatu proses pengumpulan data melalui pengamatan empiris untuk mengumpulkan informasi yang relevan dengan tujuan yang telah ditentukan. Dalam hal ini pendidik menaksir prestasi siswa dengan membaca atau mengamati apa saja yang dilakukan siswa, mengamati kinerja mereka, mendengar apa yang mereka katakan, dan menggunakan indera mereka seperti melihat, mendengar, menyentuh, mencium, dan merasakan. Menurut Zainul dan Nasution (2001) pengukuran memiliki dua karakteristik utama yaitu: 1)

penggunaan angka atau skala tertentu; 2) menurut suatu aturan atau formula tertentu.

Measurement (pengukuran) merupakan proses yang mendeskripsikan *performance* siswa dengan menggunakan suatu skala kuantitatif (system angka) sedemikian rupa sehingga sifat kualitatif dari *performance* siswa tersebut dinyatakan dengan angka-angka (Alwasilah *et al.*1996). Pernyataan tersebut diperkuat dengan pendapat yang menyatakan bahwa pengukuran merupakan pemberian angka terhadap suatu atribut atau karakter tertentu yang dimiliki oleh seseorang, atau suatu obyek tertentu yang mengacu pada aturan dan formulasi yang jelas. Aturan atau formulasi tersebut harus disepakati secara umum oleh para ahli (Zainul & Nasution, 2001). Dengan demikian, pengukuran dalam bidang pendidikan berarti mengukur atribut atau karakteristik peserta didik tertentu. Dalam hal ini yang diukur bukan peserta didik tersebut, akan tetapi karakteristik atau atributnya. Senada dengan pendapat tersebut, Secara lebih ringkas, Arikunto dan Jabar (2004) menyatakan pengertian pengukuran (*measurement*) sebagai kegiatan membandingkan suatu hal dengan satuan ukuran tertentu sehingga sifatnya menjadi kuantitatif. Pada Tabel 1. diberikan contoh standar kriteria yang digunakan untuk mengukur kemampuan siswa dalam menyusun laporan praktikum.

Measurement dapat dilakukan dengan cara tes atau non-tes. Amalia (2003) mengungkapkan bahwa tes terdiri atas tes tertulis (*paper and pencil test*) dan tes lisan. Sementara itu alat ukur non-tes terdiri atas pengumpulan kerja siswa (portofolio), hasil karya siswa (produk), penugasan (proyek), dan kinerja (*performance*). Endang Purwati (2008) mengartikan pengukuran sebagai kegiatan atau upaya yang dilakukan untuk memberikan angka-angka pada suatu gejala atau peristiwa, atau benda, sehingga hasil pengukuran akan selalu berupa angka. Masidjo (1995) mengartikan pengukuran sifat suatu objek adalah suatu kegiatan menentukan kuantitas suatu objek melalui aturan-aturan tertentu sehingga kuantitas yang diperoleh benar-benar mewakili sifat dari suatu objek yang dimaksud.

Cangelosi (1991) mengartikan pengukuran adalah proses pengumpulan data melalui pengamatan empiris yang digunakan untuk mengumpulkan informasi yang relevan dengan tujuan yang telah ditentukan.

Wiersma & Jurs (dalam Djaali & Pudji Muljono, 2007) mengatakan bahwa pengukuran adalah penilaian numeric pada fakta-fakta dari objek yang hendak diukur menurut criteria atau satuan-satuan tertentu. Jadi pengukuran bisa diartikan sebagai proses memasangkan fakta-fakta suatu objek dengan fakta-fakta satuan tertentu.