



EVALUASI

EVALUASI HASIL BELAJAR

HASIL BELAJAR

Penerbit
Universitas Hamzanwadi Press



Penulis:

**Muhammad Ropii
Muh. Fahrurrozzi**

Evaluasi Hasil Belajar

Muhammad Ropii
Muh. Fahrurrozi

Universitas Hamzanwadi Press
2017

Evaluasi Hasil Belajar

Penulis:

Muhammad Ropii
Muh. Fahrurrozi

ISBN978-602-50844-5-4

Editor :

SyukrulHamdi

Layout:

M. Marzuki

Penerbit :

Universitas Hamzanwadi Press

Redaksi:

Jln. TGKH Muhammad Zainuddin Abdul Madjid No. 132 Pancor
Selong Lombok Timur NTB 83612

Cetakan Pertama Oktober 2017

Hak Cipta © Dilindungi oleh Undang-Undang

Dilarang keras memperjualbelikan seluruh isi buku ini tanpa izin tertulis dari penerbit

KATA PENGANTAR

Puji dan Syukur sedalam-dalamnya atas segala limpahan karunia dan nikmat Allah yang tiada ternilai dan tiada pernah putus diberikan kepada hamba-hambaNya. Nikmat dan karunia itu pula yang menjadi kekuatan sehingga penyusunan buku ini dapat terselesaikan.

Penulis menyadari bahwa ide-ide, bimbingan dan sumbangan saran dari berbagai pihak telah memberikan kontribusi yang besar bagi penyusun buku ini. Oleh sebab itu dengan kerendahan hati penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan perhatian dan bantuan untuk penyelesaian buku ini.

Terima kasih penulis sampaikan kepada kemenristek dikti yang sudah memberikan bantuan hibah Revitalisasi LPTK.

Terima kasih penulis sampaikan kepada semua dosen Program Studi Pendidikan Ekonomi dan Ketua Program Studi Pendidikan Ekonomi, Dr. Muh. Fahrurrozi, MM yang telah memberikan bimbingan, wawasan pengetahuan, nuansa-nuansa baru, dan partisipasi dalam penulisan buku ini. Berkenaan dengan itu, penulis mengucapkan terima kasih kepada rekan-rekan sejawat yang telah membantu menyelesaikan buku ini.

Terima kasih penulis sampaikan kepada Rektor Universitas Hamzanwadi, Dr. Ir. Hj. Sitti Rohmi Djalilah, M. Pd., yang secara konsisten terus berupaya mengembangkan kemampuan sumber daya manusia Universitas Hamzanwadi, khususnya melalui penyediaan fasilitas penulisan buku bagi tenaga pengajar Universitas Hamzanwadi.

Terimakasih yang sedalamnya kepada Dekan FKIP Universitas Hamzanwadi, Abdullah Muzakkar, M.Si yang telah memberikan kesempatan kepada penulis untuk mengembangkan diri dalam berbagai bentuk kegiatan, salah satunya diantaranya di dalam kegiatan penulisan buku bagi dosen FKIP Universitas Hamzanwadi.

Apabila di dalam buku ini masih banyak kekurangan dan kesalahan penulis benar-benar mengakuinya. Oleh karena itu, kritik konstruktif demi perbaikan buku ini sangat diharapkan. Semua jenis bantuan dan kritikan konstruktif dari berbagai pihak yang disebutkan di atas, atau pihak-pihak lain yang tidak dapat disebutkan di dalam lembaran ini, penulis ucapkan ribuan terima kasih. Mudah-mudahan apa yang telah diberikan menjadi sesuatu yang bermanfaat dan bernilai ibadah di hadapa Allah SWT.

Akhirnya, Semoga buku ini memberikan manfaat pada kita, khususnya bagi pembaca dalam upaya meningkatkan kompetensi. Amin.

DAFTAR ISI

Halaman Judul.....	i
Halaman Copyright	ii
Kata Pengantar.....	iii
Daftar Isi	v
Daftar Gambar	vi
Daftar Tabel.....	vii

BAB I. Konsep Evaluasi Pembelajaran

A. Pengertian Pengukuran, Penilaian Dan Evaluasi	1
B. Tujuan, Fungsi Dan Prinsip Evaluasi Pembelajaran	7

BAB II. Ruang lingkup, karakteristik dan pendekatan Evaluasi Pembelajaran

A. Ruang Lingkup Evaluasi Pembelajaran Dalam Perspektif Domain Hasil Belajar	21
B. Ruang Lingkup Evaluasi Pembelajaran Dalam Perspektif Sistem Pembelajaran	24
C. Ruang Lingkup Evaluasi Pembelajaran Berbasis Penilaian Proses Dan Hasil Belajar.....	26
D. Ruang Lingkup Evaluasi Pembelajaran Dalam Perspektif Penilaian Berbasis Kelas	27
E. Karakteristik Alat Ukur Yang Baik.....	32
F. Model-Model Evaluasi	36
G. Pendekatan Evaluasi.....	43

BAB III. Prosedur Pengembangan Evaluasi Pembelajaran

A. Perencanaan Evaluasi.....	46
B. Pelaksanaan Evaluasi	57
C. Pengolahan Data	60
D. Pelaporan Hasil Evaluasi	62
E. Penggunaan Hasil Evaluasi	66

BAB IV. Pengembangan Alat Evaluasi Jenis Tes

A. Pengembangan Tes Bentuk Uraian	69
B. Pengembangan Tes Objektif	83
C. Pengembangan Tes Lisan.....	88
D. Pengembangan Tes Tindakan (Performance Test)	90

BAB V. Pengembangan Alat Evaluasi Jenis Non Tes

A. Observasi (Observation)	91
B. Wawancara (Interview)	95
C. Skala Sikap (Attitude Scale).....	97
BAB VI. Penilaian Portofolio	
A. Pengertian Penilaian Portofolio	99
B. Tujuan Dan Fungsi Penilaian Portofolio	100
C. Prinsip-Prinsip Penilaian Portofolio.....	103
D. Karakteristik Penilaian Portofolio	105
E. Kelebihan Dan Kekurangan Penilaian Portofolio.....	107
F. Jenis Penilaian Portofolio.....	108
G. Tahap-Tahap Penilaian Portofolio.....	113
H. Bahan-Bahan Penilaian Portofolio	115
BAB VII. Teknik Pengolahan Hasil Evaluasi	
A. Cara Memberi Skor Mentah Untuk Tes Uraian.....	117
B. Cara Memberi Skor Mentah Untuk Tes Objektif	119
C. Skor Total (Total Score)	121
D. Penilaian Acuan Patokan (Pap)	122
E. Penilaian Acuan Norma (Pan)	126
BAB VIII. Analisis Kualitas Tes	
A. Validitas.....	129
B. Reliabilitas	136
C. Kepraktisan	141
BAB IX. Menganalisis Kualitas Butir Soal	
A. Tingkat Kesukaran Soal (Difficulty Index)	144
B. Daya Pembeda (Discriminating Power).....	148
C. Analisis Pengecoh	150
D. Efektifitas Fungsi Opsi	151
BAB X. Teknik Penentuan Nilai Akhir Dan Penyusunan Ranking	
A. Pentingnya Memanfaatkan Hasil Evaluasi	154
B. Manfaat Hasil Evaluasi	157
C. Faktor-Faktor Yang Turut Diperhitungkan Dalam Penilaian	162
D. Teknik Penentuan Nilai Akhir	163
E. Teknik Penyusunan Ranking	165

Daftar Gambar

Gambar 1.1 Hubungan Evaluasi Penilaian dan Tes.....	6
Gambar 1.2 Fungsi Penilaian.....	15
Gambar 1.3 Prinsip-Prinsip Umum Evaluasi.....	16
Gambar 2.1 Ruang Lingkup Evaluasi Pembelajaran	31
Gambar 2.2 Karakteristik Alat Ukur Yang Baik	33
Gambar 2.3 Pendekatan Evaluasi Pembelajaran	44
Gambar 3.1 Langkah Melakukan Analisis Silabus.....	49
Gambar 6.1 Tujuan Penilaian Portofolio.....	101
Gambar 6.2 Fungsi Penilaian Portofolio	102
Gambar 6.3 Prinsip Penilaian Portofolio	103
Gambar 6.4 Karakteristik Penilaian Portofolio	106
Gambar 6.5 Jenis Portofolio	108

Daftar Tabel

Tabel 3.1 Format Laporan Peserta Didik Dalam Mata Pelajaran	65
Tabel 3.2 Tingkatan Hasil Belajar Tidak Sama Dengan Tingkat Kelas	65
Tabel 4.1 Perbedaan Antara Tugas Baku Dengan Tugas Buatan Guru	74
Table 4.2 Kartu Telaah Soal Bentuk Uraian.....	82
Table 9.1 Pengujian Efektifitas	152
Tabel 10.1 Simpel Rank Dari 20 Orang Siswa	166

BAB I

KONSEP EVALUASI PEMBELAJARAN

A. PENGERTIAN PENGUKURAN, PENILAIAN DAN EVALUASI

Evaluasi merupakan salah komponen penting dan tahap yang harus ditempuh oleh guru untuk mengetahui keefektifan pembelajaran. Hasil yang diperoleh dapat dijadikan balikan (*feed-back*) bagi guru dalam memperbaiki dan menyempurnakan program dan kegiatan pembelajaran. Di sekolah, Anda sering mendengar bahwa guru sering memberikan ulangan harian, ujian akhir semester, ujian blok, tagihan, tes tertulis, tes lisan, tes tindakan, dan sebagainya. Istilah-istilah ini pada dasarnya merupakan bagian dari sistem evaluasi itu sendiri.

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 metode psikologi, yaitu suatu cara untuk menyelidiki seseorang. Penyelidikan tersebut dilakukan mulai dari pemberian suatu tugas kepada seseorang atau untuk menyelesaikan suatu masalah tertentu. Sebagaimana dikemukakan Sax (1980: 13) bahwa "*a test may be defined as a task or series of task used to obtain systematic observations presumed to be representative of educational or psychological traits or attributes*". (tes dapat didefinisikan sebagai tugas atau serangkaian tugas yang digunakan untuk memperoleh pengamatan-pengamatan sistematis, yang dianggap mewakili ciri atau aribut pendidikan atau psikologis). Istilah tugas dapat berbentuk soal atau perintah/suruhan lain yang harus dikerjakan oleh peserta didik. Hasil kuantitatif ataupun kualitatif dari pelaksanaan tugas itu digunakan untuk menarik simpulan-simpulan tertentu terhadap peserta didik.

S. Hamid Hasan (1988: 7) menjelaskan "tes adalah alat pengumpulan data yang dirancang secara khusus. Kekhususan tes dapat terlihat dari konstruksi butir (soal) yang dipergunakan". Rumusan ini lebih terfokus kepada tes sebagai alat pengumpul data. Memang pengumpulan data bukan hanya ada dalam prosedur penelitian, tetapi juga ada dalam prosedur evaluasi. Dengan kata lain, untuk mengumpulkan data evaluasi, guru memerlukan suatu alat, antara lain tes. Tes dapat berupa pertanyaan. Oleh sebab itu, jenis pertanyaan, rumusan pertanyaan, dan pola jawaban yang disediakan harus memenuhi suatu perangkat kriteria yang ketat. Demikian pula waktu yang disediakan untuk menjawab soal-soal serta administrasi penyelenggaraan tes diatur secara khusus pula. Persyaratan-persyaratan ini berbeda dengan alat pengumpul data lainnya.

Dengan demikian, tes pada hakikatnya adalah suatu alat yang berisi serangkaian tugas yang harus dikerjakan atau soal-soal yang harus dijawab oleh peserta didik untuk mengukur suatu aspek perilaku tertentu. Artinya, fungsi tes adalah sebagai alat ukur. Dalam tes prestasi belajar, aspek perilaku yang hendak diukur adalah tingkat kemampuan peserta didik dalam menguasai materi pelajaran yang telah disampaikan.

Ahman dan Glock dalam S. Hamid Hasan (1988: 9) menjelaskan '*in the last analysis measurement is only a part, although a very substantial part of evaluation. It provides information upon which an evaluation can be based... Educational measurement is the process that attempt to obtain a quantified representation of the degree to which a trait is possessed by a pupil*'. (dalam analisis terakhir, pengukuran hanya merupakan bagian, yaitu bagian yang sangat substansial dari evaluasi. Pengukuran menyediakan informasi, di mana evaluasi dapat didasarkan ... Pengukuran pendidikan adalah proses yang berusaha untuk mendapatkan representasi secara kuantitatif tentang sejauh mana suatu ciri yang dimiliki oleh peserta didik). Pendapat yang sama dikemukakan oleh Wiersma dan Jurs (1985), bahwa "*technically, measurement is the assignment of numerals to objects or events according to rules that give numeral quantitative meaning*". (secara teknis, pengukuran adalah pengalihan dari angka ke objek atau peristiwa sesuai dengan aturan yang memberikan makna angka secara kuantitatif).

Dengan demikian, dapat dikemukakan bahwa pengukuran adalah suatu proses atau kegiatan untuk menentukan kuantitas daripada sesuatu. Kata "sesuatu" bisa berarti peserta didik, guru, gedung sekolah, meja belajar, *white board*, dan sebagainya. Dalam proses pengukuran, tentu guru harus menggunakan alat ukur (tes atau non-tes). Alat ukur tersebut harus standar, yaitu memiliki derajat validitas dan reliabilitas yang tinggi. Dalam bidang pendidikan, psikologi, maupun variabel-variabel sosial lainnya, kegiatan pengukuran biasanya menggunakan tes. Dalam sejarah perkembangannya, aturan mengenai pemberian angka ini didasarkan pada teori pengukuran psikologi yang dinamakan *psychometric*. Namun demikian, boleh saja suatu kegiatan evaluasi dilakukan tanpa melalui proses pengukuran.

Istilah penilaian merupakan alih bahasa dari istilah *assessment*, bukan dari istilah *evaluation*. Dalam proses pembelajaran, penilaian sering dilakukan guru untuk memberikan berbagai informasi secara berkesinambungan dan menyeluruh tentang proses dan hasil yang telah dicapai peserta didik. Artinya, penilaian tidak hanya ditujukan pada penguasaan salah satu bidang tertentu saja, tetapi bersifat menyeluruh yang mencakup aspek pengetahuan, keterampilan, sikap dan nilai-nilai. Sementara itu, Anthony J. Nitko (1996: 4) menjelaskan "*assessment is a broad*

term defined as a process for obtaining information that is used for making decisions about students, curricula and programs, and educational policy". (penilaian adalah suatu proses untuk memperoleh informasi yang digunakan untuk membuat keputusan tentang peserta didik, kurikulum, program, dan kebijakan pendidikan).

Dalam hubungannya dengan proses dan hasil belajar, penilaian dapat didefinisikan sebagai suatu proses atau kegiatan yang sistematis dan berkesinambungan untuk mengumpulkan informasi tentang proses dan hasil belajar peserta didik dalam rangka membuat keputusan-keputusan berdasarkan kriteria dan pertimbangan tertentu. Jika dilihat dalam konteks yang lebih luas, keputusan tersebut dapat menyangkut keputusan tentang peserta didik, keputusan tentang kurikulum dan program atau juga keputusan tentang kebijakan pendidikan.

Keputusan tentang peserta didik meliputi pengelolaan pembelajaran, penempatan peserta didik sesuai dengan jenjang atau jenis program pendidikan, bimbingan dan konseling, dan menyeleksi peserta didik untuk pendidikan lebih lanjut. Keputusan tentang kurikulum dan program meliputi keefektifan (*summative evaluation*) dan bagaimana cara memperbaikinya (*formative evaluation*). Keputusan tentang kebijakan pendidikan dapat dibuat pada tingkat lokal/daerah (kabupaten/kota), regional (provinsi), dan tingkat nasional.

Keputusan penilaian terhadap suatu hasil belajar sangat bermanfaat untuk membantu peserta didik merefleksikan apa yang mereka ketahui, bagaimana mereka belajar, dan mendorong tanggung jawab dalam belajar. Keputusan penilaian dapat dibuat oleh guru, sesama peserta didik (*peer*) atau oleh dirinya sendiri (*self-assessment*). Pengambilan keputusan perlu menggunakan pertimbangan yang berbeda-beda dan membandingkan hasil penilaian. Pengambilan keputusan harus dapat membimbing peserta didik untuk melakukan perbaikan hasil belajar.

Guba dan Lincoln (1985: 35), mendefinisikan evaluasi sebagai "*a process for describing an evaluand and judging its merit and worth*". (suatu proses untuk menggambarkan evaluasi (orang yang dievaluasi) dan menimbang makna dan nilainya). Sax (1980: 18) juga berpendapat "*evaluation is a process through which a value judgement or decision is made from a variety of observations and from the background and training of the evaluator*". (evaluasi adalah suatu proses dimana pertimbangan atau keputusan suatu nilai dibuat dari berbagai pengamatan, latar belakang serta pelatihan dari evaluator). Dari dua rumusan tentang evaluasi ini, dapat kita peroleh gambaran bahwa evaluasi adalah suatu proses yang sistematis dan berkelanjutan untuk menentukan kualitas (nilai dan arti) daripada sesuatu, berdasarkan pertimbangan dan kriteria tertentu untuk

membuat suatu keputusan. Berdasarkan pengertian ini, ada beberapa hal yang perlu kita pahami lebih lanjut, yaitu:

1. Evaluasi adalah suatu proses bukan suatu hasil (produk).

Hasil yang diperoleh dari kegiatan evaluasi adalah kualitas daripada sesuatu, baik yang menyangkut tentang nilai maupun arti. Sedangkan kegiatan untuk sampai kepada pemberian nilai dan arti itu adalah evaluasi. Jika Anda melakukan kajian tentang evaluasi, maka yang Anda lakukan adalah mempelajari bagaimana proses pemberian pertimbangan mengenai kualitas daripada sesuatu. Gambaran kualitas yang dimaksud merupakan konsekuensi logis dari proses evaluasi yang dilakukan. Proses tersebut tentu dilakukan secara sistematis dan berkelanjutan, dalam arti terencana, sesuai dengan prosedur dan aturan, dan terus menerus.

2. Tujuan evaluasi adalah untuk menentukan kualitas daripada sesuatu, terutama yang berkenaan dengan nilai dan arti.

S. Hamid Hasan (1988: 14-15) secara tegas membedakan kedua istilah tersebut sebagai berikut: Pemberian nilai dilakukan apabila seorang evaluator memberikan pertimbangannya mengenai evaluasi tanpa menghubungkannya dengan sesuatu yang bersifat dari luar. Pemberian nilai dan arti ini dalam bahasa yang dipergunakan Scriven (1967) adalah *formatif* dan *sumatif*. Jika *formatif* dan *sumatif* merupakan fungsi evaluasi, maka nilai dan arti adalah hasil kegiatan yang dilakukan oleh evaluasi.

3. Dalam proses evaluasi harus ada pemberian pertimbangan (*judgement*).

Pemberian pertimbangan ini pada dasarnya merupakan konsep dasar evaluasi. Melalui pertimbangan inilah ditentukan nilai dan arti (*worth and merit*) dari sesuatu yang sedang dievaluasi. Tanpa pemberian pertimbangan, suatu kegiatan bukanlah termasuk kategori kegiatan evaluasi.

4. Pemberian pertimbangan tentang nilai dan arti haruslah berdasarkan kriteria tertentu.

Tanpa kriteria yang jelas, pertimbangan nilai dan arti yang diberikan bukanlah suatu proses yang dapat diklasifikasikan sebagai evaluasi. 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. Jika yang dievaluasi itu adalah proses pembelajaran, maka kriteria yang dimaksud bisa saja dikembangkan dari karakteristik proses pembelajaran itu sendiri, tetapi dapat pula dikembangkan kriteria umum tentang proses pembelajaran. Kriteria ini penting dibuat oleh evaluator dengan pertimbangan (a) hasil evaluasi dapat

dipertanggung jawabkan secara ilmiah (b) evaluator lebih percaya diri (c) menghindari adanya unsur subjektivitas (d) memungkinkan hasil evaluasi akan sama sekalipun dilakukan pada waktu dan orang yang berbeda, dan (e) memberikan kemudahan bagi evaluator dalam melakukan penafsiran hasil evaluasi. Kriteria sangat diperlukan untuk menentukan pencapaian indikator hasil belajar peserta didik yang sedang diukur. Dalam pengembangan kriteria untuk menentukan kualitas jawaban peserta didik, ada beberapa hal yang harus dipertimbangkan, antara lain (a) kriteria harus meluas tetapi tidak memakan waktu, sehingga sulit dilaksanakan (b) dapat dipahami dengan jelas oleh peserta didik, orang tua dan guru (c) mencerminkan keadilan, dan (d) tidak merefleksikan variabel yang bias, latar belakang budaya, sosial-ekonomi, ras dan gender.

Berdasarkan rumusan pengertian tentang tes, pengukuran, penilaian dan evaluasi yang telah dikemukakan di atas, dapat ditarik kesimpulan bahwa ada jenis evaluasi atau penilaian yang mempergunakan tes secara intensif sebagai alat pengumpulan data, seperti penilaian hasil belajar. Walaupun dalam perkembangan terakhir tentang jenis evaluasi atau penilaian seperti ini menunjukkan bahwa tes bukan satu-satunya alat pengumpul data. Namun demikian harus diakui pula, bahwa tes merupakan alat pengumpul data evaluasi dan penilaian yang paling tua dan penting. Tes bukanlah evaluasi, bahkan bukan pula pengukuran. Tes lebih sempit ruang lingkupnya dibandingkan pengukuran, dan pengukuran lebih sempit dibandingkan evaluasi.

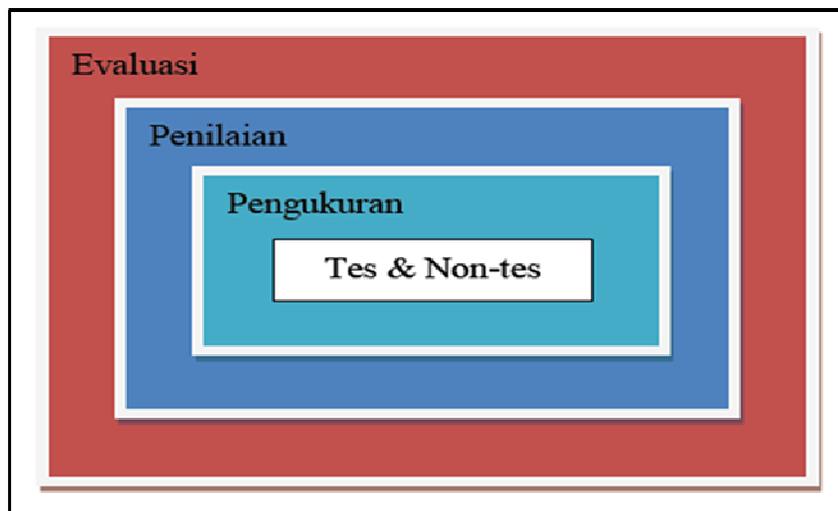
Sebagaimana telah disinggung sebelumnya bahwa tes dibangun berdasarkan teori pengukuran tertentu. Tanpa bantuan teori pengukuran, maka pembuatan tes dapat dikatakan tidak mungkin. Bagaimana Anda harus membuat pertanyaan-pertanyaan dalam suatu tes, bagaimana Anda ingin mengukur derajat validitas dan reliabilitas tes berdasarkan teori *psychometric*, mencerminkan peranan teori pengukuran yang sangat besar dan penting. Pengukuran dalam psikometrik tidak lagi merupakan bagian integral ataupun suatu langkah yang selalu harus ditempuh dalam kegiatan evaluasi. Pengukuran hanya merupakan salah satu langkah yang mungkin dipergunakan dalam kegiatan evaluasi.

1. Persamaan dan Perbedaan Evaluasi dengan Penilaian.

Persamaannya adalah keduanya mempunyai pengertian menilai atau menentukan nilai sesuatu. Di samping itu, alat yang digunakan untuk mengumpulkan datanya juga sama. Sedangkan perbedaannya terletak pada ruang lingkup (*scope*) dan pelaksanaannya. Ruang lingkup penilaian lebih sempit dan biasanya hanya terbatas pada salah satu komponen atau aspek saja, seperti prestasi belajar peserta didik.

Pelaksanaan penilaian biasanya dilakukan dalam konteks internal, yakni orang-orang yang menjadi bagian atau terlibat dalam sistem pembelajaran yang bersangkutan. Misalnya, guru menilai prestasi belajar peserta didik, supervisor menilai kinerja guru, dan sebagainya. Ruang lingkup evaluasi lebih luas, mencakup semua komponen dalam suatu sistem (sistem pendidikan, sistem kurikulum, sistem pembelajaran) dan dapat dilakukan tidak hanya pihak internal (evaluasi internal) tetapi juga pihak eksternal (evaluasi eksternal), seperti konsultan mengevaluasi suatu program.

Evaluasi dan penilaian lebih bersifat komprehensif yang meliputi pengukuran, sedangkan tes merupakan salah satu alat (*instrument*) pengukuran. Pengukuran lebih terbatas kepada gambaran yang bersifat kuantitatif (angka-angka) tentang kemajuan belajar peserta didik (*learning progress*), sedangkan evaluasi dan penilaian lebih bersifat kualitatif. Di samping itu, evaluasi dan penilaian pada hakikatnya merupakan suatu proses membuat keputusan tentang nilai suatu objek. Keputusan penilaian (*value judgement*) tidak hanya didasarkan kepada hasil pengukuran (*quantitative description*), tetapi dapat pula didasarkan kepada hasil pengamatan dan wawancara (*qualitative description*). Untuk lebih jelasnya, Anda dapat memperhatikan gambar berikut ini.



Gambar 1.1. Hubungan Evaluasi-Penilaian-Pengukuran dan Tes

Dengan demikian, pengertian evaluasi pembelajaran adalah suatu proses atau kegiatan yang sistematis, berkelanjutan dan menyeluruh dalam rangka pengendalian, penjaminan dan penetapan kualitas (nilai dan arti) pembelajaran terhadap berbagai komponen pembelajaran, berdasarkan pertimbangan dan kriteria tertentu, sebagai bentuk pertanggungjawaban guru dalam melaksanakan pembelajaran.

Sedangkan penilaian hasil belajar adalah suatu proses atau kegiatan yang sistematis, berkelanjutan dan menyeluruh dalam rangka pengumpulan dan pengolahan informasi untuk menilai pencapaian proses dan hasil belajar peserta didik.

2. Kedudukan Evaluasi Dalam Pembelajaran

Kata dasar “pembelajaran” adalah belajar. Dalam arti sempit pembelajaran dapat diartikan sebagai suatu proses atau cara yang dilakukan agar seseorang dapat melakukan kegiatan belajar. Sedangkan belajar adalah suatu proses perubahan tingkah laku karena interaksi individu dengan lingkungan dan pengalaman. Perubahan tingkah laku tersebut bukan karena pengaruh obat-obatan atau zat kimia lainnya dan cenderung bersifat permanen. Istilah “pembelajaran” (*instruction*) berbeda dengan istilah “pengajaran” (*teaching*). Kata “pengajaran” lebih bersifat formal dan hanya ada di dalam konteks guru dengan peserta didik di kelas, sedangkan kata “pembelajaran” tidak hanya ada dalam konteks guru dengan peserta didik di kelas secara formal, tetapi juga meliputi kegiatan-kegiatan belajar peserta didik di luar kelas yang mungkin saja tidak dihadiri oleh guru secara fisik.

Kata “pembelajaran” lebih menekankan pada kegiatan belajar peserta didik (*child-centered*) secara sungguh-sungguh yang melibatkan aspek intelektual, emosional, dan sosial, sedangkan kata “pengajaran” lebih cenderung pada kegiatan mengajar guru (*teacher-centered*) di kelas. Dengan demikian, kata “pembelajaran” ruang lingkungannya lebih luas daripada kata “pengajaran”. Dalam arti luas, pembelajaran adalah suatu proses atau kegiatan yang sistematis dan sistemik, yang bersifat interaktif dan komunikatif antara pendidik (guru) dengan peserta didik, sumber belajar dan lingkungan untuk menciptakan suatu kondisi yang memungkinkan terjadinya tindakan belajar peserta didik, baik di kelas maupun di luar kelas, dihadiri guru secara fisik atau tidak, untuk menguasai kompetensi yang telah ditentukan.

B. TUJUAN, FUNGSI DAN PRINSIP EVALUASI PEMBELAJARAN

1. Tujuan Evaluasi Pembelajaran

Jika kita ingin melakukan kegiatan evaluasi, terlepas dari jenis evaluasi apa yang digunakan, maka guru harus mengetahui dan memahami terlebih dahulu tentang tujuan dan fungsi evaluasi. Bila tidak, maka guru akan mengalami kesulitan merencanakan dan melaksanakan evaluasi. Hampir setiap orang yang membahas evaluasi membahas pula tentang tujuan dan fungsi evaluasi. Tujuan evaluasi pembelajaran adalah untuk mengetahui keefektifan dan efisiensi sistem

pembelajaran, baik yang menyangkut tentang tujuan, materi, metode, media, sumber belajar, lingkungan maupun sistem penilaian itu sendiri. Sedangkan tujuan khusus evaluasi pembelajaran disesuaikan dengan jenis evaluasi pembelajaran itu sendiri, seperti evaluasi perencanaan dan pengembangan, evaluasi monitoring, evaluasi dampak, evaluasi efisiensi-ekonomis, dan evaluasi program komprehensif.

Dalam konteks yang lebih luas lagi, Sax (1980: 28) mengemukakan tujuan evaluasi dan pengukuran adalah untuk *“selection, placement, diagnosis and remediation, feedback : norm-referenced and criterion-referenced interpretation, motivation and guidance of learning, program and curriculum improvement: formative and summative evaluations, and theory development”*. (seleksi, penempatan, diagnosis dan remediasi, umpan balik: penafsiran acuan-norma dan acuan-patokan, motivasi dan bimbingan belajar, perbaikan program dan kurikulum: evaluasi formatif dan sumatif, dan pengembangan teori).

Perlu diketahui bahwa evaluasi banyak digunakan dalam berbagai bidang dan kegiatan, antara lain bimbingan dan penyuluhan, supervisi, seleksi, dan pembelajaran. Setiap bidang atau kegiatan tersebut mempunyai tujuan yang berbeda. Dalam kegiatan bimbingan, tujuan evaluasi adalah untuk memperoleh informasi secara menyeluruh mengenai karakteristik peserta didik, sehingga dapat diberikan bimbingan dengan sebaik-baiknya. Begitu juga dalam kegiatan supervisi, tujuan evaluasi adalah untuk menentukan keadaan suatu situasi pendidikan atau pembelajaran, sehingga dapat diusahakan langkah-langkah perbaikan untuk meningkatkan mutu pendidikan di madrasah. Dalam kegiatan seleksi, tujuan evaluasi adalah untuk mengetahui tingkat pengetahuan, keterampilan, sikap dan nilai-nilai peserta didik untuk jenis pekerjaan, jabatan atau pendidikan tertentu.

Menurut Kellough dan Kellough dalam Swearingen (2006) tujuan penilaian adalah untuk membantu belajar peserta didik, mengidentifikasi kekuatan dan kelemahan peserta didik, menilai efektifitas strategi pembelajaran, menilai dan meningkatkan efektifitas program kurikulum, menilai dan meningkatkan efektifitas pembelajaran, menyediakan data yang membantu dalam membuat keputusan, komunikasi dan melibatkan orang tua peserta didik. Chittenden (1994) mengemukakan tujuan penilaian (*assessment purpose*) adalah *“keeping track, checking-up, finding-out, and summing-up”*.

1. *Keeping track*, yaitu untuk menelusuri dan melacak proses belajar peserta didik sesuai dengan rencana pelaksanaan pembelajaran yang telah ditetapkan. Untuk itu, guru harus mengumpulkan data dan

informasi dalam kurun waktu tertentu melalui berbagai jenis dan teknik penilaian untuk memperoleh gambaran tentang pencapaian kemajuan belajar peserta didik.

2. *Checking-up*, yaitu untuk mengecek ketercapaian kemampuan peserta didik dalam proses pembelajaran dan kekurangan-kekurangan peserta didik selama mengikuti proses pembelajaran. Dengan kata lain, guru perlu melakukan penilaian untuk mengetahui bagian mana dari materi yang sudah dikuasai peserta didik dan bagian mana dari materi yang belum dikuasai.
3. *Finding-out*, yaitu untuk mencari, menemukan dan mendeteksi kekurangan kesalahan atau kelemahan peserta didik dalam proses pembelajaran, sehingga guru dapat dengan cepat mencari alternatif solusinya.
4. *Summing-up*, yaitu untuk menyimpulkan tingkat penguasaan peserta didik terhadap kompetensi yang telah ditetapkan. Hasil penyimpulan ini dapat digunakan guru untuk menyusun laporan kemajuan belajar ke berbagai pihak yang berkepentingan.

Adapun tujuan penilaian hasil belajar adalah:

1. Untuk mengetahui tingkat penguasaan peserta didik terhadap materi yang telah diberikan.
2. Untuk mengetahui kecakapan, motivasi, bakat, minat, dan sikap peserta didik terhadap program pembelajaran.
3. Untuk mengetahui tingkat kemajuan dan kesesuaian hasil belajar peserta didik dengan standar kompetensi dan kompetensi dasar yang telah ditetapkan.
4. Untuk mendiagnosis keunggulan dan kelemahan peserta didik dalam mengikuti kegiatan pembelajaran. Keunggulan peserta didik dapat dijadikan dasar bagi guru untuk memberikan pembinaan dan pengembangan lebih lanjut, sedangkan kelemahannya dapat dijadikan acuan untuk memberikan bantuan atau bimbingan.
5. Untuk seleksi, yaitu memilih dan menentukan peserta didik yang sesuai dengan jenis pendidikan tertentu.
6. Untuk menentukan kenaikan kelas.
7. Untuk menempatkan peserta didik sesuai dengan potensi yang dimilikinya.

Pertama, dapat mengetahui kedudukan peserta didik dalam kelompoknya. Anda dapat memprakirakan apakah seorang peserta didik dalam kelompoknya dapat dimasukkan ke dalam golongan anak yang biasa atau yang luar biasa dalam arti supergenius atau lambat majunya. Anda juga dapat membuat perencanaan yang realistis mengenai masa depan anak. Hal ini penting, karena keberhasilan

peserta didik sebagai anggota masyarakat dikelak kemudian hari akan ditentukan oleh ada tidaknya perencanaan masa depan yang realistis ini.

Kedua, apabila pengetahuan tentang kemajuan peserta didik tadi digabungkan dengan pengetahuan tentang kapasitas (kemampuan dasar) peserta didik, maka ia dapat dipergunakan sebagai petunjuk mengenai kesungguhan usaha anak dalam menempuh program pendidikannya. Melalui petunjuk ini pula kita dapat membantu peserta didik sesuai dengan kompetensi yang diharapkan. Tujuan manapun yang akan dicapai, Anda tetap harus melakukan evaluasi terhadap kemampuan peserta didik dan komponen-komponen pembelajaran lainnya.

2. Fungsi Evaluasi Pembelajaran

Cronbach (1963: 236) menjelaskan *“evaluation used to improved the course while it is still fluid contributes more to improvement of education than evaluation used to appraise a product already on the market”*. Cronbach nampaknya lebih menekankan fungsi evaluasi untuk perbaikan, sedangkan Scriven (1967) membedakan fungsi evaluasi menjadi dua macam, yaitu fungsi formatif dan fungsi sumatif. Fungsi formatif dilaksanakan apabila hasil yang diperoleh dari kegiatan evaluasi diarahkan untuk memperbaiki bagian tertentu atau sebagian besar bagian kurikulum yang sedang dikembangkan. Sedangkan fungsi sumatif dihubungkan dengan penyimpulan mengenai kebaikan dari sistem secara keseluruhan. Fungsi ini baru dapat dilaksanakan jika pengembangan program pembelajaran telah dianggap selesai.

Fungsi evaluasi memang cukup luas, bergantung kepada dari sudut mana Anda melihatnya. Bila kita lihat secara menyeluruh, fungsi evaluasi adalah:

1. Secara psikologis, peserta didik selalu butuh untuk mengetahui hinggamana kegiatan yang telah dilakukan sesuai dengan tujuan yang hendak dicapai. Peserta didik adalah manusia yang belum dewasa. Mereka masih mempunyai sikap dan moral yang heteronom, membutuhkan pendapat orang-orang dewasa (seperti orang tua dan guru) sebagai pedoman baginya untuk mengadakan orientasi pada situasi tertentu. Dalam menentukan sikap dan tingkah lakunya, mereka pada umumnya tidak berpegang kepada pedoman yang berasal dari dalam dirinya, melainkan mengacu kepada norma-norma yang berasal dari luar dirinya. Dalam pembelajaran, mereka perlu mengetahui prestasi belajarnya, sehingga ia merasakan kepuasan dan ketenangan.

2. Secara sosiologis, evaluasi berfungsi untuk mengetahui apakah peserta didik sudah cukup mampu untuk terjun ke masyarakat. Mampu dalam arti peserta didik dapat berkomunikasi dan beradaptasi terhadap seluruh lapisan masyarakat dengan segala karakteristiknya. Lebih jauh dari itu, peserta didik diharapkan dapat membina dan mengembangkan semua potensi yang ada dalam masyarakat. Hal ini penting, karena mampu-tidaknya peserta didik terjun ke masyarakat akan memberikan ukuran tersendiri terhadap institusi pendidikan yang bersangkutan. Untuk itu, materi pembelajaran harus sesuai dengan kebutuhan masyarakat.
3. Secara didaktis-metodis, evaluasi berfungsi untuk membantu guru dalam menempatkan peserta didik pada kelompok tertentu sesuai dengan kemampuan dan kecakapannya masing-masing serta membantu guru dalam usaha memperbaiki proses pembelajarannya.
4. Evaluasi berfungsi untuk mengetahui kedudukan peserta didik dalam kelompok, apakah ia termasuk anak yang pandai, sedang atau kurang pandai. Hal ini berhubungan dengan sikap dan tanggung jawab orang tua sebagai pendidik pertama dan utama di lingkungan keluarga. Anda dan orang tua perlu mengetahui kemajuan peserta didik untuk menentukan langkah-langkah selanjutnya.
5. Evaluasi berfungsi untuk mengetahui taraf kesiapan peserta didik dalam menempuh program pendidikannya. Jika peserta didik sudah dianggap siap (fisik dan non-fisik), maka program pendidikan dapat dilaksanakan. Sebaliknya, jika peserta didik belum siap, maka hendaknya program pendidikan tersebut jangan dulu diberikan, karena akan mengakibatkan hasil yang kurang memuaskan.
6. Evaluasi berfungsi membantu guru dalam memberika bimbingan dan seleksi, baik dalam rangka menentukan jenis pendidikan, jurusan, maupun kenaikan kelas. Melalui evaluasi, Anda dapat mengetahui potensi peserta didik, sehingga dapat memberikan bimbingan sesuai dengan tujuan yang diharapkan. Begitu juga tentang kenaikan kelas. Jika peserta didik belum menguasai kompetensi yang ditentukan, maka peserta didik tersebut jangan dinaikkan ke kelas berikutnya atau yang lebih tinggi. Kegagalan ini merupakan hasil keputusan evaluasi, karena itu Anda perlu mengadakan bimbingan yang lebih profesional.
7. Secara administratif, evaluasi berfungsi untuk memberikan laporan tentang kemajuan peserta didik kepada orang tua, pejabat pemerintah yang berwenang, kepala sekolah, guru-guru dan peserta didik itu sendiri. Hasil evaluasi dapat memberikan gambaran secara

umum tentang semua hasil usaha yang dilakukan oleh institusi pendidikan.

Sementara itu, Stanley dalam Oemar Hamalik (1989: 6) mengemukakan secara spesifik tentang fungsi tes dalam pembelajaran yang dikategorikan ke dalam tiga fungsi yang saling berinterelasi, yakni “fungsi instruksional, fungsi administratif, dan fungsi bimbingan”.

1. Fungsi intruksional

- a. Proses konstruksi suatu tes merangsang Anda untuk menjelaskan dan merumuskan kembali tujuan-tujuan pembelajaran (kompetensi dasar) yang bermakna. Jika Anda terlibat secara aktif dalam perumusan tujuan pembelajaran (kompetensi dasar dan indikator), maka Anda akan terdorong untuk memperbaiki program pengalaman belajar bagi peserta didik, di samping akan memperbaiki alat evaluasi itu sendiri. Anda juga akan merasakan bahwa kompetensi dasar dan indikator yang telah dirumuskan itu akan bermakna bagi Anda dan peserta didik, sehingga akan memperkaya berbagai pengalaman belajar.
- b. Suatu tes akan memberikan umpan balik kepada guru. Umpan balik yang bersumber dari hasil tes akan membantu Anda untuk memberikan bimbingan belajar yang lebih bermakna bagi peserta didik. Tes yang dirancang dengan baik dapat dijadikan alat untuk mendiagnosis diri peserta didik, yakni untuk meneliti kelemahan-kelemahan yang dirasakannya sendiri.
- c. Tes-tes yang dikonstruksi secara cermat dapat memotivasi peserta didik melakukan kegiatan belajar. Pada umumnya setiap peserta didik ingin berhasil dengan baik dalam setiap tes yang ditempuhnya, bahkan ingin lebih baik dari teman-teman sekelasnya. Keinginan ini akan mendorongnya belajar lebih baik dan teliti. Artinya, ia akan bertarung dengan waktu guna menguasai materi pelajaran yang akan dievaluasi itu.
- d. Ulangan adalah alat yang bermakna dalam rangka penguasaan atau pemantapan belajar (*overlearning*). Ulangan ini dilaksanakan dalam bentuk review, latihan, pengembangan keterampilan dan konsep-konsep. Pemantapan, penguasaan dan pengembangan ingatan (*retention*) akan lebih baik jika dilakukan ulangan secara periodik dan kontinu. Kendatipun peserta didik dapat menjawab semua pertanyaan dalam tes, tetapi ulangan ini tetap besar manfaatnya, karena penguasaan materi pelajaran akan bertambah mantap.

2. Fungsi administratif

- a. Tes merupakan suatu mekanisme untuk mengontrol kualitas suatu sekolah atau suatu sistem sekolah. Norma-norma lokal maupun norma-norma nasional menjadi dasar untuk melihat untuk menilai keampuhan dan kelemahan kurikuler sekolah, apalagi jika daerah setempat tidak memiliki alat yang dapat dipergunakan untuk melaksanakan evaluasi secara periodik.
- b. Tes berguna untuk mengevaluasi program dan melakukan penelitian. Keberhasilan suatu program inovasi dapat dilihat setelah diadakan pengukuran terhadap hasil program sesuai dengan tujuan khusus yang telah ditetapkan. Percobaan metode mengajar untuk menemukan cara belajar efektif dan efisien bagi para peserta didik, baru dapat dilaksanakan setelah diadakan serangkaian kegiatan eksperimen, selanjutnya dapat diukur keberhasilannya dengan tes.
- c. Tes dapat meningkatkan kualitas hasil seleksi. Seleksi sering dilakukan untuk menentukan bakat peserta didik dan kemungkinan berhasil dalam studinya pada suatu lembaga pendidikan. Apakah seorang calon memilih keterampilan dalam mengemban tugas tertentu, apakah peserta didik tergolong anak terbelakang, dan sebagainya. Hasil seleksi sering digunakan untuk menempatkan dan mengklasifikasikan peserta didik dalam rangka program bimbingan. Anda juga dapat menggunakan hasil tes untuk menentukan apakah peserta didik perlu dibimbing, dilatih, diobati, dan diajari.
- d. Tes berguna sebagai alat untuk melakukan akreditasi, penguasaan (*mastery*), dan sertifikasi. Tes dapat dipergunakan untuk mengukur kompetensi seorang lulusan. Misalnya, seorang calon guru sudah dapat dikatakan memiliki kompetensi yang diharapkan setelah dia mampu mendemonstrasikan kemampuannya di dalam kelas. Untuk mengetahui tingkat penguasaan kompetensi, kemudian memberikan sertifikat, diperlukan pengukuran dengan alat tertentu, yaitu tes.

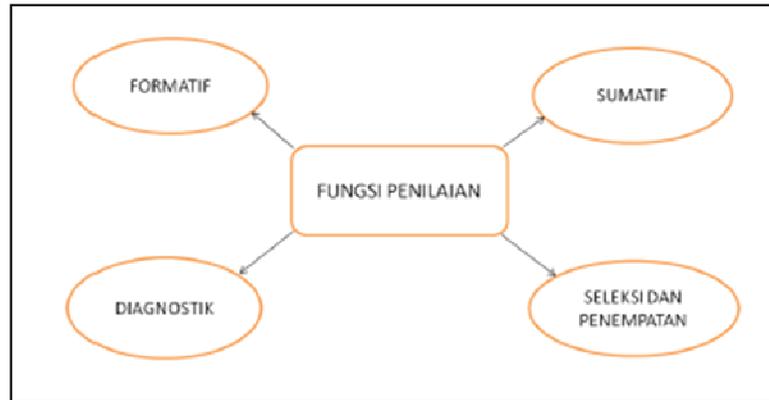
3. Fungsi bimbingan

Tes sangat penting untuk mendiagnosis bakat-bakat khusus dan kemampuan (*ability*) peserta didik. Bakat skolastik, prestasi, minat, kepribadian, merupakan aspek-aspek penting yang harus mendapat perhatian dalam proses bimbingan. Informasi dari hasil tes standar (*standardized test*) dapat membantu kegiatan bimbingan dan seleksi ke sekolah yang lebih tinggi, memilih jurusan/program studi, mengetahui kemampuan, dan sebagainya. Untuk memperoleh

informasi yang lengkap sesuai dengan kebutuhan bimbingan, maka diperlukan alat ukur yang memadai, seperti tes.

Fungsi evaluasi pembelajaran adalah: Pertama, untuk perbaikan dan pengembangan sistem pembelajaran. Sebagaimana Anda ketahui bahwa pembelajaran sebagai suatu sistem memiliki berbagai komponen, seperti tujuan, materi, metoda, media, sumber belajar, lingkungan, guru dan peserta. Dengan demikian, perbaikan dan pengembangan pembelajaran harus diarahkan kepada semua komponen pembelajaran tersebut. Kedua, untuk akreditasi. Dalam UU No. 20/2003 Bab 1 Pasal 1 Ayat 22 dijelaskan bahwa “akreditasi adalah kegiatan penilaian kelayakan program dalam satuan pendidikan berdasarkan kriteria yang telah ditetapkan”. Salah satu komponen akreditasi adalah pembelajaran. Artinya, fungsi akreditasi dapat dilaksanakan jika hasil evaluasi pembelajaran digunakan sebagai dasar akreditasi lembaga pendidikan. Sedangkan fungsi penilaian hasil belajar adalah:

1. Fungsi formatif, yaitu untuk memberikan umpan balik (*feedback*) kepada guru sebagai dasar untuk memperbaiki proses pembelajaran dan mengadakan program remedial bagi peserta didik.
2. Fungsi sumatif, yaitu untuk menentukan nilai (angka) kemajuan/hasil belajar peserta didik dalam mata pelajaran tertentu, sebagai bahan untuk memberikan laporan kepada berbagai pihak, penentuan kenaikan kelas dan penentuan lulus-tidaknya peserta didik.
3. Fungsi diagnostik, yaitu untuk memahami latar belakang (psikologis, fisik dan lingkungan) peserta didik yang mengalami kesulitan belajar, dimana hasilnya dapat digunakan sebagai dasar dalam memecahkan kesulitan-kesulitan tersebut.
4. Fungsi penempatan, yaitu untuk menempatkan peserta didik dalam situasi pembelajaran yang tepat (misalnya dalam penentuan program spesialisasi) sesuai dengan tingkat kemampuan peserta didik.



Gambar 1.2. Fungsi Penilaian

3. Prinsip-prinsip Umum Evaluasi

Untuk memperoleh hasil evaluasi yang lebih baik, Anda harus memperhatikan prinsip-prinsip umum evaluasi sebagai berikut:

1) Kontinuitas

Evaluasi tidak boleh dilakukan secara insidental, karena pembelajaran itu sendiri adalah suatu proses yang kontinu. Oleh sebab itu, Anda harus melakukan evaluasi secara kontinu. Hasil evaluasi yang diperoleh pada suatu waktu harus senantiasa dihubungkan dengan hasil-hasil pada waktu sebelumnya, sehingga dapat diperoleh gambaran yang jelas dan berarti tentang perkembangan peserta didik. Perkembangan belajar peserta didik tidak dapat dilihat dari dimensi produk saja tetapi juga dimensi proses bahkan dari dimensi input.

2) Komprehensif

Dalam melakukan evaluasi terhadap suatu objek, Anda harus mengambil seluruh objek itu sebagai bahan evaluasi. Misalnya, jika objek evaluasi itu adalah peserta didik, maka seluruh aspek kepribadian peserta didik itu harus dievaluasi, baik yang menyangkut kognitif, afektif maupun psikomotor. Begitu juga dengan objek-objek evaluasi yang lain.

3) Adil dan objektif

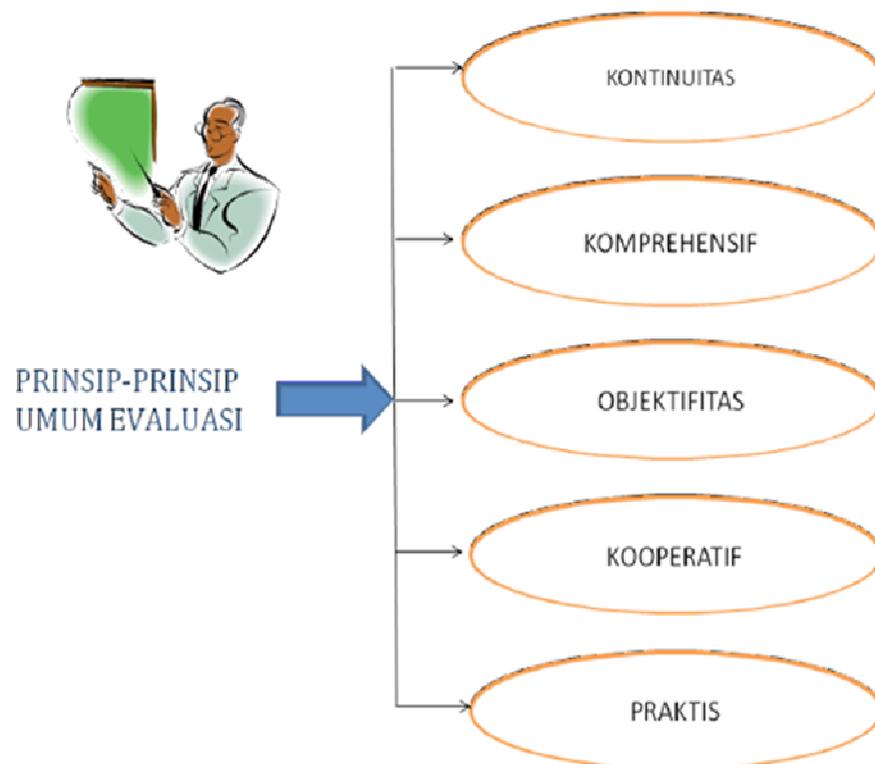
Dalam melaksanakan evaluasi, Anda harus berlaku adil tanpa pilih kasih. Semua peserta didik harus diperlakukan sama tanpa “pandang bulu”. Anda juga hendaknya bertindak secara objektif, apa adanya sesuai dengan kemampuan peserta didik. Sikap like and dislike, perasaan, keinginan, dan prasangka yang bersifat negatif harus dijauhkan. Evaluasi harus didasarkan atas kenyataan (data dan fakta) yang sebenarnya, bukan hasil manipulasi atau rekayasa.

4) Kooperatif

Dalam kegiatan evaluasi, Anda hendaknya bekerjasama dengan semua pihak, seperti orang tua peserta didik, sesama guru, kepala sekolah, termasuk dengan peserta didik itu sendiri. Hal ini dimaksudkan agar semua pihak merasa puas dengan hasil evaluasi, dan pihak-pihak tersebut merasa dihargai.

5) Praktis

Praktis mengandung arti mudah digunakan, baik bagi Anda sendiri yang menyusun alat evaluasi maupun orang lain yang akan menggunakan alat tersebut. Untuk itu, Anda harus memperhatikan bahasa dan petunjuk mengerjakan soal.



Gambar 1.3. Prinsip-prinsip Umum Evaluasi

Dalam konteks hasil belajar, Depdiknas (2003 : 7) mengemukakan prinsip-prinsip umum penilaian adalah mengukur hasil-hasil belajar yang telah ditentukan dengan jelas dan sesuai dengan kompetensi serta tujuan pembelajaran; mengukur sampel tingkah laku yang representatif dari hasil belajar dan bahan-bahan yang tercakup dalam pengajaran; mencakup jenis-jenis instrumen penilaian yang paling sesuai untuk mengukur hasil belajaryang diinginkan; direncanakan sedemikian rupa agar hasilnya sesuai dengan yang digunakan secara khusus; dibuat dengan reliabilitas yang sebesar-besarnya dan harus ditafsirkan secara hati-hati; dan dipakai untuk memperbaiki proses dan hasil belajar.

Dalam penilaian hasil belajar, harus memperhatikan pula hal-hal sebagai berikut:

- 1) Penilaian hendaknya dirancang sedemikian rupa, sehingga jelas abilitas yang harus dinilai, materi yang akan dinilai, alat penilaian dan interpretasi hasil penilaian.
- 2) Penilaian harus menjadi bagian integral dalam proses pembelajaran.
- 3) Untuk memperoleh hasil yang objektif, penilaian harus menggunakan berbagai alat (instrumen), baik yang berbentuk tes maupun non-tes.
- 4) Pemilihan alat penilaian harus sesuai dengan kompetensi yang ditetapkan.
- 5) Alat penilaian harus mendorong kemampuan penalaran dan kreatifitas peserta didik, seperti: tes tertulis esai, tes kinerja, hasil karya peserta didik, proyek, dan portofolio.
- 6) Objek penilaian harus mencakup aspek pengetahuan, keterampilan, sikap dan nilai-nilai.
- 7) Penilaian harus mengacu kepada prinsip diferensiasi, yaitu memberikan peluang kepada peserta didik untuk menunjukkan apa yang diketahui, apa yang dipahami dan apa yang dapat dilakukan.
- 8) Penilaian tidak bersifat diskriminatif. Artinya, guru harus bersikap adil dan jujur kepada semua peserta didik, serta bertanggung jawab kepada semua pihak.
- 9) Penilaian harus diikuti dengan tindak lanjut.
- 10) Penilaian harus berorientasi kepada kecakapan hidup dan bersifat mendidik.

4. Jenis Evaluasi Pembelajaran

Dilihat dari pengertian, tujuan, fungsi, prosedur dan sistem pembelajaran, maka pada hakikatnya pembelajaran adalah suatu program. Artinya, evaluasi yang digunakan dalam pembelajaran adalah evaluasi program, bukan penilaian hasil belajar. Penilaian hasil belajar hanya merupakan bagian dari evaluasi pembelajaran. Sebagai suatu program, evaluasi pembelajaran dibagi menjadi lima jenis, yaitu :

- a. Evaluasi perencanaan dan pengembangan. Hasil evaluasi ini sangat diperlukan untuk mendisain program pembelajaran. Sasaran utamanya adalah memberikan bantuan tahap awal dalam penyusunan program pembelajaran. Persoalan yang disoroti menyangkut tentang kelayakan dan kebutuhan. Hasil evaluasi ini dapat meramalkan kemungkinan implementasi program dan tercapainya keberhasilan program pembelajaran. Pelaksanaan

evaluasi dilakukan sebelum program sebenarnya disusun dan dikembangkan.

- b. Evaluasi monitoring, yaitu untuk memeriksa apakah program pembelajaran mencapai sasaran secara efektif dan apakah program pembelajaran terlaksana sebagaimana mestinya. Hasil evaluasi ini sangat baik untuk mengetahui kemungkinan pemborosan sumber-sumber dan waktu pelaksanaan pembelajaran, sehingga dapat dihindarkan.
- c. Evaluasi dampak, yaitu untuk mengetahui dampak yang ditimbulkan oleh suatu program pembelajaran. Dampak ini dapat diukur berdasarkan kriteria keberhasilan sebagai indikator ketercapaian tujuan program pembelajaran.
- d. Evaluasi efisiensi-ekonomis, yaitu untuk menilai tingkat efisiensi program pembelajaran. Untuk itu, diperlukan perbandingan antara jumlah biaya, tenaga dan waktu yang diperlukan dalam program pembelajaran dengan program lainnya yang memiliki tujuan yang sama.
- e. Evaluasi program komprehensif, yaitu untuk menilai program pembelajaran secara menyeluruh, seperti pelaksanaan program, dampak program, tingkat keefektifan dan efisiensi.

Sedangkan penilaian proses dan hasil belajar, dapat dibagi menjadi empat jenis, yaitu penilaian formatif, penilaian sumatif, penilaian diagnostik, dan penilaian penempatan.

1. Penilaian Formatif (*formative assessment*)

Penilaian formatif dimaksudkan untuk memantau kemajuan belajar peserta didik selama proses belajar berlangsung, untuk memberikan balikan (*feedback*) bagi penyempurnaan program pembelajaran, serta untuk mengetahui kelemahan-kelemahan yang memerlukan perbaikan, sehingga hasil belajar peserta didik dan proses pembelajaran guru menjadi lebih baik. Soal-soal penilaian formatif ada yang mudah dan adapula yang sukar, bergantung kepada tugas-tugas belajar (*learning tasks*) dalam program pembelajaran yang akan dinilai. Tujuan utama penilaian formatif adalah untuk memperbaiki proses pembelajaran, bukan untuk menentukan tingkat kemampuan peserta didik. Penilaian formatif sesungguhnya merupakan penilaian acuan patokan (*criterion-referenced assessment*). Apa yang dimaksudkan dengan penilaian formatif seperti yang diberikan pada akhir satuan pelajaran sesungguhnya bukan sebagai penilaian formatif lagi, sebab data-data yang diperoleh akhirnya digunakan untuk menentukan tingkat hasil belajar peserta didik. Kiranya lebih tepat jika penilaian pada akhir satuan pelajaran itu dipandang sebagai penilaian sub-

sumatif. Jika dimaksudkan untuk perbaikan proses pembelajaran, maka maksud itu baru terlaksana pada jangka panjang, yaitu pada saat penyusunan program tahun berikutnya.

2. Penilaian Sumatif (*summative assessment*)

Istilah “sumatif” berasal dari kata “sum” yang berarti “*total obtained by adding together items, numbers or amounts*”. Penilaian sumatif berarti penilaian yang dilakukan jika satuan pengalaman belajar atau seluruh materi pelajaran dianggap telah selesai. Contohnya adalah ujian akhir semester dan ujian nasional. Penilaian sumatif diberikan dengan maksud untuk mengetahui apakah peserta didik sudah dapat menguasai standar kompetensi yang telah ditetapkan atau belum. Tujuan penilaian sumatif adalah untuk menentukan nilai (angka) berdasarkan tingkatan hasil belajar peserta didik yang selanjutnya dipakai sebagai angka rapor. Hasil penilaian sumatif juga dapat dimanfaatkan untuk perbaikan proses pembelajaran secara keseluruhan. Sejak diberlakukannya Kurikulum 2004 dan sekarang KTSP, penilaian sumatif termasuk penilaian acuan patokan / PAP (*criterion-referenced assessment*), dimana kemampuan peserta didik dibandingkan dengan sebuah kriteria, dalam hal ini kompetensi. Cakupan materinya lebih luas dan soal-soalnya meliputi tingkat mudah, sedang, dan sulit. Adapun fungsi utama penilaian sumatif adalah (a) untuk menentukan nilai akhir peserta didik dalam periode tertentu Misalnya, akhir catur wulan, akhir semester, akhir tahun, atau akhir suatu sekolah. Nilai tersebut biasanya dilaporkan dalam buku laporan pendidikan atau Surat Tanda Tamat Belajar (STTB). Dengan demikian, guru akan mengetahui kedudukan seorang peserta didik dibandingkan dengan peserta didik lain dalam hal prestasi belajarnya, (b) untuk memberikan informasi tentang kecakapan atau keterampilan peserta didik dalam periode tertentu, dan (c) untuk memprakirakan berhasil tidaknya peserta didik dalam pelajaran berikutnya yang lebih tinggi. Agar fungsi memprakirakan ini dapat berjalan dengan baik, maka Anda perlu memperhatikan hal-hal berikut. Pertama, pelajaran berikutnya harus mempunyai hubungan dengan pelajaran yang sudah ditempuhnya. Kedua, pelajaran berikutnya masih berhubungan dengan karakteristik peserta didik. Ketiga, dapat dipergunakan untuk menentukan bahan pelajaran berikutnya. Keempat, sebagai bahan pertimbangan untuk menyempurnakan urutan (*sequence*) dan ruang lingkup (*scope*) materi pelajaran, termasuk metode, media dan sumber belajar yang dipergunakan dalam serangkaian kegiatan pembelajaran.

3. Penilaian Penempatan (*placement assessment*)

Pada umumnya penilaian penempatan dibuat sebagai prates (*pretest*). Tujuan utamanya adalah untuk mengetahui apakah peserta didik telah memiliki keterampilan-keterampilan yang diperlukan untuk mengikuti suatu program pembelajaran dan hinggamana peserta didik telah menguasai kompetensi dasar sebagaimana yang tercantum dalam silabus dan Rencana Pelaksanaan Pembelajaran (RPP). Tujuan yang pertama masalahnya berkaitan dengan kesiapan peserta didik menghadapi program baru, sedangkan tujuan yang kedua berkaitan dengan kesesuaian program pembelajaran dengan kemampuan peserta didik. Luas bahan prates lebih terbatas dan tingkat kesukaran soalnya relatif rendah. Hal ini berdasarkan kenyataan bahwa prates digunakan untuk menentukan apakah peserta didik telah memiliki kemampuan-kemampuan minimal untuk mempelajari suatu unit materi pelajaran atau belum sama sekali. Prates seperti ini adalah *criterion-referenced assessment* yang fungsi utamanya adalah untuk mengidentifikasi ada-tidaknya *prerequisite skills*. Prates dibuat untuk menentukan hinggamana peserta didik telah menguasai materi pelajaran atau memperoleh pengalaman belajar seperti tercantum dalam program pembelajaran, dan sebenarnya tidak berbeda dengan tes hasil belajar. Dalam hal seperti itu prates dibuat sebagai *norm-referenced assessment*.

4. Penilaian Diagnostik (*diagnostic assessment*)

Penilaian diagnostik dianggap penting agar Anda dapat mengetahui kesulitan belajar peserta didik berdasarkan hasil penilaian formatif sebelumnya. Untuk itu, Anda memerlukan sejumlah soal untuk satu bidang yang diperkirakan merupakan kesulitan bagi peserta didik. Soal-soal tersebut bervariasi dan difokuskan pada kesulitan. Penilaian diagnostik biasanya dilaksanakan sebelum suatu pelajaran dimulai. Tujuannya adalah untuk menjajagi pengetahuan dan keterampilan yang telah dikuasai oleh peserta didik. Dengan kata lain, apakah peserta didik sudah mempunyai pengetahuan dan keterampilan tertentu untuk dapat mengikuti materi pelajaran lain. Penilaian diagnostik semacam ini disebut juga *test of entering behavior*.

BAB II

RUANG LINGKUP, KARAKTERISTIK DAN PENDEKATAN EVALUASI PEMBELAJARAN

A. RUANG LINGKUP EVALUASI PEMBELAJARAN DALAM PERSPEKTIF DOMAIN HASIL BELAJAR

Menurut Benyamin S. Bloom, dkk (1956) hasil belajar dapat dikelompokkan ke dalam tiga domain, yaitu kognitif, afektif dan psikomotor. Setiap domain disusun menjadi beberapa jenjang kemampuan, mulai dari hal yang sederhana sampai dengan hal yang kompleks, mulai dari hal yang mudah sampai dengan hal yang sukar, dan mulai dari hal yang konkrit sampai dengan hal yang abstrak. Adapun rincian domain tersebut adalah sebagai berikut:

1. Domain kognitif (*cognitive domain*). Domain ini memiliki enam jenjang kemampuan, yaitu :
 - a. Pengetahuan (*knowledge*), yaitu jenjang kemampuan yang menuntut peserta didik untuk dapat mengenali atau mengetahui adanya konsep, prinsip, fakta atau istilah tanpa harus mengerti atau dapat menggunakannya. Kata kerja operasional yang dapat digunakan diantaranya: mendefinisikan, memberikan, mengidentifikasi, memberi nama, menyusun daftar, mencocokkan, menyebutkan, membuat garis besar, menyatakan, dan memilih.
 - b. Pemahaman (*comprehension*), yaitu jenjang kemampuan yang menuntut peserta didik untuk memahami atau mengerti tentang materi pelajaran yang disampaikan guru dan dapat memanfaatkannya tanpa harus menghubungkannya dengan hal-hal lain. Kemampuan ini dijabarkan lagi menjadi tiga, yakni menterjemahkan, menafsirkan, dan mengekstrapolasi. Kata kerja operasional yang dapat digunakan diantaranya: mengubah, mempertahankan, membedakan, memprakirakan, menjelaskan, menyimpulkan, memberi contoh, meramalkan, dan meningkatkan.
 - c. Penerapan (*application*), yaitu jenjang kemampuan yang menuntut peserta didik untuk menggunakan ide-ide umum, tata cara ataupun metode, prinsip dan teori-teori dalam situasi baru dan konkrit. Kata kerja operasional yang dapat digunakan diantaranya: mengubah, menghitung, mendemonstrasikan, mengungkapkan, mengerjakan dengan teliti, menjalankan, memanipulasikan, menghubungkan, menunjukkan, memecahkan, menggunakan.

- d. Analisis (*analysis*), yaitu jenjang kemampuan yang menuntut peserta didik untuk menguraikan suatu situasi atau keadaan tertentu ke dalam unsur-unsur atau komponen pembentuknya. Kemampuan analisis dikelompokkan menjadi tiga, yaitu analisis unsur, analisis hubungan, dan analisis prinsip-prinsip yang terorganisasi. Kata kerja operasional yang dapat digunakan diantaranya: mengurai, membuat diagram, memisah-misahkan, menggambarkan kesimpulan, membuat garis besar, menghubungkan, merinci.
 - e. Sintesis (*synthesis*), yaitu jenjang kemampuan yang menuntut peserta didik untuk menghasilkan sesuatu yang baru dengan cara menggabungkan berbagai faktor. Hasil yang diperoleh dapat berupa tulisan, rencana atau mekanisme. Kata kerja operasional yang dapat digunakan diantaranya: menggolongkan, menggabungkan, memodifikasi, menghimpun, menciptakan, merencanakan, merekonstruksikan, menyusun, membangkitkan, mengorganisir, merevisi, menyimpulkan, menceritakan.
 - f. Evaluasi (*evaluation*), yaitu jenjang kemampuan yang menuntut peserta didik untuk dapat mengevaluasi suatu situasi, keadaan, pernyataan atau konsep berdasarkan kriteria tertentu. Hal penting dalam evaluasi ini adalah menciptakan kondisi sedemikian rupa, sehingga peserta didik mampu mengembangkan kriteria atau patokan untuk mengevaluasi sesuatu. Kata kerja operasional yang dapat digunakan diantaranya: menilai, membandingkan, mempertentangkan, mengeritik, membeda-bedakan, mempertimbangkan kebenaran, menyokong, menafsirkan, dan menduga.
2. Domain afektif (*affective domain*), yaitu internalisasi sikap yang menunjuk ke arah pertumbuhan batiniah dan terjadi bila peserta didik menjadi sadar tentang nilai yang diterima, kemudian mengambil sikap sehingga menjadi bagian dari dirinya dalam membentuk nilai dan menentukan tingkah laku. Domain afektif terdiri atas beberapa jenjang kemampuan, yaitu:
- a. Kemauan menerima (*receiving*), yaitu jenjang kemampuan yang menuntut peserta didik untuk peka terhadap eksistensi fenomena atau rangsangan tertentu. Kepekaan ini diawali dengan kesadaran kemampuan untuk menerima dan memperhatikan. Kata kerja operasional yang dapat digunakan diantaranya: menanyakan, memilih, menggambarkan, mengikuti, memberikan, berpegang teguh, menjawab, menggunakan.
 - b. Kemauan menanggapi/jawab (*responding*), yaitu jenjang kemampuan yang menuntut peserta didik untuk tidak hanya peka

pada suatu fenomena tetapi juga bereaksi terhadap salah satu cara. Penekanannya pada kemauan peserta didik untuk menjawab secara sukarela, membaca tanpa ditugaskan. Kata kerja operasional yang dapat digunakan diantaranya: menjawab, membantu, memperbincangkan, memberi nama, menunjukkan, mempraktikkan, mengemukakan, membaca, melaporkan, menuliskan, memberitahu, dan mendiskusikan.

- c. Menilai (*valuing*), yaitu jenjang kemampuan yang menuntut peserta didik untuk menilai suatu objek, fenomena atau tingkah laku tertentu secara konsisten. Kata kerja operasional yang digunakan diantaranya: melengkapi, menerangkan, membentuk, mengusulkan, mengambil bagian, dan memilih.
 - d. Organisasi (*organization*), yaitu jenjang kemampuan yang menuntut peserta didik untuk menyatukan nilai-nilai yang berbeda, memecahkan masalah, membentuk suatu sistem nilai. Kata kerja operasional yang dapat digunakan diantaranya: mengubah, mengatur, menggabungkan, membandingkan, mempertahankan, menggeneralisasikan, memodifikasi.
3. Domain psikomotor (*psychomotor domain*), yaitu kemampuan peserta didik yang berkaitan dengan gerakan tubuh atau bagian-bagiannya, mulai dari gerakan yang sederhana sampai dengan gerakan yang kompleks. Perubahan pola gerakan memakan waktu sekurang-kurangnya 30 menit. Kata kerja operasional yang digunakan harus sesuai dengan kelompok keterampilan masing-masing, yaitu:
- a. *Muscular or motor skill*, yang meliputi: mempertontonkan gerak, menunjukkan hasil, melompat, menggerakkan, menampilkan.
 - b. *Manipulations of materials or objects*, yang meliputi: mereparasi, menyusun, membersihkan, menggeser, memindahkan, membentuk.
 - c. *Neuromuscular coordination*, yang meliputi: mengamati, menerapkan, menghubungkan, menggandeng, memadukan, memasang, memotong, menarik dan menggunakan.

Berdasarkan taksonomi Bloom di atas, maka kemampuan peserta didik dapat diklasifikasikan menjadi dua, yaitu tingkat tinggi dan tingkat rendah. Kemampuan tingkat rendah terdiri atas pengetahuan, pemahaman, dan aplikasi, sedangkan kemampuan tingkat tinggi meliputi analisis, sintesis, evaluasi, dan kreatifitas. Dengan demikian, kegiatan peserta didik dalam menghafal termasuk kemampuan tingkat rendah. Dilihat cara berpikir, makakemampuan berpikir tingkat tinggi dibagi menjadi dua, yaitu berpikir kritis dan berpikir kreatif. Berpikir kreatif adalah kemampuan melakukan generalisasi dengan menggabungkan, mengubah atau mengulang kembali keberadaan ide-ide tersebut.

Sedangkan kemampuan berpikir kritis merupakan kemampuan memberikan rasionalisasi terhadap sesuatu dan mampu memberikan penilaian terhadap sesuatu tersebut. Rendahnya kemampuan peserta didik dalam berpikir, bahkan hanya dapat menghafal, tidak terlepas dari kebiasaan guru dalam melakukan evaluasi atau penilaian yang hanya mengukur tingkat kemampuan yang rendah saja melalui paper and pencil test. Peserta didik tidak akan mempunyai kemampuan berpikir tingkat tinggi jika tidak diberikan kesempatan untuk mengembangkannya dan tidak diarahkan untuk itu.

B. RUANG LINGKUP EVALUASI PEMBELAJARAN DALAM PERSPEKTIF SISTEM PEMBELAJARAN

Sebagaimana telah disinggung sebelumnya bahwa ruang lingkup evaluasi pembelajaran hendaknya bertitik tolak dari tujuan evaluasi pembelajaran itu sendiri. Hal ini dimaksudkan agar apa yang dievaluasi relevan dengan apa yang diharapkan. Tujuan evaluasi pembelajaran adalah untuk mengetahui keefektifan dan efisiensi sistem pembelajaran, baik yang menyangkut tentang tujuan, materi, metode, media, sumber belajar, lingkungan, guru dan peserta didik serta sistem penilaian itu sendiri. Secara keseluruhan, ruang lingkup evaluasi pembelajaran adalah :

1. Program pembelajaran, yang meliputi:

- a. Tujuan pembelajaran umum atau kompetensi dasar, yaitu target yang harus dikuasai peserta didik dalam setiap pokok bahasan/topik. Kriteria yang digunakan untuk mengevaluasi tujuan pembelajaran umum atau kompetensi dasar ini adalah keterkaitannya dengan tujuan kurikuler atau standar kompetensi dari setiap bidang studi/mata pelajaran dan tujuan kelembagaan, kejelasan rumusan kompetensi dasar, kesesuaiannya dengan tingkat perkembangan peserta didik, pengembangannya dalam bentuk hasil belajar dan indikator, penggunaan kata kerja operasional dalam indikator, dan unsur-unsur penting dalam kompetensi dasar, hasil belajar dan indikator.
- b. Isi/materi pembelajaran, yaitu isi kurikulum yang berupa topik/pokok bahasan dan sub topik/sub pokok bahasan beserta rinciannya dalam setiap bidang studi atau mata pelajaran. Isi kurikulum tersebut memiliki tiga unsur, yaitu logika (pengetahuan benar salah, berdasarkan prosedur keilmuan), etika (baik-buruk), dan estetika (keindahan). Materi pembelajaran dapat dikelompokkan menjadi enam jenis, yaitu fakta, konsep/teori, prinsip, proses, nilai dan keterampilan. Kriteria yang digunakan, antara lain: kesesuaiannya dengan kompetensi dasar dan hasil belajar, ruang

lingkup materi, urutan logis materi, kesesuaiannya dengan tingkat perkembangan dan kebutuhan peserta didik, waktu yang tersedia dan sebagainya.

- c. Metode pembelajaran, yaitu cara guru menyampaikan materi pelajaran, seperti metode ceramah, tanya jawab, diskusi, pemecahan masalah, dan sebagainya. Kriteria yang digunakan, antara lain: kesesuaiannya dengan kompetensi dasar dan hasil belajar, kesesuaiannya dengan kondisi kelas/sekolah, kesesuaiannya dengan tingkat perkembangan peserta didik, kemampuan guru dalam menggunakan metode, waktu, dan sebagainya.
 - d. Media pembelajaran, yaitu alat-alat yang membantu untuk mempermudah guru dalam menyampaikan isi/materi pelajaran. Media dapat dibagi tiga kelompok, yaitu media audio, media visual, dan media audio-visual. Kriteria yang digunakan sama seperti komponen metode.
 - e. Sumber belajar, yang meliputi: pesan, orang, bahan, alat, teknik, dan latar. Sumber belajar dapat dibedakan menjadi dua jenis, yaitu sumber belajar yang dirancang (*resources by design*) dan sumber belajar yang digunakan (*resources by utilization*). Kriteria yang digunakan sama seperti komponen metode.
 - f. Lingkungan, terutama lingkungan sekolah dan lingkungan keluarga. Kriteria yang digunakan, antara lain: hubungan antara peserta didik dengan teman sekelas/sekolah maupun di luar sekolah, guru dan orang tua; kondisi keluarga dan sebagainya.
 - g. Penilaian proses dan hasil belajar, baik yang menggunakan tes maupun non-tes. Kriteria yang digunakan, antara lain: kesesuaiannya dengan kompetensi dasar, hasil belajar, dan indikator; kesesuaiannya dengan tujuan dan fungsi penilaian, unsur-unsur penting dalam penilaian, aspek-aspek yang dinilai, kesesuaiannya dengan tingkat perkembangan peserta didik, jenis dan alat penilaian.
2. Proses pelaksanaan pembelajaran:
- a. Kegiatan, yang meliputi: jenis kegiatan, prosedur pelaksanaan setiap jenis kegiatan, sarana pendukung, efektifitas dan efisiensi, dan sebagainya.
 - b. Guru, terutama dalam hal: menyampaikan materi, kesulitan-kesulitan guru, menciptakan suasana pembelajaran yang kondusif, menyiapkan alat-alat dan perlengkapan yang diperlukan, membimbing peserta didik, menggunakan teknik penilaian, menerapkan disiplin kelas, dan sebagainya.

- c. Peserta didik, terutama dalam hal: peranserta peserta didik dalam kegiatan belajar dan bimbingan, memahami jenis kegiatan, mengerjakan tugas-tugas, perhatian, keaktifan, motivasi, sikap, minat, umpan balik, kesempatan melaksanakan praktik dalam situasi yang nyata, kesulitan belajar, waktu belajar, istirahat, dan sebagainya.
3. Hasil pembelajaran, baik untuk jangka pendek (sesuai dengan pencapaian indikator), jangka menengah (sesuai dengan target untuk setiap bidang studi/mata pelajaran), dan jangka panjang (setelah peserta didik terjun ke masyarakat).

C. RUANG LINGKUP EVALUASI PEMBELAJARAN BERBASIS PENILAIAN PROSES DAN HASIL BELAJAR

1. Sikap:
 - a. Apakah sikap peserta didik sudah sesuai dengan apa yang diharapkan?
 - b. Bagaimanakah sikap peserta didik terhadap guru, mata pelajaran, orang tua, suasana sekolah, lingkungan, metoda dan media pembelajaran?
 - c. Bagaimana sikap dan tanggung jawab peserta didik terhadap tugas-tugas yang diberikan oleh guru di sekolah?
 - d. Bagaimana sikap peserta didik terhadap tata tertib sekolah dan kepemimpinan kepala sekolah?
2. Pengetahuan dan pemahaman peserta didik terhadap bahan pelajaran:
 - a. Apakah peserta didik sudah mengetahui dan memahami tugas-tugasnya sebagai warga negara, warga masyarakat, warga sekolah, dan sebagainya?
 - b. Apakah peserta didik sudah mengetahui dan memahami tentang materi yang telah diajarkan?
3. Kecerdasan peserta didik:
 - a. Apakah peserta didik sampai taraf tertentu sudah dapat memecahkan masalah-masalah yang dihadapi, khususnya dalam pelajaran?
 - b. Bagaimana upaya guru meningkatkan kecerdasan peserta didik?
4. Perkembangan jasmani/kesehatan:
 - B. Apakah jasmani peserta didik sudah berkembang secara harmonis?
 - C. Apakah peserta didik sudah mampu menggunakan anggota-anggotabadaannya dengan cekatan?
 - D. Apakah peserta didik sudah memiliki kecakapan dasar dalam olahraga?

- E. Apakah prestasi peserta didik dalam olahraga sudah memenuhi syarat-syarat yang ditentukan?
 - F. Apakah peserta didik sudah dapat membiasakan diri hidup sehat?
5. Keterampilan:
- a. Apakah peserta didik sudah terampil membaca, menulis, dan berhitung?
 - b. Apakah peserta didik sudah terampil menggunakan tangannya untuk menggambar, olah raga, dan sebagainya?

Dengan demikian, hasil belajar merupakan gambaran tentang apa yang harus digali, dipahami, dan dikerjakan peserta didik. Hasil belajar ini merefleksikan keluasan, kedalaman, dan kerumitan (secara bergradasi). Hasil belajar harus digambarkan secara jelas dan dapat diukur dengan teknik-teknik penilaian tertentu. Perbedaan antara kompetensi dengan hasil belajar terdapat pada batasan dan patokan-patokan kinerja peserta didik yang dapat diukur.

Indikator hasil belajar dapat digunakan sebagai dasar penilaian terhadap peserta didik dalam mencapai pembelajaran dan kinerja yang diharapkan. Indikator hasil belajar merupakan uraian kemampuan yang harus dikuasai peserta didik dalam berkomunikasi secara spesifik serta dapat dijadikan ukuran untuk menilai ketercapaian hasil pembelajaran. Peserta didik diberi kesempatan untuk menggunakan pengetahuan, keterampilan, sikap dan nilai-nilai yang sudah mereka kembangkan selama pembelajaran dan dalam menyelesaikan tugas-tugas yang sudah ditentukan. Selama proses ini, guru dapat menilai apakah peserta didik telah mencapai suatu hasil belajar yang ditunjukkan dengan pencapaian beberapa indikator dari hasil belajar tersebut. Apabila hasil belajar peserta didik dapat direfleksikan dalam kebiasaan berpikir dan bertindak, berarti peserta didik tersebut telah mencapai suatu kompetensi.

D.RUANG LINGKUP EVALUASI PEMBELAJARAN DALAM PERSPEKTIF PENILAIAN BERBASIS KELAS

Sesuai dengan petunjuk pengembangan Kurikulum Berbasis Kompetensi yang dikeluarkan oleh Departemen Pendidikan Nasional (2004), maka ruang lingkup penilaian berbasis kelas adalah sebagai berikut:

1. Kompetensi Dasar Mata Pelajaran

Kompetensi dasar pada hakikatnya adalah pengetahuan, keterampilan, sikap dan nilai-nilai yang direfleksikan dalam kebiasaan berfikir dan bertindak setelah peserta didik menyelesaikan suatu aspek atau subjek mata pelajaran tertentu. Kompetensi dasar ini merupakan standar kompetensi minimal mata pelajaran. Kompetensi dasar

merupakan bagian dari kompetensi tamatan. Untuk mencapai kompetensi dasar, perlu adanya materi pembelajaran yang harus dipelajari oleh peserta didik. Bertitik tolak dari materi pelajaran inilah dikembangkan alat penilaian.

2. Kompetensi Rumpun Pelajaran

Rumpun pelajaran merupakan kumpulan dari mata pelajaran atau disiplin ilmu yang lebih spesifik. Dengan demikian, kompetensi rumpun pelajaran pada hakikatnya merupakan pengetahuan, keterampilan, sikap dan nilai-nilai yang direfleksikan dalam kebiasaan berfikir dan bertindak yang seharusnya dicapai oleh peserta didik setelah menyelesaikan rumpun pelajaran tersebut. Misalnya, rumpun mata pelajaran IPS merupakan kumpulan dari disiplin ilmu Ekonomi, Geografi, Sejarah dan Sosiologi. Penilaian kompetensi rumpun pelajaran dilakukan dengan mengukur hasil belajar tamatan. Hasil belajar tamatan merupakan ukuran kompetensi rumpun pelajaran.

Hasil belajar mencerminkan keluasan dan kedalaman serta kerumitan kompetensi yang dirumuskan dalam pengetahuan, perilaku, keterampilan, sikap dan nilai-nilai yang dapat diukur dengan menggunakan berbagai teknik penilaian. Perbedaan hasil belajar dan kompetensi terletak pada batasan dan patokan-patokan kinerja peserta didik yang dapat diukur. Setiap hasil belajar memiliki seperangkat indikator. Anda harus menggunakan indikator sebagai acuan penilaian terhadap peserta didik, apakah hasil pembelajaran sudah tercapai sesuai dengan kinerja yang diharapkan. Setiap rumpun pelajaran menentukan hasil belajar tamatan yang dapat dijadikan acuan dalam pengembangan alat penilaian pada setiap kelas.

3. Kompetensi Lintas Kurikulum

Kompetensi lintas kurikulum merupakan kompetensi yang harus dicapai melalui seluruh rumpun pelajaran dalam kurikulum. Kompetensi lintas kurikulum pada hakikatnya merupakan pengetahuan, keterampilan, sikap dan nilai-nilai yang direfleksikan dalam kebiasaan berfikir dan bertindak, baik mencakup kecakapan belajar sepanjang hayat maupun kecakapan hidup yang harus dicapai oleh peserta didik melalui pengalaman belajar secara berkesinambungan. Penilaian ketercapaian kompetensi lintas kurikulum ini dilakukan terhadap hasil belajar dari setiap rumpun pelajaran dalam kurikulum.

Kompetensi lintas kurikulum yang diharapkan dikuasai peserta didik adalah:

- a. Menjalankan hak dan kewajiban secara bertanggungjawab terutama dalam menjamin perasaan aman dan menghargai sesama.

- b. Menggunakan bahasa untuk berinteraksi dan berkomunikasi dengan orang lain.
- c. Memilih, memadukan dan menerapkan konsep-konsep dan teknik-teknik numerik dan spasial, serta mencari dan menyusun pola, struktur dan hubungan.
- d. Menemukan pemecahan masalah-masalah baru berupa prosedur maupun produk teknologi melalui penerapan dan penilaian pengetahuan, konsep, prinsip dan prosedur yang telah dipelajari, serta memilih, mengembangkan, memanfaatkan, mengevaluasi, dan mengelola teknologi komunikasi/ informasi.
- e. Berpikir kritis dan bertindak secara sistematis dalam setiap pengambilan keputusan berdasarkan pemahaman dan penghargaan terhadap dunia fisik, makhluk hidup, dan teknologi.
- f. Berwawasan kebangsaan dan global, terampil serta aktif berpartisipasi dalam kehidupan berbangsa dan bernegara dilandasi dengan pemahaman terhadap nilai-nilai dan konteks budaya, geografi dan sejarah.
- g. Beradab, berbudaya, bersikap religius, bercitarasa seni, susila, kreatif dengan menampilkan dan menghargai karya artistik dan intelektual, serta meningkatkan kematangan pribadi.
- h. Berpikir terarah/terfokus, berpikir lateral, memperhitungkan peluang dan potensi, serta luwes untuk menghadapi berbagai kemungkinan.
- i. Percaya diri dan komitmen dalam bekerja, baik secara mandiri maupun bekerjasama.

4. Kompetensi Tamatan

Kompetensi tamatan merupakan pengetahuan, keterampilan, sikap dan nilai-nilai yang direfleksikan dalam kebiasaan berfikir dan bertindak setelah peserta didik menyelesaikan jenjang pendidikan tertentu. Kompetensi tamatan ini merupakan batas dan arah kompetensi yang harus dimiliki peserta didik setelah mengikuti pembelajaran suatu pelajaran tertentu. Untuk meluluskan tamatan diperlukan kompetensi lulusan. Kompetensi lulusan suatu jenjang sekolah dapat dijabarkan dari visi dan misi yang ditetapkan sekolah. Acuan untuk merumuskan kompetensi lulusan adalah struktur keilmuan mata pelajaran, perkembangan psikologi peserta didik, dan persyaratan yang ditentukan oleh pengguna lulusan (jenjang sekolah selanjutnya dan atau dunia kerja).

Sejalan dengan tujuan pendidikan nasional, kompetensi yang diharapkan dimiliki oleh lulusan atau tamatan sekolah dapat dirumuskan sebagai berikut:

- a. Berkenaan dengan aspek afektif, peserta didik memiliki keimanan dan ketakwaan terhadap Tuhan Yang Maha Esa sesuai dengan ajaran agama Ruang Lingkup, Karakteristik dan Pendekatan Evaluasi Pembelajaran dan kepercayaan masing-masing yang tercermin dalam perilaku sehari-hari, memiliki nilai-nilai etika dan estetika, serta mampu mengamalkan dan mengekspresikannya dalam kehidupan sehari-hari, memiliki nilai-nilai demokrasi, toleransi, dan humaniora, serta menerapkannya dalam kehidupan bermasyarakat, berbangsa dan bernegara, baik dalam lingkup nasional maupun global.
- b. Berkenaan dengan aspek kognitif, peserta didik dapat menguasai ilmu, teknologi dan kemampuan akademik untuk melanjutkan pendidikan ke jenjang yang lebih tinggi.
- c. Berkenaan dengan aspek psikomotorik, peserta didik memiliki keterampilan berkomunikasi, keterampilan hidup, dan mampu beradaptasi dengan perkembangan lingkungan sosial, budaya dan lingkungan alam, baik lokal, regional, maupun global; memiliki kesehatan jasmani dan rohani yang bermanfaat untuk melaksanakan tugas/kegiatan sehari-hari.

5. Pencapaian Keterampilan Hidup

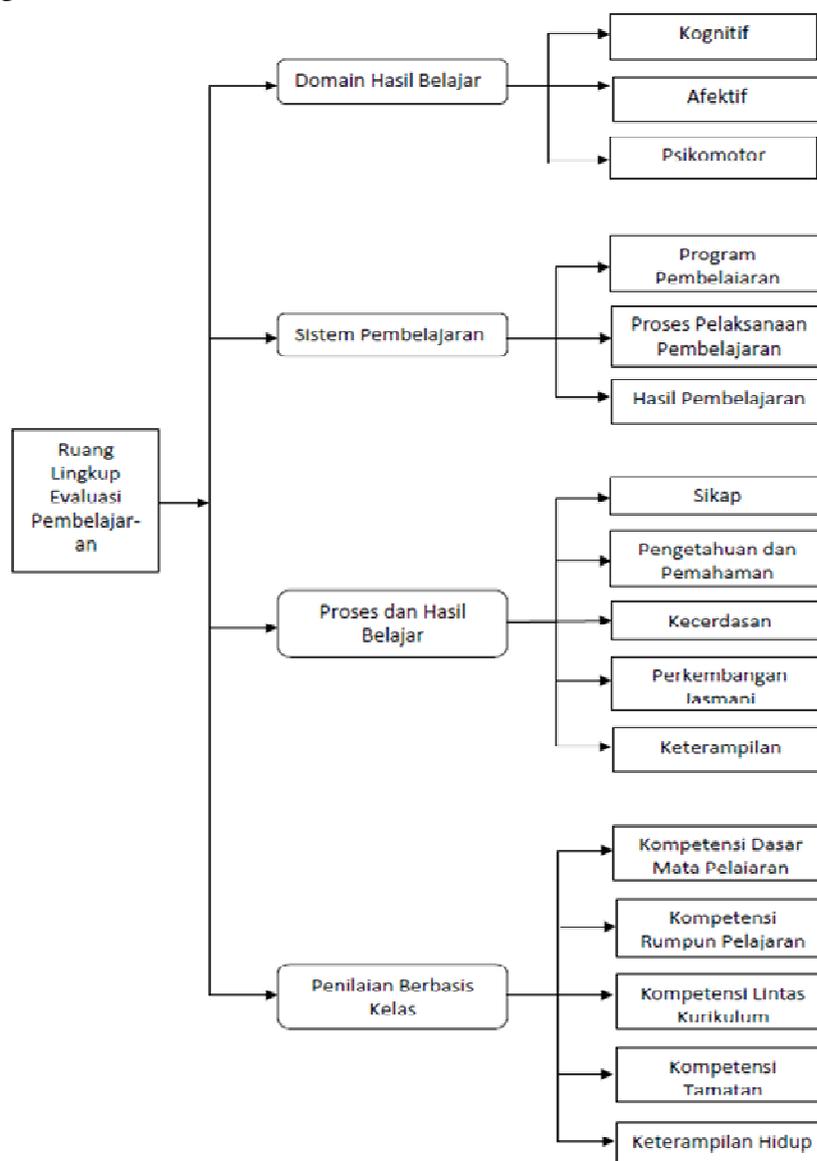
Penguasaan berbagai kompetensi dasar, kompetensi lintas kurikulum, kompetensi rumpun pelajaran dan kompetensi tamatan melalui berbagai pengalaman belajar dapat memberikan efek positif (*nurturan effects*) dalam bentuk kecakapan hidup (*life skills*). Kecakapan hidup yang dimiliki peserta didik melalui berbagai pengalaman belajar ini, juga perlu Anda nilai sejauhmana kesesuaiannya dengan kebutuhan mereka untuk dapat bertahan dan berkembang dalam kehidupannya di lingkungan keluarga, sekolah dan masyarakat. Jenis-jenis kecakapan hidup yang perlu Anda nilai antara lain:

- a. Keterampilan diri (keterampilan personal) yang meliputi: penghayatan diri sebagai makhluk Tuhan Yang Maha Esa, motivasi berprestasi, komitmen, percaya diri, dan mandiri.
- b. Keterampilan berpikir rasional, yang meliputi: berpikir kritis dan logis, berpikir sistematis, terampil menyusun rencana secara sistematis, dan terampil memecahkan masalah secara sistematis.
- c. Keterampilan sosial, yang meliputi: keterampilan berkomunikasi lisan dan tertulis; keterampilan bekerjasama, kolaborasi, lobi; keterampilan berpartisipasi; keterampilan mengelola konflik; dan keterampilan mempengaruhi orang lain.
- d. Keterampilan akademik, yang meliputi: keterampilan merancang, melaksanakan, dan melaporkan hasil penelitian ilmiah; keterampilan

membuat karya tulis ilmiah; keterampilan mentransfer dan mengaplikasikan hasil-hasil penelitian untuk memecahkan masalah, baik berupa proses maupun produk.

- e. Keterampilan vokasional, yang meliputi: keterampilan menemukan algoritma, model, prosedur untuk mengerjakan suatu tugas; keterampilan melaksanakan prosedur; dan keterampilan mencipta produk dengan menggunakan konsep, prinsip, bahan dan alat yang telah dipelajari.

Secara keseluruhan, dapat melihat ruang lingkup evaluasi pembelajaran pada gambar berikut ini:



Gambar 2.1. Ruang Lingkup Evaluasi Pembelajaran

E. KARAKTERISTIK ALAT UKUR YANG BAIK

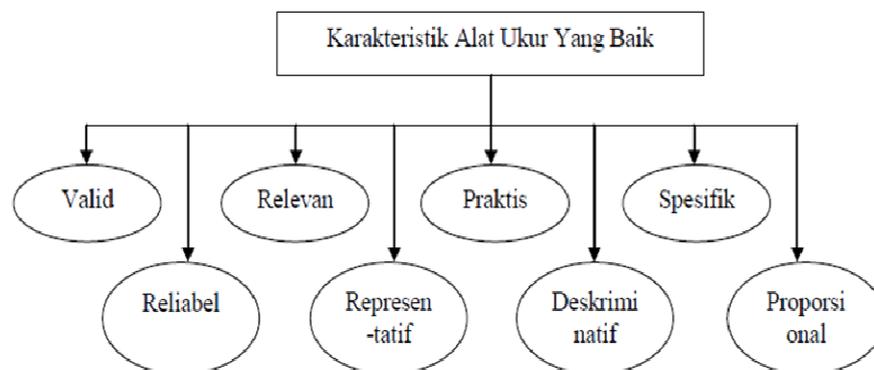
Evaluasi sangat berguna untuk meningkatkan kualitas proses dan hasil pembelajaran. Pentingnya evaluasi dalam pembelajaran, dapat dilihat dari tujuan dan fungsi evaluasi maupun sistem pembelajaran itu sendiri. Evaluasi tidak dapat dipisahkan dari pembelajaran, sehingga guru mau tidak mau harus melakukan evaluasi pembelajaran. Melalui evaluasi, Anda dapat melihat tingkat kemampuan peserta didik, baik secara kelompok maupun individual. Anda juga dapat melihat berbagai perkembangan hasil belajar peserta didik, baik yang menyangkut domain kognitif, afektif maupun psikomotor. Pada akhirnya, guru akan memperoleh gambaran tentang keefektifan proses pembelajaran. Setelah Anda memahami pentingnya evaluasi dalam kegiatan pembelajaran di sekolah, tentunya Anda juga perlu tahu apa karakteristik dari alat ukur yang baik.

Pemahaman tentang alat ukur ini menjadi penting karena dalam praktik evaluasi atau penilaian di sekolah, pada umumnya guru melakukan proses pengukuran. Dalam pengukuran tentu harus ada alat ukur (instrumen), baik yang berbentuk tes maupun nontes. Alat ukur tersebut ada yang baik, ada pula yang kurang baik. Alat ukur yang baik adalah alat ukur yang memenuhi syarat-syarat atau kaidah-kaidah tertentu, dapat memberikan data yang akurat sesuai dengan fungsinya, dan hanya mengukur sampel perilaku tertentu. Secara sederhana, Zainal Arifin (2011: 69) mengemukakan karakteristik instrumen evaluasi yang baik adalah “valid, reliabel, relevan, representatif, praktis, deskriptif, spesifik dan proporsional”.

1. Valid, artinya suatu alat ukur dapat dikatakan valid jika betul-betul mengukur apa yang hendak diukur secara tepat. Misalnya, alat ukur matapelajaran Ekonomi, maka alat ukur tersebut harus betul-betul dan hanya mengukur kemampuan peserta didik dalam mempelajari Ekonomi, tidak boleh dicampuradukkan dengan materi pelajaran yang lain. Validitas suatu alat ukur dapat ditinjau dari berbagai segi, antara lain validitas ramalan (*predictive validity*), validitas bandingan (*concurrent validity*), dan validitas isi (*content validity*), validitas konstruk (*construct validity*), dan lain-lain. Penjelasan tentang validitas ini dapat Anda baca uraian berikutnya.
2. Reliabel, artinya suatu alat ukur dapat dikatakan reliabel atau handal jika ia mempunyai hasil yang taat asas (*consistent*). Misalnya, suatu alat ukur diberikan kepada sekelompok peserta didik saat ini, kemudian diberikan lagi kepada sekelompok peserta didik yang sama pada saat yang akan datang, dan ternyata hasilnya sama atau mendekati sama,

maka dapat dikatakan alat ukur tersebut mempunyai tingkat reliabilitas yang tinggi.

3. Relevan, artinya alat ukur yang digunakan harus sesuai dengan standar kompetensi, kompetensi dasar, dan indikator yang telah ditetapkan. Alat ukur juga harus sesuai dengan domain hasil belajar, seperti domain kognitif, afektif, dan psikomotor. Jangan sampai ingin mengukur domain kognitif menggunakan alat ukur non-tes. Hal ini tentu tidak relevan.
4. Representatif, artinya materi alat ukur harus betul-betul mewakili dari seluruh materi yang disampaikan. Hal ini dapat dilakukan bila guru menggunakan silabus sebagai acuan pemilihan materi tes. Guru juga harus memperhatikan proses seleksi materi, mana materi yang bersifat aplikatif dan mana yang tidak, mana yang penting dan mana yang tidak.
5. Praktis, artinya mudah digunakan. Jika alat ukur itu sudah memenuhi syarat tetapi sukar digunakan, berarti tidak praktis. Kepraktisan ini bukan hanya dilihat dari pembuat alat ukur (guru), tetapi juga bagi orang lain yang ingin menggunakan alat ukur tersebut.
6. Deskriminatif, artinya adalah alat ukur itu harus disusun sedemikian rupa, sehingga dapat menunjukkan perbedaan-perbedaan yang sekecil apapun. Semakin baik suatu alat ukur, maka semakin mampu alat ukur tersebut menunjukkan perbedaan secara teliti. Untuk mengetahui apakah suatu alat ukur cukup deskriminatif atau tidak, biasanya didasarkan atas uji daya pembeda alat ukur tersebut.
7. Spesifik, artinya suatu alat ukur disusun dan digunakan khusus untuk objek yang diukur. Jika alat ukur tersebut menggunakan tes, maka jawaban tes jangan menimbulkan ambivalensi atau spekulasi.
8. Proporsional, artinya suatu alat ukur harus memiliki tingkat kesulitan yang proporsional antara sulit, sedang dan mudah. Begitu juga ketika menentukan jenis alat ukur, baik tes maupun non-tes.



Gakbar 2.2. Karakteristik Alat Ukur yang Baik

Dalam buku *Successful Teaching* karangan J. Mursell yang diterjemahkan ke dalam bahasa Indonesia oleh J. Mursell dan S. Nasution (tanpa tahun: 23) dikemukakan bahwa ciri-ciri evaluasi yang baik adalah “evaluasi dan hasil langsung, evaluasi dan transfer, dan evaluasi langsung dari proses belajar”.

1. Evaluasi dan hasil Langsung.

Dalam proses pembelajaran, guru sering melakukan kegiatan evaluasi, baik ketika proses pembelajaran sedang berlangsung maupun ketika sesudah proses pembelajaran selesai. Jika evaluasi diadakan ketika proses pembelajaran sedang berlangsung, maka guru ingin mengetahui keefektifan dan kesesuaian strategi pembelajaran dengan tujuan yang ingin dicapai. Jika evaluasi dilakukan sesudah proses pembelajaran selesai, berarti guru ingin mengetahui hasil atau prestasi belajar yang diperoleh peserta didik.

2. Evaluasi dan transfer.

Hal penting yang berkenaan dengan proses belajar adalah kemungkinan mentransfer hasil yang dipelajari ke dalam situasi yang fungsional. Dasar pemikiran ini merupakan asas psikologis yang logis dan rasional. Peserta didik tidak dapat disebut telah menguasai materi ekonomi (misalnya), jika ia belum dapat menggunakannya dalam kehidupan. Apabila suatu hasil belajar tidak dapat ditransfer dan hanya dapat digunakan dalam satu situasi tertentu saja, maka hasil belajar itu disebut hasil belajar palsu. Sebaliknya, jika suatu hasil belajar dapat ditransfer kepada penggunaan yang aktual, maka hasil belajar itu disebut hasil belajar otentik. Jadi, evaluasi yang baik harus mengukur hasil belajar yang otentik dan kemungkinan dapat ditransfer.

Dalam penelitian sering ditemui hasil-hasil pembelajaran yang dicapai tampaknya baik, tetapi sebenarnya hasil itu palsu. Peserta didik dapat mengucapkan kata-kata yang dihafalkan dari buku pelajarannya, tetapi mereka tidak dapat menggunakannya dalam situasi baru. Penguasaan materi pelajaran seperti ini tidak lebih dari “penguasaan beo”. Evaluasi yang menekankan pada hasil-hasil palsu, baik untuk informasi bagi peserta didik maupun untuk tujuan lain, berarti evaluasi itu palsu. Jika peserta didik hanya memiliki pengetahuan yang bersifat informatif, belum tentu menjamin pemahaman dan pengertiannya. Oleh karena itu, penekanan pada pengetahuan yang bersifat informatif tidak akan menghasilkan pola berpikir yang baik. Ada dua sebab mengapa hasil pembelajaran yang mengakibatkan dan berhubungan dengan proses transfer menjadi penting artinya dalam proses evaluasi. *Pertama*, hasil-hasil itu menyatakan secara khusus dan sejelas-jelasnya kepada guru mengenai apa yang sebenarnya terjadi ataupun

tidak terjadi, dan sampai dimana pula telah tercapai hasil belajar yang penuh makna serta otentik sifatnya. *Kedua*, hasil belajar sangat erat hubungannya dengan tujuan peserta didik belajar, sehingga mempunyai efek yang sangat kuat terhadap pembentukan pola dan karakter belajar yang dilakukan peserta didik. Oleh karena itu, belajar hendaknya dilakukan untuk mendapatkan hasil-hasil yang dapat ditransfer dan setiap waktu dapat digunakan menurut keperluannya.

3. Evaluasi langsung dari proses belajar.

Di samping harus mengetahui hasil belajar, Anda juga harus menilai proses belajar. Hal ini dimaksudkan agar proses belajar dapat diorganisasi sedemikian rupa, sehingga dapat mencapai hasil yang optimal. Anda dapat mengetahui proses apa yang dilalui peserta didik dalam mempelajari sesuatu. Misalnya, apakah peserta didik dalam mempelajari ekonomi cukup sekedar ataukah ia mempelajari seluruh materi untuk memahami dan memecahkan masalah-masalah kehidupan.

Penelitian tentang proses belajar yang diikuti oleh peserta didik merupakan suatu hal yang sangat penting. Anda akan mengetahui dimana letak kesulitan peserta didik, kemudian mencari alternatif bagaimana mengatasi kesulitan tersebut. Di samping itu, penelitian tentang proses belajar bermanfaat juga bagi peserta didik itu sendiri. Peserta didik akan melihat kelemahannya, kemudian berusaha memperbaikinya, dan akhirnya dapat mempertinggi hasil belajarnya. Meneliti proses belajar seorang anak bukan pekerjaan yang mudah. Hal ini memerlukan waktu, tenaga, pemikiran, dan pengalaman. Anda dapat menggunakan suatu metode untuk menilai proses belajar dengan memperhatikan prinsip konteks, vokalisasi, sosialisasi, individualisasi, dan urutan (*sequence*).

Seorang peserta didik tidak dapat belajar dengan baik, karena ia tidak menggunakan konteks yang baik. Ia tidak menggunakan bermacam-macam sumber dan tidak menggunakan situasi-situasi yang konkrit. Peserta didik tidak dapat belajar dengan baik, karena tidak mempunyai fokus tertentu, misalnya tidak melihat masalah-masalah pokok yang harus dipecahkannya, atau mungkin pula tidak sesuai dengan bakat dan minatnya (individualisasi) serta tidak mendiskusikannya dengan orang lain (sosialisasi). Dalam evaluasi pembelajaran, Anda jangan terfokus kepada hasil belajar saja, tetapi juga harus memperhatikan transfer hasil belajar dan proses belajar yang dijalani oleh peserta didik.

F. MODEL-MODEL EVALUASI

Pada tahun 1949, Tyler pernah mengemukakan model evaluasi *black box*. Model ini banyak digunakan oleh orang-orang yang melakukan kegiatan evaluasi. Studi tentang evaluasi belum begitu menarik perhatian orang banyak, karena kurang memiliki nilai praktis. Baru sekitar tahun 1960-an studi evaluasi mulai berdiri sendiri menjadi salah satu program studi di perguruan tinggi, tidak hanya di jenjang sarjana (S.1) dan magister (S.2) tetapi juga pada jenjang doctor (S.3). Sekitar tahun 1972, model evaluasi mulai berkembang. Taylor dan Cowley, misalnya, berhasil mengumpulkan berbagai pemikiran tentang model evaluasi dan menerbitkannya dalam suatu buku. Model evaluasi yang dikembangkan lebih banyak menggunakan pendekatan positivisme yang berakar pada teori psikometrik. Dalam model tersebut, pengukuran dan tes masih sangat dominan, sekalipun tidak lagi diidentikkan dengan evaluasi. Penggunaan desain eksperimen seperti yang dikemukakan Campbell dan Stanley (1963) menjadi ciri utama dari model evaluasi. Berkembangnya model evaluasi pada tahun 70-an tersebut diawali dengan adanya pandangan alternatif dari *paraexpert*. Pandangan alternatif yang dilandasi sebuah paradigma fenomenologi banyak menampilkan model evaluasi.

Dari sekian banyak model-model evaluasi yang dikemukakan, tes dan pengukuran tidak lagi menempati posisi yang menentukan. Penggunaannya hanya untuk tujuan-tujuan tertentu saja, bukan lagi menjadi suatu keharusan, seperti ketika model pertama ditampilkan. Tes dan pengukuran tidak lagi menjadi parameter kualitas suatu studi evaluasi yang dilakukan. Perkembangan lain yang menarik dalam model evaluasi ini adalah adanya suatu upaya untuk bersikap eklektik dalam penggunaan pendekatan positivisme maupun fenomenologi yang oleh Patton (1980) disebut *paradigm of choice*. Walaupun usaha ini tidak melahirkan model dalam pengertian terbatas tetapi memberikan alternatif baru dalam melakukan evaluasi.

Dalam studi tentang evaluasi, banyak sekali dijumpai model-model evaluasi dengan format atau sistematika yang berbeda, sekalipun dalam beberapa model ada juga yang sama. Misalnya saja, Said Hamid Hasan (2009) mengelompokkan model evaluasi sebagai berikut :

1. Model evaluasi kuantitatif, yang meliputi: model *Tyler*, model teoritik Taylor dan Maguire, model pendekatan sistem Alkin, model Countenance Stake, model CIPP, model ekonomi mikro.
2. Model evaluasi kualitatif, yang meliputi: model studi kasus, model iluminatif, dan model responsif.

Sementara itu, Kaufman dan Thomas dalam Suharsimi Arikunto dan Cepi Safruddin AJ (2007: 24) membedakan model evaluasi menjadi delapan, yaitu:

1. *Goal Oriented Evaluation Model*, dikembangkan oleh Tyler.
2. *Goal Free Evaluation Model*, dikembangkan oleh Scriven.
3. *Formatif Sumatif Evaluation Model*, dikembangkan oleh Michael Scriven
4. *Countenance Evaluation Model*, dikembangkan oleh Stake.
5. *Responsive Evaluation Model*, dikembangkan oleh Stake.
6. *CSE-UCLA Evaluation Model*, menekankan pada “kapan” evaluasi dilakukan.
7. *CIPP Evaluation Model*, yang dikembangkan oleh Stufflebeam.
8. *Discrepancy Model*, yang dikembangkan oleh Provus.

Ada juga model evaluasi yang dikelompokkan Nana Sudjana dan R.Ibrahim (2007: 234) yang membagi model evaluasi menjadi empat model utama, yaitu “*measurement, congruence, educational system, dan illumination*”. Dari beberapa model evaluasi di atas, beberapa diantaranya akan dikemukakan secara singkat sebagai berikut:

1. Model Tyler

Nama model ini diambil dari nama pengembangnya yaitu Tyler. Dalam buku *Basic Principles of Curriculum and Instruction*, Tyler banyak mengemukakan ide dan gagasannya tentang evaluasi. Salah satu bab dari buku tersebut diberinya judul *how can the the effectiveness of learning experience be evaluated?* Model ini dibangun atas dua dasar pemikiran. *Pertama*, evaluasi ditujukan kepada tingkah laku peserta didik. *Kedua*, evaluasi harus dilakukan pada tingkah laku awal peserta didik sebelum melaksanakan kegiatan pembelajaran dan sesudah melaksanakan kegiatan pembelajaran (hasil). Dasar pemikiran yang kedua ini menunjukkan bahwa seorang evaluator harus dapat menentukan perubahan tingkah laku apa yang terjadi setelah peserta didik mengikuti pengalaman belajar tertentu, dan menegaskan bahwa perubahan yang terjadi merupakan perubahan yang disebabkan oleh pembelajaran.

Penggunaan model Tyler memerlukan informasi perubahan tingkah laku terutama pada saat sebelum dan sesudah terjadinya pembelajaran. Istilah yang populer dikalangan guru adalah tes awal (*pre-test*) dan tes akhir (*post-test*). Model ini mensyaratkan validitas informasi pada tes akhir. Untuk menjamin validitas ini maka perlu adanya kontrol dengan menggunakan disain eksperimen. Model Tyler disebut juga model black box karena model ini sangat menekankan adanya tes awal dan tes akhir. Dengan demikian, apa yang terjadi dalam proses tidak perlu diperhatikan. Dimensi proses ini dianggap sebagai kotak hitam yang

menyimpan segala macam teka-teki. Menurut Tyler, ada tiga langkah pokok yang harus dilakukan, yaitu:

- a. Menentukan tujuan pembelajaran yang akan dievaluasi.
- b. Menentukan situasi dimana peserta didik memperoleh kesempatan untuk menunjukkan tingkah laku yang berhubungan dengan tujuan.
- c. Menentukan alat evaluasi yang akan dipergunakan untuk mengukur tingkah laku peserta didik.

2. Model yang Berorientasi pada Tujuan

Sebelum KBK 2004, kita mungkin pernah mengenal adanya tujuan pembelajaran umum dan tujuan pembelajaran khusus. Model evaluasi ini menggunakan kedua tujuan tersebut sebagai kriteria untuk menentukan keberhasilan. Evaluasi diartikan sebagai proses pengukuran hingga mana tujuan pembelajaran telah tercapai. Model ini banyak digunakan oleh guru-guru karena dianggap lebih praktis untuk menentukan hasil yang diinginkan dengan rumusan yang dapat diukur. Dengan demikian, terdapat hubungan yang logis antara kegiatan, hasil dan prosedur pengukuran hasil. Tujuan model ini adalah membantu merumuskan tujuan dan menjelaskan hubungan antara tujuan dengan kegiatan. Jika rumusan tujuan pembelajaran dapat diobservasi (*observable*) dan dapat diukur (*measurable*), maka kegiatan evaluasi pembelajaran akan menjadi lebih praktis dan simpel.

Model ini dapat membantu kita menjelaskan rencana pelaksanaan pembelajaran dengan proses pencapaian tujuan. Instrumen yang digunakan bergantung kepada tujuan yang ingin diukur. Hasil evaluasi akan menggambarkan tingkat keberhasilan tujuan program pembelajaran berdasarkan kriteria program khusus. Kelebihan model ini terletak pada hubungan antara tujuan dengan kegiatan dan menekankan pada peserta didik sebagai aspek penting dalam program pembelajaran. Kekurangannya adalah memungkinkan terjadinya proses evaluasi melebihi konsekuensi yang tidak diharapkan.

3. Model Pengukuran

Model pengukuran (*measurement model*) banyak mengemukakan pemikiran-pemikiran dari R.Thorndike dan R.L.Ebel. Sesuai dengan namanya, model ini sangat menitikberatkan pada kegiatan pengukuran. Pengukuran digunakan untuk menentukan kuantitas suatu sifat (*attribute*) tertentu yang dimiliki oleh objek, orang maupun peristiwa, dalam bentuk unit ukuran tertentu. Kita dapat menggunakan model ini untuk mengungkap perbedaan-perbedaan individual maupun kelompok dalam hal kemampuan, minat dan sikap. Hasil evaluasi digunakan untuk keperluan seleksi peserta didik, bimbingan, dan perencanaan pendidikan. Objek evaluasi dalam model ini adalah

tingkah laku peserta didik, mencakup hasil belajar (kognitif), pembawaan, sikap, minat, bakat, dan juga aspek-aspek kepribadian peserta didik. Untuk itu, instrumen yang digunakan pada umumnya adalah tes tertulis (*paper and pencil test*) dalam bentuk tes objektif, yang cenderung dibakukan. Oleh sebab itu, dalam menganalisis soal sangat memperhatikan *difficulty index* dan *index of discrimination*. Model ini menggunakan pendekatan Penilaian Acuan Norma (*norm-referenced assessment*).

4. Model Kesesuaian (Ralph W.Tyler, John B.Carrol, and Lee J.Cronbach)

Menurut model ini, evaluasi adalah suatu kegiatan untuk melihat kesesuaian (*congruence*) antara tujuan dengan hasil belajar yang telah dicapai. Hasil evaluasi dapat digunakan untuk menyempurnakan sistem bimbingan peserta didik dan untuk memberikan informasi kepada pihak-pihak yang memerlukan. Objek evaluasi adalah tingkah laku peserta didik, yaitu perubahan tingkah laku yang diinginkan (*intended behaviour*) pada akhir kegiatan pendidikan, baik yang menyangkut aspek kognitif, afektif maupun psikomotor. Teknik evaluasi yang dapat digunakan tidak hanya tes (tulisan, lisan, dan perbuatan), tetapi juga non-tes (observasi, wawancara, skala sikap, dan sebagainya). Model evaluasi ini memerlukan informasi perubahan tingkah laku pada dua tahap, yaitu sebelum dan sesudah kegiatan pembelajaran. Berdasarkan konsep ini, kita perlu melakukan pre-test and post-test. Adapun langkah-langkah yang harus ditempuh dalam model evaluasi ini adalah merumuskan tujuan tingkah laku (*behavioural objectives*), menentukan situasi dimana peserta didik dapat memperlihatkan tingkah laku yang akan dievaluasi, menyusun alat evaluasi, dan menggunakan hasil evaluasi. Oleh sebab itu, model ini menekankan pada pendekatan penilaian acuan patokan (PAP).

5. *Educational System Evaluation Model* (Daniel L.Stufflebeam, Michael Scriven, Robert E.Stake, dan Malcolm M.Provus)

Menurut model ini, evaluasi berarti membandingkan *performance* dari berbagai dimensi (tidak hanya dimensi hasil saja) dengan sejumlah kriteria, baik yang bersifat mutlak/interen maupun relatif/ekstern. Model yang menekankan sistem sebagai suatu keseluruhan ini sebenarnya merupakan penggabungan dari beberapa model, sehingga objek evaluasinya pun diambil dari beberapa model, yaitu (1) model countenance dari Stake, yang meliputi: keadaan sebelum kegiatan pembelajaran berlangsung (*antecedents*), kegiatan yang terjadi dan saling mempengaruhi (*transactions*), hasil yang diperoleh (*outcomes*), (2) model CIPP dari Stufflebeam, yang meliputi *Context, Input, Process*, dan *Product*, (3) model Scriven yang meliputi *instrumental evaluation and*

consequential evaluation, (4) model Provus yang meliputi: *design, operation program, interim products*, dan *terminal products*. Dari keempat model yang tergabung dalam *educational system model*, akan dijelaskan secara singkat tentang dua model, yaitu model *countenance* model CIPP.

Model Stake menitikberatkan evaluasi pada dua hal pokok, yaitu *description* dan *judgement*. Setiap hal tersebut terdiri atas tiga dimensi, seperti telah dijelaskan di atas, yaitu *antecedents (context)*, *transaction (process)*, dan *outcomes (output)*. *Description* terdiri atas dua aspek, yaitu *intents (goals)* dan *observation (effects)* atau yang sebenarnya terjadi. Sedangkan *judgement* terdiri atas dua aspek, yaitu *standard* dan *judgement*. Dalam model ini, evaluasi dilakukan dengan membandingkan antara satu program dengan program lain yang dianggap standar. Stake mengatakan *description* berbeda dengan *judgement* atau menilai. Dalam ketiga dimensi di atas (*antecedents, transaction, outcomes*), perbandingan data tidak hanya untuk menentukan apakah ada perbedaan tujuan dengan keadaan yang sebenarnya tetapi juga dibandingkan dengan standar yang absolut untuk menilai manfaat program. Menurut Stake, suatu hasil penelitian tidak dapat diandalkan jika tidak dilakukan evaluasi.

Model CIPP berorientasi kepada suatu keputusan (*a decision oriented evaluation approach structured*). Tujuannya membantu kepala sekolah dan guru di dalam membuat keputusan. Evaluasi diartikan sebagai suatu proses menggambarkan, memperoleh dan menyediakan informasi yang berguna untuk menilai alternatif keputusan. Sesuai dengan nama modelnya, model ini membagi empat jenis kegiatan evaluasi, yaitu:

- a. *Context evaluation to serve planning decision*, yaitu konteks evaluasi untuk membantu merencanakan keputusan, menentukan kebutuhan yang akan dicapai oleh program pembelajaran, dan merumuskan tujuan program pembelajaran.
- b. *Input evaluation, structuring decision*. Kegiatan evaluasi bertujuan untuk membantu mengatur keputusan, menentukan sumber-sumber yang ada, alternatif apa yang akan diambil, apa rencana dan strategi untuk mencapai tujuan, dan bagaimana prosedur kerja untuk mencapainya.
- c. *Process evaluation, to serve implementing decision*. Kegiatan evaluasi ini bertujuan untuk membantu melaksanakan keputusan. Pertanyaan yang harus dijawab adalah hinggmana suatu rencana telah dilaksanakan, apakah rencana tersebut sesuai dengan prosedur kerja, dan apa yang harus diperbaiki.

d. *Product evaluation, to serve recycling decision.* Kegiatan evaluasi ini bertujuan untuk membantu keputusan selanjutnya. Pertanyaan yang harus Anda jawab adalah hasil apa yang telah dicapai dan apa yang dilakukan setelah program berjalan.

Proses evaluasi tidak hanya berakhir dengan suatu deskripsi mengenai keadaan sistem yang bersangkutan, tetapi harus sampai pada judgment sebagai simpulan dari hasil evaluasi. Model ini menuntut agar hasil evaluasi digunakan sebagai input untuk decision making dalam rangka penyempurnaan sistem secara keseluruhan. Pendekatan yang digunakan adalah penilaian acuan norma (PAN) dan penilaian acuan patokan (PAP).

6. *Illuminative Model* (Malcolm Parlett dan Hamilton)

Jika model *measurement* dan *congruence* lebih berorientasi pada evaluasi kuantitatif-terstruktur, maka model ini lebih menekankan pada evaluasi kualitatif-terbuka (*open-ended*). Kegiatan evaluasi dihubungkan dengan learning milieu, dalam konteks madrasah sebagai lingkungan material dan psiko-sosial, dimana guru dan peserta didik dapat berinteraksi. Tujuan evaluasi adalah untuk mempelajari secara cermat dan hati-hati terhadap pelaksanaan sistem pembelajaran, faktor-faktor yang mempengaruhinya, kelebihan dan kekurangan sistem, dan pengaruh sistem terhadap pengalaman belajar peserta didik. Hasil evaluasi lebih bersifat deskriptif dan interpretasi, bukan pengukuran dan prediksi. Model ini lebih banyak menggunakan judgment. Fungsi evaluasi adalah sebagai input untuk kepentingan pengambilan keputusan dalam rangka penyesuaian dan penyempurnaan sistem pembelajaran yang sedang dikembangkan.

Objek evaluasi model ini mencakup latar belakang dan perkembangan sistem pembelajaran, proses pelaksanaan sistem pembelajaran, hasil belajar peserta didik, kesukaran-kesukaran yang dialami dari perencanaan sampai dengan pelaksanaan, termasuk efek samping dari sistem pembelajaran itu sendiri. Pendekatan yang digunakan lebih menyerupai pendekatan yang diterapkan dalam bidang antropologi sosial, psikiatri, dan sosiologi. Cara-cara yang digunakan tidak bersifat standard, melainkan bersifat fleksibel dan selektif. Berdasarkan tujuan dan pendekatan evaluasi dalam model ini, maka ada tiga fase evaluasi yang harus ditempuh, yaitu: *observe, inquiry further, dan seek to explain.*

7. Model Responsif

Sebagaimana model illuminatif, model ini juga menekankan pada pendekatan kualitatif-naturalistik. Evaluasi tidak diartikan sebagai pengukuran melainkan pemberian makna atau melukiskan sebuah

realitas dari berbagai perspektif orang-orang yang terlibat, berminat dan berkepentingan dengan program pembelajaran. Tujuan evaluasi adalah untuk memahami semua komponen program pembelajaran melalui berbagai sudut pandangan yang berbeda. Sesuai dengan pendekatan yang digunakan, maka model ini kurang percaya terhadap hal-hal yang bersifat kuantitatif. Instrumen yang digunakan pada umumnya mengandalkan observasi langsung maupun tak langsung dengan interpretasi data yang impresionistik. Langkah-langkah kegiatan evaluasi meliputi observasi, merekam hasil wawancara, mengumpulkan data, mengecek pengetahuan awal (*preliminary understanding*) peserta didik dan mengembangkan disain atau model. Berdasarkan langkah-langkah ini, evaluator mencoba responsif terhadap orang-orang yang berkepentingan pada hasil evaluasi. Hal yang penting dalam model responsif adalah pengumpulan dan sintesis data.

Kelebihan model ini adalah peka terhadap berbagai pandangan dan kemampuannya mengakomodasi pendapat yang ambigu serta tidak fokus. Sedangkan kekurangannya antara lain (1) pembuat keputusan sulit menentukan prioritas atau penyederhanaan informasi (2) tidak mungkin menampung semua sudut pandangan dari berbagai kelompok (3) membutuhkan waktu dan tenaga. Evaluator harus dapat beradaptasi dengan lingkungan yang diamati. Untuk mempelajari lebih jauh tentang model ini, silahkan membaca buku Stake (1975) atau Lincoln dan Guba (1985).

Keberhasilan suatu evaluasi pembelajaran secara keseluruhan bukan hanya dipengaruhi penggunaan yang tepat pada sebuah model evaluasi melainkan juga dipengaruhi oleh berbagai faktor. Pertama, tujuan pembelajaran, baik tujuan pembelajaran umum maupun tujuan pembelajaran khusus (*instructional objective*). Seringkali kedua tujuan pembelajaran ini saling bertentangan satu sama lain dilihat dari kebutuhan sekolah, kurikulum, guru, peserta didik, lingkungan, dan sebagainya. Bahkan, kadang-kadang guru sendiri mempunyai tujuan sendiri-sendiri. Semuanya harus dipertimbangkan agar terdapat keseimbangan dan keserasian.

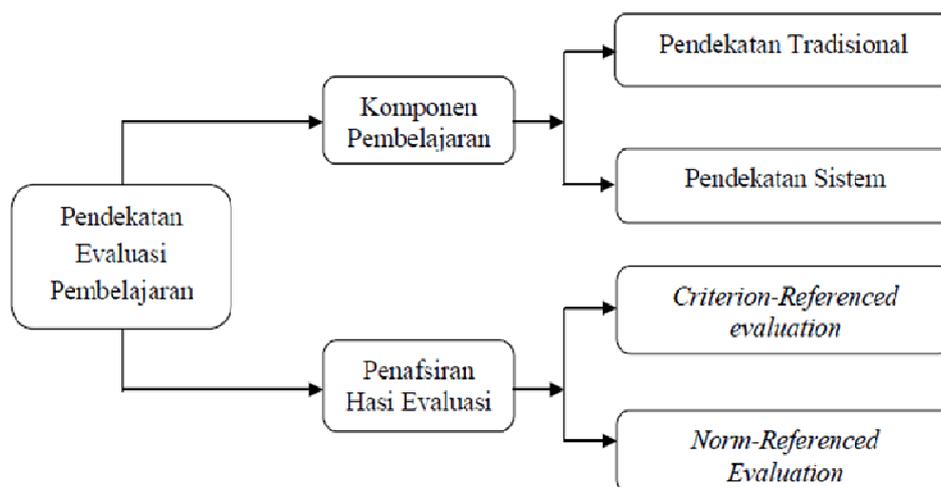
Kedua, sistem sekolah. Faktor ini perlu dipertimbangkan dengan matang dan hati-hati karena melibatkan berbagai komponen yang saling berinteraksi dan ketergantungan. Mengingat kompleksnya sistem sekolah, maka fungsi sekolah juga menjadi ganda. Di satu pihak, sekolah ingin mewariskan kebudayaan masalampau dengan sistem norma, nilai dan adat yang dianggap terbaik untuk generasi muda. Di pihak lain, sekolah berkewajiban mempersiapkan peserta

didik menghadapi masa depan, memperoleh keterampilan dan kemampuan untuk berinovasi, bahkan menghasilkan perubahan. Jadi, sekolah sekaligus bersikap konservatif-radikal serta reaksioner-progresif. Oleh sebab itu, peranan evaluasi menjadi sangat penting. Tujuannya adalah untuk melihat dan mempertimbangkan hal-hal apa yang perlu diberikan di sekolah. Begitu juga bentuk kurikulum dan silabus mata pelajaran sangat bergantung pada evaluasi yang dilaksanakan oleh guru-guru di sekolah, sehingga timbul masalah lainnya yaitu teknik evaluasi apa yang akan digunakan untuk mencapai tujuan itu.

Ketiga, pembinaan guru. Banyak program pembinaan guru yang belum menyentuh secara langsung tentang evaluasi. Program pembinaan guru lebih banyak difokuskan kepada pengembangan kurikulum dan metodologi pembelajaran. Hal ini pula yang menyebabkan perbaikan sistem evaluasi pembelajaran menjadi kurang efektif. Guru juga sering dihadapkan dengan beragam kegiatan, seperti membuat persiapan mengajar, mengikuti kegiatan ekstra kurikuler, penyesuaian diri, dan kegiatan administratif lainnya. Artinya, bagaimana mungkin kualitas sistem evaluasi pembelajaran di sekolah dapat ditingkatkan, bila fokus pembinaan guru hanya menyentuh domain-domain tertentu saja, ditambah lagi dengan kesibukan-kesibukan guru di luar tugas pokoknya sebagai pengajar.

G. PENDEKATAN EVALUASI

Pendekatan merupakan sudut pandang seseorang dalam mempelajari sesuatu. Dengan demikian, pendekatan evaluasi merupakan sudut pandang seseorang dalam menelaah atau mempelajari evaluasi. Dilihat dari komponen pembelajaran, pendekatan evaluasi dapat dibagi dua, yaitu pendekatan tradisional dan pendekatan sistem. Dilihat dari penafsiran hasil evaluasi, pendekatan evaluasi dibagi menjadi dua, yaitu *criterion-referenced evaluation* dan *norm referenced evaluation*. Lihat gambar berikut ini.



Gambar 2.3. Pendekatan Evaluasi Pembelajaran

1. Pendekatan tradisional

Pendekatan ini berorientasi kepada praktik evaluasi yang telah berjalan selama ini di sekolah yang ditujukan kepada perkembangan aspek intelektual peserta didik. Aspek-aspek keterampilan dan pengembangan sikap kurang mendapat perhatian yang serius. Peserta didik hanya dituntut untuk menguasai mata pelajaran. Kegiatan-kegiatan evaluasi juga lebih difokuskan kepada komponen produk saja, sementara komponen proses cenderung diabaikan. Hasil kajian Spencer cukup memberikan gambaran betapa pentingnya evaluasi pembelajaran. Ia mengemukakan sejumlah isi pendidikan yang dapat dijadikan dasar pertimbangan untuk merumuskan tujuan pendidikan secara komprehensif dan pada gilirannya menjadi acuan dalam membuat perencanaan evaluasi. Namun demikian, tidak sedikit guru mengalami kesulitan untuk mengembangkan sistem evaluasi di sekolah karena bertentangan dengan tradisi yang selama ini sudah berjalan. Misalnya, ada tradisi bahwa target kuantitas kelulusan setiap sekolah harus di atas 95 %, begitu juga untuk kenaikan kelas. Ada juga tradisi bahwa dalam mata pelajaran tertentu nilai peserta didik dalam buku raport harus minimal enam. Seharusnya, kebijakan evaluasi lebih menekankan kepada target kualitas yaitu kepentingan dan kebermaknaan pendidikan bagi anak.

2. Pendekatan sistem

Sistem adalah totalitas dari berbagai komponen yang saling berhubungan dan ketergantungan. Jika pendekatan sistem dikaitkan dengan evaluasi, maka pembahasan lebih difokuskan kepada komponen evaluasi, yang meliputi: komponen kebutuhan dan feasibility, komponen input, komponen proses, dan komponen produk. Dalam bahasa *Stufflebeam* disingkat CIPP, yaitu *context, input, process*

dan *prudent*. Komponen-komponen ini harus menjadi landasan pertimbangan dalam evaluasi pembelajaran secara sistematis. Berbeda dengan pendekatan tradisional yang hanya menyentuh komponen produk saja, yaitu perubahan perilaku apa yang terjadi pada peserta didik setelah mengikuti proses pembelajaran. Pendekatan ini tentu tidak salah, hanya tidak sistematis. Padahal, kita juga tahu bahwa hasil belajar tidak akan ada bila tidak melalui proses, dan proses tidak bisa berjalan bila tidak ada masukan dan guru yang melaksanakan.

Dalam literatur modern tentang evaluasi, terdapat dua pendekatan yang dapat digunakan untuk menafsirkan hasil evaluasi, yaitu penilaian acuan patokan (*criterion-referenced evaluation*) dan penilaian acuan norma (*norm-referenced evaluation*). Artinya, setelah kita memperoleh skor mentah dari setiap peserta didik, maka langkah selanjutnya adalah mengubah skor mentah menjadi nilai dengan menggunakan pendekatan tertentu.

1. Penilaian Acuan Patokan (PAP)

Pendekatan ini sering juga disebut penilaian norma absolut. Jika Anda ingin menggunakan pendekatan ini, berarti Anda harus membandingkan hasil yang diperoleh peserta didik dengan sebuah patokan atau kriteria yang secara absolut atau mutlak telah ditetapkan oleh guru. Kita juga dapat menggunakan langkah-langkah tertentu untuk menggunakan PAP, seperti menentukan skor ideal, mencari rata-rata dan simpangan baku ideal, kemudian menggunakan pedoman konversi skala nilai. Pendekatan ini cocok digunakan dalam evaluasi atau penilaian formatif yang berfungsi untuk perbaikan proses pembelajaran. Umumnya, seorang guru yang menggunakan PAP sudah dapat menyusun pedoman konversi skor menjadi skor standar sebelum kegiatan evaluasi dimulai. Oleh sebab itu, hasil pengukuran dari waktu ke waktu dalam kelompok yang sama atau berbeda dapat dipertahankan keajegannya. PAP dapat menggambarkan prestasi belajar peserta didik secara objektif apabila alat ukur yang digunakan adalah alat ukur yang standar.

2. Penilaian Acuan Norma (PAN)

Salah satu perbedaan PAP dengan PAN adalah penggunaan tolak ukur hasil/skor sebagai pembanding. Pendekatan ini membandingkan skor setiap peserta didik dengan teman satu kelasnya. Makna nilai dalam bentuk angka maupun kualifikasi memiliki sifat relatif. Artinya, jika Anda sudah menyusun pedoman konversi skor untuk suatu kelompok, maka pedoman itu hanya berlaku untuk kelompok itu saja dan tidak berlaku untuk kelompok yang lain, karena distribusi skor peserta didik sudah berbeda.

BAB III

PROSEDUR PENGEMBANGAN EVALUASI PEMBELAJARAN

A. PERENCANAAN EVALUASI

Langkah pertama yang perlu dilakukan dalam kegiatan evaluasi adalah membuat perencanaan. Perencanaan ini penting karena akan mempengaruhi langkah-langkah selanjutnya, bahkan mempengaruhi keefektifan prosedur evaluasi secara menyeluruh. W. James Popham (1974: 159) mengemukakan maksud perencanaan evaluasi adalah *"to facilitate gathering data, thereby making possible valid statements about the effect or out comes of the program, practice, or policy under study"*.

Robert H. Davis, dkk. (1974: 81-82) mengemukakan tiga kegunaan dari perencanaan evaluasi, yaitu:

1. *Evaluation plan helps you to determine whether or not you have stated your objective in behavioral terms. If the conditions, behavior or standards or objective have been stated ambiguosly, you will have difficulty designing a test to measure student achievement.*
2. *Evaluation plan early in the design process is that you will be prepared to collect the information you need when it is available.*
3. *Evaluation plan is that it provides sufficient time for test design. To design a good test requires careful preparation, and the quality of a test usually improves if it can be designed in a leisurely fashion.*

Implikasinya adalah perencanaan evaluasi harus dirumuskan secara jelas dan spesifik, terurai dan komprehensif, sehingga perencanaan tersebut bermakna dalam menentukan langkah-langkah selanjutnya. Berdasarkan perencanaan evaluasi yang matang inilah, dapat menetapkan tujuan-tujuan tingkah laku (*behavioral objective*) atau indikator yang akan dicapai, dapat mempersiapkan pengumpulan data dan informasi yang dibutuhkan serta dapat menggunakan waktu yang tepat.

Jika di dalam evaluasi itu jelas-jelas akan menggunakan tes, maka ada baiknya kita simak pendapat Norman E. Gronlund (1985) tentang langkah-langkah yang harus ditempuh dalam perencanaan suatu tes sebagai berikut:

1. Menentukan tujuan tes (*detrtermine the purpose of the test*).
2. Mengidentifikasi hasil belajar yang akan diukur melalui tes (*identify the learning outcomes to be measured by the test*).

3. Merumuskan hasil belajar dalam bentuk perilaku yang spesifik dan dapat diamati (*define the learning outcomes in the terms of specific, observable behavior*).
4. Menyusun garis besar materi pelajaran yang akan diukur melalui tes (*outline the subject matter to be measured by the test*).
5. Menyiapkan suatu tabel yang spesifik atau kisi-kisi (*prepare a table of specifications*).
6. Menggunakan tabel spesifik sebagai dasar untuk persiapan tes (*use the table of specifications as basis for preparing test*).

Berdasarkan uraian di atas, maka dalam perencanaan evaluasi, ada beberapa hal yang harus diperhatikan, seperti: tujuan, kisi-kisi, menulis soal, uji-coba dan analisis soal, revisi dan merakit soal.

1. Menentukan Tujuan Evaluasi

Dalam melaksanakan evaluasi, tentu mempunyai maksud atau tujuan tertentu. Tujuan evaluasi jangan terlalu umum, karena tidak dapat menuntun dalam menyusun soal. Misalnya, tujuan evaluasi adalah untuk mengetahui tingkat pencapaian suatu program pembelajaran atau untuk mengetahui tingkat penguasaan peserta didik terhadap materi pelajaran yang disampaikan. Tujuan evaluasi dapat jugadirumuskan untuk mengetahui kesulitan belajar peserta didik dalam proses pembelajaran. Dengan kata lain, tujuan evaluasi harus dirumuskan sesuai dengan jenis evaluasi yang akan dilakukan, seperti formatif, sumatif, diagnostik, penempatan atau seleksi. Dalam penilaian hasil belajar, tujuan harus memperhatikan domain hasil belajar. Menurut Bloom, dkk. (1956) hasil belajar dapat dikelompokkan dalam tiga domain, yaitu:

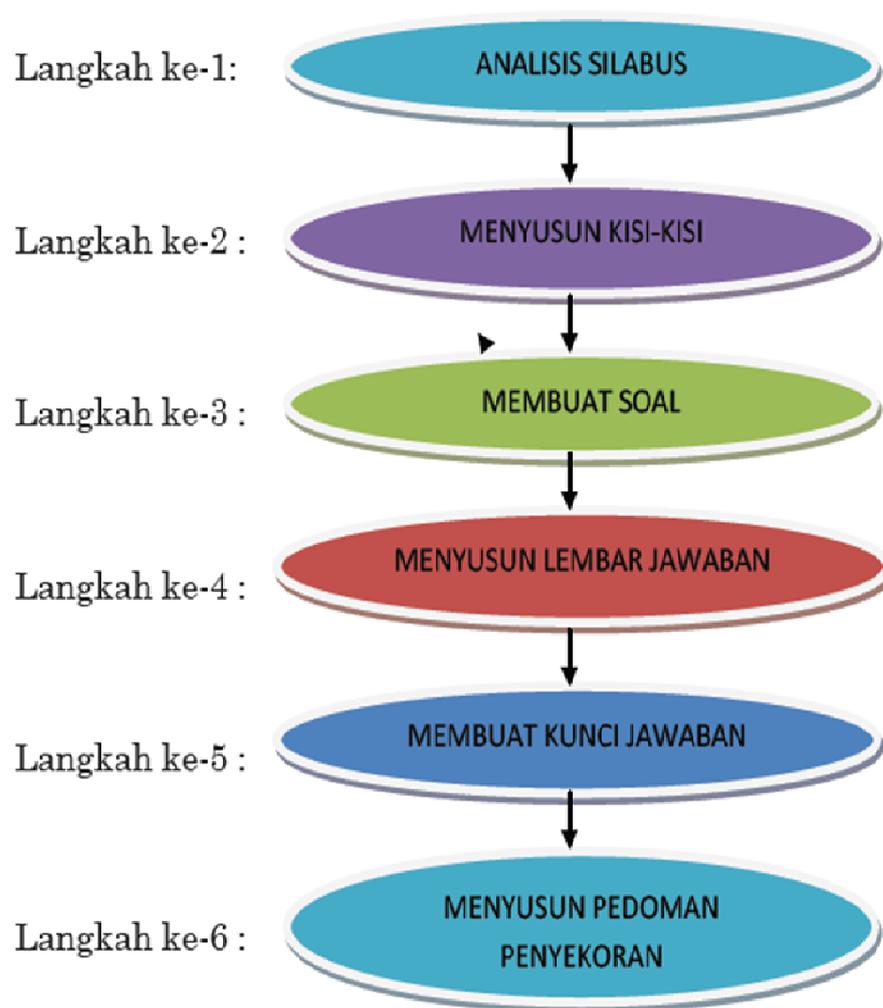
- a) Domain kognitif (*cognitif domain*)
 - 1) Pengetahuan (*knowledge*)
 - 2) Pemahaman (*comprehension*)
 - 3) Aplikasi (*aplication*)
 - 4) Analisis (*analysis*)
 - 5) Sintesis (*synthesis*)
 - 6) Evaluasi (*evaluation*)
- b) Domain afektif (*affective domain*)
 - 1) Penerimaan (*recieving*)
 - 2) Respons (*responding*)
 - 3) Penilaian (*valuing*)
 - 4) Organisasi (*organization*)
 - 5) Karakterisasi (*characterization by a value or value-complex*)
- c) Domain psikomotor (*psychomotor domain*)
 - 1) Persepsi (*perception*)

- 2) Kesiapan melakukan sesuatu pekerjaan (*set*)
- 3) Respons terbimbing (*guided response*)
- 4) Kemahiran (*complex overt response*)
- 5) Adaptasi (*adaptation*)
- 6) Orijinasi (*origination*)

2. Menyusun Kisi-kisi

Penyusunan kisi-kisi dimaksudkan agar materi evaluasi betul-betul representatif dan relevan dengan materi pelajaran yang sudah diberikan oleh guru kepada peserta didik. Jika materi evaluasi tidak relevan dengan materi pelajaran yang telah diberikan, maka akan berakibat hasil evaluasi itu kurang baik. Begitu juga jika materi evaluasi terlalu banyak dibandingkan dengan materi pelajaran, maka akan berakibat sama. Untuk melihat apakah materi evaluasi relevan dengan materi pelajaran atau apakah materi evaluasi terlalu banyak atau kurang, kita harus menyusun kisi-kisi (*lay-out atau blue-print atau table of specifications*).

Kisi-kisi adalah format pemetaan soal yang menggambarkan distribusi item untuk berbagai topik atau pokok bahasan berdasarkan jenjang kemampuan tertentu. Fungsi kisi-kisi adalah sebagai pedoman untuk menulis soal atau merakit soal menjadi perangkat tes. Jika memiliki kisi-kisi yang baik, maka akan memperoleh perangkat soal yang relatif sama sekalipun penulis soalnya berbeda. Dalam konteks penilaian hasil belajar, kisi-kisi disusun berdasarkan silabus setiap mata pelajaran. Jadi, harus melakukan analisis silabus terlebih dahulu. Perhatikan langkah-langkah berikut ini:



Gambar 3.1 Langkah Melakukan Analisis Silabus

Dalam praktiknya, seringkali guru di madrasah membuat soal langsung dari buku sumber. Hal ini jelas sangat keliru, karena buku sumber belum tentu sesuai dengan silabus. Kisi-kisi ini menjadi penting dalam perencanaan evaluasi, karena didalamnya terdapat sejumlah indikator sebagai acuan dalam menulis soal. Kisi-kisi soal yang baik harus memenuhi persyaratan tertentu, antara lain:

- a. Representatif, yaitu harus betul-betul mewakili isi kurikulum yang akan dievaluasi.
- b. Komponen-komponennya harus terurai/rinci, jelas, dan mudah dipahami.
- c. Soalnya dapat dibuat sesuai dengan indikator dan bentuk soal yang ditetapkan.

Sebenarnya, format kisi-kisi tidak ada yang baku, karena itu banyak model format yang dikembangkan para pakar evaluasi. Namun

demikian, sekedar untuk memperoleh gambaran, format kisi-kisi dapat dibagi menjadi dua komponen pokok, yaitu komponen identitas dan komponen matriks. Komponen identitas ditulis di bagian atas matriks, sedangkan komponen matriks dibuat dalam bentuk kolom yang sesuai. Komponen identitas meliputi jenis/jenjang sekolah, jurusan/program studi (bila ada), bidang studi/mata pelajaran, tahun ajaran dan semester, kurikulum acuan, alokasi waktu, jumlah soal keseluruhan, dan bentuk soal. Sedangkan komponen matriks terdiri atas kompetensi dasar, materi, jumlah soal, jenjang kemampuan, indikator, dan nomor urut soal. Contoh:

Kisi-Kisi Soal Ujian Akhir Semester

Nama Sekolah :

Mata Pelajaran :

Jurusan/Program Studi :

Kurikulum Acuan :

Alokasi Waktu :

Jumlah Soal :

Standar Kompetensi :

No	Kompetensi Dasar	Materi	Indikator	Jenjang Kemampuan	Bentuk Soal	Nomor Soal

Catatan: apabila bentuk soal yang akan digunakan lebih dari satu, sebaiknya dimasukkan ke dalam komponen matriks.

Salah satu unsur penting dalam komponen matriks adalah indikator. Indikator adalah rumusan pernyataan sebagai bentuk ukuran spesifik yang menunjukkan ketercapaian kompetensi dasar dengan menggunakan kata kerja operasional (KKO). Perhatikan contoh-contoh berikut ini:

- a. Menyebutkan empat komponen dalam sistem komputer.
- b. Menjelaskan fungsi monitor dalam pesawat komputer.
- c. Membedakan antara *hard-ware* dengan *soft-ware*.

Perhatikan juga indikator dalam matriks berikut ini:

Mata Pelajaran : IPS

Kelas/Semester : I/I

Standar Kompetensi	Kompetensi Dasar	Indikator
Memahami identitas diri dan keluarga, serta sikap saling menghormati dalam kemajemukan keluarga	Mengidentifikasi identitas diri, keluarga, dan kerabat. 1.2 Menceritakan pengalaman diri	Menyebutkan identitas diri secara lisan di depan teman-temannya. Menceritakan pengalamannya dalam bentuk karangan sederhana.

Dalam praktiknya, penggunaan kata kerja operasional untuk setiap indikator harus disesuaikan dengan domain dan jenjang kemampuan yang diukur. Berikut contoh rumusan kata kerja operasional.

a. Domain kognitif :

- 1) Pengetahuan/ingatan: mendefinisikan, memberikan, mengidentifikasi, memberi nama, menyusun daftar, mencocokkan, menyebutkan, membuat garis besar, menyatakan kembali, memilih, menyatakan, dan sebagainya.
- 2) Pemahaman: mengubah, mempertahankan, membedakan, memprakirakan, menjelaskan, menyatakan secara luas, menyimpulkan, memberi contoh, melukiskan kata-kata sendiri, meramalkan, menuliskan kembali, meningkatkan, dan sebagainya.
- 3) Penerapan: menghitung, mendemonstrasikan, mengungkapkan, mengerjakan dengan teliti, menjalankan, menghubungkan, menunjukkan, memecahkan, menggunakan, dan sebagainya.
- 4) Analisa: mengurai, membuat diagram, memisah-misahkan, menggambarkan kesimpulan, membuat garis besar, menghubungkan, merinci, dan sebagainya.
- 5) Sintesa: menggolongkan, menggabungkan, menghimpun, menciptakan, merencanakan, menjelaskan, membangkitkan,

mengorganisir, merevisi, menyimpulkan, menceritakan, dan sebagainya.

6) Evaluasi: menilai, membandingkan, mempertentangkan, mengeritik, membeda-bedakan, mempertimbangkan kebenaran, menyokong, dan sebagainya.

b. Domain afektif :

1) Kemauan menerima: bertanya, memilih, menggambarkan, mengikuti, memberi, berpegang teguh, menjawab, menggunakan, dan sebagainya.

2) Kemauan menanggapi: menjawab, membantu, memperbincangkan, memberi nama, menunjukkan, mempraktikkan, mengemukakan, membaca, melaporkan, menuliskan, memberitahu, dan sebagainya.

3) Berkeyakinan: melengkapi, menggambarkan, membeda-bedakan, mengusulkan, bekerjasama, mencoba, dan sebagainya.

4) Ketekunan, ketelitian: merevisi, melaksanakan, memeriksa kebenaran, melayani, dan sebagainya.

c. Domain psikomotor:

Menirukan, menggunakan, artikulasi (mengucapkan dengan nyata, menyatukan dengan menyambung), mewujudkan, membina, menukar, membersihkan, menyusun, menghubungkan, melatih, mengikuti, membuat bagan, melokalisir, mengikat, mencampur, mengasah/menajamkan, mengaduk, mengerjakan denganteliti, memulai, memanaskan, mengidentifikasi, dan sebagainya.

Lebih jauh, S. J. Montage dan J. J Koran (1969) mendefinisikan tujuan tingkah laku sebagai *"a goal for or desired outcome of learning wich is expresed in terms of observable behavior or performance of the leaner"*. Tujuan tingkah laku adalah tujuan atau hasil belajar yang diharapkan dan dinyatakan dalam bentuk tingkah laku atau kinerja peserta didik yang dapat diamati. Dalam kegiatan evaluasi, fungsi utama tujuan tingkah laku adalah sebagai alat yang sistematis untuk merancang cara-cara melakukan evaluasi terhadap tingkah laku peserta didik.

Manfaat adanya indikator adalah (a) dapat memilih materi, metode, media, dan sumber belajar yang tepat, sesuai dengan kompetensi yang telah ditetapkan, dan (b) sebagai pedoman dan pegangan untuk menyusun soal atau instrument penilaian lain yang tepat, sesuai dengan standar kompetensi dan kompetensi dasar yang telah ditetapkan.

Hubungan Indikator Dengan Soal

Mata Pelajaran :

Kelas :

Semester :

Standar Kompetensi :

Kompetensi Dasar :

No	Jenjang Kemampuan	Indikator	Soal-soal		No. Naskah						
			No	Rumusan Soal	I	II	III	IV	V	VI	VII
1	2	3	4	5	6	7	8	9	10	11	12

Keterangan:

Kolom 1: diisi dengan nomor urut indikator. Tiap lembar sebaiknya hanya untuk satu nomor indikator.

Kolom 2: diisi dengan jenjang kemampuan, baik dalam domain kognitif (pengetahuan, pemahaman, aplikasi, analisis, sintesis, evaluasi) maupun domain afektif dan psikomotor.

Kolom 3 : diisi dengan rumusan indikator

Kolom 4: diisi dengan nomor urut soal untuk setiap indikator. Satu indikator dapat disusun untuk beberapa soal.

Kolom 5: diisi dengan rumusan soal

Kolom 6: diisi dengan nomor soal yang bersangkutan pada naskah ujian/tes ke satu.

Kolom 7, 8, 9, dan seterusnya: diisi sama dengan kolom 6.

Setelah dirumuskan tujuan atau kompetensi secara rinci, kita perlu menentukan ruang lingkup materi yang hendak diukur dan perbandingannya. Ruang lingkup materi yang hendak diukur harus sesuai dengan silabus/kurikulum yang digunakan agar derajat keesuaian dapat diperoleh secara optimal. Berat-ringannya bobot bergantung kepada urgensi materi dan kompetensi yang harus dikuasai peserta didik. Di samping itu, juga harus menyusun bentuk soal secara bervariasi agar kelemahan setiap bentuk soal dapat ditutupi oleh bentuk soal yang lain.

Dalam kisi-kisi, harus memperhatikan domain yang akan diukur, seperti telah dikemukakan sebelumnya. Ada pula sistematika yang lebih sederhana yaitu aspek recall, komprehensi, dan aplikasi. Aspek recall berkenaan dengan aspek-aspek pengetahuan tentang istilah-istilah, definisi, fakta, konsep, metode dan prinsip-prinsip. Aspek komprehensi berkenaan dengan

kemampuan-kemampuan antara lain: menjelaskan, menyimpulkan suatu informasi, menafsirkan fakta (grafik, diagram, tabel, dll), mentransferkan pernyataan dari suatu bentuk ke dalam bentuk yang lain (misalnya dari pernyataan verbal kepada non-verbal atau dari verbal ke dalam bentuk rumus), memprakirakan akibat atau konsekuensi logis dari suatu situasi. Aspek aplikasi meliputi kemampuan-kemampuan antara lain: menerapkan hukum/prinsip/teori dalam suasana yang sesungguhnya, memecahkan masalah, membuat (grafik, diagram, dan lain-lain), mendemonstrasikan penggunaan suatu metode, prosedur, dan lain-lain.

Tingkat kesukaran soal juga harus diperhatikan agar dapat mengetahui dan menetapkan berapa jumlah soal yang termasuk sukar, sedang dan mudah. Adapun besar-kecilnya jumlah soal untuk tiap-tiap tingkat kesukaran tidak ada yang mutlak. Biasanya, jumlah soal sedang lebih banyak daripada jumlah soal mudah dan sukar, sedangkan jumlah soal mudah dan soal sukar sama banyaknya. Misalnya, soal mudah ditentukan 30%, sedang 40%, dan sukar 30 %.

Contoh:

Kisi-Kisi Soal

Nama Sekolah :
 Mata Pelajaran :
 Kelas/Semester :
 Kurikulum Acuan :
 Alokasi Waktu :
 Jumlah Soal :

Materi	BS 50				PGN30				M 20			
	Peg (30)	Pem (30)	Ap (40%)	Jlh	Peng (30%)	Pem (30%)	Ap (40%)	Jlh	Peng (30%)	Pem (30%)	Ap (40%)	Jlh
A (40%)	6	6	8	20	3	4	5	12	2	2	4	8
B (40%)	6	6	8	20	3	4	5	12	2	2	4	8
C (20%)	3	3	4	10	2	2	2	6	1	1	2	4
Jlh	15	15	20	50	8	10	12	30	5	5	10	20

Penjelasan :

- a. Misalnya, jumlah soal keseluruhan adalah 100, terdiri atas 50 soal bentuk benar-salah, 30 soal bentuk pilihan-ganda, dan 20 soal bentuk menjodohkan. Selanjutnya, tentukan pula persentase soal untuk masing-masing materi, misalnya 40 %, 40 %, dan 20 %.

Untuk soal bentuk B – S = 50, maka jumlah soal untuk setiap materi adalah :

Materi A = 40 % x 50 = 20 soal

Materi B = 40 % x 50 = 20 soal

Materi C = 20 % x 50 = 10 soal

Untuk bentuk P – G = 30, maka jumlah soal untuk setiap materi adalah:

Materi A = 40 % x 30 = 12 soal

Materi B = 40 % x 30 = 12 soal

Materi C = 20 % x 30 = 6 soal

Untuk bentuk Menjodohkan = 20, maka jumlah soal untuk setiap materi adalah:

Materi A = 40 % x 20 = 8 soal

Materi B = 40 % x 20 = 8 soal

Materi C = 20 % x 20 = 4 soal

- b. Selanjutnya, menghitung jumlah soal untuk setiap jenjang kemampuan, yaitu persentase pada setiap jenjang kemampuan dikalikan dengan jumlah soal untuk setiap bentuk soal. Misalnya:

Pengetahuan : 30 % x 20 = 6 soal

Pemahaman : 30 % x 20 = 6 soal

Aplikasi : 40 % x 20 = 8 soal

Demikian seterusnya.

Pada kisi-kisi di atas belum tampak tingkat kesukaran soal (mudah, sedang, sukar serta perbandingannya). Untuk menghitung tingkat kesukaran soal, maka pada setiap jenjang kemampuan/aspek yang diukur (pengetahuan, pemahaman, dan aplikasi) harus dibagi menjadi tiga kolom, yakni untuk kolom mudah, sedang, dan sukar dengan perbandingan (misalnya) 30 %, 40 %, dan 30 %. Dengan demikian, jumlah soal untuk masing-masing tingkat kesukaran pada setiap jenjang kemampuan dapat dihitung seperti berikut:

Untuk jenjang kemampuan pengetahuan :

Mudah : 30 % x 6 = 1,8 dihitung 2 soal.

Sedang : 40 % x 6 = 2,4 dihitung 2 soal.

Sukar : 30 % x 8 = 1,8 dihitung 2 soal.

Demikian seterusnya, sehingga melahirkan tabel yang lebih terurai.

3. Menulis Soal

Penulisan soal merupakan salah satulangkah penting untuk dapat menghasilkan alat ukur atau tes yang baik. Penulisan soal adalah penjabaran indikator menjadi pertanyaan-pertanyaan yang karakteristiknya sesuai dengan pedoman kisi-kisi. Setiap pertanyaan harus jelas dan terfokus serta menggunakan bahasa yang efektif, baik bentuk pertanyaan maupun bentuk jawabannya. Kualitas butirsoal akan menentukan kualitas tes secara keseluruhan. Setelah semua soal ditulis, sebaiknya soal tersebut dibaca lagi, jika perlu didiskusikan

kembali dengan tim penelaah soal, baik dari ahli bahasa, ahli bidang studi, termasuk ahli evaluasi.

4. Uji Coba dan Analisis Soal

Jika semua soal sudah disusun dengan baik, maka perlu diujicobakan terlebih dahulu dilapangan. Tujuannya untuk melihat soal-soal mana yang perlu diubah, diperbaiki, bahkan dibuang sama sekali, serta soal-soal mana yang baik untuk dipergunakan selanjutnya. Soal yang baik adalah soal yang sudah mengalami beberapa kali uji-coba dan revisi, yang didasarkan atas analisis empiris dan rasional. Analisis empiris dimaksudkan untuk mengetahui kelemahan-kelemahan setiap soal yang digunakan. Informasi empirik pada umumnya menyangkut segala hal yang dapat mempengaruhi validitas soal, seperti aspek-aspek keterbacaan soal, tingkat kesukaran soal, bentuk jawaban, daya pembeda soal, pengaruh kultur, dan sebagainya. Sedangkan analisis rasional dimaksudkan untuk memperbaiki kelemahan-kelemahan setiap soal.

Dalam melaksanakan uji-coba soal, ada beberapa hal yang harus kita perhatikan, antar lain: (a) ruangan tempatnya tes hendaknya diusahakan seterang mungkin, jika perlu dibuat papan pengumuman diluar agar orang lain tahu bahwa ada tes yang sedang berlangsung, (b) perlu disusun tata tertib pelaksanaan tes, baik yang berkenaan dengan peserta didik itu sendiri, guru, pengawas, maupun teknis pelaksanaan tes, (c) para pengawas tes harus mengontrol pelaksanaan tes dengan ketat, tetapi tidak mengganggu suasana tes. Peserta didik yang melanggar tata tertib tes dapat dikeluarkan dari ruang tes, (d) waktu yang digunakan harus sesuai dengan banyaknya soal yang diberikan, sehingga peserta didik dapat bekerja dengan baik, (e) peserta didik harus benar-benar patuh mengerjakan semua petunjuk dan perintah dari penguji. Sikap ini harus tetap dipelihara meskipun diberikan kesempatan kepada peserta didik untuk mengajukan pertanyaan bila ada soal yang tidak dimengerti atau kurang jelas. Tanggung jawab penguji dalam hal ini adalah memberikan petunjuk dengan sikap yang bersifat lugas, jujur, adil dan jelas. Namun demikian, antara penguji dan peserta didik hendaknya dapat menciptakan suasana yang kondusif, dan (f) hasil uji coba hendaknya diolah, dianalisis, dan diadministrasikan dengan baik, sehingga dapat diketahui soal-soal mana yang lemah untuk selanjutnya dapat diperbaiki kembali.

5. Revisi dan Merakit Soal

Setelah soal diuji-coba dan dianalisis, kemudian direvisi sesuai dengan proporsi tingkat kesukaran soal dan daya pembeda. Dengan demikian, ada soal yang masih dapat diperbaiki dari segi bahasa, ada

juga soal yang harus direvisi total, baik yang menyangkut pokok soal (*stem*) maupun alternatif jawaban (*option*), bahkan ada soal yang harus dibuang atau disisihkan. Berdasarkan hasil revisi soal ini, barulah kita merakit soal menjadi suatu alat ukur yang terpadu. Semua hal yang dapat mempengaruhi validitas skortes, seperti nomor urut soal, pengelompokan bentuk soal, penataan soal, dan sebagainya haruslah diperhatikan.

B. PELAKSANAAN EVALUASI

Pelaksanaan evaluasi artinya bagaimana cara melaksanakan suatu evaluasi sesuai dengan perencanaan evaluasi, baik menggunakan tes (tes tertulis, tes lisan dan tes perbuatan) maupun non-tes. Dalam pelaksanaan tes maupun non-tes tersebut akan berbeda satu dengan lainnya, sesuai dengan tujuan dan fungsinya masing-masing.

Dalam pelaksanaan tes lisan, harus memperhatikan tempat atau ruangan tes yang akan digunakan. Tempat ini harus terang, enak dipandang dan tidak menyheramkan, sehingga peserta didik tidak takut dan gugup. Harus dapat menciptakan suasana yang kondusif dan komunikatif, tetapi bukan berarti menciptakan suasana tes lisan menjadi suasana diskusi, debat atau ngobrol santai. Komunikatif dimaksudkan agar dapat mengarahkan jawaban peserta didik, terutama bila jawaban itu tidak sesuai dengan apa yang kita maksudkan, sebaliknya bukan dengan membentak-bentak peserta didik. Mengarahkan berbeda dengan membantu. Mengarahkan berarti memberi pengarahan secara umum untuk mencapai tujuan, sedangkan membantu berarti ada kecenderungan untuk memberi bunyi jawaban kepada peserta didik, karena ada rasa simpati, kasihan, dan sebagainya.

Dalam pelaksanaan tes lisan, tidak boleh membentak-bentak peserta didik dan dilarang memberikan kata-kata yang merupakan kunci jawaban. Ada baiknya, sebelum tes lisan dimulai, menyiapkan pokok-pokok materi yang akan ditanyakan, sehingga tidak terkecoh oleh jawaban peserta didik yang simpang siur. Ketika peserta didik masuk dan duduk di tempat ujian, hendaknya tidak langsung memberikan pertanyaan-pertanyaan, karena yakinlah bahwa siapapun yang menghadapi ujian atau tes lisan pasti ada perasaan gugup. Oleh sebab itu, pada waktu mulai tes lisan (lebih kurang 2-3 menit), harus dapat menciptakan kondisi peserta didik agar tidak gugup, seperti menanyakan identitas pribadi, pengalaman, kegiatan sehari-hari, dan sebagainya.

Pelaksanaan nontes dimaksudkan untuk mengetahui sikap dan tingkah laku peserta didik sehari-hari dengan menggunakan instrumen khusus, seperti pedoman observasi, pedoman wawancara, skala sikap,

skala minat, daftar cek, *rating scale*, *anecdotal records*, *sosiometri*, *home visit*, dan sebagainya. Dituntut tidak hanya mampu membuat dan melaksanakan tes yang baik, tetapi juga harus dapat membuat alat-alat khusus dalam nontes dan melaksanakannya dengan baik sesuai dengan prinsip-prinsip evaluasi.

Untuk mengetahui tingkat penguasaan kompetensi peserta didik, selain menggunakan tes tertulis (*pencil and paper test*), juga dapat menggunakan tes kinerja (*performance test*). Di samping itu, dapat menilai hasil kerja peserta didik dengan cara memberikan tugas atau proyek dan menganalisis semua hasil kerja dalam bentuk portofolio. Diharapkan tidak hanya menilai kognitif peserta didik, tetapi juga non-kognitif, seperti pengembangan pribadi, kreatifitas, dan keterampilan interpersonal, sehingga dapat diperoleh gambaran yang komprehensif dan utuh.

Realitas menunjukkan bahwa tidak ada satu teknik dan bentuk evaluasi yang dapat mengumpulkan data tentang keefektifan pembelajaran, prestasi dan kemajuan belajar peserta didik secara sempurna. Pengukuran tunggal tidak cukup untuk memberikan gambaran atau informasi tentang keefektifan pembelajaran dan tingkat penguasaan kompetensi (pengetahuan, keterampilan, sikap dan nilai-nilai) peserta didik. Hasil evaluasi juga tidak mutlak dan tidak abadi, karena sistem belajar dan pembelajaran terus berkembang sesuai dengan perkembangan ilmu pengetahuan dan teknologi serta pengalaman belajar peserta didik. Penetapan salah satu teknik dan bentuk evaluasi (misalnya hanya tes objektif) dapat menghambat penguasaan kompetensi peserta didik secara utuh, sehingga tidak memberikan umpan balik dalam rangka diagnosis atau memodifikasi pengalaman belajar. Tujuan pelaksanaan evaluasi adalah untuk mengumpulkan data dan informasi mengenai keseluruhan aspek kepribadian dan prestasi belajar peserta didik yang meliputi:

1. Data pribadi (*personal*) peserta didik, seperti nama, tempat dan tanggal lahir, jenis kelamin, golongan darah, alamat, dan lain-lain.
2. Data tentang kesehatan peserta didik, seperti: penglihatan, pendengaran, penyakit yang sering diderita, kondisi fisik dan sebagainya.
3. Data tentang prestasi belajar (*achievement*) peserta didik di sekolah.
4. Data tentang sikap (*attitude*) peserta didik, seperti sikap terhadapsesamatemannya, sikap terhadap kegiatan pembelajaran, sikap terhadap guru dan kepala sekolah, sikap terhadap lingkungan sosial, dan lain-lain.

5. Data tentang bakat (*aptitude*) peserta didik, seperti ada tidaknya bakat di bidang olah raga, keterampilan mekanis, manajemen, kesenian, keguruan, dan sebagainya.
6. Persoalan penyesuaian (*adjustment*), seperti kegiatan anak dalam organisasi di sekolah, forum ilmiah, olah raga, kependuan, dan sebagainya.
7. Data tentang minat (*intrest*) peserta didik.
8. Data tentang rencana masa depan peserta didik yang dibantu oleh guru dan orang tua sesuai dengan kesanggupan anak.
9. Data tentang latar belakang keluarga peserta didik, seperti pekerjaan orang tua, penghasilan tetap tiap bulan, kondisi lingkungan, hubungan peserta didik dengan orang tua dan saudara-saudaranya, dan sebagainya.

Dari jenis-jenis data di atas jelas kiranya bahwa banyak data yang harus dikumpulkan dari lapangan melalui kegiatan evaluasi. Pengumpulan data ini harus diperhitungkan dengan cermat dan matang serta berpedoman kepada prinsip dan fungsi evaluasi itu sendiri. Ada kecenderungan pelaksanaan evaluasi selama ini kurang begitu memuaskan (terutama) bagi peserta didik. Hal ini dapat dilihat dari berbagai segi, antara lain: (a) proses dan hasil evaluasi kurang memberi keuntungan pada peserta didik, baik secara langsung maupun tidak langsung, (b) penggunaan teknik dan prosedur evaluasi yang kurang tepat berdasarkan apa yang sudah dipelajari peserta didik, (c) prinsip-prinsip umum evaluasi kurang dipertimbangkan dan pemberian skor cenderung tidak adil dan tidak objektif, dan (d) cakupan evaluasi kurang memperhatikan aspek-aspek penting dari pembelajaran.

Jika semua data sudah dikumpulkan, maka data itu harus diseleksi dengan teliti, sehingga kita dapat memperoleh data yang baik dan benar. Sebaliknya, bila data yang terkumpul tidak diseleksi lagi, maka ada kemungkinan data itu tidak sesuai dengan tujuan yang telah ditetapkan, bahkan mungkin pula bertentangan, sehingga mengakibatkan keaburan atau kekurangjelasan dari apa yang diharapkan. Data yang harus diseleksi tidak hanya data dari hasil evaluasi, tetapi juga data yang diperoleh dari pihak lain tentang peserta didik. Namun demikian, tidak semua data yang diperoleh pasti mempunyai kesalahan. Jika sendiri melaksanakan evaluasi, tentu akan lebih berhati-hati dalam memilih dan menggunakan teknik dan alat evaluasi.

Ada beberapa hal yang memungkinkan timbulnya kesalahan-kesalahan dalam pengumpulan data, yaitu:

1. Kesalahan-kesalahan yang mungkin ditimbulkan karena kurang sempurna alat-alat evaluasi. Misalnya, pada data yang berupa skor tes,

mungkin tes yang dipergunakan kurang baik, tidak valid, tidak reliabel, tidak praktis, dan sebagainya. Pada data yang berupa hasil-hasil observasi, mungkin pedoman observasinya kurang jelas, data hasil observasi kurang lengkap atau tidak melukiskan variabel yang harus diobservasi. Prosedur verifikasinya adalah meneliti kembali alat-alat evaluasi yang digunakan dalam pengumpulan data. Apakah alat-alat evaluasi tersebut sudah cukup baik atau belum. Jika berdasarkan hasil pemeriksaan ternyata tidak ada kekeliruan, berarti kesalahannya bukan bersumber dari alat evaluasi yang digunakan. Oleh sebab itu, pemeriksaan harus dilanjutkan pada sumber kesalahan yang lain.

2. Kesalahan-kesalahan yang mungkin ditimbulkan oleh kurang sempurnanya prosedur pelaksanaan evaluasi yang dilakukan. Misalnya, pada data yang berupa skor tes, mungkin pada waktu pelaksanaan tes tersebut terjadi peristiwa-peristiwa yang berlawanan dengan kelaziman-kelaziman yang biasa, pengawasan kurang ketat, kondisi tempat pelaksanaan tes kurang nyaman, cahaya kurang terang, dan sebagainya. Prosedur verifikasinya adalah meninjau kembali komponen-komponen yang terkait dalam pelaksanaan evaluasi, syarat-syarat pelaksanaan evaluasi, dan faktor-faktor yang menghambat pelaksanaan evaluasi. Jika disini tidak ditemukan sebab-sebab yang menimbulkan kesalahan, maka pemeriksaan harus dilanjutkan pada sumber kesesatan yang lain.
3. Kesalahan yang mungkin ditimbulkan oleh kurang sempurnanya cara pencatatan hasil evaluasi. Misalnya, pada data yang berupa skor tes kemungkinan kita sudah menjumlahkan skor yang dicapai peserta didik. Prosedur verifikasinya adalah meneliti kembali pencatatan skor yang telah dilakukan, seperti ada tidaknya kekeliruan pada waktu mencatat hasil evaluasi, ada tidaknya kekeliruan dalam pemberian skor, dan ada tidaknya kekeliruan dalam menjumlahkan skor setiap peserta didik. Jika disinipun tidak ditemukan kesalahan, berarti data yang dikumpulkan itu tidak mengandung kesalahan. Hal-hal semacam inilah yang diperlukan dalam menyeleksi dan meneliti data yang diperoleh.

C. PENGOLAHAN DATA

Dalam pengolahan data biasanya sering digunakan analisis statistik. Analisis statistik digunakan jika ada data kuantitatif, yaitu data-data yang berbentuk angka-angka, sedangkan untuk data kualitatif, yaitu data yang berbentuk kata-kata, tidak dapat diolah dengan statistik. Jika data kualitatif itu akan diolah dengan statistik, maka data tersebut harus diubah terlebih dahulu menjadi data kuantitatif (kuantifikasi data).

Meskipun demikian, tidak semua data kualitatif dapat diubah menjadi data kuantitatif, sehingga tidak mungkin diolah dengan statistik. Ada empat langkah pokok dalam mengolah hasil evaluasi, yaitu:

1. Menskor, yaitu memberikan skor pada hasil evaluasi yang dapat dicapai oleh peserta didik. Untuk menskor atau memberikan angka diperlukan tiga jenis alat bantu, yaitu: kunci jawaban, kunci skoring, dan pedoman konversi.
2. Mengubah skor mentah menjadi skor standar sesuai dengan norma tertentu.
3. Mengkonversikan skor standar ke dalam nilai, baik berupa hurup atau angka.
4. Melakukan analisis soal (jika diperlukan) untuk mengetahui derajat validitas dan reliabilitas soal, tingkat kesukaran soal (*difficulty index*), dan daya pembeda.

Jika data sudah diolah dengan aturan-aturan tertentu, langkah selanjutnya adalah menafsirkan data itu, sehingga memberikan makna. Langkah penafsiran data sebenarnya tidak dapat dilepaskan dari pengolahan data itu sendiri, karena setelah mengolah data dengan sendirinya akan menafsirkan hasil pengolahan itu. Memberikan penafsiran maksudnya adalah membuat pernyataan mengenai hasil pengolahan data yang didasarkan atas kriteria tertentu yang disebut norma. Norma dapat ditetapkan terlebih dahulu secara rasional dan sistematis sebelum kegiatan evaluasi dilaksanakan, tetapi dapat pula dibuat berdasarkan hasil-hasil yang diperoleh dalam melaksanakan evaluasi. Sebaliknya, bila penafsiran data itu tidak berdasarkan kriteria atau norma tertentu, maka ini termasuk kesalahan besar. Misalnya, seorang peserta didik naik kelas. Kenaikan kelas itu kadang-kadang tidak berdasarkan kriteria-kriteria yang disepakati, tetapi hanya berdasarkan pertimbangan pribadi dan kemanusiaan, maka keputusan ini termasuk keputusan yang tidak objektif dan merugikan semua pihak.

Dalam kegiatan pembelajaran, biasanya kriteria bersumber pada tujuan setiap mata pelajaran (standar kompetensi dan kompetensi dasar). Kompetensi ini tentu masih bersifat umum, karena itu harus dijabarkan menjadi indikator yang dapat diukur dan dapat diamati. Jika kriteria ini sudah dirumuskan dengan jelas, maka baru kita menafsirkan angka-angka yang sudah diolah itu berupa kata-kata atau pernyataan. Dalam menyusun kata-kata ini sering guru mengalami kesulitan. Kesulitan itu antara lain penyusunan kata-kata sering melampaui batas-batas kriteria yang telah ditentukan, bahkan tidak didukung oleh data-data yang ada. Hal ini disebabkan adanya kecenderungan pada guru untuk menonjolkan kelebihan suatu sekolah dibandingkan dengan sekolah yang lain.

Kesulitan yang sering terjadi adalah penyusunan rumusan tafsiran atau pernyataan yang berlebihan (*overstatement*) di luar batas-batas kebenaran. Kesalahan semacam ini sebenarnya tidak hanya terjadi karena kekurangtelitian dalam menafsirkan data saja, tetapi mungkin pula sudah muncul pada langkah-langkah sebelumnya.

Ada dua jenis penafsiran data, yaitu penafsiran kelompok dan penafsiran individual.

1. Penafsiran kelompok adalah penafsiran yang dilakukan untuk mengetahui karakteristik kelompok berdasarkan data hasil evaluasi, seperti prestasi kelompok, rata-rata kelompok, sikap kelompok terhadap guru dan materi pelajaran yang diberikan, dan distribusi nilai kelompok. Tujuan utamanya adalah sebagai persiapan untuk melakukan penafsiran kelompok, untuk mengetahui sifat-sifat tertentu pada suatu kelompok, dan untuk mengadakan perbandingan antar kelompok.
2. Penafsiran individual adalah penafsiran yang hanya dilakukan secara perorangan. Misalnya, dalam kegiatan bimbingan dan penyuluhan atau situasi klinis lainnya. Tujuan utamanya adalah untuk melihat tingkat kesiapan peserta didik (*readiness*), pertumbuhan fisik, kemajuan belajar, dan kesulitan-kesulitan yang dihadapinya.

Dalam melakukan penafsiran data, baik secara kelompok maupun individual, kita harus menggunakan norma-norma yang standar, sehingga data yang diperoleh dapat dibandingkan dengan norma-norma tersebut. Berdasarkan penafsiran ini dapat diputuskan bahwa peserta didik mencapai tarap kesiapan yang memadai atau tidak, ada kemajuan yang berarti atau tidak, ada kesulitan atau tidak. Jika ingin menggambarkan pertumbuhan peserta didik, penyebaran skor, dan perbandingan antar kelompok, maka perlu menggunakan garis (kurva), grafik, atau dalam beberapa hal diperlukan profil, dan bukan dengan daftar angka-angka. Daftar angka-angka biasanya digunakan untuk melukiskan posisi atau kedudukan peserta didik, baik secara perorangan maupun kelompok.

D. PELAPORAN HASIL EVALUASI

Semua hasil evaluasi harus dilaporkan kepada berbagai pihak yang berkepentingan, seperti orang tua/wali, atasan, pemerintah, dan peserta didik itu sendiri sebagai akuntabilitas publik. Hal ini dimaksudkan agar proses dan hasil yang dicapai peserta didik termasuk perkembangannya dapat diketahui oleh berbagai pihak, sehingga orang tua/wali (misalnya) dapat menentukan sikap yang objektif dan mengambil langkah-langkah yang pasti sebagai tindak lanjut dari laporan tersebut. Sebaliknya, jika

hasil evaluasi itu tidak dilaporkan, orang tua peserta didik tidak dapat mengetahui kemajuan belajar yang dicapai anaknya, karena itu pula mungkin orang tua peserta didik tidak mempunyai sikap dan rencana yang pasti terhadap anaknya, baik dalam rangka pemilihan minat dan bakat, bimbingan maupun untuk melanjutkan studi yang lebih tinggi.

Hasil evaluasi juga perlu dilaporkan kepada pemerintah, dalam hal ini Kementerian Pendidikan di tingkat provinsi dan kabupaten/kota. Tujuannya adalah untuk melihat kemajuan-kemajuan peserta didik, baik secara kelompok maupun individual, yang pada gilirannya akan memberikan penilaian tersendiri pada sekolah yang bersangkutan. Misalnya, dalam satu laporan dikatakan bahwa peserta didik kelas VI di sekolah "X" lulus 99%, maka sekolah tersebut dianggap masyarakat baik atau sekolah favorit. Sebaliknya, jika peserta didik Sekolah tersebut lulus 70%, maka dianggap sekolah tersebut tidak bermutu. Semakin tinggi persentase kelulusan, maka makin tinggi pula penilaian yang diberikan oleh masyarakat terhadap Sekolah tersebut, sekalipun persentase kelulusan tidak menjamin berkualitasnya suatu sekolah. Laporan juga penting bagi peserta didik itu sendiri agar mereka mengetahui tingkat kemampuan yang dimilikinya dan dapat menentukan sikap serta tindakan yang harus dilakukan selanjutnya.

Laporan kemajuan belajar peserta didik merupakan sarana komunikasi antara sekolah, peserta didik, dan orang tua dalam upaya mengembangkan dan menjaga hubungan kerja sama yang baik diantara mereka. Untuk itu, harus memperhatikan beberapa hal sebagai berikut:

1. Konsisten dengan pelaksanaan penilaian di sekolah.
2. Memuat rincian hasil belajar peserta didik berdasarkan kriteria yang telah ditentukan dan dikaitkan dengan penilaian yang bermanfaat bagi pengembangan peserta didik.
3. Menjamin orang tua akan informasi permasalahan peserta didik dalam belajar.
4. Mengandung berbagai cara dan strategi komunikasi.
5. Memberikan informasi yang benar, jelas, komprehensif, dan akurat.

Laporan kemajuan belajar peserta didik yang selama ini dilakukan oleh pihak madrasah cenderung hanya bersifat kuantitatif, sehingga kurang dapat dipahami maknanya. Misalnya, seorang peserta didik mendapat nilai 5 dalam buku rapot pada mata pelajaran IPS-Ekonomi. Jika hanya angka yang disajikan, maka peserta didik maupun orang tua akan sulit menafsirkan nilai tersebut, apakah nilai "kurang" tersebut berkaitan dengan bidang pengetahuan dan pemahaman, praktik, sikap atau semuanya. Oleh karena itu, bentuk laporan kemajuan peserta didik harus disajikan secara sederhana, mudah dibaca dan dipahami,

komunikatif, dan menampilkan profil atau tingkat kemajuan peserta didik, sehingga peran serta masyarakat, orang tua, dan stakeholder dalam dunia pendidikan semakin meningkat. Paling tidak, pihak-pihak terkait dapat dengan mudah mengidentifikasi kompetensi-kompetensi yang sudah dan belum dikuasai peserta didik serta kompetensi mana yang harus ditingkatkan. Bagi peserta didik sendiri dapat mengetahui keunggulan dan kelemahan dirinya serta pada aspek mana ia harus belajar lebih banyak.

Untuk sekedar gambaran, isi laporan hendaknya memuat hal-hal seperti: profil belajar peserta didik di sekolah (akademik, fisik, sosial dan emosional), peran serta peserta didik dalam kegiatan di sekolah (aktif, cukup, kurang atau tidak aktif), kemajuan hasil belajar peserta didik selama kurun waktu belajar tertentu (meningkat, biasa-biasa saja atau menurun), himbauan terhadap orang tua. Isi laporan tersebut hendaknya mudah dipahami orang tua. Untuk itu, kita harus menggunakan bahasa yang komunikatif, menitikberatkan pada proses dan hasil yang telah dicapai peserta didik, memberikan perhatian terhadap pengembangan dan pembelajaran peserta didik, dan memberikan hasil penilaian yang tepat dan akurat.

Dalam dokumen kurikulum berbasis kompetensi, Pusat Kurikulum Balitbang Depdiknas (2002: 35) menjelaskan "laporan kemajuan siswa dapat dikategorikan menjadi dua jenis, yaitu laporan prestasi dalam mata pelajaran dan laporan pencapaian".

1. Laporan Prestasi Mata Pelajaran

Laporan prestasi mata pelajaran berisi informasi tentang pencapaian kompetensi dasar yang telah ditetapkan dalam kurikulum. Pada masa lalu, prestasi belajar peserta didik dalam setiap mata pelajaran dilaporkan dalam bentuk angka. Bagi peserta didik dan orang tua, angka ini kurang memberi informasi tentang kompetensi dasar dan pengetahuan apa yang telah dimiliki peserta didik, sehingga sulit menentukan jenis bantuan apa yang harus diberikan kepada peserta didik agar mereka menguasai kompetensi dasar yang telah ditetapkan. Laporan prestasi belajar hendaknya menyajikan prestasi belajar peserta didik dalam menguasai kompetensi mata pelajaran tertentu dan tingkat penguasaannya. Sebaliknya, orang tua dapat membaca catatan guru tentang pencapaian kompetensi tertentu sebagai masukan kepada peserta didik dan orang tua untuk membantu meningkatkan kinerjanya.

Contoh:

Tabel 3.1. Format Laporan Prestasi Peserta Didik Dalam Mata Pelajaran

No	Kemampuan Dasar	Nilai				Deskripsi Pencapaian
		A	B	C	D	

Catatan Kompetensi (contoh):

1. Peserta didik menunjukkan kemahiran di dalam Tetapi memerlukan bantuan dalam hal.
2. Secara umum peserta didik telah berhasil menguasai dari kompetensi.

Dengan demikian, isi laporan prestasi belajar sebaiknya disajikan secara kualitatif atau menggabungkan antara angka (kuantitatif) dengan deskripsi (kualitatif).

2. Laporan Pencapaian

Laporan pencapaian merupakan laporan yang menggambarkan kualitas pribadi peserta didik sebagai internalisasi dan kristalisasi setelah peserta didik belajar melalui berbagai kegiatan, baik intra, ekstra maupun ko kurikuler pada kurun waktu tertentu. Dalam kurikulum berbasis kompetensi, hasil belajar peserta didik dibandingkan antara kemampuan sebelum dan sesudah kegiatan pembelajaran berdasarkan kriteria yang telah ditetapkan dalam kurikulum. Tingkat pencapaian hasil belajar yang ditetapkan dalam kurikulum dibagi menjadi delapan tingkatan (*level*) yang dirinci ke dalam rumusan kemampuan dari yang paling dasar secara bertahap gradasinya mencapai tingkat yang paling tinggi. Delapan tingkatan hasil belajar tidak sama dengan tingkat kelas dalam satuan pendidikan. Tingkat pencapaian hasil belajar peserta didik tidak selalu sama dengan peserta didik yang lain untuk setiap mata pelajaran. Kesetaraan antara tingkat pencapaian hasil belajar dengan prestasi belajar peserta normal sebagai berikut.

Tabel 3.2. Tingkatan Hasil Belajar Tidak Sama Dengan Tingkat Kelas

Tingkat (level)	Pada umumnya dicapai anak di kelas
0	0 (TK atau Pradasar)
1	1 – 2
2	3 – 4
3	5 – 6
4	7 – 8
4a	9
5	10 – 11
6	11 – 12

Berikut contoh format tingkat pencapaian hasil belajar peserta didik untuk beberapa mata pelajaran.

Laporan Pencapaian Hasil Belajar

Nama :

Kelas :

Semester :

Mata Pelajaran	Level								Keterangan
	0	1	2	3	4	4a	5	6	
1. Ekonomi									
2. Sejarah									
3. Geografi									
4. Sosiologi									
5. dst.									

Catatan:

Penetapan tingkat pencapaian peserta didik dalam rentang skala 0 – 6 berdasarkan penilaian hasil belajar peserta didik dibandingkan dengan kriteria yang telah ditetapkan dalam buku kurikulum dan hasil belajar. Rincian tingkat kompetensi tiap mata pelajaran juga dapat dilihat pada buku Kurikulum dan Hasil Belajar Rumpun Pelajaran.

E. PENGGUNAAN HASIL EVALUASI

Tahap akhir dari prosedur evaluasi adalah penggunaan hasil evaluasi. Salah satu penggunaan hasil evaluasi adalah laporan. Laporan dimaksudkan untuk memberikan *feedback* kepada semua pihak yang terlibat dalam pembelajaran, baik secara langsung maupun tidak langsung. Pihak-pihak yang dimaksud antara lain: peserta didik, guru, kepala sekolah, orang tua, penilik, dan pemakai lulusan. Sedangkan penggunaan hasil evaluasi, Remmer (1967) mengatakan '*we discuss here the use of test results to help students understand them selves better, explain pupil growth and development to parents and assist the teacher in planning instruction*'. Dengan demikian, dapat menggunakan hasil evaluasi untuk membantu pemahaman peserta didik menjadi lebih baik, menjelaskan pertumbuhan dan perkembangan peserta didik kepada orang tua, dan membantu guru dalam menyusun perencanaan pembelajaran.

Sehubungan dengan hal tersebut, Julian C. Stanley dalam Dimiyati dan Mudjiono (1994: 206) mengemukakan '*just what is to be done, of course, depends on the purpose of the program*'. Dengan demikian, apa yang harus dilakukan terhadap hasil-hasil evaluasi yang kita peroleh bergantung kepada tujuan program evaluasi itu sendiri yang tentunya sudah dirumuskan sebelumnya. Berdasarkan penjelasan di atas, maka dapat dikemukakan beberapa jenis penggunaan hasil evaluasi sebagai berikut:

1. Untuk keperluan laporan pertanggungjawaban

Asumsinya adalah banyak pihak yang berkepentingan dengan hasil evaluasi. Misalnya, orang tua perlu mengetahui kemajuan atau perkembangan hasil belajar anaknya, sehingga dapat menentukan langkah-langkah berikutnya. Oleh sebab itu, Kita harus membuat laporan ke berbagai pihak sebagai bentuk akuntabilitas publik, sebagaimana telah penulis kemukakan pada uraian sebelumnya.

2. Untuk keperluan seleksi

Asumsinya adalah setiap awal dan akhir tahun ada peserta didik yang mau masuk sekolah dan ada peserta didik yang mau menamatkan sekolah pada jenjang pendidikan tertentu. Hasil evaluasi dapat digunakan untuk menyeleksi, baik ketika peserta didik mau masuk Sekolah/jenjang atau jenis pendidikan tertentu, selama mengikuti program pendidikan, pada saat mau menyelesaikan jenjang pendidikan, maupun ketika masuk dunia kerja. Ketika peserta didik mengikuti program pendidikan, terkadang dari pihak sekolah dan komite sekolah membuat kelas-kelas unggulan. Untuk itu diperlukan seleksi melalui tindakan evaluasi.

3. Untuk keperluan promosi

Asumsinya adalah pada akhir tahun pelajaran, ada peserta didik yang naik kelas atau lulus. Bagi peserta didik yang lulus dari jenjang pendidikan tertentu akan diberikan ijazah atau sertifikat, sebagai bukti fisik kelulusan. Begitu juga jika peserta didik memperoleh prestasi belajar yang baik, maka mereka akan naik ke kelas berikutnya. Kegiatan ini semua merupakan salah satu bentuk promosi. Dengan demikian, promosi itu diberikan setelah dilakukan kegiatan evaluasi. Jika promosi itu untuk kenaikan kelas, maka kriteria yang digunakan adalah kriteria kenaikan kelas, yaitu aspek ketercapaian kompetensi dasar mata pelajaran yang telah ditetapkan dalam kurikulum. Peserta didik yang dinyatakan naik kelas adalah peserta didik yang sudah menguasai kompetensi pada kelas tertentu dan diprediksi mampu mengikuti program pendidikan pada kelas berikutnya.

Sesuai dengan prinsip peningkatan mutu pendidikan, maka kriteria peserta didik yang dinyatakan naik kelas atau lulus harus dirinci lebih operasional. Misalnya, peserta didik dinyatakan naik kelas bila menguasai minimal 60% kompetensi yang menyangkut beberapa mata pelajaran atau peserta didik dinyatakan lulus bila menguasai minimal 60% dari keseluruhan kompetensi untuk semua mata pelajaran di kelas tersebut. Kita juga dapat mempertimbangkan kriteria yang menyangkut perilaku atau kinerja peserta didik. Rincian kriteria kenaikan kelas atau kelulusan sesuai dengan prinsip manajemen

berbasis madrasah perlu disusun bersama antara Kementerian Pendidikan kabupaten/kota, Dewan Pendidikan, sekolah dan komite sekolah.

4. Untuk keperluan diagnosis

Asumsinya adalah hasil evaluasi menunjukkan ada peserta didik yang kurang mampu menguasai kompetensi sesuai dengan kriteria yang telah ditetapkan. Atas dasar asumsi ini, maka perlu melakukan diagnosis terhadap peserta didik yang dianggap kurang mampu tersebut. Kita harus mencari faktor-faktor penyebab bagi peserta didik yang kurang mampu dalam menguasai kompetensi tertentu, sehingga dapat diberikan bimbingan atau pembelajaran remedial. Bagi peserta didik yang mampu menguasai kompetensi lebih cepat dari peserta didik yang lain, mereka juga berhak mendapatkan pelayanan tindak lanjut untuk mengoptimalkan laju perkembangan mereka. Sekolah diharapkan menyediakan alternatif program bagi mereka berupa kegiatan yang dapat memperkaya pengetahuan dan keterampilannya di suatu bidang tertentu ataupun suatu sistem percepatan belajar, sehingga memungkinkan mereka dapat menyelesaikan/tamat sekolah lebih cepat. Untuk menetapkan kebijakan suatu jenis perlakuan kepada peserta didik dan teknik pelaksanaannya perlu melibatkan peran serta masyarakat melalui komite sekolah.

5. Untuk memprediksi masa depan peserta didik

Hasil evaluasi perlu dianalisis oleh setiap guru mata pelajaran. Tujuannya untuk mengetahui sikap, bakat, minat dan aspek-aspek kepribadian lainnya dari peserta didik, serta dalam hal apa peserta didik dianggap paling menonjol sesuai dengan indikator keunggulan. Apapun dan bagaimanapun bentuk hasil belajar peserta didik, kita harus menyampaikannya kepada guru bimbingan dan penyuluhan (BP) agar hasil belajar tersebut dapat dianalisis dan dijadikan dasar untuk pengembangan peserta didik dalam memilih jenjang pendidikan, profesi atau karir di masa yang akan datang.

BAB IV

PENGEMBANGAN ALAT EVALUASI JENIS TES

A. PENGEMBANGAN TES BENTUK URAIAN

1. Jenis Tes

Tes merupakan suatu teknik atau cara yang digunakan dalam rangka melaksanakan kegiatan pengukuran, yang didalamnya terdapat berbagai pertanyaan, pernyataan atau serangkaian tugas yang harus dikerjakan atau dijawab oleh peserta didik untuk mengukur aspek perilaku peserta didik. Dalam rumusan ini terdapat beberapa unsur penting, yaitu: Pertama, tes merupakan suatu cara atau teknik yang disusun secara sistematis dan digunakan dalam rangka kegiatan pengukuran. Kedua, di dalam tes terdapat berbagai pertanyaan dan pernyataan atau serangkaian tugas yang harus dijawab dan dikerjakan oleh peserta didik. Ketiga, tes digunakan untuk mengukur suatu aspek perilaku peserta didik. Keempat, hasil tes peserta didik perlu diberi skor dan nilai.

Tes dapat dibedakan atas beberapa jenis, dan pembagian jenis-jenis ini dapat ditinjau dari berbagai sudut pandang. Heaton (1988), misalnya, membagi tes menjadi empat bagian, yaitu tes prestasi belajar (*achievement test*), tes penguasaan (*proficiency test*), tes bakat (*aptitude test*), dan tes diagnostik (*diagnostic test*). Untuk melengkapi pembagian jenis tes tersebut, Brown (2004) menambahkan satu jenis tes lagi yang disebut tes penempatan (*placement test*). Dalam bidang psikologi, tes dapat diklasifikasikan menjadi empat bagian, yaitu:

- 1) Tes intelegensia umum, yaitu tes untuk mengukur kemampuan umum seseorang.
- 2) Tes kemampuan khusus, yaitu tes untuk mengukur kemampuan potensial dalam bidang tertentu.
- 3) Tes prestasi belajar, yaitu tes untuk mengukur kemampuan aktual sebagai hasil belajar.
- 4) Tes kepribadian, yaitu tes untuk mengukur karakteristik pribadi seseorang.

Berdasarkan jumlah peserta didik, tes hasil belajar dapat dibedakan atas dua jenis, yaitu tes kelompok dan tes perorangan. Tes kelompok, yaitu tes yang diadakan secara kelompok. Disini guru akan berhadapan dengan sekelompok peserta didik. Tes perorangan yaitu tes yang dilakukan secara perorangan (*individual*). Disini guru akan berhadapan dengan seorang peserta didik. Dilihat dari cara penyusunannya, tes dapat dibedakan atas dua jenis, yaitu:

a. Tes Buatan Guru (*teacher-made test*)

Tes buatan guru adalah tes yang disusun sendiri oleh guru yang akan mempergunakan tes tersebut. Tes ini biasanya digunakan untuk ulangan harian, formatif, dan ulangan umum (sumatif). Tes buatan guru ini dimaksudkan untuk mengukur tingkat penguasaan peserta didik terhadap materi pelajaran yang sudah disampaikan. Untuk itu, Anda harus membuat soal secara logis dan rasional mengenai pokok-pokok materi apa saja yang patut dan seharusnya ditanyakan sebagai bahan pengetahuan penting untuk diketahui dan dipahami oleh peserta didiknya. Kualitas tes atau tingkat kesahihan dan keandalannya masih belum menjamin keobjektifannya, sebab hanya diberikan kepada sekelompok peserta didik, kelas, dan madrasah tertentu saja. Jadi, masih bersifat sektoral, karena belum diujicobakan kepada sekelompok besar tes, sehingga pengukurannya masih belum meyakinkan.

Begitu juga tingkat kesukaran itemnya tidak didasarkan atas sifat-sifat atau karakteristik peserta didiknya. Mereka dianggap memiliki taraf berpikir dan taraf penguasaan materi yang sama. Padahal, setiap peserta didik secara psikologis mempunyai kemampuan yang berbeda. Oleh sebab itu, sebaiknya item-item tes disusun secara cermat berdasarkan tingkat kemampuan individu yang heterogen, sedangkan penjelasan-penjelasan yang bersifat umum bisa sama. Tes buatan guru bersifat temporer, artinya hanya berlaku pada saat tertentu dan situasi tertentu pula. Pada kesempatan lain belum tentu tes tersebut dapat digunakan, karena mungkin berubah, baik bentuk itemnya maupun kapasitas peserta didiknya.

Ada tes buatan guru yang bersifat hafalan semata, dan ada pula yang bersifat analitis. Sebagai guru yang profesional tentu akan menyusun soal yang berimbang dari kedua sifat tersebut. Hal ini dimaksudkan agar peserta didik dapat mengetahui siapa yang mempunyai kemampuan yang mantap dalam mengingat atau menghafal sesuatu, dan siapa pula yang mempunyai daya pikir yang kritis, analitis, luas dan asosiatif. Situasi terakhir inilah yang harus diciptakan guru.

b. Tes yang Dibakukan (*standardized test*)

Tes yang dibakukan atau tes baku adalah tes yang sudah memiliki derajat validitas dan reliabilitas yang tinggi berdasarkan percobaan-percobaan terhadap sampel yang cukup besar dan representatif. Tes baku adalah tes yang dikaji berulang-ulang kepada sekelompok besar peserta didik, dan item-itemnya relevan serta

mempunyai daya pembeda yang tinggi. Di samping itu, tes baku telah diklasifikasikan sesuai dengan tingkat usia dan kelasnya. Tes baku biasanya telah dianalisis secara statistik dan diuji secara empiris oleh para ahli (pakar), karena itu dapat dinyatakan sah (valid) untuk digunakan secara umum. Pengolahan secara statistik dimaksudkan untuk mencari derajat kesahihan dan keandalan serta daya pembeda yang tinggi dari setiap item, sehingga soal itu betul-betul tepat diberikan dan dapat dijadikan alat pengukur kemampuan setiap orang secara umum. Sedangkan pengujian secara empiris dimaksudkan untuk mengetahui kelemahan-kelemahan setiap item. Tes baku bertujuan untuk mengukur kemampuan peserta didik dalam tiga aspek, yaitu kedudukan belajar, kemajuan belajar, dan diagnostik.

Untuk mengetahui kedudukan belajar, setiap peserta didik dibandingkan dengan teman-teman sekelasnya, setingkat Sekolah, atau setingkat dari beberapa Sekolah. Tes ini dilakukan pada tingkat tertentu dan waktu tertentu saja. Tes baku juga digunakan untuk mengukur kemajuan belajar peserta didik dalam mata pelajaran tertentu. Artinya, jika guru telah selesai membahas satu atau beberapa pokok bahasan dari mata pelajaran tertentu, guru bisa memberikan ulangan harian atau ulangan umum pada setiap semester. Adakalanya tes itu diberikan beberapa kali, sehingga kemajuan atau kemunduran belajar peserta didik dapat diketahui. Tes untuk kemajuan belajar inilah yang paling sering dan umum dilakukan oleh setiap guru dalam kegiatan pembelajaran, baik untuk laporan kemajuan belajar peserta didik maupun untuk keperluan seleksi. Adapun pelaksanaannya dapat dilakukan secara tertulis, lisan dan perbuatan, bergantung kepada tujuan dan materinya.

Tes baku bertujuan untuk mengetahui kelebihan dan kekurangan peserta didik dalam menguasai materi pelajaran tertentu secara luas. Tes ini berisi materi-materi yang disusun dari yang termudah sampai yang tersukar serta terdiri atas cakupan yang luas. Dewasa ini tes diagnostik telah banyak dilakukan pada semua sekolah untuk semua tingkatan. Tes diagnostik biasanya dilakukan serempak pada beberapa sekolah dalam waktu yang sama dengan materi tes yang sama. Hasil tes diagnostik akan menunjukkan kelebihan dan kekurangan tertentu dari sekolah tertentu. Beberapa hal yang harus diperhatikan dalam pengembangan tes baku, antara lain:

- a) Aspek yang hendak diukur
Dalam keterangan tes tersebut dijelaskan aspek apa saja yang hendak diukur, misalnya kemampuan membaca, perbendaharaan pengetahuan umum, sikap, minat, kepribadian.
- b) Pihak penyusun
Nama orang, baik secara individual maupun kelompok ataupun organisasi yang merancang tes itu, perlu dicantumkan dalam tes tersebut. Misalnya, tes bahasa Inggris yang diselenggarakan oleh *Modern Language Association* (TEOFL) oleh *College Entrance Examination Board and Educationaal Testing Service*, tes masuk perguruan tinggi negeri yang sekarang kita kenal dengan istilah SNM-PTN. Nama pihak penyusun tes akan memberikan jaminan mutu dan kesahihan tesnya.
- c) Tujuan penggunaan tes
Tujuan penggunaan tes perlu dirumuskan dengan jelas dan tegas, sehingga tidak mengaburkan tester dalam mengambil kesimpulan tentang peserta didik. Ada tujuan tes untuk diagnostik, ada pula untuk mengetahui hasil belajar peserta didik. Semua itu harus dicantumkan dalam keterangan tentang tes tersebut. Jika tujuan penggunaan tes tidak diketahui atau diabaikan, maka fungsi tes tersebut akan hilang dan tidak akan mencapai apa yang diharapkan. Dengan demikian, tester akan memperoleh gambaran yang keliru tentang testi, akhirnya kesimpulan yang ditarik daripadanya akan salah pula.
- d) Sampel
Dalam tes itu disebutkan pula sampel yang akan digunakan dan variasi heterogenitasnya untuk dikenai tes tersebut. Selain itu dinyatakan pula lamanya waktu yang dibutuhkan dalam pelaksanaan tes itu dan berapa kali tes itu dapat dicobakan kepada testi yang sama atau berlainan. Jika ketentuan tentang sampel, waktu, dan frekuensi pelaksanaan ini kurang ditaati, fungsi tes itu akan kurang meyakinkan.
- e) Kesahihan dan keandalan
Agar tes tersebut sah (valid) dan andal (reliabel), maka ketentuan-ketentuan tentang cara atau langkah-langkah yang ditempuh harus dipatuhi, baik oleh tester maupun oleh testi, sehingga pelaksanaannya dapat berjalan dengan lancar tanpa mengalami kesulitan yang berarti. Hal ini sangat berpengaruh terhadap kesahihan dan keandalan suatu tes.
- f) Pengadministrasian

Ketentuan-ketentuan pokok mengenai pengadministrasian suatu tes perlu disusun secara teratur dan baik sesuai dengan fungsi administrasi itu sendiri, mulai dari perencanaan, pelaksanaan, sampai pada penilaian. Dalam perencanaan perlu dimuat waktu, bahan atau materi, tujuan dan cara pelaksanaannya. Sedangkan dalam pelaksanaan perlu dimuat tempat atau ruangan dimana tes itu dilaksanakan, pengawas tes, dan jumlah peserta didik yang mengikuti tes tersebut. Dalam penilaian perlu dimuat teknik atau prosedur mengolah data, sehingga data tersebut dapat memberikan makna bagi semua pihak. Oleh sebab itu, kita harus membuat laporan untuk orang tua, pemerintah, kepala madrasah dan peserta didik itu sendiri.

g) Cara menskor

Setelah tes dilaksanakan dan data sudah terkumpul, selanjutnya perlu diolah. Dalam pengolahan harus diperhatikan pendekatan penilaian yang digunakan, standar norma, *passing grade*, dan peringkat (*ranking*). Untuk pendekatan penilaian dapat digunakan penilaian acuan patokan (*criterion-referenced assessment*) atau penilaian acuan norma (*norm-referenced assessment*). Hal ini bergantung kepada tujuan dan maksud evaluasi itu sendiri. Begitu juga dengan standar norma, ada standar 0 – 4, 0 – 10 dan 0 – 100. Standar norma yang digunakan harus disesuaikan dengan kebutuhan. Di samping itu, perlu pula ditentukan batas lulus (*passing grade*) dan peringkat (*ranking*) dari keseluruhan testi agar guru dapat mengetahui kedudukan seorang peserta didik dibandingkan dengan peserta didik lainnya. Semua catatan dan keterangan mengenai skoring tes ini harus didokumentasikan dalam suatu berkas dan dibuat laporan pemeriksaan untuk dijadikan bahan pedoman dalam pelaksanaan tes berikutnya. Dokumen ini harus dirahasiakan bagi siapapun. Pada zaman modern sekarang ini, ketika teknologi sudah semakin canggih, pelaksanaan penskoran dan penentuan batas lulus dapat dilakukan dengan cepat dan tepat oleh pesawat komputer di samping secara manual.

h) Kunci jawaban

Biasanya pada lembaran soal dilampirkan kunci jawaban sekalian untuk dijadikan dasar dalam pemeriksaan. Ada kalanya lembar kunci jawaban ini disatukan dengan petunjuk pelaksanaan, skoring, dan tata tertib tes. Pada tes tertulis berbentuk esai, kunci jawabannya hanya memuat pokok-pokok materi yang penting saja yang harus dicantumkan oleh testi sebagai syarat dalam

tesnya. Sedangkan dalam tes tertulis berbentuk objektif, kunci jawabannya memuat jawaban yang pasti. Di samping itu, ditetapkan pula ketentuan-ketentuan mengenai cara menggunakan kunci jawaban agar tidak salah penggunaannya.

i) Tabel skor mentah (*raw score*) dan skor terjabar

Selain lampiran-lampiran peraturan mengenai pelaksanaan tes, disertakan pula tabel-tabel yang diperlukan untuk pengolahan skor mentah ke dalam skor terjabar serta petunjuk pelaksanaannya.

j) Penafsiran

Akhirnya, setelah seluruh tes itu rampung dikerjakan sampailah kepada penafsiran tentang hasil tes itu. Kecenderungan apa yang dapat kita temukan dan bagaimana keputusan serta kesimpulannya, akan diperoleh setelah diadakan penafsiran data.

Ketentuan-ketentuan di atas merupakan ketentuan pokok yang harus diperhatikan dalam melaksanakan suatu tes, sehingga hasil tes dapat memenuhi standar yang kita harapkan. Dalam penyelenggaraan suatu tes hendaknya dibentuk suatu panitia dengan beberapa staf anggotanya serta pembagian kerjanya (*job description*). Di samping itu, disusun pula jadwal kerja panitia, dan yang tidak kurang pentingnya adalah tersedianya dana untuk pembiayaan tes tersebut.

Tabel 4.1. Perbedaan Antara Tes Baku Dengan Tes Buatan Guru

Tes Baku	Tes Buatan Guru
<ul style="list-style-type: none"> • Berdasarkan isi dan tujuan-tujuan yang bersifat umum. • Mencakup pengetahuan dan kecakapan yang luas. • Dikembangkan oleh tenaga yang berkompeten dan profesional. • Item-item sudah diujicobakan, dianalisis, dan direvisi. • Memiliki derajat kesahihan dan keandalan yang tinggi. • Memiliki ukuran-ukuran bermacam-macam kelompok yang secara luas mewakili performance seluruh daerah. 	<ul style="list-style-type: none"> • Berdasarkan isi dan tujuan-tujuan yang bersifat khusus. • Mencakup pengetahuan dan kecakapan yang khusus. • Dikembangkan oleh seorang guru tanpa bantuan dari luar. • Item-item jarang diujicobakan sebelum menjadi bagian tes tersebut. • Memiliki derajat kesahihan dan keandalan yang rendah. • Biasanya terbatas pada kelas atau satu sekolah sebagai suatu kelompok pemakainya.

Berdasarkan aspek pengetahuan dan keterampilan, maka tes dapat dibagi menjadi dua jenis, yaitu tes kemampuan dan tes kecepatan.

1. Tes Kemampuan (*power test*)

Prinsip tes kemampuan adalah tidak adanya batasan waktu di dalam pengerjaan tes. Jika waktu tes tidak dibatasi, maka hasil tes dapat mengungkapkan kemampuan peserta didik yang sebenarnya. Sebaliknya, jika waktu pelaksanaan tes dibatasi, maka ada kemungkinan kemampuan peserta didik tidak dapat diungkapkan secara utuh. Artinya, skor yang diperoleh bukan menggambarkan kemampuan peserta didik yang sebenarnya. Namun demikian, bukan berarti peserta didik yang paling lambat harus ditunggu sampai selesai. Tes kemampuan menghendaki agar sebagian peserta didik dapat menyelesaikan tes dalam waktu yang disediakan. Implikasinya adalah guru harus menghitung waktu pelaksanaan tes yang logis, rasional, dan proporsional ketika menyusun kisi-kisi tes.

2. Tes Kecepatan (*speed test*)

Aspek yang diukur dalam tes kecepatan adalah kecepatan peserta didik dalam mengerjakan sesuatu pada waktu atau periode tertentu. Pekerjaan tersebut biasanya relatif mudah, karena aspek yang diukur benar-benar kecepatan bekerja atau kecepatan berpikir peserta didik, bukan kemampuan lainnya. Misalnya, guru ingin mengetes kecepatan berlari, kecepatan membaca, kecepatan mengendarai kendaraan, dan sebagainya dalam waktu yang telah ditentukan.

Tes tertulis atau sering disebut *paper and pencil test* adalah tes yang menuntut jawaban dari peserta didik dalam bentuk tertulis. Tes tertulis ada yang bersifat formal dan ada pula yang bersifat nonformal. Tes yang bersifat formal meliputi jumlah testi yang cukup besar yang diselenggarakan oleh suatu panitia resmi yang diangkat oleh pemerintah. Tes formal mempunyai tujuan yang lebih luas dan didasarkan atas standar tertentu yang berlaku umum. Sedangkan tes nonformal berlaku untuk tujuan tertentu dan lingkungan terbatas yang diselenggarakan langsung oleh pihak pelaksana dalam situasi setengah resmi tanpa melalui institusi resmi. Tes tertulis ada dua bentuk, yaitu bentuk uraian (*essay*) dan bentuk objektif (*objective*).

Menurut sejarah, yang ada lebih dahulu adalah bentuk uraian. Mengingat bentuk uraian ini banyak kelemahannya, maka orang berusaha untuk menyusun tes dalam bentuk yang lain, yaitu tes objektif. Namun demikian, tidak berarti bentuk uraian ditinggalkan sama sekali. Bentuk uraian dapat digunakan untuk mengukur kegiatan-

kegiatan belajar yang sulit diukur oleh bentuk objektif. Disebut bentuk uraian, karena menuntut peserta didik untuk menguraikan, mengorganisasikan dan menyatakan jawaban dengan kata-katanya sendiri dalam bentuk, teknik, dan gaya yang berbeda satu dengan lainnya. Bentuk uraian sering juga disebut bentuk subjektif, karena dalam pelaksanaannya sering dipengaruhi oleh faktor subjektifitas guru. Dilihat dari luas- sempitnya materi yang ditanyakan, maka tes bentuk uraian ini dapat dibagi menjadi dua bentuk, yaitu uraian terbatas (*restricted respons items*) dan uraian bebas (*extended respons items*).

1. Uraian Terbatas

Dalam menjawab soal bentuk uraian terbatas ini, peserta didik harus mengemukakan hal-hal tertentu sebagai batas-batasnya. Walaupun kalimat jawaban peserta didik itu beraneka ragam, tetap harus ada pokok-pokok penting yang terdapat dalam sistematika jawabannya sesuai dengan batas-batas yang telah ditentukan dan dikehendaki dalam soalnya.

2. Uraian Bebas

Dalam bentuk ini peserta didik bebas untuk menjawab soal dengan cara dan sistematika sendiri. Peserta didik bebas mengemukakan pendapat sesuai dengan kemampuannya. Oleh karena itu, setiap peserta didik mempunyai cara dan sistematika yang berbeda-beda. Namun demikian, guru tetap harus mempunyai acuan atau patokan dalam mengoreksi jawaban peserta didik nanti.

Sehubungan dengan kedua bentuk uraian di atas, Depdikbud sering menyebutnya dengan istilah lain, yaitu Bentuk Uraian Objektif (BUO) dan Bentuk Uraian Non Objektif (BUNO). Kedua bentuk ini sebenarnya merupakan bagian dari bentuk uraian terbatas, karena pengelompokan tersebut hanya didasarkan pada pendekatan/cara pemberian skor. Perbedaan BUO dan BUNO terletak pada kepastian pemberian skor. Pada soal BUO, kunci jawaban dan pedoman penskorannya lebih pasti. Kunci jawaban pada soal BUNO, pedoman penskoran dinyatakan dalam rentangan (0 – 4 atau 0 – 10), sehingga pemberian skor dapat dipengaruhi oleh unsur subjektif. Untuk mengurangi unsur subjektifitas ini, Anda dapat melakukannya dengan cara membuat pedoman penskoran secara rinci dan jelas, sehingga pemberian skor dapat relatif sama.

1. Bentuk Uraian Objektif (BUO)

Bentuk uraian seperti ini memiliki sehimpunan jawaban dengan rumusan yang relatif lebih pasti, sehingga dapat dilakukan penskoran secara objektif. Sekalipun pemeriksa berbeda tetapi dapat

menghasilkan skor yang relatif sama. Soal bentuk ini memiliki kunci jawaban yang pasti, sehingga jawaban benar bisa diberi skor 1 dan jawaban salah 0. Anthony J. Nitko (1996) menjelaskan bentuk uraian terbatas dapat digunakan untuk menilai hasil belajar yang kompleks, yaitu berupa kemampuan-kemampuan : menjelaskan hubungan sebab-akibat, melukiskan pengaplikasian prinsip-prinsip, mengajukan argumentasi-argumentasi yang relevan, merumuskan hipotesis dengan tepat, merumuskan asumsi yang tepat, melukiskan keterbatasan data, merumuskan kesimpulan secara tepat, menjelaskan metoda dan prosedur, dan hal-hal sejenis yang menuntut kemampuan peserta didik untuk melengkapi jawabannya. Dalam penskoran bentuk soal uraian objektif, skor hanya dimungkinkan menggunakan dua kategori, yaitu benar atau salah. Untuk setiap kata kunci yang benar diberi skor 1 (satu) dan untuk kata kunci yang dijawab salah atau tidak dijawab diberi skor 0 (nol). Dalam satu rumusan jawaban dapat mengandung lebih dari satu kata kunci, sehingga skor maksimum jawaban dapat lebih dari satu. Kata kunci tersebut dapat berupa kalimat, kata, bilangan, simbol, gambar, grafik, ide, gagasan atau pernyataan. Diharapkan dengan pembagian yang tegas seperti ini, unsur subjektifitas dapat dihindari atau dikurangi. Adapun langkah-langkah pemberian skor soal bentuk uraian objektif adalah:

- a. Tuliskan semua kata kunci atau kemungkinan jawaban benar secara jelas untuk setiap soal.
 - b. Setiap kata kunci yang dijawab benar diberi skor 1. Tidak ada skor setengah untuk jawaban yang kurang sempurna. Jawaban yang diberi skor 1 adalah jawaban sempurna, jawaban lainnya adalah 0.
 - c. Jika satu pertanyaan memiliki beberapa sub pertanyaan, rincilah kata kunci dari jawaban soal tersebut menjadi beberapa kata kunci sub jawaban dan buat skoranya.
 - d. Jumlahkan skor dari semua kata kunci yang telah ditetapkan pada soal tersebut. Jumlah skor ini disebut skor maksimum.
2. Bentuk Uraian Non-Objektif (BUNO)

Bentuk soal seperti ini memiliki rumusan jawaban yang sama dengan rumusan jawaban uraian bebas, yaitu menuntut peserta didik untuk mengingat dan mengorganisasikan (menguraikan dan memadukan) gagasan-gagasan pribadi atau hal-hal yang telah dipelajarinya dengan cara mengemukakan atau mengekspresikan gagasan tersebut dalam bentuk uraian tertulis sehingga dalam penskorannya sangat memungkinkan adanya unsur subjektifitas.

Bentuk uraian bebas dapat digunakan untuk menilai hasil belajar yang bersifat kompleks, seperti kemampuan menghasilkan, menyusun dan menyatakan ide-ide, memadukan berbagai hasil belajar dari berbagai bidang studi, merekayasa bentuk-bentuk orisinal (seperti mendisain sebuah eksperimen), dan menilai arti atau makna suatu ide. Dalam penyekoran soal bentuk uraian non-objektif, skor dijabarkan dalam rentang. Besarnya rentang skor ditetapkan oleh kompleksitas jawaban, seperti 0 – 2, 0 – 4, 0 – 6, 0 – 8, 0 – 10 dan lain-lain. Skor minimal harus 0, karena peserta didik yang tidak menjawab pun akan memperoleh skor minimal tersebut. Sedangkan skor maksimum ditentukan oleh penyusun soal dan keadaan jawaban yang dituntut dalam soal tersebut. Adapun langkah-langkah pemberian skor untuk soal bentuk uraian non-objektif adalah:

- a. Tulislah garis-garis besar jawaban sebagai kriteria jawaban untuk dijadikan pegangan dalam pemberian skor.
- b. Tetapkan rentang skor untuk setiap kriteria jawaban.
- c. Pemberian skor pada setiap jawaban bergantung pada kualitas jawaban yang diberikan oleh peserta didik.
- d. Jumlahkan skor-skor yang diperoleh dari setiap kriteria jawaban sebagai skor peserta didik. Jumlah skor tertinggi dari setiap kriteria jawaban disebut skor maksimum dari suatu soal.
- e. Periksa soal untuk setiap nomor dari semua peserta didik sebelum pindah ke nomor soal yang lain. Tujuannya untuk menghindari pemberian skor berbeda terhadap jawaban yang sama.
- f. Jika setiap butir soal telah selesai diskor, hitunglah jumlah skor perolehan peserta didik untuk setiap soal. Kemudian hitunglah nilai tiap soal dengan rumus :

$$\text{Nilai Tiap Soal} = \frac{\text{Skor perolehan peserta didik}}{\text{Skor maksimum tiap butir soal}} \times \text{Bobot soal}$$
- g. Jumlahkan semua nilai yang diperoleh dari semua soal. Jumlah nilai ini disebut nilai akhir dari suatu perangkat tes yang diberikan.

2. Metode Pengoreksian Soal Bentuk Uraian

Untuk mengoreksi soal bentuk uraian dapat dilakukan dengan tiga metode, yaitu “metode per nomor (*whole method*), metode per lembar (*separated method*), dan metode bersilang (*cross method*)” (Zainal Arifin, 1991: 30).

- a) Metode per nomor. Di sini kita mengoreksi hasil jawaban peserta didik untuk setiap nomor. Misalnya, kita mengoreksi nomor satu

untuk seluruh peserta didik, kemudian nomor dua untuk seluruh peserta didik, dan seterusnya. Kebaikannya adalah pemberian skor yang berbeda atas dua jawaban yang kualitasnya sama hampir tidak akan terjadi, karena jawaban peserta didik yang satu selalu dibandingkan dengan jawaban peserta didik yang lain. Sedangkan kelemahannya adalah pelaksanaannya terlalu berat dan memakan waktu banyak.

- b) Metode per lembar. Di sini kita mengoreksi setiap lembar jawaban peserta didik mulai dari nomor satu sampai dengan nomor terakhir. Kebaikannya adalah relatif lebih murah dan tidak memakan waktu banyak. Sedangkan kelemahannya adalah guru sering memberi skor yang berbeda atas dua jawaban yang sama kualitasnya, atau sebaliknya.
- c) Metode bersilang. Disini kita mengoreksi jawaban peserta didik dengan jalan menukarkan hasil koreksi dari seorang korektor kepada korektor yang lain. Jika telah selesai dikoreksi oleh seorang korektor, lalu dikoreksi kembali oleh korektor yang lain. Kelebihannya adalah faktor subjektif dapat dikurangi. Sedangkan kelemahannya adalah membutuhkan waktu dan tenaga yang banyak.

Dalam pelaksanaan pengoreksian, kita boleh memilih salah satu diantara ketiga metode tersebut, atau mungkin kita menggunakannya secara bervariasi. Hal ini harus disesuaikan dengan kebutuhan. Misalnya, kita menghendaki hasil jawaban yang betul-betul objektif, maka lebih tepat bila kita menggunakan metode bersilang. Sebaliknya, bila ada waktu luang, kita dapat menggunakan metode per nomor atau metode per lembar.

Selanjutnya, Zainal Arifin (1991: 30) mengemukakan “di samping metode-metode di atas, ada juga metode lain untuk mengoreksi jawaban soal bentuk uraian, yaitu *“analytical method dan sorting method”*”.

1. *Analytical method*, yaitu suatu cara untuk mengoreksi jawaban peserta didik dan guru sudah menyiapkan sebuah model jawaban, kemudian dianalisis menjadi beberapa langkah atau unsur yang terpisah, dan setiap langkah disediakan skor-skor tertentu. Setelah satu model jawaban tersusun, maka jawaban masing-masing peserta didik dibandingkan dengan model jawaban tersebut, kemudian diberi skor sesuai dengan tingkat kebenarannya.

Sorting method, yaitu metode memilih yang dipergunakan untuk memberi skor terhadap jawaban-jawaban yang tidak dibagi-bagi menjadi unsur-unsur, juga dapat menggunakan metode lain untuk pemberian skor soal bentuk uraian, yaitu:

1. *Point method*, yaitu setiap jawaban dibandingkan dengan jawaban ideal yang telah ditetapkan dalam kunci jawaban dan skor yang diberikan untuk setiap jawaban akan bergantung kepada derajat kepadanannya dengan kunci jawaban. Metode ini sangat cocok digunakan untuk bentuk uraian terbatas, karena setiap jawaban sudah dibatasi dengan kriteria tertentu.
2. *Rating method*, yaitu setiap jawaban peserta didik ditetapkan dalam salah satu kelompok yang sudah dipilah-pilah berdasarkan kualitasnya selagi jawaban tersebut dibaca. Kelompok-kelompok tersebut menggambarkan kualitas dan menentukan berapa skor yang akan diberikan kepada setiap jawaban. Misalnya, sebuah soal akan diberi skor maksimum 8, maka bagi soal tersebut dapat dibuat 9 kelompok jawaban dari 8 sampai 0. Metode ini sangat cocok digunakan untuk bentuk uraian bebas.

Setiap bentuk soal tentu mempunyai kelebihan dan kekurangan. Begitu juga bentuk uraian. Kebaikan tes bentuk uraian antara lain (1) menyusunnya relatif mudah (2) guru dapat menilai peserta didik mengenai kreatifitas, menganalisa dan mensintesa suatu soal. Hal ini berarti memberikan kebebasan yang luas kepada peserta didik untuk menyatakan tanggapannya (3) guru dapat memperoleh data-data mengenai kepribadian peserta didik (4) peserta didik tidak dapat menerka-nerka (5) derajat ketepatan dan kebenaran peserta didik dapat dilihat dari ungkapan kalimat-kalimatnya (6) sangat cocok untuk mengukur dan menilai hasil belajar yang kompleks, yang sukar diukur dengan mempergunakan bentuk objektif.

Kelemahan tes bentuk uraian antara lain (1) sukar sekali menilai jawaban peserta didik secara tepat dan komprehensif (2) ada kecenderungan guru untuk memberikan nilai seperti biasanya (3) menghendaki respon-respon yang relatif panjang (4) untuk mengoreksi jawaban diperlukan waktu yang lama (5) guru sering terkecoh dalam memberikan nilai, karena keindahan kalimat dan tulisan, bahkan juga oleh lembar jawaban (6) hanya terbatas pada guru-guru yang menguasai materi yang dapat mengoreksi jawaban peserta didik, sehingga kurang praktis bila jumlah peserta didik cukup banyak. Dalam menyusun soal bentuk uraian, ada baiknya kita ikuti petunjuk praktis berikut ini.

1. Materi yang akan diujikan hendaknya materi yang kurang cocok diukur dengan menggunakan bentuk objektif.
2. Setiap pertanyaan hendaknya menggunakan petunjuk dan rumusan yang jelas dan mudah dipahami, sehingga tidak menimbulkan kebingungan pada peserta didik.

3. Jangan memberikan kesempatan kepada peserta didik untuk memilih beberapa soal dari sejumlah soal yang diberikan, sebab cara demikian tidak memungkinkan untuk memperoleh skor yang dapat dibandingkan.
4. Persoalan yang terkandung dalam tes bentuk uraian hendaknya difokuskan pada hal-hal seperti: menelaah persoalan, melukiskan persoalan, menjelaskan persoalan, membandingkan dua hal atau lebih, mengemukakan kritik terhadap sesuatu, menyelesaikan suatu persoalan seperti menghitung, membuat contoh mengenai suatu pengertian, memecahkan suatu persoalan dengan jalan mengaplikasikan prinsip-prinsip yang telah dikuasainya, dan menyusun suatu konsepsi.

3. Analisis Soal Bentuk Uraian

Ada dua cara yang dapat dilakukan untuk menganalisis soal bentuk uraian. Pertama, secara rasional yang dilakukan sebelum tes itu digunakan/ diujicobakan seperti menggunakan kartu telaah. Contoh.

Tabel 4.2 Kartu Telaah Soal Bentuk Uraian

No soal:		Perangkat:	
No	Aspek yang Ditelaah	Ya	Tidak
A. Materi			
1	Soal sesuai dengan indikator.		
2	Batasan pertanyaan dan jawaban yang diharapkan jelas.		
3	Isi materi sesuai dengan tujuan tes.		
4	Isi materi sesuai dengan jenjang, jenis sekolah, dan kelas.		
B. Konstruksi			
5	Rumusan kalimat soal atau pertanyaan harus menggunakan kata tanya atau perintah yang menuntut jawaban terurai.		
6	Ada petunjuk yang jelas tentang cara mengerjakan soal.		
7	Ada pedoman penskoran.		
8	Gambar, grafik, tabel, diagram, dan sejenisnya disajikan dengan jelas dan terbaca.		
C. Bahasa			
9	Rumusan kalimat soal komunikatif.		
10	Butir soal menggunakan bahasa Indonesia		

	yang baik dan benar.		
11	Rumusan soal tidak menggunakan kata/kalimat yang menimbulkan penafsiran ganda atau salah pengertian.		
12	Tidak menggunakan bahasa lokal/daerah.		
13	Rumusan soal tidak mengandung kata-kata yang dapat menyinggung perasaan peserta didik.		
	Catatan :		

Kedua, secara empiris yaitu menganalisis hasil ujian atau hasil uji-coba secara kuantitatif. Untuk itu, ada dua hal yang harus dipelajari:

1. Daya Pembeda Soal

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara peserta didik yang pandai (menguasai materi) dengan peserta didik yang kurang pandai (kurang/tidak menguasai materi). Logikanya adalah peserta didik yang pandai tentu akan lebih mampu menjawab dibandingkan dengan peserta didik yang kurang pandai. Indeks daya pembeda biasanya dinyatakan dengan proporsi. Semakin tinggi proporsi itu, maka semakin baik soal tersebut membedakan antara peserta didik yang pandai dengan peserta didik yang kurang pandai. Untuk menguji daya pembeda (DP) ini, perlu menempuh langkah-langkah sebagai berikut:

- Menghitung jumlah skor total tiap peserta didik.
- Mengurutkan skor total mulai dari skor terbesar sampai dengan skor terkecil.
- Menetapkan kelompok atas dan kelompok bawah. Jika jumlah peserta didik banyak (di atas 30) dapat ditetapkan 27 % .
- Menghitung rata-rata skor untuk masing-masing kelompok (kelompok atas maupun kelompok bawah).
- Menghitung daya pembeda soal dengan rumus:

$$DP = \frac{\bar{X}_{KA} + \bar{X}_{KB}}{\text{Skor Maks}}$$

Keterangan:

DP = daya pembeda

\bar{X}_{KA} = rata-rata kelompok atas

\bar{X}_{KB} = rata-rata kelompok bawah

Skor Maks = skor maksimum

- Membandingkan daya pembeda dengan kriteria seperti berikut:

0,40 ke atas = sangat baik

0,30 – 0,39 = baik

- 0,20 – 0,29 = cukup, soal perlu perbaikan
 0,19 ke bawah = kurang baik, soal harus dibuang

2. Tingkat Kesukaran Soal

Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasa dinyatakan dengan indeks. Indeks ini biasa dinyatakan dengan proporsi yang besarnya antara 0,00 sampai dengan 1,00. Semakin besar indeks tingkat kesukaran berarti soal tersebut semakin mudah. Untuk menghitung tingkat kesukaran soal bentuk uraian, dapat menggunakan langkah-langkah sebagai berikut:

- a. Menghitung rata-rata skor untuk tiap butir soal dengan rumus:

$$\text{Rata-rata} = \frac{\text{Jumlah skor peserta didik tiap soal}}{\text{Jumlah peserta didik}}$$

- b. Menghitung tingkat kesukaran dengan rumus:

$$\text{Tingkat kesukaran} = \frac{\text{Rata-rata}}{\text{Skor maksimum tiap soal}}$$

- c. Membandingkan tingkat kesukaran dengan kriteria berikut:

- 0,00 – 0,30 = sukar
 0,31 – 0,70 = sedang
 0,71 – 1,00 = mudah

- d. Membuat penafsiran tingkat kesukaran dengan cara membandingkan koefisien tingkat kesukaran (poin b) dengan kriteria (poin c).

B. PENGEMBANGAN TES OBJEKTIF

1. Benar-Salah (*true-false, or yes-no*)

Bentuk tes benar-salah (B – S) adalah pernyataan yang mengandung dua kemungkinan jawaban, yaitu benar atau salah. Peserta didik diminta untuk menentukan pilihannya mengenai pertanyaan-pertanyaan atau pernyataan-pernyataan dengan cara seperti yang diminta dalam petunjuk mengerjakan soal. Salah satu fungsi bentuk soal benar-salah adalah untuk mengukur kemampuan peserta didik dalam membedakan antara fakta dengan pendapat. Agar soal dapat berfungsi dengan baik, maka materi yang ditanyakan hendaknya homogen dari segi isi. Bentuk soal seperti ini lebih banyak digunakan untuk mengukur kemampuan mengidentifikasi informasi berdasarkan hubungan yang sederhana. Jika akan digunakan untuk mengukur kemampuan yang lebih tinggi, paling juga untuk kemampuan menghubungkan antara dua hal yang homogen. Dalam penyusunan soal bentuk benar-salah tidak hanya menggunakan kalimat pertanyaan atau pernyataan tetapi juga dalam bentuk gambar, tabel dan diagram.

Bentuk soal benar-salah dapat juga digunakan untuk mengukur kemampuan tentang sebab-akibat. S.Surapranata (2004: 96) menjelaskan “soal semacam ini biasanya mengandung dua hal benar dalam satu pernyataan ataupun pertanyaan dan peserta didik diminta untuk memutuskan benar-salahnya hubungan antara dua hal tersebut”.

Kebaikan tes bentuk B – S antara lain (1) mudah disusun dan dilaksanakan, karena itu banyak digunakan (2) dapat mencakup materi yang lebih luas. Namun demikian, tidak semua materi dapat diukur dengan bentuk benar-salah (3) dapat dinilai dengan cepat dan objektif (4) banyak digunakan untuk mengukur fakta-fakta dan prinsip-prinsip. Sedangkan kelemahan tes bentuk B – S antara lain (1) ada kecenderungan peserta didik menjawab coba-coba (2) pada umumnya memiliki derajat validitas dan reliabilitas yang rendah, kecuali jika itemnya banyak sekali (3) sering terjadi kekaburan, karena itu sukar untuk menyusun item yang benar-benar jelas (4) dan terbatas mengukur aspek pengetahuan saja. Beberapa petunjuk praktis dalam menyusun soal bentuk B – S:

- a. Dalam menyusun item bentuk benar-salah ini hendaknya jumlah item cukup banyak, sehingga dapat dipertanggungjawabkan. Sebaliknya, jika jumlah item kurang dari 50, kiranya kurang dapat dipertanggungjawabkan.
- b. Jumlah item yang benar dan salah hendaknya sama.
- c. Berilah petunjuk cara mengerjakan soal yang jelas dan memakai kalimat yang sederhana.
- d. Hindarkan pernyataan yang terlalu umum, kompleks, dan negatif.
- e. Hindarkan penggunaan kata yang dapat memberi petunjuk tentang jawaban yang dikehendaki.

Kelemahan yang paling menyolok dari bentuk tes benar-salah ini adalah sangat mudahnya ditebak tanpa dapat diketahui oleh korektor. Untuk menghilangkan kelemahan ini, maka orang menambahkan pada item benar-salah ini dengan “koreksi”. Di sini peserta didik tidak hanya dituntut memilih benar atau salah dari setiap item, tetapi harus dapat memberikan koreksi jika item tersebut dinyatakan salah oleh peserta didik yang bersangkutan.

Jika pernyataannya benar, maka tidak perlu dikoreksi lagi, artinya peserta didik langsung menyilang huruf B (benar). Sebaliknya, jika pernyataannya salah, peserta didik harus membenarkan bagian kalimat yang dicetak miring atau digarisbawahi dan menempatkannya pada titik-titik atau garis kosong yang terletak di belakang item yang bersangkutan. Adapun bagian kalimat yang dicetak miring itu harus merupakan inti persoalannya. Jadi, tidak boleh sembarangan kata saja.

2. Pilihan-Ganda (*multiple-choice*)

Soal tes bentuk pilihan-ganda dapat digunakan untuk mengukur hasil belajar yang lebih kompleks dan berkenaan dengan aspek ingatan, pengertian, aplikasi, analisis, sintesis dan evaluasi. Soal tes bentuk pilihan-ganda terdiri atas pembawa pokok persoalan dan pilihan jawaban. Pembawa pokok persoalan dapat dikemukakan dalam bentuk pertanyaan dan dapat pula dalam bentuk pernyataan (*statement*) yang belum sempurna yang sering disebut *stem*. Sedangkan pilihan jawaban itu mungkin berbentuk perkataan, bilangan atau kalimat dan sering disebut *option*. Pilihan jawaban terdiri atas jawaban yang benar atau yang paling benar, selanjutnya disebut kunci jawaban dan kemungkinan jawaban salah yang dinamakan pengecoh (*distractor* atau *decoy* atau *fails*) namun memungkinkan seseorang memilihnya apabila tidak menguasai materi yang ditanyakan dalam soal.

Mengenai jumlah alternatif jawaban sebenarnya tidak ada aturan baku. Anda bisa membuat 3, 4 atau 5 alternatif jawaban. Semakin banyak semakin bagus. Hal ini dimaksudkan untuk mengurangi faktor menebak (*chance of guessing*), sehingga dapat meningkatkan validitas dan reliabilitas soal. Semakin banyak alternatif jawaban, semakin kecil kemungkinan peserta didik menerka. Adapun kemampuan yang dapat diukur oleh bentuk soal pilihan-ganda, antara lain: mengenal istilah, fakta, prinsip, metode, dan prosedur; mengidentifikasi penggunaan fakta dan prinsip; menafsirkan hubungan sebab-akibat; dan menilai metode dan prosedur. Ada beberapa jenis tes bentuk pilihan-ganda, yaitu:

- a. *Distracters*, yaitu setiap pertanyaan atau pernyataan mempunyai beberapa pilihan jawaban yang salah, tetapi disediakan satu pilihan jawaban yang benar. Tugas peserta didik adalah memilih satu jawaban yang benar itu.
- b. Analisis hubungan antara hal, yaitu bentuk soal yang dapat digunakan untuk melihat kemampuan peserta didik dalam menganalisis hubungan antara pernyataan dengan alasan (sebab-akibat).
- c. Variasi negatif, yaitu setiap pertanyaan atau pernyataan mempunyai beberapa pilihan jawaban yang benar tetapi disediakan satu kemungkinan jawaban yang salah. Tugas peserta didik adalah memilih jawaban yang salah tersebut.
- d. Variasi berganda, yaitu memilih beberapa kemungkinan jawaban yang semuanya benar, tetapi ada satu jawaban yang paling benar. Tugas peserta didik adalah memilih jawaban yang paling benar.

- e. Variasi yang tidak lengkap, yaitu pertanyaan atau pernyataan yang memiliki beberapa kemungkinan jawaban yang belum lengkap. Tugas peserta didik adalah mencari satu kemungkinan jawaban yang benar dan melengkapinya.

Kebaikan soal bentuk pilihan-ganda antara lain (1) cara penilaian dapat dilakukan dengan mudah, cepat, dan objektif (2) kemungkinan peserta didik menjawab dengan terkaan dapat dikurangi (3) dapat digunakan untuk menilai kemampuan peserta didik dalam berbagai jenjang kemampuan kognitif (4) dapat digunakan berulang-ulang (5) sangat cocok untuk jumlah peserta tes yang banyak. Adapun kelemahan tes bentuk pilihan-ganda antara lain (1) tidak dapat digunakan untuk mengukur kemampuan verbal dan pemecahan masalah (2) penyusunan soal yang benar-benar baik membutuhkan waktu lama (3) sukar menentukan alternatif jawaban yang benar-benar homogen, logis, dan berfungsi. Beberapa petunjuk praktis dalam menyusun soal bentuk pilihan-ganda:

- a. Harus mengacu kepada kompetensi dasar dan indikator soal.
 - b. Berilah petunjuk mengerjakannya dengan jelas.
 - c. Jangan memasukkan materi soal yang tidak relevan dengan apa yang sudah dipelajari peserta didik.
 - d. Pernyataan pada soal seharusnya merumuskan persoalan yang jelas dan berarti.
 - e. Pernyataan dan pilihan hendaknya merupakan kesatuan kalimat yang tidak terputus.
 - f. Alternatif jawaban harus berfungsi, homogen dan logis.
 - g. Panjang pilihan pada suatu soal hendaknya lebih pendek daripada itemnya.
 - h. Usahakan agar pernyataan dan pilihan tidak mudah diasosiasikan.
 - i. Alternatif jawaban yang betul hendaknya jangan sistematis.
 - j. Harus diyakini benar bahwa hanya ada satu jawaban yang benar.
3. Menjodohkan (*matching*)

Soal tes bentuk menjodohkan sebenarnya masih merupakan bentuk pilihan-ganda. Perbedaannya dengan bentuk pilihan-ganda adalah pilihan-ganda terdiri atas *stem* dan *option*, kemudian peserta didik tinggal memilih salah satu *option* yang dianggap paling tepat. Sedangkan bentuk menjodohkan terdiri atas kumpulan soal dan kumpulan jawaban yang keduanya dikumpulkan pada dua kolom yang berbeda, yaitu kolom sebelah kiri menunjukkan kumpulan persoalan, dan kolom sebelah kanan menunjukkan kumpulan jawaban. Jumlah pilihan jawaban dibuat lebih banyak dari jumlah persoalan.

Bentuk soal menjodohkan sangat baik untuk mengukur kemampuan peserta didik dalam mengidentifikasi informasi berdasarkan hubungan yang sederhana dan kemampuan menghubungkan antara dua hal. Semakin banyak hubungan antara premis dengan respon dibuat, maka semakin baik soal yang disajikan.

Kebaikan soal bentuk menjodohkan antara lain (1) relatif mudah disusun (2) penyekorannya mudah, objektif dan cepat (3) dapat digunakan untuk menilai teori dengan penemunya, sebab dan akibatnya, istilah dan definisinya (4) materi tes cukup luas. Adapun kelemahan soal bentuk menjodohkan yaitu (1) ada kecenderungan untuk menekankan ingatan saja (2) kurang baik untuk menilai pengertian guna membuat tafsiran. Beberapa petunjuk praktis dalam menyusun soal bentuk menjodohkan :

- a. Buatlah petunjuk tes dengan jelas, singkat, dan mudah dipahami.
 - b. Harus sesuai dengan kompetensi dasar dan indikator.
 - c. Hendaknya kumpulan soal diletakkan di sebelah kiri sedangkan jawabannya di sebelah kanan.
 - d. Jumlah alternatif jawaban hendaknya lebih banyak daripada jumlah soal.
 - e. Susunlah item-item dan alternatif jawaban dengan sistematika tertentu. Misalnya, sebelum pada pokok persoalan, didahului dengan *stem*, atau bisa juga langsung pada pokok persoalan.
 - f. Hendaknya seluruh kelompok soal dan jawaban hanya terdapat dalam satu halaman.
 - g. Gunakan kalimat yang singkat dan langsung terarah pada pokok persoalan.
4. Jawaban Singkat (*short answer*) dan Melengkapi (*completion*)

Kedua bentuk tes ini masing-masing menghendaki jawaban dengan kalimat dan atau angka-angka yang hanya dapat dinilai benar atau salah. Soal tes bentuk jawaban singkat biasanya dikemukakan dalam bentuk pertanyaan. Dengan kata lain, soal tersebut berupa suatu kalimat bertanya yang dapat dijawab dengan singkat, berupa kata, frase, nama, tempat, nama tokoh, lambang, dan lain-lain.

Kebaikan tes bentuk jawaban singkat dan melengkapi antara lain (1) relatif mudah disusun (2) sangat baik untuk menilai kemampuan peserta didik yang berkenaan dengan fakta-fakta, prinsip-prinsip, dan terminologi (3) menuntut peserta didik untuk mengemukakan pendapatnya secara singkat dan jelas (4) pemeriksaan lembar jawaban dapat dilakukan dengan objektif. Adapun kelemahannya adalah (1) pada umumnya hanya berkenaan dengan kemampuan mengingat saja, sedangkan kemampuan yang lain agak terabaikan (2) pada soal bentuk

melengkapi, jika titik-titik kosong yang harus diisi terlalu banyak, para peserta didik sering terkecoh (3) dalam memeriksa lembar jawaban dibutuhkan waktu yang cukup banyak.

Beberapa petunjuk praktis dalam menyusun soal bentuk jawaban singkat dan melengkapi:

- a. Hendaknya tidak menggunakan soal yang terbuka, sehingga ada kemungkinan peserta didik menjawab secara terurai.
- b. Untuk soal tes bentuk melengkapi hendaknya tidak mengambil pernyataan langsung dari buku (*textbook*).
- c. Titik-titik kosong sebagai tempat jawaban hendaknya diletakkan pada akhir atau dekat akhir kalimat daripada pada awal kalimat.
- d. Jangan menyediakan titik-titik kosong terlalu banyak. Pilihlah untuk masalah yang urgen saja.
- e. Pernyataan hendaknya hanya mengandung satu alternatif jawaban.
- f. Jika perlu dapat digunakan gambar-gambar sehingga dapat dipersingkat dan jelas.

Kebaikan tes objektif antara lain (1) seluruh ruang lingkup (*scope*) yang diajarkan dapat dinyatakan pada item-item tes objektif (2) kemungkinan jawaban spekulatif dalam ujian dapat dihindarkan (3) jawaban bersifat mutlak, karena itu penilaian dapat dilakukan secara objektif (4) pengoreksian dapat dilakukan oleh siapa saja, sekalipun tidak mengetahui dan menguasai materinya (5) pemberian skor dapat dilakukan dengan mudah dan cepat (6) korektor tidak akan terpengaruh oleh baik-buruknya tulisan (7) tidak mungkin terjadi dua orang peserta didik yang jawabannya sama, tetapi mendapat skor yang berbeda. Sedangkan kelemahannya adalah (1) mengkontruksi soalnya sangat sulit (2) membutuhkan waktu yang lama (3) ada kemungkinan peserta didik mencontoh jawaban orang lain dan berpikir pasif (4) umumnya hanya mampu mengukur proses-proses mental yang dangkal.

C. PENGEMBANGAN TES LISAN

Tes lisan adalah tes yang menuntut jawaban dari peserta didik dalam bentuk lisan. Peserta didik akan mengucapkan jawaban dengan kata-katanya sendiri sesuai dengan pertanyaan atau perintah yang diberikan. Tes lisan dapat berbentuk seperti berikut:

1. Seorang guru menilai seorang peserta didik.
2. Seorang guru menilai sekelompok peserta didik.
3. Sekelompok guru menilai seorang peserta didik.
4. Sekelompok guru menilai sekelompok peserta didik.

Kebaikan tes lisan antara lain (1) dapat mengetahui langsung kemampuan peserta didik dalam mengemukakan pendapatnya secara lisan (2) tidak perlu menyusun soal-soal secara terurai, tetapi cukup mencatat pokok-pokok permasalahannya saja (3) kemungkinan peserta didik akan menerka-nerka jawaban dan berspekulasi dapat dihindari. Sedangkan kelemahannya adalah (1) memakan waktu yang cukup banyak, apalagi jika jumlah peserta-didiknya banyak (2) sering muncul unsur subjektifitas bilamana dalam suasana ujian lisan itu hanya ada seorang guru dan seorang peserta didik. Beberapa petunjuk praktis dalam pelaksanaan tes lisan adalah:

1. Jangan terpengaruh oleh faktor-faktor subjektifitas, misalnya dilihat dari kecantikan, kekayaan, anak pejabat atau bukan, hubungan keluarga.
2. Berikanlah skor bagi setiap jawaban yang dikemukakan oleh peserta didik. Biasanya kita memberikan penilaian setelah tes itu selesai. Cara ini termasuk cara yang kurang baik, akibatnya penilaian akan dipengaruhi oleh jawaban-jawaban yang terakhir.
3. Catatlah hal-hal atau masalah yang akan ditanyakan dan ruang lingkup jawaban yang diminta untuk setiap pertanyaan. Hal ini dimaksudkan agar jangan sampai pertanyaan yang diajukan menyimpang dari permasalahan dan tak sesuai dengan jawaban peserta didik.
4. Ciptakan suasana ujian yang menyenangkan. Hal ini dimaksudkan agar peserta didik tidak ketakutan menghadapi ujian lisan tersebut. Kadang-kadang ada juga guru yang sampai berbuat tidak wajar seperti membentak-bentak peserta didik, dan mungkin pula bertindak berlebihan. Tindakan ini harus dihindari, karena dapat mengakibatkan proses pemikiran peserta didik menjadi terhambat, sehingga apa yang dikemukakan oleh mereka tidak mencerminkan kemampuan yang sesungguhnya.
5. Jangan mengubah suasana ujian lisan menjadi suasana diskusi atau suasana ngobrol santai atau juga menjadi suasana pembelajaran.

Demikianlah beberapa kelebihan dan kelemahan tes lisan berikut petunjuk praktisnya. Petunjuk ini dapat dijadikan pegangan atau pedoman bagi guru dalam menyelenggarakan tes lisan. Petunjuk-petunjuk praktis untuk suatu ujian biasanya telah dimuat sebagai pedoman seperti yang telah disebutkan tadi. Jadi, Anda harus mempelajari petunjuk praktis itu sebaik-baiknya sebelum kegiatan tes dimulai.

D. PENGEMBANGAN TES TINDAKAN (*PERFORMANCE TEST*)

Tes perbuatan atau tes praktik adalah tes yang menuntut jawaban peserta didik dalam bentuk perilaku, tindakan, atau perbuatan. Lebih jauh Stiggins (1994: 375) mengemukakan “tes tindakan adalah suatu bentuk tes dimana peserta didik diminta untuk melakukan kegiatan khusus di bawah pengawasan penguji yang akan mengobservasi penampilannya dan membuat keputusan tentang kualitas hasil belajar yang didemonstrasikan”. Peserta didik bertindak sesuai dengan apa yang diperintahkan dan ditanyakan.

Tes tindakan sangat bermanfaat untuk memperbaiki kemampuan/perilaku peserta didik, karena secara objektif kesalahan-kesalahan yang dibuat oleh peserta didik dapat diamati dan diukur, sehingga menjadi dasar pertimbangan untuk praktik selanjutnya. Sebagaimana jenis tes yang lain, tes tindakanpun mempunyai kelebihan dan kekurangan. Kelebihan tes tindakan adalah (1) satu-satunya teknik tes yang dapat digunakan untuk mengetahui hasil belajar dalam bidang keterampilan, (2) sangat baik digunakan untuk mencocokkan kesesuaian antara pengetahuan teori dengan keterampilan praktik, sehingga hasil penilaian menjadi lengkap (3) dalam pelaksanaannya tidak memungkinkan peserta didik untuk menyontek (4) guru dapat mengenal lebih dalam tentang karakteristik masing-masing peserta didik sebagai dasar tindak lanjut hasil penilaian, seperti pembelajaran remedial.

Adapun kelemahan/kekurangan tes tindakan adalah (1) memakan waktu yang lama (2) dalam hal tertentu membutuhkan biaya yang besar (3) cepat membosankan (4) jika tes tindakan sudah menjadi sesuatu yang rutin, maka ia tidak mempunyai arti apa-apa lagi (5) memerlukan syarat-syarat pendukung yang lengkap, baik waktu, tenaga maupun biaya. Jika syarat-syarat tersebut tidak terpenuhi, maka hasil penilaian tidak dapat dipertanggungjawabkan dengan baik.

BAB V

PENGEMBANGAN ALAT EVALUASI JENIS NON TES

A. OBSERVASI (*OBSERVATION*)

Sebenarnya observasi merupakan suatu proses yang alami, dimana kita semua sering melakukannya, baik secara sadar maupun tidak sadar di dalam kehidupan sehari-hari. Di dalam kelas, kita sering melihat, mengamati dan melakukan interpretasi. Dalam kehidupan sehari-haripun kita sering mengamati orang lain. Pentingnya observasi dalam kegiatan evaluasi pembelajaran mengharuskan guru untuk memahami lebih jauh tentang *judgement*, bertindak secara reflektif, dan menggunakan komentar orang lain sebagai informasi untuk membuat *judgement* yang lebih reliabel.

Hal yang harus dipahami adalah bahwa tidak semua apa yang dilihat disebut observasi. Observasi yang dilakukan di kelas tidak cukup dengan hanya duduk dan melihat melainkan harus dilakukan secara sistematis, sesuai dengan aspek-aspek tertentu, dan berdasarkan tujuan yang jelas. Untuk memperoleh hasil observasi yang baik, maka kemampuan dalam melakukan pengamatan harus sering dilatih, mulai dari hal-hal yang sederhana sampai dengan hal-hal yang kompleks.

Observasi merupakan salah satu alat evaluasi jenis nontes yang dilakukan dengan jalan pengamatan dan pencatatan secara sistematis, logis, objektif dan rasional mengenai berbagai fenomena, baik dalam situasi yang sebenarnya maupun dalam situasi buatan untuk mencapai tujuan tertentu. Observasi tidak hanya digunakan dalam kegiatan evaluasi, tetapi juga dalam bidang penelitian, terutama penelitian kualitatif (*qualitative research*). Tujuan utama observasi adalah (1) untuk mengumpulkan data dan informasi mengenai suatu fenomena, baik yang berupa peristiwa maupun tindakan, baik dalam situasi yang sesungguhnya maupun dalam situasi buatan, (2) untuk mengukur perilaku kelas, interaksi antara peserta didik dengan guru, dan faktor-faktor yang dapat diamati lainnya, terutama kecakapan sosial (*social skills*). Dalam evaluasi, observasi dapat digunakan untuk menilai proses dan hasil belajar peserta didik, seperti tingkah laku peserta didik pada waktu belajar, berdiskusi, mengerjakan tugas, dan lain-lain. Jika Anda ingin menggunakan observasi sebagai alat evaluasi, maka harus memahami tentang:

1. Konsep dasar observasi, mulai dari pengertian, tujuan, fungsi, peranan, karakteristik, prinsip-prinsip sampai dengan prosedur observasi.
2. Perencanaan observasi, seperti menentukan kegiatan apa yang akan diobservasi, siapa yang akan melakukan observasi, rencana sampling, menyusun pedoman observasi, melatih pihak-pihak yang akan melakukan observasi dalam menggunakan pedoman observasi.
3. Prosedur observasi, mulai dari perencanaan, pelaksanaan, pengolahan dan penafsiran sampai dengan pelaporan hasil observasi.

Observasi mempunyai beberapa karakteristik, antara lain (1) mempunyai arah dan tujuan yang jelas. Hal ini dimaksudkan agar pelaksanaan observasi tidak menyimpang dari permasalahan. Oleh sebab itu, dalam pelaksanaannya harus ada pedoman observasi. (2) bersifat ilmiah, yaitu dilakukan secara sistematis, logis, kritis, objektif dan rasional (3) terdapat berbagai aspek-aspek yang akan diobservasi dan (4) praktis penggunaannya. Selanjutnya Good dkk. mengemukakan enam ciri observasi, yaitu:

1. Observasi mempunyai arah yang khusus, bukan secara tidak teratur melihat sekeliling untuk mencari kesan-kesan umum.
2. Observasi ilmiah tentang tingkah laku adalah sistematis, bukan secara sesuka hati dan untung-untungan mendekati situasi.
3. Observasi bersifat kuantitatif, mencatat jumlah peristiwa tentang tipe-tipe tingkah laku tertentu.
4. Observasi mengadakan pencatatan dengan segera; pencatatan-pencatatan dilakukan secepat-cepatnya, bukan menyandarkan diri pada ingatan.
5. Observasi meminta keahlian, dilakukan oleh seseorang yang memang telah terlatih untuk melakukannya.
6. Hasil-hasil observasi dapat dicek dan dibuktikan untuk menjamin keadaan dan kesahihan (C.V. Good, A.S. Barr, and D.E. Scates, 1936: 404-406).

Ciri-ciri observasi yang dikemukakan oleh Good dkk. mempunyai kelemahan, antara lain (1) dalam penyelidikan yang bersifat eksploitatif, justru yang bersifat kuantitatif kebanyakan dikesampingkan (2) dalam observasi partisipan tidak dapat dilakukan pencatatan dengan segera. Oleh sebab itu, observasi harus dilakukan dengan hati-hati dan terencana. Dilihat dari kerangka kerjanya, observasi dibedakan menjadi dua jenis, yaitu:

1. Observasi berstruktur, yaitu semua kegiatan guru sebagai observer telah ditetapkan terlebih dahulu berdasarkan kerangka kerja yang berisi faktor-faktor yang telah diatur kategorisasinya. Isi dan luas materi observasi telah ditetapkan dan dibatasi dengan jelas dan tegas.

2. Observasi tak berstruktur, yaitu semua kegiatan guru sebagai observer tidak dibatasi oleh suatu kerangka kerja yang pasti. Kegiatan observer hanya dibatasi oleh tujuan observasi itu sendiri.

Sedangkan bila dilihat dari teknis pelaksanaannya, observasi dapat ditempuh melalui tiga cara, yaitu:

1. Observasi langsung, yaitu observasi yang dilakukan secara langsung terhadap objek yang diselidiki.
2. Observasi tak langsung, yaitu observasi yang dilakukan melalui perantara, baik teknik maupun alat tertentu.
3. Observasi partisipasi, yaitu observasi yang dilakukan dengan cara ikut ambil bagian atau melibatkan diri dalam situasi objek yang diteliti.

Selanjutnya, Sutrisno Hadi (1981: 141) mengemukakan ada tiga jenis observasi yang masing-masing hanya cocok untuk keadaan-keadaan tertentu, yaitu "observasi partisipan-observasi nonpartisipan, observasi sistematis-observasi nonsistematis, dan observasi eksperimental-observasi noneksperimental".

Observasi partisipan adalah observasi yang dilakukan dimana observer turut ambil bagian dalam peri kehidupan orang atau objek-objek yang diobservasi. Sedangkan observasi dengan pura-pura disebut *quasi participant observation*. Jika unsur-unsur partisipasi sama sekali tidak terdapat didalamnya, maka disebut *nonparticipant observation*. Observasi sistematis (*systematic observation*) disebut juga observasi berstruktur (*structured observation*). Ciri pokok observasi ini adalah adanya kerangka yang memuat faktor-faktor yang telah diatur kategorisasinya lebih dahulu, dan ciri-ciri khusus dari tiap-tiap faktor dalam kategori-kategori itu. Sedangkan observasi yang tidak menggunakan kerangka disebut observasi non-sistematis. Kadang-kadang observasi sistematis menggunakan beberapa macam alat pencatat mekanis (*mechanical recording devices*) seperti film, kamera, *tape recorder*. Keuntungannya adalah kita dapat memutarinya kembali setiap waktu bila diperlukan, sehingga dapat dianalisis lebih lanjut. Kelemahannya antara lain membutuhkan biaya yang besar dan tenaga yang profesional.

Dalam peristiwa-peristiwa tertentu mungkin tidak terlibat dalam dinamika dan kompleksitas situasi yang diselidikinya, tetapi perlu mengendalikan unsur-unsur penting dalam situasi tertentu, sehingga situasi itu dapat diatur sesuai dengan tujuan observasi dan dapat dikendalikan untuk menghindari bahaya timbulnya faktor-faktor yang tak diharapkan. Observasi yang dilakukan dalam situasi seperti itu disebut observasi eksperimental atau observasi dalam situasi tes. Observasi eksperimental biasanya tidak memerlukan observer yang banyak. Faktor-faktor lain yang dapat mempengaruhi tingkah laku observasi (yang

diobservasi) telah dikontrol secermat-cermatnya, sehingga observasi ini dipandang orang sebagai suatu alat penilaian yang relatif murni untuk mengamati pengaruh kondisi-kondisi tertentu terhadap tingkah laku peserta didik. Kebaikan observasi eksperimental antara lain:

1. Tersedianya kesempatan bagi guru untuk mengamati sifat-sifat tertentu dari peserta didik yang jarang sekali timbul dalam keadaan normal. Misalnya, keberanian, reaksi-reaksi terhadap frustrasi, dan ketidakjujuran.
2. Observasi ini merupakan observasi yang dibakukan secermat-cermatnya.

Observasi kelas merupakan sumber informasi yang penting di dalam evaluasi. Untuk mempermudah proses pengamatan dan mencatat apa yang terjadi di dalam kelas, dapat menggunakan selembar kertas yang cukup lebar dan selanjutnya menuliskan nama-nama peserta didik yang disusun dalam sebuah daftar. Selembar kertas ini selanjutnya disebut pedoman observasi. Melalui pedoman observasi ini, kita dapat mengetahui apa yang terjadi di kelas dan apa yang dilakukan oleh setiap peserta didik dalam mengikuti kegiatan pembelajaran. Pedoman atau lembar observasi ini harus terus diisi oleh guru dengan catatan baru, sehingga perkembangan peserta didik dari waktu ke waktu dapat diketahui.

Sebagaimana alat evaluasi yang lain, observasi secara umum mempunyai kelebihan dan kekurangan. Kebaikan observasi antara lain (1) observasi merupakan alat untuk mengamati berbagai macam fenomena (2) observasi cocok untuk mengamati peserta didik yang sedang melakukan suatu kegiatan (3) banyak hal yang tidak dapat diukur dengan tes, tetapi justru lebih tepat dengan observasi (4) tidak terikat dengan laporan pribadi. Adapun kelemahannya adalah (1) sering kali pelaksanaan observasi terganggu oleh keadaan cuaca, bahkan ada kesan yang kurang menyenangkan dari observer ataupun dari observasi itu sendiri (2) biasanya masalah pribadi sulit diamati (3) jika proses yang diamati memakan waktu lama, maka observer sering menjadi jenuh. Untuk menyusun pedoman observasi, sebaiknya mengikuti langkah-langkah sebagai berikut:

1. Merumuskan tujuan observasi.
2. Membuat *lay-out* atau kisi-kisi observasi.
3. Menyusun pedoman observasi.
4. Menyusun aspek-aspek yang akan diobservasi, baik yang berkenaan dengan proses belajar peserta didik maupun kepribadiannya.
5. Melakukan uji-coba pedoman observasi untuk melihat kelemahan-kelemahan pedoman observasi.

6. Merevisi pedoman observasi berdasarkan hasil uji-coba.
7. Melaksanakan observasi pada saat kegiatan berlangsung.
8. Mengolah dan menafsirkan hasil observasi.

B. WAWANCARA (*INTERVIEW*)

Wawancara merupakan salah satu bentuk alat evaluasi jenis non-tes yang dilakukan melalui percakapan dan tanya-jawab, baik langsung maupun tidak langsung dengan peserta didik. Pengertian wawancara langsung adalah wawancara yang dilakukan secara langsung antara pewawancara (*interviewer*) atau guru dengan orang yang diwawancarai (*interviewee*) atau peserta didik tanpa melalui perantara. Sedangkan wawancara tidak langsung artinya pewawancara atau guru menanyakan sesuatu kepada peserta didik melalui perantara orang lain atau media. Jadi, tidak menemui langsung kepada sumbernya. Tujuan wawancara adalah:

1. Untuk memperoleh informasi secara langsung guna menjelaskan suatu situasi dan kondisi tertentu.
2. Untuk melengkapi suatu penyelidikan ilmiah.
3. Untuk memperoleh data agar dapat mempengaruhi situasi atau orang tertentu.

Wawancara mempunyai beberapa kelebihan dan kekurangan. Kelebihan wawancara antara lain (1) dapat berkomunikasi secara langsung kepada peserta didik, sehingga informasi yang diperoleh dapat diketahui objektivitasnya (2) dapat memperbaiki proses dan hasil belajar (3) pelaksanaan wawancara lebih fleksibel, dinamis dan personal. Sedangkan kelemahan wawancara adalah (1) jika jumlah peserta didik cukup banyak, maka proses wawancara banyak menggunakan waktu, tenaga, dan biaya (2) adakalanya terjadi wawancara yang berlarut-larut tanpa arah, sehingga data kurang dapat memenuhi apa yang diharapkan (3) sering timbul sikap yang kurang baik dari peserta didik yang diwawancarai dan sikap *overaction* dari guru sebagai pewawancara, karena itu perlu adanya adaptasi diri antara pewawancara dengan orang yang diwawancarai. Pertanyaan wawancara dapat menggunakan bentuk seperti berikut:

1. Bentuk pertanyaan berstruktur, yaitu pertanyaan yang menuntut jawaban agar sesuai dengan apa yang terkandung dalam pertanyaan tersebut. Pertanyaan semacam ini biasanya digunakan jika masalahnya tidak terlalu kompleks dan jawabannya sudah konkret.
2. Bentuk pertanyaan tak berstruktur, yaitu pertanyaan yang bersifat terbuka dimana peserta didik secara bebas menjawab pertanyaan

tersebut. Pertanyaan semacam ini tidak memberi struktur jawaban kepada peserta didik, karena jawaban dalam pertanyaan itu bebas.

3. Bentuk pertanyaan campuran, yaitu pertanyaan yang menuntut jawaban campuran, ada yang berstruktur ada pula yang bebas.

Untuk menyusun pedoman wawancara, sebaiknya mengikuti langkah-langkah sebagai berikut:

1. Merumuskan tujuan wawancara
2. Membuat kisi-kisi atau *layout* dan pedoman wawancara.
3. Menyusun pertanyaan sesuai dengan data yang diperlukan dan bentuk pertanyaan yang diinginkan. Untuk itu perlu diperhatikan kata-kata yang digunakan, cara bertanya, dan jangan membuat peserta didik bersikap defensif.
4. Melaksanakan uji-coba untuk melihat kelemahan-kelemahan pertanyaan yang disusun, sehingga dapat diperbaiki lagi.
5. Melaksanakan wawancara dalam situasi yang sebenarnya.

Contoh:

Kisi-kisi Pedoman Wawancara

No.	Masalah	Tujuan	Pertanyaan	Bentuk Pertanyaan

Format Pedoman Wawancara

No.	Aspek-aspek yang diwawancara	Ringkasan Jawaban	Ket.
1.			
2.			
3.			
4.			
5.			
Dst.			

Dalam melaksanakan wawancara, harus memperhatikan hal-hal sebagai berikut:

1. Hubungan baik antara pewawancara dengan orang yang diwawancarai perlu dipupuk dan dibina, sehingga akan tampak hubungan yang sehat dan harmonis.
2. Dalam wawancara jangan terlalu kaku, tunjukkan sikap yang bebas, ramah, terbuka, dan adaptasikan diri dengannya.
3. Perlakukan responden itu sebagai sesama manusia secara jujur.

4. Hilangkan prasangka-prasangka yang kurang baik, sehingga pertanyaan-pertanyaan yang diajukan bersifat netral.
5. Pertanyaan hendaknya jelas, tepat, dengan bahasa yang sederhana.

C. SKALA SIKAP (*ATTITUDE SCALE*)

Sikap merupakan suatu kecenderungan tingkah laku untuk berbuat sesuatu dengan cara, metode, teknik dan pola tertentu terhadap dunia sekitarnya, baik berupa orang-orang maupun berupa objek-objek tertentu. Sikap mengacu kepada perbuatan atau perilaku seseorang, tetapi tidak berarti semua perbuatan identik dengan sikap. Perbuatan seseorang mungkin saja bertentangan dengan sikapnya. Perlu mengetahui norma-norma yang ada pada peserta didik, bahkan sikap peserta didik terhadap dunia sekitarnya, terutama terhadap mata pelajaran dan lingkungan madrasah. Jika terdapat sikap peserta didik yang negatif, perlu mencari suatu cara atau teknik tertentu untuk menempatkan atau mengubah sikap negatif itu menjadi sikap yang positif.

Dalam mengukur sikap, hendaknya memperhatikan tiga komponen sikap, yaitu (1) kognisi, yaitu berkenaan dengan pengetahuan peserta didik tentang objek, (2) afeksi, yaitu berkenaan dengan perasaan peserta didik terhadap objek, dan (3) konasi, yaitu berkenaan dengan kecenderungan berperilaku peserta didik terhadap objek. Juga harus memilih salah satu model skala sikap. Adapun model-model skala sikap yang biasa digunakan untuk menilai sikap peserta didik terhadap suatu objek, antara lain:

1. Menggunakan bilangan untuk menunjukkan tingkat-tingkat dari objek sikap yang dinilai, seperti 1, 2, 3, 4 dan seterusnya.
2. Menggunakan frekuensi terjadinya atau timbulnya sikap itu, seperti: selalu, seringkali, kadang-kadang, pernah dan tidak pernah.
3. Menggunakan istilah-istilah yang bersifat kualitatif, seperti: bagus sekali, baik, sedang, dan kurang. Ada juga istilah-istilah lain, seperti: sangat setuju, setuju, ragu-ragu (tidak punya pendapat), tidak setuju, dan sangat tidak setuju.
4. Menggunakan istilah-istilah yang menunjukkan status/kedudukan, seperti: sangat rendah, di bawah rata-rata, di atas rata-rata, dan sangat tinggi.
5. Menggunakan kode bilangan atau huruf, seperti: selalu (diberi kode 5), kadang-kadang (4), jarang (3), jarang sekali (2), dan tidak pernah (1).

Salah satu model untuk mengukur sikap, yaitu dengan menggunakan skala sikap yang dikembangkan oleh Likert. Dalam skala Likert, peserta didik tidak disuruh memilih pernyataan-pernyataan yang positif saja, tetapi memilih juga pernyataan-pernyataan yang negatif. Tiap

item dibagi ke dalam lima skala, yaitu sangat setuju, setuju, tidak tentu, tidak setuju, dan sangat tidak setuju. Setiap pernyataan positif diberi bobot 4, 3, 2, 1, dan 0, sedangkan pernyataan negatif diberi bobot sebaliknya, yaitu 0, 1, 2, 3, dan 4. Untuk menyusun skala Likert, perlu mengikuti langkah-langkah sebagai berikut:

1. Memilih variabel afektif yang akan diukur.
2. Membuat beberapa pernyataan tentang variabel afektif yang akan diukur.
3. Mengklasifikasikan pernyataan positif dan negatif.
4. Menentukan jumlah gradual dan frase atau angka yang dapat menjadi alternatif pilihan.
5. Menyusun pernyataan dan pilihan jawaban menjadi sebuah alat penilaian.
6. Melakukan uji-coba.
7. Membuang butir-butir pernyataan yang kurang baik.
8. Melaksanakan penilaian.

Options pada skala Likert tidak disusun secara berurutan, tetapi dicampuradukkan. Hal ini dimaksudkan untuk menghindari adanya jawaban yang mempunyai kecenderungan untuk memilih tempat yang sama, seperti selalu ingin memilih option nomor 3, 5, dan nomor-nomor tengah lainnya. Kecenderungan untuk memilih nomor-nomor ini karena option yang disediakan sudah diurutkan. Skala Likert biasanya menggunakan skala dengan lima kategori, tetapi dalam hal tertentu kita bisa menggunakan kategori-kategori yang lain dengan jumlah kategori ganjil, misalnya 3, 5, 7, 9, 11 dan seterusnya, sehingga ada kategori tengah-tengah yang merupakan kategori netral. Sebenarnya, kita dapat mengembangkan sendiri model skala sikap.

Di samping itu, ada juga model lain yang dapat digunakan dalam pengukuran atau penilaian sikap sosial, yaitu model yang dipergunakan oleh Allport-Vernon- Lindezey. Studi ini disebut studi tentang nilai-nilai (*study of values*). Studi ini bermaksud untuk melihat apakah seseorang lebih mengutamakan nilai-nilai politik, nilai-nilai sosial, nilai-nilai ekonomi, nilai-nilai keagamaan, atau nilai-nilai kebenaran.

BAB VI

PENILAIAN PORTOFOLIO

A. PENGERTIAN PENILAIAN PORTOFOLIO

Popham (1995: 163) menjelaskan, portofolio adalah sekumpulan sistematis tentang pekerjaan seseorang. Dalam pendidikan, portofolio mengacu pada kumpulan sistematis mengenai pekerjaan siswa. Hampir semua dengan pendapat Popham, menurut Genesee & Upshur (1997: 99), portofolio adalah sekumpulan pekerjaan siswa yang dapat menunjukkan kepada mereka (juga bagi yang lain) atas usaha, kemajuan, dan pencapaian mereka dalam bidang studi tertentu. Ditegaskan oleh Epstein (2003) bahwa portofolio, dalam konteks kelas, adalah kumpulan koleksi pekerjaan siswa yang menunjukkan penguasaan pengetahuan, keterampilan, dan sikap siswa. Pekerjaan siswa dalam portofolio sering mengacu pada benda atau barang. Sementara itu menurut Costantino & Lorenzo (1988), portofolio adalah suatu seleksi dari artefak (benda/barang hasil kecerdasan manusia) dan catatan reflektif yang mempresentasikan pengalaman-pengalaman profesional, kompetensi, dan perkembangan dalam ukuran waktu tertentu. Dikemukakan oleh mereka, bahwa portofolio dapat berupa sebuah *file folder*, kotak, atau wadah. Berdasarkan beberapa pendapat di atas dapat dinyatakan bahwa portofolio adalah kumpulan hasil karya, pekerjaan, atau tugas siswa. Karya-karya yang disusun berdasarkan urutan kategori kegiatan siswa itu dipilih dan nilai, sehingga dapat menggambarkan perkembangan kemampuan atau kompetensi siswa. Dengan kata lain, portofolio sangat sesuai untuk mengetahui perkembangan unjuk kerja siswa.

Penilaian portofolio merupakan penilaian berkelanjutan yang didasarkan pada kumpulan informasi yang menunjukkan perkembangan kemampuan peserta didik dalam satu periode tertentu. Informasi tersebut dapat berupa karya peserta didik dari proses pembelajaran yang dianggap terbaik oleh peserta didik, hasil tes (bukan nilai) atau bentuk informasi lain yang terkait dengan kompetensi tertentu dalam satu mata pelajaran (Depdiknas, 2006: 15-16).

Penilaian portofolio, pada dasarnya, menilai karya-karya siswa secara individu pada satu periode untuk suatu mata pelajaran. Akhir suatu periode hasil karya tersebut dikumpulkan dan dinilai oleh guru dan peserta didik sendiri. Berdasarkan informasi perkembangan tersebut, guru dan peserta didik sendiri dapat menilai perkembangan kemampuan peserta didik dan terus melakukan perbaikan. Dengan demikian,

portofolio dapat memperlihatkan perkembangan kemajuan belajar peserta didik melalui karyanya, antara lain: karangan, puisi, surat, komposisi muasik, gambar, foto, lukisan, resensi buku/literature, laporan penelitian, synopsis, dan sebagainya.

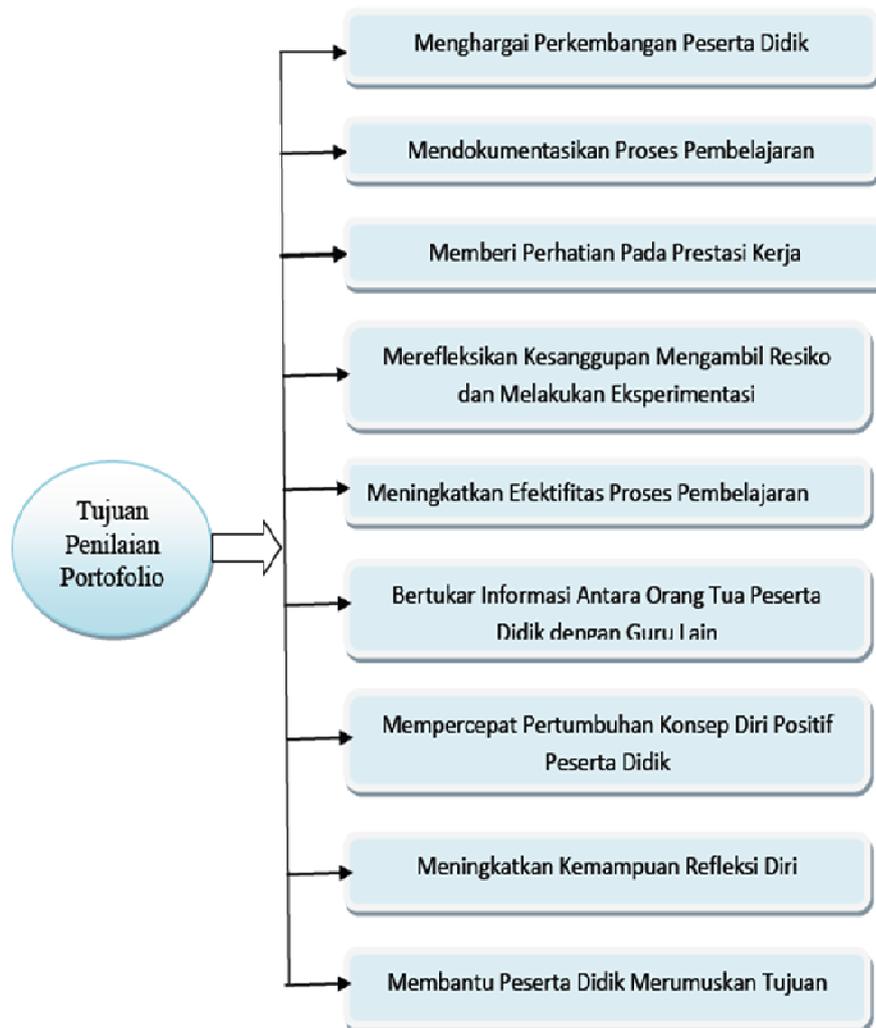
Memperhatikan pendapat di atas, nyata sekali bawa penilaian portofolio memfasilitas siswa untuk semakin mengembangkan kemampuannya serta mengembangkan kemampuan refleksinya. Lebih dari itu, siswa akan tumbuh menjadi manusia yang lebih dewasa dan bertanggung jawab. Hal itu dimungkinkan, karena siswa dituntut untuk menilai karya yang telah dibuat atau disusunnya. Dengan penilaian itu siswa akan mengetahui kelebihan/keunggulan atau kekurangan/kelemahannya. Pengetahuan itulah yang pada akhirnya diharapkan dapat dipakai oleh siswa untuk senantiasa meningkatkan kompetensinya.

B. TUJUAN DAN FUNGSI PENILAIAN PORTOFOLIO

Penilaian portofolio dapat digunakan sebagai alat formatif maupun sumatif. Portofolio sebagai alat formatif digunakan untuk memantau kemajuan peserta didik dari hari ke hari dan mendorong peserta didik dalam merefeksi pembelajaran mereka sendiri. Portofolio seperti ini difokuskan pada proses perkembangan peserta didik dan digunakan untuk tujuan formatif dan diagnostik. Penilaian portofolio ditujukan juga untuk penilaian sumatif pada akhir semester atau akhir tahun pelajaran. Hasil penilaian portofolio sebagai alat sumatif ini dapat digunakan untuk mengisi angka rapor peserta didik, yang menunjukkan prestasi peserta didik dalam mata pelajaran tertentu.

1. Tujuan Penilaian Portofolio

Pada hakikatnya tujuan penilaian portofolio adalah untuk memberikan informasi kepada orang tua tentang perkembangan peserta didik secara lengkap dengan dukungan data dan dokumen yang akurat. Rapor merupakan bentuk laporan prestasi peserta didik dalam belajar dalam kurun waktu tertentu. Portofolio merupakan lampiran dari rapor, dengan demikian rapor tetap harus dibuat.



Gambar 6.1. Tujuan Penilaian Portopolio

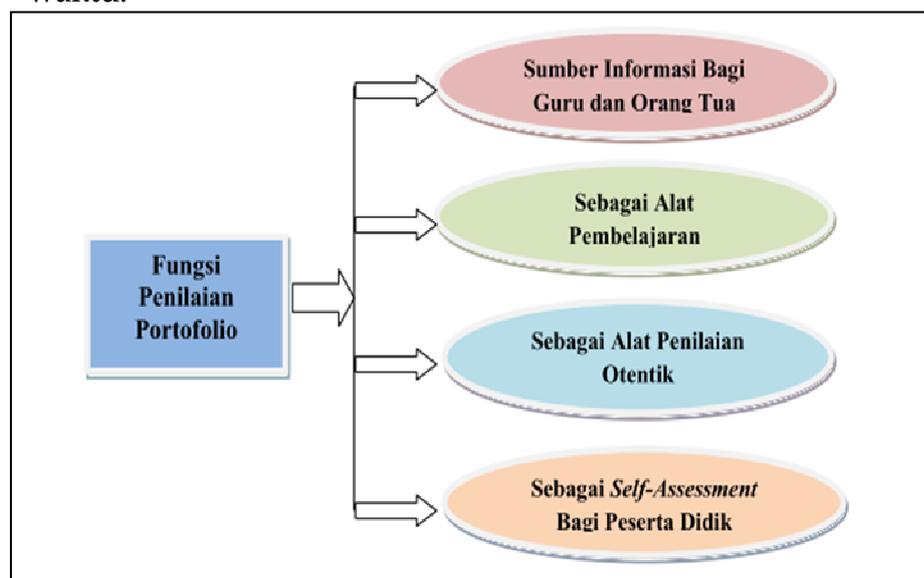
Tujuan portofolio ditetapkan oleh apa yang harus dikerjakan dan siapa yang akan menggunakan penilaian portofolio tersebut. Fakta yang paling penting dalam portofolio adalah digunakannya tes tertulis (*paper and pencil test*), project, product, dan catatan kemampuan (*records of performance*). S. Surapranata dan M. Hatta (2004: 76) mengemukakan penilaian portofolio dapat digunakan untuk mencapai beberapa tujuan, yaitu “menghargai perkembangan yang dialami peserta didik, mendokumentasikan proses pembelajaran yang berlangsung, memberi perhatian pada prestasi kerja peserta didik yang terbaik, merefleksikan kesanggupan mengambil resiko dan melakukan eksperimentasi, meningkatkan efektifitas proses pengajaran, bertukar informasi dengan orang tua/wali peserta didik dan guru lain, membina dan mempercepat pertumbuhan konsep diri positif pada peserta didik, meningkatkan

kemampuan melakukan refleksi diri, membantu peserta didik dalam merumuskan tujuan”.

2. Fungsi Penilaian Portofolio

Sebagaimana telah disinggung sebelumnya bahwa portofolio merupakan kumpulan karya peserta didik yang disimpan dalam sebuah *file*. Namun demikian, bukan berarti portofolio hanya merupakan tempat penyimpanan hasil pekerjaan peserta didik melainkan juga sebagai sumber informasi bagi guru, orang tua dan peserta didik itu sendiri. Portofolio dapat dijadikan sebagai bahan tindak lanjut dari suatu pekerjaan yang telah dilakukan peserta didik, sehingga guru dan orang tua mempunyai kesempatan untuk mengembangkan kemampuan peserta didik. Fungsi penilaian portofolio dapat kita lihat dari berbagai segi, yaitu:

- a) Portofolio sebagai sumber informasi bagi guru dan orang tua untuk mengetahui pertumbuhan dan perkembangan kemampuan peserta didik, tanggung jawab dalam belajar, perluasan dimensi belajar, dan inovasi pembelajaran.
- b) Portofolio sebagai alat pembelajaran merupakan komponen kurikulum, karena portofolio mengharuskan peserta didik untuk mengoleksi dan menunjukkan hasil kerja mereka.
- c) Portofolio sebagai alat penilaian otentik (*authentic assessment*).
- d) Portofolio sebagai sumber informasi bagi peserta didik untuk melakukan *self-assessment*. Maksudnya, peserta didik mempunyai kesempatan yang banyak untuk menilai diri sendiri dari waktu ke waktu.



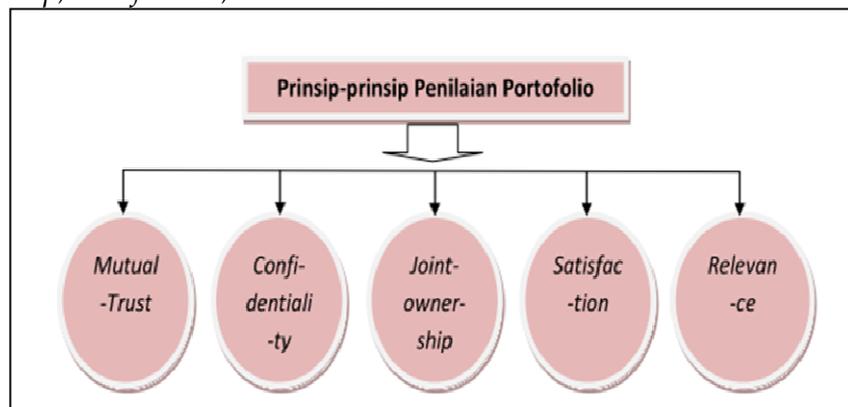
Gambar 6.2 .Fungsi Penilaian Portofolio

Selanjutnya, Direktorat PLP-Ditjen Dikdasmen-Depdiknas (2003: 123) mengemukakan bahwa penilaian portofolio dapat digunakan untuk:

- a) Memperlihatkan perkembangan pemikiran atau pemahaman siswa pada periode waktu tertentu.
- b) Menunjukkan suatu pemahaman dari beberapa konsep, topik, dan isu yang diberikan.
- c) Mendemonstrasikan perbedaan bakat.
- d) Mendemonstrasikan kemampuan untuk memproduksi atau mengkreasi suatu pekerjaan baru secara orisinal.
- e) Mendokumentasikan kegiatan selama periode waktu tertentu.
- f) Mendemonstrasikan kemampuan menampilkan suatu karya seni.
- g) Mendemonstrasikan kemampuan mengintegrasikan teori dan praktik.
- h) Merefleksikan nilai-nilai individual atau pandangan dunia secara lebih luas.

C. PRINSIP-PRINSIP PENILAIAN PORTOFOLIO

Proses penilaian portofolio menuntut terjadi interaksi multi arah, yaitu dari guru ke peserta didik, dari peserta didik ke guru, dan antar peserta didik. Direktorat PLP Ditjen Dikdasmen Depdiknas (2003: 124) mengemukakan pelaksanaan penilaian portofolio hendaknya memperhatikan prinsip-prinsip "*mutual trust, confidentiality, joint ownership, satisfaction, and relevance*".



Gambar 6.3. Prinsip Penilaian Portofolio

1. *Mutual trust* (saling mempercayai), artinya jangan ada saling mencurigai antara kita dengan peserta didik maupun antar peserta didik. Mereka harus sama-sama saling percaya, saling membutuhkan, saling membantu, terbuka, jujur, dan adil, sehingga dapat membangun suasana penilaian yang lebih kondusif. Juga hendaknya dapat

menciptakan suasana penilaian yang kondusif, wajar dan alami, sehingga hasil penilaian yang diperoleh betul-betul menggambarkan kemampuan peserta didik yang sesungguhnya.

2. *Confidentiality* (kerahasiaan bersama), artinya harus menjaga kerahasiaan semua hasil pekerjaan peserta didik dan dokumen yang ada, baik perorangan maupun kelompok, tidak boleh diberikan atau diperlihatkan kepada siapapun sebelum diadakan pameran. Hal ini dimaksudkan agar peserta didik yang mempunyai kelemahan tidak merasa dipermalukan. Menjaga kerahasiaan bersama ini juga mempunyai arti lain, yaitu memotivasi peserta didik untuk memperbaiki hasil pekerjaannya dan meningkatkan kepercayaan peserta didik kepada guru.
3. *Joint Ownership* (milik bersama), artinya semua hasil pekerjaan peserta didik dan dokumen yang ada harus menjadi milik bersama antara guru dan peserta didik, karena itu harus jaga bersama, baik penyimpanannya maupun penempatannya. Berikan kemudahan kepada peserta didik untuk melihat, menyimpan dan mengambil kembali portofolio mereka. Hal ini dimaksudkan juga untuk menumbuhkan rasa tanggung jawab peserta didik.
4. *Satisfaction* (kepuasan), artinya semua dokumen dalam rangka pencapaian standar kompetensi, kompetensi dasar dan indikator harus dapat memuaskan semua pihak, baik guru, orang tua maupun peserta didik, karena dokumen tersebut merupakan bukti karya terbaik peserta didik sebagai hasil pembinaan guru.
5. *Relevance* (kesesuaian), artinya dokumen yang ada harus sesuai dengan standar kompetensi, kompetensi dasar, indikator yang diharapkan. Kesesuaian ini pada gilirannya berkaitan dengan prinsip kepuasan.

Di samping prinsip-prinsip tersebut di atas, S. Surapranata dan M. Hatta (2004: 79) menambahkan tiga prinsip, yaitu “penciptaan budaya mengajar, refleksi bersama, serta proses dan hasil”. Penilaian portofolio hanya dapat dilakukan jika pembelajarannya pun menggunakan pendekatan portofolio. Artinya, jika guru dalam pembelajaran hanya menuntut peserta didik untuk menghafal pengetahuan atau fakta pada tingkat rendah, maka penilaian portofolio tidak akan bermakna. Penilaian portofolio akan efektif jika pembelajarannya menuntut peserta didik untuk menunjukkan kemampuan yang nyata dan menggambarkan pengembangan aspek pengetahuan, keterampilan, sikap dan nilai pada taraf yang lebih tinggi.

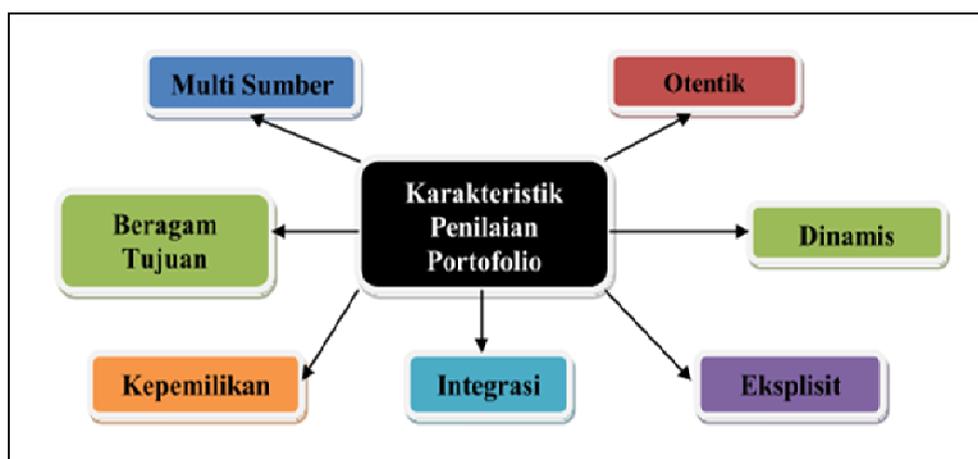
Prinsip penilaian portofolio yang lain adalah memberikan kesempatan kepada peserta didik untuk melakukan refleksi bersama-sama. Peserta didik dapat merefleksikan tentang proses berpikir mereka

sendiri, kemampuan pemahaman mereka sendiri, pemecahan masalah, dan pengambilan keputusan. Tidak hanya itu, penilaian portofolio juga harus diarahkan untuk menilai proses belajar peserta didik, seperti: catatan perilaku harian, sikap dan motivasi belajar, antusias tidaknya dalam mengikuti pelajaran, baik dalam kegiatan belajar kelompok maupun individual. Bukan hanya proses belajar, tetapi juga harus menilai hasil akhir suatu tugas yang diberikan oleh guru.

D. KARAKTERISTIK PENILAIAN PORTOFOLIO

Sebagaimana telah penulis kemukakan bahwa penilaian portofolio dilakukan sesuai dengan kegiatan pembelajaran yang berbasis portofolio (*portfolio-based instruction*). Kalau guru menggunakan model pembelajaran tradisional, tentu guru akan kesulitan melakukan penilaian portofolio, terutama dalam mengembangkan instrumen penilaiannya. Dengan demikian, kita harus melaksanakan kegiatan pembelajaran portofolio yang tidak hanya terjadi di dalam kelas tetapi juga di luar kelas. Implikasinya adalah bahwa hasil pekerjaan peserta didik yang dinilai melalui penilaian portofolio adalah hasil pekerjaan peserta didik yang dilakukan baik di kelas maupun di luar kelas sesuai dengan tuntutan kompetensi dasarnya, tidak hanya dalam dimensi proses tapi juga dimensi produk.

Di samping itu, melalui penilaian portofolio, peserta didik dapat memantau perkembangan kemampuannya secara mandiri, menunjukkan cara belajar yang berbeda antara seorang peserta didik dengan peserta didik lainnya, menunjukkan kualitas hasil pekerjaannya, menunjukkan kelebihan yang mereka miliki, mengembangkan kemampuan bersosialisasi, dan memotivasi dirinya untuk lebih giat melakukan kegiatan belajar, memberikan peluang yang besar bagi peserta didik untuk melakukan dialog dengan guru dan orang tuanya secara intensif tentang kelebihan dan kekurangannya. Menurut Barton dan Collins dalam S. Surapranata dan M. Hatta (2004: 82) terdapat beberapa karakteristik esensial penilaian portofolio, yaitu 'multi sumber, otentik, dinamis, eksplisit, integrasi, kepemilikan, dan beragam tujuan'.



Gambar 6.4. Karakteristik Penilaian Portofolio

Multi sumber dimaksudkan bahwa pelaksanaan penilaian portofolio harus dilakukan dari berbagai sumber, seperti peserta didik, guru, orang tua, masyarakat, dan *evidence* lainnya, seperti gambar, lukisan, jurnal, audio, video tape dan sebagainya, baik secara tertulis maupun tindakan. *Evidence* yang dimaksud haruslah otentik dan berhubungan dengan program pembelajaran, kegiatan, standar kompetensi, kompetensi dasar dan indikator yang hendak dicapai. Penilaian portofolio menuntut adanya pertumbuhan dan perkembangan dari setiap peserta didik. Oleh sebab itu, sebaiknya setiap *evidence* dari waktu ke waktu harus dikumpulkan dan didokumentasikan. Seandainya *evidence* tersebut akan dipilih, maka pilihlah secara selektif. Penilaian portofolio juga harus jelas, baik jenis, teknik, prosedur maupun kompetensi yang akan diukur. Kejelasan yang dimaksud bukan hanya untuk guru tapi juga peserta didik. Dalam pelaksanaannya, antara kegiatan peserta didik di kelas dengan kehidupan nyata haruslah terintegrasi.

Penilaian portofolio tidak terlepas dari kehidupan sehari-hari, sehingga peserta didik tidak jauh dari apa yang mereka alami. Peserta didik juga dapat dengan mudah mengaitkan antara kemampuan yang diperolehnya dengan kenyataan sehari-hari. Hal yang sangat penting dalam penilaian portofolio adalah adanya rasa memiliki bagi setiap peserta didik terhadap semua *evidence* yang dikumpulkan guru, sehingga peserta didik dapat menjaga dengan baik semua *evidence*. Pelaksanaan penilaian portofolio bukan hanya mengacu kepada kompetensi yang harus dikuasai oleh peserta didik tetapi juga tujuan-tujuan lain yang bermanfaat bagi program pembelajaran, seperti keefektifan program, perkembangan peserta didik, dan dapat dijadikan alat komunikasi peserta didik ke berbagai pihak yang berkepentingan.

E. KELEBIHAN DAN KEKURANGAN PENILAIAN PORTOFOLIO

Setiap konsep atau model penilaian tentu ada kelebihan dan kekurangannya. Begitu juga dengan model penilaian portofolio. Kelebihan model penilaian portofolio antara lain:

- a) Dapat melihat pertumbuhan dan perkembangan kemampuan peserta didik dari waktu ke waktu berdasarkan feed-back dan refleksi diri.
- b) Membantu guru melakukan penilaian secara adil, objektif, transparan dan dapat dipertanggungjawabkan tanpa mengurangi kreatifitas peserta didik di kelas.
- c) Mengajak peserta didik untuk belajar bertanggung jawab terhadap apa yang telah mereka kerjakan, baik di kelas maupun di luar kelas dalam rangka implementasi program pembelajaran.
- d) Meningkatkan peranserta peserta didik secara aktif dalam kegiatan pembelajaran dan penilaian.
- e) Memberi kesempatan kepada peserta didik untuk meningkatkan kemampuan mereka.
- f) Membantu guru mengklarifikasi dan mengidentifikasi program pembelajaran.
- g) Terlibatnya berbagai pihak, seperti orang tua, guru, komite sekolah dan masyarakat lainnya dalam melihat pencapaian kemampuan peserta didik.
- h) Memungkinkan peserta didik melakukan penilaian diri (*self-assessment*), refleksi dan mengembangkan kemampuan berpikir kritis (*critical thinking*).
- i) Memungkinkan guru melakukan penilaian secara fleksibel tetapi tetap mengacu kepada kompetensi dasar dan indikator hasil belajar yang ditentukan.
- j) Guru dan peserta didik sama-sama bertanggung jawab untuk merancang dan menilai kemajuan belajar.
- k) Dapat digunakan untuk menilai kelas yang heterogen antara peserta didik yang pandai dan kurang pandai.
- l) Memungkinkan guru memberikan hadiah terhadap setiap usaha belajar peserta didik.

Adapun kekurangan penilaian portofolio antara lain:

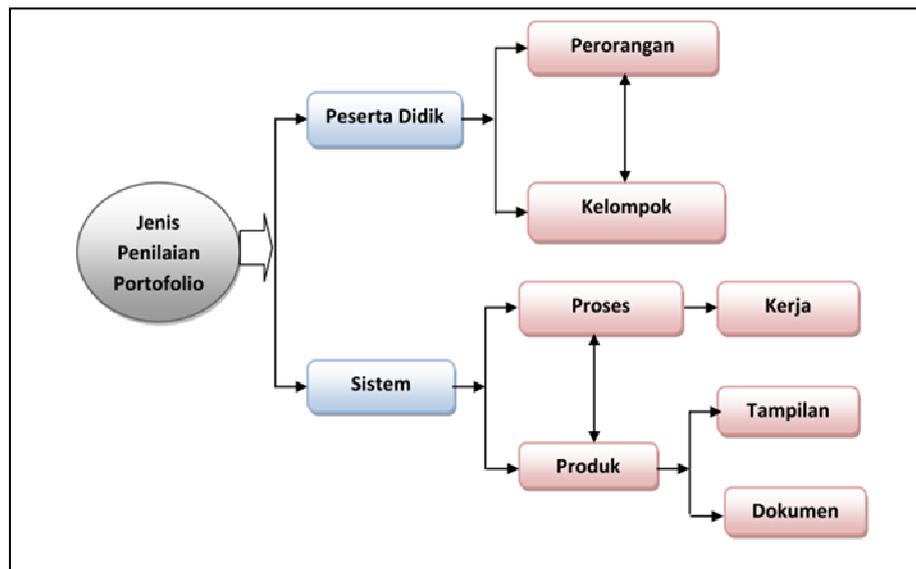
- a) Membutuhkan waktu dan kerja ekstra.
- b) Penilaian portofolio dianggap kurang reliabel dibandingkan dengan bentuk penilaian yang lain.
- c) Ada kecenderungan guru hanya memperhatikan pencapaian akhir, sehingga proses penilaian kurang mendapat perhatian.
- d) Jika guru melaksanakan proses pembelajaran yang bersifat *teacher-oriented*, kemungkinan besar inisiatif dan kreatifitas peserta didik akan

terbelenggu, sehingga penilaian portofolio tidak dapat dilaksanakan dengan baik.

- e) Orang tua peserta didik sering berpikir skeptis karena laporan hasil belajar anaknya tidak berbentuk angka.
- f) Penilaian portofolio masih relatif baru, sehingga banyak guru, orang tua dan peserta didik yang belum mengetahui dan memahaminya.
- g) Tidak tersedianya kriteria penilaian yang jelas.
- h) Analisis terhadap penilaian portofolio agak sulit dilakukan sebagai akibat dikurangnya penggunaan angka.
- i) Sulit dilakukan terutama menghadapi ujian dalam skala nasional.
- j) Dapat menjebak peserta didik jika terlalu sering menggunakan format yang lengkap dan detail.

F. JENIS PENILAIAN PORTOFOLIO

Apabila dilihat dari jumlah peserta didik, maka penilaian portofolio dapat dibagi menjadi dua jenis, yaitu portofolio perorangan dan portofolio kelompok. Jika dilihat dari sistem, portofolio dapat dibagi dua jenis, yaitu portofolio proses dan portofolio produk.



Gambar 6.5. Jenis Portofolio

Portofolio perorangan merupakan kumpulan hasil kerja peserta didik secara perorangan, dan portofolio kelompok merupakan kumpulan hasil karya sekelompok peserta didik atau kelas tertentu.

3. Portofolio Proses

Jenis portofolio proses menunjukkan tahapan belajar dan menyajikan catatan perkembangan peserta didik dari waktu ke waktu.

Portofolio proses menunjukkan kegiatan pembelajaran untuk mencapai standar kompetensi, kompetensi dasar, dan sekumpulan indikator yang telah ditetapkan dalam kurikulum, serta menunjukkan semua hasil dari awal sampai dengan akhir selama kurun waktu tertentu. Tujuan menggunakan portofolio proses adalah untuk membantu peserta didik mengidentifikasi tujuan pembelajaran, perkembangan hasil belajar dari waktu ke waktu, dan menunjukkan pencapaian hasil belajar. Pendekatan ini lebih menekankan pada bagaimana peserta didik belajar, berkreasi, termasuk mulai dari draft awal, bagaimana proses awal itu terjadi, dan tentunya sepanjang peserta didik dinilai.

Dalam portofolio proses, dapat menyajikan berbagai macam tugas yang setara atau yang berbeda kepada peserta didik. Dengan kata lain, peserta didik boleh memilih tugas-tugas yang dianggapnya cocok untuk mereka. Untuk dapat memutuskan apa yang harus dikerjakan peserta didik atau peserta didik diajak bekerjasama dengan peserta didik lain dalam mengerjakan tugas tertentu. Biasanya, portofolio proses digunakan untuk melihat proses pembuatan suatu karya atau suatu pekerjaan yang menuntut adanya proses diskusi antara peserta didik dengan guru atau sesama peserta didik. Berdasarkan proses kegiatan tersebut, guru dapat membantu peserta didik untuk mengidentifikasi kelebihan dan kekurangannya.

Salah satu bentuk portofolio proses adalah portofolio kerja (*working portfolio*) yaitu bentuk yang digunakan untuk memilih koleksi evidence peserta didik, memantau kemajuan atau perkembangan, dan menilai peserta didik dalam mengelola kegiatan belajar mereka sendiri. Peserta didik mengumpulkan semua hasil kerja termasuk coretan-coretan (sketsa), buram, catatan, kumpulan untuk rangsangan, buram setengah jadi, dan pekerjaan yang sudah selesai. Portofolio kerja bermanfaat bagi peserta didik terutama untuk memberikan informasi tentang bagaimana mengorganisasikan dan mengelola kerja, merefleksi dari pencapaiannya, memantau perkembangan, dan menetapkan tujuan dan arahan.

Informasi ini dapat digunakan untuk diskusi antara peserta didik dengan guru. Melalui portofolio kerja ini, guru dapat membantu peserta didik mengidentifikasi kekuatan dan kelemahan masing-masing. Untuk itu, kerjasama yang efektif antara guru dengan peserta didik sangat diperlukan. Di samping itu, informasi ini dapat digunakan juga oleh guru untuk memperbaiki cara belajar peserta didik. Namun demikian, keberhasilan portofolio kerja sangat bergantung kepada kemampuan peserta didik untuk merefleksikan dan mendokumentasikan kemajuan proses pembelajaran.

Dalam portofolio kerja ini yang dinilai adalah cara kerja (pengorganisasian) dan hasil kerja. Adapun kriterianya antara lain:

- a) Adakah pembagian kerja diantara anggota kelompok ?
 - b) Apakah masing-masing anggota telah bekerja sesuai dengan tugasnya ?
 - c) Berapa besar kontribusi kerja kelompok terhadap hasil yang dicapai kelompok ?
 - d) Adakah bukti tanggung jawab bersama ?
 - e) Apakah kelengkapan data yang diperoleh telah sesuai dengan tugas anggota kelompok masing-masing ?
 - f) Apakah informasi yang diperoleh akurat?
 - g) Apakah portofolio telah disusun dengan baik?
4. Portofolio Produk

Jenis penilaian portofolio ini hanya menekankan pada penguasaan (*masteri*) dari tugas yang dituntut dalam standar kompetensi, kompetensi dasar, dan sekumpulan indikator pencapaian hasil belajar, serta hanya menunjukkan *evidence* yang paling baik, tanpa memperhatikan bagaimana dan kapan *evidence* tersebut diperoleh. Tujuan portofolio produk adalah untuk mendokumentasikan dan merefleksikan kualitas prestasi yang telah dicapai. Contoh portofolio produk adalah portofolio tampilan (*showportfolio*) dan portofolio dokumentasi (*documentaryportfolio*).

a. Portofolio Tampilan

Portofolio bentuk ini merupakan sekumpulan hasil karya peserta didik atau dokumen terseleksi yang dipersiapkan untuk ditampilkan kepada umum. Misalnya, mempertanggungjawabkan suatu proyek, menyelenggarakan pameran, atau mempertahankan suatu konsep. Portofolio ini sangat bermanfaat jika guru ingin mengetahui kemampuan peserta didik yang sesungguhnya dan bagaimana ketepatan isi portofolio mengacu kepada kompetensi yang telah ditetapkan. Bentuk ini biasanya digunakan untuk tujuan pertanggungjawaban (*accountability*). Syarat pokok yang harus dipenuhi oleh peserta didik dalam portofolio tampilan adalah keaslian *evidence*. Untuk itu, ada beberapa hal yang harus diperhatikan oleh peserta didik dan guru. Pertama, peserta didik harus menandatangani lembar pernyataan keaslian. Kedua, pesertadidik memberikan penghargaan kepada semua sumber yang telah membantu, termasuk identitasnya serta bentuk bantuan yang diberikan. Ketiga, guru harus melihat perencanaan, draft pekerjaan peserta didik dan catatan selama proses berlangsung. Keempat, guru harus betul-betul mengamati bagaimana peserta didik menampilkan

hasil pekerjaan mereka. Aspek yang dinilai dalam bentuk portofolio tampilan adalah :

- 1) Signifikansi materi, yaitu apakah materi yang dipilih benar-benar merupakan materi yang penting bermakna untuk diketahui dan dipecahkan ? atau seberapa besar tingkat kebermaknaan informasi yang dipilih berkaitan dengan topik yang dibahasnya ? Apakah materi yang dipilih sesuai dengan standar kompetensi, kompetensi dasar, dan indikator pencapaian hasil belajar ?
- 2) Pemahaman, yaitu seberapa baik tingkat pemahaman peserta didik terhadap hakikat dan lingkup masalah, kebijakan, atau langkah-langkah yang dirumuskan.
- 3) Argumentasi, yaitu apakah peserta didik dalam mempertahankan argumentasinya sudah cukup memadai, sistematis, dan relevan?
- 4) Responsifness (kemampuan memberikan respon), yaitu seberapa besar tingkat kesesuaian antara respon yang diberikan dengan pertanyaan? Dalam memberikan respon, adakah bukti-bukti fisik yang ditunjukkan?
- 5) Kerjasama kelompok, yaitu apakah anggota kelompok turut berpartisipasi secara aktif dalam penyajian? Adakah bukti yang menunjukkan tanggung jawab anggota dalam kelompok? Apakah para penyaji menghargai pendapat orang lain? Adakah kekompakan kerja diantara para anggota kelompok?

Lembar Penilaian Penampilan

Judul Penampilan :

Kelas/Kelompok :

Petunjuk Penilaian :

1. Setiap kriteria diberi skor dalam skala 5 (1 – 5)
2. Skor 1 = rendah; 2 = cukup; 3 = rata-rata; 4 = baik; 5 = istimewa

No	Kriteria Penilaian	Nilai	Catatan
01	Signifikansi: 2. Seberapa besar tingkat kesesuaian atau kebermaknaan informasi yang diberikan dengan topik yang dibahas.		
02	Pemahaman: 3. Seberapa baik tingkat pemahaman peserta Didik terhadap hakikat dan ruang lingkup masalah yang disajikan.		
03	Argumentasi: 4. Seberapa baik alasan yang diberikan peserta didik terkait dengan permasalahan yang dibicarakan.		
04	Responsifness : 5. Seberapa besar kesesuaian jawaban		

	yang diberikan peserta didik dengan pertanyaan yang muncul.		
05	Kerjasama kelompok : 6. Seberapa besar anggota kelompok berpartisipasi dalam penyajian. 7. Bagaimana setiap anggota merasa bertanggung jawab atas permasalahan kelompok. 8. Bagaimana para penyaji menghargai pendapat orang lain.		

Penilai,

.....

b. Portofolio Dokumen

Portofolio dokumen menyediakan informasi baik proses maupun produk yang dihasilkan oleh peserta didik. Portofolio ini digunakan untuk memilih koleksi evidence peserta didik yang sesuai dengan kompetensi dan akan dijadikan dasar penilaian. Evidence peserta didik yang digunakan dalam portofolio dokumentasi dapat berasal dari catatan guru atau kombinasi antara catatan guru dengan kegiatan peserta didik. Model portofolio ini bermanfaat bagi peserta didik dan orang tua untuk mengetahui kemajuan hasil belajar, kelebihan dan kekurangan peserta didik dalam belajar secara perorangan. Berdasarkan dokumen ini, baik peserta didik, orang tua maupun guru dapat melihat:

- 1) Proses apa yang telah diikuti ?
- 2) Kerja apa yang telah dilakukan ?
- 3) Dokumen apa yang telah dihasilkan ?
- 4) Apakah hal-hal pokok telah terdokumentasikan ?
- 5) Apakah dokumen disusun berdasarkan sumber-sumber data masing-masing ?
- 6) Apakah dokumen berkaitan dengan yang akan disajikan ?
- 7) Standar kompetensi mana yang telah dikuasai sampai pada pekerjaan terakhir ?

Indikator untuk penilaian dokumen itu antara lain: kelengkapan, kejelasan, akurasi informasi yang didapat, dukungan data, kebermaknaan data grafis, dan kualifikasi dokumen. Untuk menilai suatu dokumen dapat dibuatkan model format penilaiannya.

Lembar Penilaian Dokumen

Judul Dokumen :

Kelas/ Kelompok :

Petunjuk Penilaian :

1. Setiap kriteria diberi skor dalam skala 5 (1 – 5)
2. Skor 1 = Rendah; 2 = Cukup; 3 = Rata-rata; 4 = Baik; 5 = Istimewa

No	Kriteria Penilaian	Nilai	Catatan
1	Kelengkapan: 1. Apakah dokumen lengkap untuk menjawab suatu permasalahan		
2	Kejelasan: 2. Tersusun dengan baik 3. Tertulis dengan baik 4. Mudah dipahami		
3	Informasi: 5. Akurat 6. Memadai 7. Penting		
4	Dukungan: 8. Memuat contoh untuk hal-hal yang utama 9. Memuat alasan yang baik		
5	Data Grafis: 10. Berkaitan dengan isi setiap bagian 11. Diberi judul yang tepat 12. Memberikan informasi 13. Meningkatkan pemahaman		
6	Bagian Dokumentasi: 14. Cukup memadai 15. Dapat dipercaya 16. Berkaitan dengan hal yang dijelaskan 17. Terpilih (terseleksi)		
Jumlah Skor Kualifikasi Penilaian			

Penilai:

.....

G. TAHAP-TAHAP PENILAIAN PORTOFOLIO

Menurut Anthoni J. Nitko (1996: 281), ada enam tahap untuk menggunakan sebuah sistem portofolio (*six steps for crafting a portfolio system*), yaitu “mengidentifikasi tujuan dan fokus portofolio, mengidentifikasi isi materi umum yang akan dinilai, mengidentifikasi pengorganisasian portofolio, menggunakan portofolio dalam praktik, evaluasi pelaksanaan portofolio, dan evaluasi portofolio secara umum”. Tahap pertama akan merupakan dasar bagi penentuan tahap selanjutnya. Oleh sebab itu, jawablah semua pertanyaan pada tahap pertama tersebut sebelum lanjut pada tahap berikutnya. Dalam tulisan ini, tahap-tahap penilaian portofolio yang disarankan adalah:

3. Menentukan tujuan dan fokus portofolio.

Hal ini dapat dilakukan dengan menjawab pertanyaan-pertanyaan sebagai berikut:

- a) Mengapa portofolio itu akan dilakukan ?
- b) Tujuan pembelajaran dan tujuan kurikulum (dalam hal ini kompetensi dasar) apa yang akan dicapai ?
- c) Alat penilaian yang bagaimana yang tepat untuk menilai tujuan tersebut ?
- d) Apakah portofolio akan difokuskan pada hasil pekerjaan yang baik, pertumbuhan dan kemajuan belajar, atau keduanya ?
- e) Apakah portofolio itu akan digunakan untuk formatif, sumatif, diagnostik atau semuanya ?
- f) Siapa yang akan dilibatkan dalam menentukan tujuan, fokus, dan pengaturan (*organization*) portofolio ?

4. Menentukan isi portofolio

Setelah menentukan tujuan, langkah selanjutnya adalah menentukan isi portofolio. Dengan demikian, isi portofolio tentunya harus sesuai dengan tujuan portofolio. Isi portofolio harus menunjukkan kemampuan peserta didik sesuai dengan kompetensi yang diharapkan. Untuk itu, semua kegiatan pembelajaran, baik di kelas maupun di luar kelas harus selalu diamati dan dinilai.

5. Mengembangkan kriteria penilaian.

Kriteria penilaian harus dirumuskan dengan jelas, baik yang berhubungan dengan proses pembelajaran maupun hasil belajar yang diharapkan. Kriteria penilaian sangat bergantung kepada kompetensi, cara menilai dan *evidence* yang dinilai.

6. Menyusun format penilaian

Sebagaimana isi dan kriteria penilaian, maka format penilaianpun harus mengacu kepada tujuan. Format penilaian banyak modelnya. Salah satunya bisa menggunakan model skala dengan lima kriteria, seperti: sangat baik, baik, cukup, kurang, dan sangat kurang, Contoh.

Penilaian Portofolio

Kompetensi Dasar:	Nama :		
	Tanggal :		
Indikator	Penilaian		
	Baik	Cukup	Kurang
1.			
2.			
3.			
Dicapai melalui:	Komentar guru:		
- Bantuan guru			
- Seluruh kelas			
- Perorangan			
Komentar orang tua:			

7. Mengidentifikasi pengorganisasian portofolio. Siapa yang akan terlibat dalam portofolio tersebut?
8. Menggunakan portofolio dalam praktik.
9. Menilai pelaksanaan portofolio.
10. Menilai portofolio secara umum.

H. BAHAN-BAHAN PENILAIAN PORTOFOLIO

Pada prinsipnya, setiap tindakan belajar peserta didik harus diberikan penghargaan. Tujuannya adalah untuk memberikan penguatan dan semangat belajar. Penghargaan tersebut dapat berbentuk tulisan atau lisan. Semua penghargaan tersebut dapat dijadikan bahan penilaian portofolio. Bahan penilaian portofolio bisa juga diambil dari hasil pekerjaan peserta didik, seperti Lembar Kerja Siswa, hasil rangkuman, gambar, klipping, hasil kerja kelompok, hasil tes, buku catatan dan hal-hal yang menyangkut pribadi peserta didik. Di samping itu, bahan penilaian portofolio dapat diperoleh dari alat-alat audio-visual, video atau disket. Dengan demikian, bahan-bahan penilaian portofolio dapat dikelompokkan menjadi beberapa bagian, yaitu:

1. Penghargaan yang diperoleh peserta didik, baik tertulis maupun lisan, seperti sertifikat hasil lomba atau catatan guru tentang penghargaan lisan yang pernah diberikan kepada peserta didik dalam kurun waktu tertentu.
2. Hasil pekerjaan peserta didik, seperti Lembar Kerja Siswa (LKS), klipping, gambar, hasil ulangan, hasil kerja kelompok, hasil rangkuman, dan lain-lain.
3. Catatan/laporan dari orang tua peserta didik atau teman sekelas.

4. Catatan pribadi peserta didik, seperti: bukti kehadiran, hasil presentasi dari tugas-tugas yang selesai dikerjakan, catatan-catatan kejadian khusus (*anecdotal records*), daftar kehadiran, dan lain-lain.
5. Bahan-bahan lain yang relevan, yaitu (a) bahan yang dapat memberikan informasi tentang perkembangan yang dialami peserta didik, dan (b) bahan yang dapat dijadikan sebagai bahan pertimbangan dalam pengambilan keputusan tentang kurikulum dan pembelajaran.
6. Alat-alat audio-visual, video atau disket.

Setelah semua bahan penilaian portofolio dikumpulkan, kemudian disusun dan disimpan dalam sebuah dokumen. Dalam rangka penataan sebuah dokumen, guru hendaknya memperhatikan hal-hal sebagai berikut: Pertama, setiap dokumen harus dibuat identitas peserta didik, seperti nama, nomor induk, kelas, dan nama sekolah. Kedua, untuk mempermudah pengecekan isi dokumen, maka setiap dokumen harus dibuat daftar isi dokumen. Ketiga, isi dokumen harus dimasukkan ke dalam satu map atau folder dan disusun secara sistematis sesuai dengan kompetensi yang telah ditetapkan. Keempat, isi dokumen hendaknya dikelompokkan sesuai dengan mata pelajaran dan setiap mata pelajaran diberikan warna yang berbeda. Kelima, setiap isi dokumen harus ada catatan atau komentar dari guru dan orang tua. Keenam, isi dokumen hendaknya tidak ditentukan sepihak oleh guru, tetapi harus melibatkan peserta didik melalui proses diskusi. Hal ini dimaksudkan agar peserta didik tidak hanya dijadikan sebagai objek penilaian tetapi juga subjek penilaian.

Di samping itu, guru hendaknya memperhatikan prinsip-prinsip dokumentasi portofolio, antara lain kelengkapan dan ketepatan data, ketepatan waktu, tingkat keterbacaan, praktis, sistematis, dan relevan.

BAB VII

TEKNIK PENGOLAHAN HASIL EVALUASI

A. CARA MEMBERI SKOR MENTAH UNTUK TES URAIAN

Dalam bentuk uraian biasanya skor mentah dicari dengan menggunakan sistem bobot. Sistem bobot ada dua cara, yaitu: Pertama, bobot dinyatakan dalam skor maksimum sesuai dengan tingkat kesukarannya. Misalnya, untuk soal yang mudah skor maksimumnya adalah 6, untuk soal sedang skor maksimumnya adalah 7, dan untuk soal sukar skor maksimumnya adalah 10. Cara ini tidak memungkinkan peserta didik mendapat skor maksimum sepuluh. Kedua, bobot dinyatakan dalam bilangan-bilangan tertentu sesuai dengan tingkat kesukaran soal. Misalnya, soal yang mudah diberi bobot 3, soal sedang diberi bobot 4, dan soal sukar diberi bobot 5. Cara ini memungkinkan peserta didik mendapat skor sepuluh.

Contoh 1: seorang peserta didik diberi tiga soal dalam bentuk uraian. Setiap soal diberi skor (X) maksimum dalam rentang 1-10 sesuai dengan kualitas jawaban peserta didik.

Perhitungan Skor dengan Sistem Bobot Pertama

No Soal	Tingkat Kesukaran	Jawaban	Skor (X)
1	Mudah	Betul	6
2	Sedang	Betul	7
3	Sukar	Betul	10
	Jumlah		23

Rumus: $Skor = \frac{\sum X}{\sum s}$

Keterangan:

$\sum X$ = jumlah skor

$\sum s$ = jumlah soal

Jadi, skor peserta didik $A = \frac{23}{3} = 7,67$

Contoh 2: seorang peserta didik dites dengan tiga soal dalam bentuk uraian. Masing-masing soal diberi bobot sesuai dengan tingkat kesukarannya, yaitu bobot 5 untuk soal yang sukar, 4 untuk soal sedang, dan 3 untuk soal yang mudah. Tiap-tiap soal diberikan skor (X) dengan rentang 1-10 sesuai dengan kualitas jawaban yang betul. Kemudian skor (X) yang dicapai oleh setiap peserta didik dikalikan dengan bobot setiap soal. Hasil perhitungannya adalah sebagai berikut:

Perhitungan Skor dengan Sistem Bobot Kedua

No Soal	Tingkat Kesukaran	Jawaban	Skor (X)	Bobot (B)	XB
1	Mudah	Betul	10	3	30
2	Sedang	Betul	10	4	40
3	Sukar	Betul	10	5	50
Jumlah				12	120

Rumus: $Skor = \frac{\sum XB}{\sum B}$

Keterangan:

TK = tingkat kesukaran

X = skor setiap soal

B = bobot sesuai dengan tingkat kesukaran soal

$\sum XB$ = jumlah hasil perkalian X dengan B

Jadi, skor yang diperoleh peserta didik : $\frac{120}{12} = 10$

Untuk memudahkan pemberian skor, ada baiknya digunakan sistem yang kedua. Sistem bobot diberikan kepada soal bentuk uraian dengan maksud untuk memberikan skor secara adil kepada peserta didik berdasarkan kemampuannya masing-masing dalam menjawab soal-soal yang berbeda tingkat kesukarannya. Agaknya kurang adil apabila peserta didik yang sanggup menjawab soal yang sukar itu diberi skor yang sama dengan peserta didik yang hanya sanggup menjawab soal yang mudah saja.

Pedoman penyekoran di atas hanya dapat digunakan untuk bentuk uraian biasa, yaitu uraian bebas dan uraian terbatas. Untuk bentuk uraian objektif dan non-objektif harus menggunakan pedoman seperti berikut:

Contoh 3: Indikator: Dapat menentukan harga satuan dari campuran dua jenis bahan yang berbeda harga. Soal: Bu Adi, seorang pedagang beras eceran membeli 30 kg beras jenis X seharga Rp.180.000,00 kemudian dicampur dengan 10 kg beras jenis Y seharga Rp. 40.000,00. Pertanyaan:

1. Berapa kg berat seluruh beras yang dibeli bu Anggi ?
2. Berapa rupiah uang yang dibayarkan bu Anggi ?
3. Berapa harga beras campuran tiap kg ?

Pedoman Penyekoran Bentuk Uraian Objektif (BUO)

Kunci Jawaban	Skor
a) 30 kg + 10 kg = 40 kg.	1
b) Rp.180.000,00 + Rp.40.000,00. = Rp.220.000,00.	1
c) Harga beras campuran tiap kg adalah : <u>Rp. 220.000,00</u> 40 kg	1

= Rp 5.500,00	1
Skor Maksimum	4

B. CARA MEMBERI SKOR MENTAH UNTUK TES OBJEKTIF

Dalam memberikan skor pada soal tes bentuk objektif ini dapat menggunakan cara, yaitu:

1. Tanpa rumus tebakan (*non-guessing formula*)

Biasanya digunakan apabila soal belum diketahui tingkat kebaikannya. Caranya ialah menghitung jumlah jawaban yang betul saja. Setiap jawaban yang betul diberi skor 1, dan jawaban yang salah diberi skor 0. Jadi, skor = jumlah jawaban yang betul.

2. Menggunakan rumus tebakan (*guessing formula*)

Biasanya rumus ini digunakan apabila soal-soal tes itu sudah pernah diujicobakan dan dilaksanakan, sehingga dapat diketahui tingkat kebenarannya. Penggunaan rumus tebakan ini bukan karena guru sudah mengetahui bahwa peserta didik itu menebak, tetapi tes bentuk objektif ini memang sangat memungkinkan peserta didik untuk menebak. Adapun rumus-rumus tebakan tersebut adalah sebagai berikut:

a. Untuk item bentuk benar-salah (*true-false*)

$$\text{Rumus: } S = \sum B - \sum S$$

Keterangan:

S = skor yang dicari

$\sum B$ = jumlah jawaban yang benar

$\sum S$ = jumlah jawaban yang salah

Contoh: Seorang peserta didik dites dengan soal bentuk B – S sebanyak 30 soal. Ternyata, peserta didik tersebut dapat menjawab soal dengan betul 25 butir soal, berarti jumlah jawaban yang salah ada 5 soal. Dengan demikian, skor peserta didik yang bersangkutan adalah:

$$\text{Skor} = 25 - 5 = 20.$$

b. Untuk item bentuk pilihan-ganda (*multiple choice*)

$$\text{Rumus: } S = \sum B - \frac{\sum S}{n-1}$$

Keterangan:

S = skor yang dicari

$B \sum$ = jumlah jawaban yang benar

$S \sum$ = jumlah jawaban yang salah

n = jumlah alternatif jawaban yang disediakan

1 = bilangan tetap

Contoh: Seorang peserta didik A dites dengan soal bentuk pilihan-ganda sebanyak 10 soal. Ternyata, peserta didik A dapat menjawab

soal dengan betul sebanyak 7 butir soal, berarti jumlah jawaban yang salah adalah 3 soal. Jumlah alternatif jawaban (*option*) = 4. Dengan demikian, skor peserta didik A adalah: $Skor = 7 - \frac{3}{4-1} = 6$

Di samping rumus di atas, Anda juga bisa menggunakan rumus lain. Menurut Ainur Rofieq (2008) cara penskoran tes bentuk pilihan ganda ada tiga macam, yaitu “penskoran tanpa ada koreksi jawaban, penskoran ada koreksi jawaban, dan penskoran dengan butir beda bobot”.

1) Penskoran tanpa koreksi, yaitu penskoran dengan cara setiap butir soal yang dijawab benar mendapat nilai satu (tergantung dari bobot butir soal). Skor peserta didik diperoleh dengan cara menghitung banyaknya butir soal yang dijawab benar.

$$Rumus: s = \frac{B}{N} \times 100 \quad (\text{skala } 0 - 100)$$

Keterangan:

B = Jumlah jawaban benar

N = Jumlah soal

Contoh: Berdasarkan contoh soal di atas, dimana jumlah soal ada 10, jumlah jawaban benar ada 7, maka skor yang diperoleh peserta didik A adalah: $Skor = \frac{7}{10} \times 100 = 70$

2) Penskoran ada koreksi jawaban, yaitu pemberian skor dengan memberikan pertimbangan pada butir soal yang dijawab salah dan tidak dijawab. Adapun rumusnya adalah:

$$Skor = \left[\left(B - \frac{S}{P-1} \right) / N \right] \times 100$$

Keterangan:

B = jumlah soal yang dijawab benar

S = jumlah soal yang dijawab salah

P = jumlah pilihan jawaban tiap soal

1 = bilangan tetap

N = jumlah soal

Catatan: Soal yang tidak dijawab diberi skor 0

Contoh: Berdasarkan contoh soal di atas, dimana jumlah soal ada 10, jumlah jawaban benar ada 7, jumlah jawaban salah = 3, dan jumlah pilihan jawaban = 4, maka skor yang diperoleh peserta didik adalah:

$$Skor = \left[\left(7 - \frac{3}{4-1} \right) / 10 \right] \times 100 = 60$$

3) Penskoran dengan butir beda bobot, yaitu pemberian skor dengan memberikan bobot berbeda untuk sejumlah soal. Biasanya bobot butir soal menyesuaikan dengan tingkatan kognitif (pengetahuan, pemahaman, aplikasi, analisis, sintesis dan evaluasi) yang telah ditetapkan guru.

$$Skor = \sum \frac{(B \times b)}{Si} \times 100\%$$

Keterangan:

B = jumlah soal yang dijawab benar

b = bobot setiap soal

Si = skor ideal

Contoh: Siswa mengikuti ujian akhir semester mata pelajaran IPS. Jumlah soal 50, terdiri atas enam jenjang domain kognitif yang diberi bobot sebagai berikut: pengetahuan dengan bobot 1, pemahaman dengan bobot 2, aplikasi dengan bobot 3, analisis dengan bobot 4, sintesis dengan bobot 5, dan evaluasi dengan bobot 6. Ardi dapat menjawab benar 10 soal dari 10 soal pada jenjang pengetahuan, 8 soal dari 10 soal pada jenjang pemahaman, 10 soal dari 15 soal pada jenjang aplikasi, 4 soal dari 6 soal pada jenjang analisis, 5 soal dari 7 soal pada jenjang sintesis, dan 1 soal dari 2 soal pada jenjang evaluasi. Pertanyaan: Berapa skor yang diperoleh Siswa? Untuk mempermudah perhitungan skor, Anda perlu menyusun tabel seperti berikut:

Perhitungan Skor Pada Soal Berbeda Bobot

Jenjang Dominan Kognitif	Jumlah Soal	Bobot (b)	Jumlah Soal x Bobot (b)	B
Pengetahuan	10	1	10	10
Pemahaman	10	2	20	8
Aplikasi	15	3	45	10
Analisis	6	4	24	4
Sintesis	7	5	35	5
Evaluasi	2	6	12	1
Jumlah	50		146	38

$$Skor = \sum \frac{(10 \times 1) + (8 \times 2) + (10 \times 3) + (4 \times 4) + (5 \times 5) + (6 \times 1)}{146} \times 100\% = 70,55$$

Artinya, Siswa dapat menguasai materi IPS sebesar 70,55 %.

C. SKOR TOTAL (TOTAL SCORE)

Skor total adalah jumlah skor yang diperoleh dari seluruh bentuk soal setelah diolah dengan rumus tebakan (*guessing formula*). Jika kita mengambil contoh-contoh di atas, maka skor total siswa adalah $20 + 6 + 5 + 7 = 38$. Skor ini selanjutnya disebut skor mentah (*raw score*). Setelah dihitung skor mentah setiap peserta didik, langkah selanjutnya adalah mengolah skor mentah tersebut menjadi nilai-nilai jadi. Pengolahan skor

dimaksudkan untuk menetapkan batas lulus (*passing grade*) dan untuk mengubah skor mentah menjadi skor terjabar (*drived score*) atau skor standar. Untuk menentukan batas lulus, terlebih dahulu harus menghitung rata-rata (*mean*) dan simpangan baku (*standard deviation*), kemudian mengubah skor mentah menjadi skor terjabar atau skor standar berdasarkan kriteria atau norma tertentu.

Sebelum mempelajari lebih lanjut tentang perhitungan rata-rata dan simpangan baku, sebaiknya kita memahami dahulu tentang pendekatan penafsiran hasil evaluasi, karena perhitungan kedua teknik statistik dasar tersebut berkaitan dengan pendekatan yang digunakan.

D. PENILAIAN ACUAN PATOKAN (PAP)

Pendekatan ini lebih menitikberatkan pada apa yang dapat dilakukan oleh peserta didik. Dengan kata lain, kemampuan-kemampuan apa yang telah dicapai peserta didik sesudah menyelesaikan satu bagian kecil dari suatu keseluruhan program. Jadi, penilaian acuan patokan meneliti apa yang dapat dikerjakan oleh peserta didik, dan bukan membandingkan seorang peserta didik dengan teman sekelasnya, melainkan dengan suatu kriteria atau patokan yang spesifik. Kriteria yang dimaksud adalah suatu tingkat pengalaman belajar yang diharapkan tercapai sesudah selesai kegiatan belajar atau sejumlah kompetensi dasar yang telah ditetapkan terlebih dahulu sebelum kegiatan belajar berlangsung. Misalnya, kriteria yang digunakan 75% atau 80%. Bagi peserta didik yang kemampuannya di bawah kriteria yang telah ditetapkan dinyatakan tidak berhasil dan harus mendapatkan *remedial*.

Tujuan penilaian acuan patokan adalah untuk mengukur secara pasti tujuan atau kompetensi yang ditetapkan sebagai kriteria keberhasilannya. Penilaian acuan patokan sangat bermanfaat dalam upaya meningkatkan kualitas hasil belajar, sebab peserta didik diusahakan untuk mencapai standar yang telah ditentukan, dan hasil belajar peserta didik dapat diketahui derajat pencapaiannya. Untuk menentukan batas lulus (*passing grade*) dengan pendekatan ini, setiap skor peserta didik dibandingkan dengan skor ideal yang mungkin dicapai oleh peserta didik. Misalnya, dalam suatu tes ditetapkan skor idealnya adalah 100, maka peserta didik yang memperoleh skor 85 sama dengan memperoleh nilai 8,5 dalam skala 0 – 10. Demikian seterusnya.

Contoh: Diketahui skor 52 orang peserta didik sebagai berikut:

32 20 35 24 17 30 36 27 37 50 36 35 50 43 31 25 44 36 30 40
27 36 37 32 21 22 42 39 47 28 50 27 43 17 42 34 38 37 31 32
22 31 38 46 50 38 50 21 29 33 34 29

Pedoman konversi yang digunakan dalam mengubah skor mentah menjadi skor standar pada norma absolut skala lima adalah:

Tingkat Penguasaan	Skor Standar
90 % - 100 %	A
80 % - 89 %	B
70 % - 79 %	C
60 % - 69 %	D
> 59 %	E

Jika skor maksimum ditetapkan berdasarkan kunci jawaban = 60, maka penguasaan 90 % = $0,90 \times 60 = 55$, penguasaan 80 % = $0,80 \times 60 = 48$, penguasaan 70 % = $0,70 \times 60 = 42$, penguasaan 60 % = $0,60 \times 60 = 36$. Dengan demikian, diperoleh tabel konversi sebagai berikut:

Skor Mentah	Skor Standar
54 – 60	A
48 – 53	B
42 – 47	C
36 – 41	D
> 35	E

Jadi, peserta didik yang memperoleh skor 50 berarti nilainya B, skor 35 nilainya E (tidak lulus), skor 44 nilainya C, dan seterusnya. Jika dikehendaki standar sepuluh, maka skor peserta didik dapat dikonversi dengan pedoman sebagai berikut:

Tingkat Penguasaan	Skor Standar
95 % - 100 %	10
85 % - 94 %	9
75 % - 84 %	8
65 % - 74 %	7
55 % - 64 %	6
45 % - 54 %	5
35 % - 44 %	4
25 % - 34 %	3
15 % - 24 %	2
05 % - 14 %	1

Selanjutnya, persentase tingkat penguasaan terlebih dahulu diubah dalam bentuk tabel konversi. Caranya sama dengan skala lima di atas, setiap batas bawah tingkat penguasaan dikalikan dengan skor maksimum. Contoh: penguasaan 95 % = $0,95 \times 60 = 57$, penguasaan 85 % = $0,85 \times 60 = 51$, penguasaan 75 % = $0,75 \times 60 = 45$, dan seterusnya. Dengan demikian, tabel konversinya adalah:

Skor Mentah	Skor Standar
57 – 60	10

51 – 56	9
45 – 50	8
39 – 44	7
33 – 38	6
27 – 32	5
21 – 26	4
15 – 20	3
09 – 14	2
03 – 08	1

Berdasarkan tabel di atas, maka peserta didik yang memperoleh skor 47 nilainya 8, skor 35 nilainya 6, skor 24 nilainya 4, dan seterusnya. Penafsiran dengan pendekatan PAP dapat juga menggunakan langkah-langkah sebagai berikut:

- Mencari skor ideal, yaitu skor yang mungkin dicapai oleh peserta didik, jika semua soal dapat dijawab dengan betul.
- Mencari rata-rata (\bar{X}) ideal dengan rumus:

$$X \text{ ideal} = \frac{1}{2} \times \text{skor ideal}$$
- Mencari simpangan baku (s) ideal dengan rumus:

$$\text{Skor ideal} = \frac{1}{3} \times \bar{X} \text{ ideal}$$
- Menyusun pedoman konversi sesuai dengan kebutuhan.

Berdasarkan langkah-langkah di atas, maka pengolahannya adalah:

- Mencari skor ideal, yaitu 60.
- Mencari rata-rata ideal, yaitu $\frac{1}{2} \times 60 = 30$.
- Mencari simpangan baku ideal, yaitu $\frac{1}{3} \times 30 = 10$.
- Menyusun pedoman konversi:

a. Skala lima

$\bar{X} + 1,5 (s) = 30 + 1,5 (10) = 45$	—————→	A
$\bar{X} + 0,5 (s) = 30 + 0,5 (10) = 35$	—————→	B
$\bar{X} - 0,5 (s) = 30 - 0,5 (10) = 25$	—————→	C
$\bar{X} - 1,5 (s) = 30 - 1,5 (10) = 15$	—————→	D
	—————→	E

Dengan demikian, skor 32 nilainya C, skor 20 nilainya D, skor 35 nilainya C, skor 24 nilainya D, dan skor 17 nilainya D.

b. Skala Sepuluh

	—————→	10
--	--------	----

$\bar{X} + 2,25 (s) = 30 + 2,25 (10) = 52,5$	→	9
$\bar{X} + 1,75 (s) = 30 + 1,75 (10) = 47,5$	→	8
$\bar{X} + 1,25 (s) = 30 + 1,25 (10) = 42,5$	→	7
$\bar{X} + 0,75 (s) = 30 + 0,75 (10) = 37,5$	→	6
$\bar{X} + 0,25 (s) = 30 + 0,25 (10) = 32,5$	→	5
$\bar{X} - 0,25 (s) = 30 - 0,25 (10) = 27,5$	→	4
$\bar{X} - 0,75 (s) = 30 - 0,75 (10) = 22,5$	→	3
$\bar{X} - 1,25 (s) = 30 - 1,25 (10) = 17,5$	→	2
$\bar{X} - 1,75 (s) = 30 - 1,75 (10) = 12,5$	→	1
$\bar{X} - 2,25 (s) = 30 - 2,25 (10) = 7,5$	→	0

Dengan demikian, skor 32 nilainya 5, skor 20 nilainya 3, skor 35 nilainya 6, skor 24 nilainya 4, dan skor 17 nilainya 2.

c. Skala 0 – 100 (T – skor) :

$$\text{Rumus: T - Skor} = 50 + \left(\frac{X - \bar{X}}{s} \right) 10$$

Keterangan:

50 dan 10 = bilangan tetap

X = skor mentah yang diperoleh setiap peserta didik

\bar{X} = rata-rata

s = simpangan baku

Contoh: Peserta didik A memperoleh skor mentah 35. Rata-rata = 60 dan simpangan baku = 20. Dengan demikian, nilai yang diperoleh peserta didik A dalam skala nilai 0 -100 adalah: $50 + \frac{35-60}{20} 10 = 37,5$.

d. Konversi dengan Z – score :

Z – score adalah suatu ukuran yang menunjukkan berapa besarnya simpangan baku seseorang berada di bawah atau di atas rata-rata dalam kelompok tersebut.

$$\text{Rumus: } Z = \frac{x - \bar{x}}{s}$$

Contoh: Diketahui : skor (X) = 35; rata-rata (\bar{X}) = 60; simpangan baku = 20. Jadi, Z - Skor = $\frac{35-60}{20} = -1,25$

e. Peringkat (*ranking*)

Menafsirkan skor mentah dapat pula dilakukan dengan cara menyusun peringkat. Caranya adalah dengan mengurutkan skor, mulai dari skor terbesar sampai dengan skor terkecil. Skor terbesar diberi peringkat 1, begitu seterusnya sampai dengan skor terkecil. Skor-skor yang sama harus diberi peringkat yang sama pula.

Contoh: Diketahui 5 (lima) orang peserta didik memperoleh skor dalam bidang studi Pendidikan Ekonomi sebagai berikut: 20, 35, 25, 25, dan 30. Untuk memberikan peringkat terhadap skor-skor tersebut dapat diikuti langkah-langkah berikut: Pertama, mengurutkan skor-skor tersebut dari yang terbesar sampai dengan terkecil dengan diberi nomor urut sesuai dengan jumlah data, yaitu: 35 30 25 25 20. Kedua, memberi peringkat berdasarkan nomor urut, tetapi untuk skor yang sama harus diberi peringkat yang sama.

Skor	Peringkat
35	1
30	2
25	3,5
25	3,5
20	5

Peringkat untuk skor 25 adalah 3,5 yang diperoleh dari $(3 + 4) : 2 = 3,5$. Skor selanjutnya diberi peringkat sesuai dengan nomor urut selanjutnya. Dalam kurikulum berbasis kompetensi, prestasi belajar peserta didik ditentukan oleh perbandingan antara pencapaian sebelum dan sesudah pembelajaran serta kriteria penguasaan kompetensi yang ditentukan. Oleh karena itu, dalam penilaian berbasis kelas lebih tepat apabila menggunakan penilaian acuan patokan (PAP).

E. PENILAIAN ACUAN NORMA (PAN)

Dalam penilaian acuan norma, makna angka (skor) seorang peserta didik ditemukan dengan cara membandingkan hasil belajarnya dengan hasil belajar peserta didik lainnya dalam satu kelompok/kelas. Peserta didik dikelompokkan berdasarkan jenjang hasil belajar, sehingga dapat diketahui kedudukan relatif seorang peserta didik dibandingkan dengan teman sekelasnya. Tujuan

penilaian acuan norma adalah untuk membedakan peserta didik atas kelompok-kelompok tingkat kemampuan, mulai dari yang terendah sampai dengan tertinggi. Secara ideal, pendistribusian tingkat kemampuan dalam satu kelompok menggambarkan suatu kurva normal.

Pada umumnya, penilaian acuan norma dipergunakan untuk seleksi. Soal tes dalam pendekatan ini dikembangkan dari bagian bahan yang dianggap oleh guru urgen sebagai sampel dari bahan yang telah disampaikan. Kita berwenang untuk menentukan bagian mana yang lebih urgen. Untuk itu, harus dapat membatasi jumlah soal yang diperlukan, karena tidak semua materi yang disampaikan kepada peserta didik dapat dimunculkan soal-soalnya secara lengkap. Soal-soal harus dibuat dengan tingkat kesukaran yang bervariasi, mulai dari yang mudah sampai dengan sukar, sehingga memberikan kemungkinan jawaban peserta didik bervariasi, soal dapat menyebar, dan dapat membandingkan peserta didik yang satu dengan lainnya.

Peringkat dan klasifikasi anak yang didasarkan pada penilaian acuan norma lebih banyak mendorong kompetisi daripada membangun semangat kerjasama. Lagi pula tidak menolong sebagian besar peserta didik yang mengalami kegagalan. Dengan kata lain, keberhasilan peserta didik hanya ditentukan oleh kelompoknya. Dalam Kurikulum Berbasis Kompetensi, prestasi peserta didik ditentukan oleh perbandingan antara pencapaian sebelum dan sesudah pembelajaran, serta kriteria penguasaan kompetensi yang ditentukan.

Penilaian acuan norma biasanya digunakan pada akhir unit pembelajaran untuk menentukan tingkat hasil belajar peserta didik. Pedoman konversi yang digunakan dalam pendekatan PAN sama dengan pendekatan PAP. Perbedaannya hanya terletak dalam menghitung rata-rata dan simpangan baku. Dalam pendekatan PAN, rata-rata dan simpangan baku dihitung dengan rumus statistik sesuai dengan skor mentah yang diperoleh peserta didik. Adapun langkah-langkah pengolahannya adalah:

- a) Mencari skor mentah setiap peserta didik.
- b) Menghitung rata-rata (\bar{X}) aktual dengan rumus:

$$\bar{X}_{\text{aktual}} = Md + \left\{ \frac{\sum fd}{n} \right\} i$$

Keterangan:

Md = *mean* duga

f = frekuensi

d = deviasi

fd = frekuensi kali deviasi

n = jumlah sampel

i = interval

c) Menghitung simpangan baku (s) aktual dengan rumus:

$$S = i \sqrt{\frac{n(\sum f a^2) - (\sum f a)^2}{n - (n-1)}}$$

d) Menyusun pedoman konversi.

BAB VIII

ANALISIS KUALITAS TES

A. VALIDITAS

Sebelum menggunakan suatu tes, hendaknya mengukur terlebih dahulu derajat validitasnya berdasarkan kriteria tertentu. Dengan kata lain, untuk melihat apakah tes tersebut valid (sahih), maka harus membandingkan skor peserta didik yang didapat dalam tes dengan skor yang dianggap sebagai nilai baku. Misalnya, nilai ujian akhir semester peserta didik dalam salah satu mata pelajaran dibandingkan dengan nilai ujian akhir semester pada mata pelajaran yang lain. Semakin mendekati kedua skor tersebut, maka semakin soal ujian akhir tadi dapat dikatakan valid. Validitas suatu tes erat kaitannya dengan tujuan penggunaan tes tersebut. Namun demikian, tidak ada validitas yang berlaku secara umum. Artinya, jika suatu tes dapat memberikan informasi yang sesuai dan dapat digunakan untuk mencapai tujuan tertentu, maka tes itu valid untuk tujuan tersebut.

Ada dua unsur penting dalam validitas ini. Pertama, validitas menunjukkan suatu derajat, ada yang sempurna, ada yang sedang, dan ada pula yang rendah. Kedua, validitas selalu dihubungkan dengan suatu putusan atau tujuan yang spesifik. Sebagaimana pendapat R. L. Thorndike dan H. P. Hagen (1977: 56) bahwa "*validity is always in relation to a specific decision or use*". Sementara itu, Gronlund (1985: 79-81) mengemukakan ada tiga faktor yang mempengaruhi validitas hasil tes, yaitu "faktor instrumen evaluasi, faktor administrasi evaluasi dan penskoran, dan faktor dari jawaban peserta didik".

1. Faktor instrumen evaluasi

Mengembangkan instrumen evaluasi memang tidaklah mudah, apalagi jika seorang evaluator tidak atau kurang memahami prosedur dan teknik evaluasi itu sendiri. Jika instrumen evaluasi kurang baik, maka dapat berakibat hasil evaluasi menjadi kurang baik. Untuk itu, dalam mengembangkan instrumen evaluasi, seorang evaluator harus memperhatikan hal-hal yang mempengaruhi validitas instrumen dan berkaitan dengan prosedur penyusunan instrumen, seperti silabus, kisi-kisi soal, petunjuk mengerjakan soal dan pengisian lembar jawaban, kunci jawaban, penggunaan kalimat efektif, bentuk alternatif jawaban, tingkat kesukaran, daya pembeda, dan sebagainya.

2. Faktor administrasi evaluasi dan penskoran

Dalam administrasi evaluasi dan penskoran, banyak sekali terjadi penyimpangan atau kekeliruan, seperti: alokasi waktu untuk

pengerjaan soal yang tidak proporsional, memberikan bantuan kepada peserta didik dengan berbagai cara, peserta didik saling menyontek ketika ujian, kesalahan penskoran, termasuk kondisi fisik dan psikis peserta didik yang kurang menguntungkan.

3. Faktor jawaban dari peserta didik

Dalam praktiknya, faktor jawaban peserta didik justru lebih banyak berpengaruh daripada dua faktor sebelumnya. Faktor ini meliputi kecenderungan peserta didik untuk menjawab secara cepat tetapi tidak tepat, keinginan melakukan coba-coba, dan penggunaan gaya bahasa tertentu dalam menjawab soal bentuk uraian.

Selanjutnya, Kerlinger (1986) mengemukakan "validitas instrumen tidak cukup ditentukan oleh derajat ketepatan instrumen untuk mengukur apa yang seharusnya diukur, tetapi perlu juga dilihat dari tiga kriteria yang lain, yaitu *appropriateness*, *meaningfulness*, dan *usefulness*". *Appropriateness* menunjukkan kelayakan dari tes sebagai suatu instrumen, yaitu seberapa jauh instrumen dapat menjangkau keragaman aspek perilaku peserta didik. *Meaningfulness* menunjukkan kemampuan instrumen dalam memberikan keseimbangan soal-soal pengukurannya berdasar tingkat kepentingan dari setiap fenomena. *Usefulness to inferences* menunjukkan sensitif tidaknya instrumen dalam menangkap fenomena perilaku dan tingkat ketelitian yang ditunjukkan dalam membuat kesimpulan.

Dalam literatur modern tentang evaluasi, banyak dikemukakan tentang jenis-jenis validitas, antara lain: validitas permukaan (*face validity*), validitas isi (*content validity*), validitas empiris (*empirical validity*), dan validitas konstruk (*construct validity*), dan validitas faktor (*factorial validity*).

1. Validitas permukaan

Validitas ini menggunakan kriteria yang sangat sederhana, karena hanya melihat dari sisi muka atau tampak dari instrumen itu sendiri. Artinya, jika suatu tes secara sepintas telah dianggap baik untuk mengungkap fenomena yang akan diukur, maka tes tersebut sudah dapat dikatakan memenuhi syarat validitas permukaan, sehingga tidak perlu lagi adanya *judgement* yang mendalam.

2. Validitas isi

Validitas isi sering digunakan dalam pengukuran hasil belajar. Tujuan utamanya adalah untuk mengetahui hinggmana peserta didik menguasai materi pelajaran yang telah disampaikan, dan perubahan-perubahan psikologis apa yang timbul pada diri peserta didik tersebut setelah mengalami proses pembelajaran tertentu. Jika dilihat dari segi kegunaannya dalam penilaian hasil belajar, validitas isi ini sering

disebut juga validitas kurikuler dan validitas perumusan. Validitas kurikuler berkenaan dengan pertanyaan apakah materi tes relevan dengan kurikulum yang sudah ditentukan. Pertanyaan ini timbul karena sering terjadi materi tes tidak mencakup keseluruhan aspek-aspek yang akan diukur, baik aspek kognitif, afektif, maupun psikomotorik, tetapi hanya pengetahuan yang bersifat fakta-fakta pelajaran tertentu. Diharapkan dengan validitas kurikuler ini timbul ketelitian yang jelas dan totalitas dengan menjelajahi semua aspek yang tercakup dalam kisi-kisi dan Rencana Pelaksanaan Pembelajaran (RPP) yang bersangkutan. Validitas kurikuler ini dapat dilakukan dengan beberapa cara, antara lain mencocokkan materi tes dengan silabus dan kisi-kisi, melakukan diskusi dengan sesama pendidik, atau mencermati kembali substansi dari konsep yang akan diukur. Validitas perumusan berkenaan dengan pertanyaan apakah aspek-aspek dalam soal-soal itu betul-betul tercakup dalam perumusan tentang apa yang hendak diukur. Di samping itu, validitas isi dapat juga disebut validitas rasional atau validitas logis. Sebagaimana dikemukakan oleh R.L. Thorndike dan H.P. Hagen (1977: 58) bahwa "*scientific analysis is essentially a rational and judgmental one, this is sometimes spoken of as rational or logical validity*". Pernyataan ini memang ada benarnya, karena pengujian validitas harus dilakukan secara rasional dan logis, sehingga suatu tes hasil belajar dapat memiliki validitas yang sempurna.

3. Validitas empiris

Validitas ini biasanya menggunakan teknik statistik, yaitu analisis korelasi. Hal ini disebabkan validitas empiris mencari hubungan antara skor tes dengan suatu kriteria tertentu yang merupakan suatu tolok ukur di luar tes yang bersangkutan. Namun, kriteria itu harus relevan dengan apa yang akan diukur. Validitas empiris disebut juga validitas yang dihubungkan dengan kriteria (*criterion-related validity*) atau validitas statistik (*statistical validity*). Ada tiga macam validitas empiris, yaitu:

- a. Validitas prediktif (*predictive validity*)
- b. Validitas kongkuren (*concurrent validity*)
- c. Validitas sejenis (*congruent validity*)

Validitas prediktif ialah jika kriteria standar yang digunakan adalah untuk meramalkan prestasi belajar murid di masa yang akan datang. Dengan kata lain, validitas prediktif bermaksud melihat hinggamana suatu tes dapat memprakirakan perilaku peserta didik pada masa yang akan datang. Sedangkan validitas konkuren ialah jika kriteria standarnya berlainan.

Ada beberapa hal yang harus diperhatikan dalam menginterpretasikan koefisien validitas, antara lain data mengenai karakteristik sampel validitas, prosedur-prosedur dalam pengukuran validitas, dan pola kriteria khusus yang dikorelasikan dengan hasil tes itu. Sehubungan dengan kriteria khusus, Anastasi dalam Conny Semiawan Stamboel (1986: 50), mengemukakan ada delapan kriteria sebagai bahan bandingan untuk merumuskan apa yang hendak diselidiki oleh suatu tes, yaitu “diferensiasi umur, kemajuan akademis, kriteria dalam pelaksanaan latihan khusus, kriteria dalam pelaksanaan kerja, penilaian, kelompok yang dipertentangkan, korelasi dengan tes lain, dan konsistensi internal”.

a. Diferensiasi umur

Kriteria yang paling utama dalam validitas tes intelegensi adalah umur. Kebanyakan tes intelegensi, baik yang dipakai di Sekolah maupun tes pra-madrasah, senantiasa dibandingkan dengan umur kronologis untuk menentukan apakah angka bertambah dengan bertambahnya umur. Jika suatu tes dianggap valid, maka nilai tes bagi peserta didik akan naik dengan bertambahnya umur. Namun, anggapan ini tidak berlaku bagi perkembangan semua fungsi dalam hubungannya dengan bertambahnya umur secara konsisten (ini terbukti dari beberapa tes kepribadian). Suatu hal yang juga perlu dicermati adalah corak kondisi lingkungan tempat tes itu dibakukan. Kriteria peningkatan umur tidak bersifat universal tetapi tidak dapat juga dikatakan bahwa ini berlaku bagi corak masing-masing kebudayaan.

b. Kemajuan akademis

Pada umumnya tes intelegensi divalidkan dengan kemajuan akademis. Juga sering dikatakan bahwa makin lama seseorang belajar di madrasah, makin tinggi pendidikannya, makin tinggi pula kemajuan akademisnya. Padahal, setiap jenis dan jenjang pendidikan itu bersifat selektif. Bagi peserta didik yang tak sanggup meneruskan, biasanya termasuk *dropout*. Namun demikian, banyak pula faktor non-intelektual yang ikut mempengaruhi keberhasilan pendidikan seorang peserta didik. Dengan kata lain, berhasil tidaknya pendidikan seseorang tidak hanya dilihat dari faktor intelektual tetapi juga faktor non-intelektual. Untuk memperoleh gambaran yang komprehensif dan holistik tentang hal ini perlu diadakan penyelidikan yang lebih jauh.

c. Kriteria dalam pelaksanaan latihan khusus

Corak kriteria dalam pengembangan tes bakat khusus didasarkan atas prestasi dalam latihan tertentu secara khusus. Beberapa tes bakat profesi (*professional aptitude test*) telah divalidkan dengan tes hasil belajar dalam bidang-bidang tersebut. Misalnya, tes untuk memasuki profesi kedokteran, hukum, dan sebagainya. Ada beberapa tes untuk

memasuki profesi tertentu yang disebut *tailor-made test*, yaitu tes yang telah dibuat khusus untuk keperluan tersebut, seperti tes penerbangan.

d. Kriteria dalam pelaksanaan kerja

Dalam validitas tes kepribadian dan validitas tes bakat khusus banyak digunakan kriteria yang didasarkan atas kinerja dalam pelaksanaan kerja (*on the job performance*). Mengingat masing-masing pekerjaan memiliki kekhasan sendiri dan berbeda-beda tingkat, bentuk, maupun coraknya, maka untuk masing-masing pekerjaan diciptakan tes yang terkenal dengan istilah *tailor-made test*.

e. Penilaian

Pengertian penilaian disini adalah teknik untuk memperoleh informasi tentang kemajuan belajar peserta didik di Sekolah. Selain itu, juga mencakup pekerjaan yang memerlukan latihan khusus ataupun sukses dalam penilaian pribadi oleh seorang pengamat terhadap berbagai fungsi psikologis. Misalnya, kondisi-kondisi, orijinalitas, kepemimpinan, atau kejujuran. Jika kondisi-kondisi pengenalan dalam situasi tempat kemampuan yang khusus itu dinyatakan, maka perlu disertai skala penilaian yang dipersiapkan secara teliti.

f. Kelompok yang dipertentangkan

Konsep validitas melalui kelompok yang dipertentangkan menyelidiki pengaruh kehidupan sehari-hari yang tak disengaja. Kriteria ini didasarkan atas kelebihan suatu kelompok tertentu dihadapkan kepada kelompok yang lain dalam menjalankan suatu tes tertentu. Misalnya, suatu tes bakat musik dicobakan dalam suatu sekolah musik maupun dalam suatu Sekolah. Kriteria itu didasarkan atas faktor yang menyolok, yang diperoleh dari hasil nilai kedua kelompok tersebut dalam menjalankan tes itu.

g. Korelasi dengan tes lain

Korelasi antara tes baru dengan tes lama merupakan perbandingan kriteria dalam menyelidiki perilaku yang sama. Dalam hal ini suatu tes verbal tertulis bisa dibandingkan dengan tes individual atau tes kelompok. Untuk mengukur apakah suatu tes yang baru memiliki validitas dan bebas dari pengaruh faktor lain, maka dipergunakan tes jenis lain dalam membandingkannya. Jadi, kadang-kadang tes kepribadian dikorelasikan dengan tes internal atau tes hasil belajar.

h. Konsistensi internal

Kriteria konsistensi internal adalah skor total yang diperoleh peserta didik dalam suatu tes. Kriteria ini terutama digunakan dalam bidang tes kepribadian. Kadang-kadang untuk keperluan ini juga digunakan percobaan tes dengan dua kelompok, yaitu antara kelompok berhasil dan kelompok kurang berhasil. Skor setiap soal tes dari kelompok yang

berhasil dibandingkan dengan skor setiap soal tes dari kelompok yang kurang berhasil. Soal-soal yang gagal menunjukkan perbedaan antara kelompok yang berhasil dengan kelompok yang kurang berhasil harus diperbaiki, atau dibuang. Kriteria konsistensi internal ini menghasilkan indeks homogenitas soal, tetapi tidak dapat dianggap sepenuhnya sebagai pengganti validitas.

Untuk menguji validitas empiris dapat dihitung dengan korelasi *product-moment*, korelasi perbedaan peringkat, dan diagram pencar.

a. Korelasi *product-moment* dengan angka simpangan

$$\text{Rumus } r_{xy} = \frac{\Sigma xy}{\sqrt{(x^2)(y^2)}}$$

Keterangan:

r = koefisien korelasi

Σxy = jumlah produk x dan y

Contoh: 10 orang peserta didik kelas 11 SMA mendapat nilai dalam mata pelajaran Ekonomi dan Akuntansi seperti berikut. Nilai 10 Orang Peserta Didik Kelas 11 dalam Mata Pelajaran Ekonomi dan Akuntansi

No	Nama	Ekonomi	Akuntansi
1	A	5	6
2	B	7	8
3	C	8	7
4	D	5	5
5	E	6	7
6	F	7	7
7	G	4	5
8	H	5	7
9	I	8	8
10	J	6	6

Langkah-langkah penyelesaian:

a) Buat tabel persiapan seperti berikut:

No	X	Y	x	y	x ²	y ²	xy

b) Masukkan nilai masing-masing mata pelajaran, dimana nilai Ekonomi sebagai variabel X dan nilai Akuntansi sebagai variabel Y.

c) Jumlahkan semua nilai yang ada dalam variabel X dan variabel Y, kemudian hitung rata-rata X dan rata-rata Y.

- d) Cari nilai pada kolom x dengan jalan nilai tiap-tiap peserta didik dalam kolom X dikurangi dengan rata-rata X .
- e) Cari nilai pada kolom y dengan jalan nilai tiap-tiap peserta didik dalam kolom Y dikurangi dengan rata-rata Y .
- f) Cari nilai pada kolom x^2 dengan jalan mengkuadratkan masing-masing nilai dalam kolom x .
- g) Cari nilai pada kolom y^2 dengan jalan mengkuadratkan masing-masing nilai dalam kolom y .
- h) Cari nilai pada kolom xy dengan jalan mengalikan tiap-tiap nilai dalam kolom x dengan nilai-nilai dalam kolom y .

Berdasarkan langkah-langkah di atas dapat dihitung koefisien korelasi sebagai berikut. Perhitungan Korelasi Product-Moment dengan Angka Simpangan.

No	Nilai Ekonomi (X)	Nilai Akuntansi (Y)	x	y	x^2	y^2	xy
1	5	6	-1,1	-0,6	1,21	0,36	0,66
2	7	8	0,9	1,4	0,81	1,96	1,26
3	8	7	1,9	0,4	3,61	0,16	0,76
4	5	5	-1,1	-1,6	1,21	2,56	1,76
5	6	7	-0,1	0,4	0,01	0,16	-0,04
6	7	7	0,9	0,4	0,81	0,16	0,36
7	4	5	-2,1	-1,6	4,41	2,56	3,36
8	5	7	-1,1	0,4	1,21	0,16	-0,04
9	8	8	1,9	1,4	3,61	1,96	2,66
10	6	6	-0,1	-0,6	0,01	0,36	0,06
Σ	61	66			16,9	10,4	10,4
\bar{X}	6,1	6,6					

Jadi, nilai r_{xy} adalah 0,784 berdasarkan hasil perhitunga rumus di atas.

4. Validitas konstruk

Konstruk adalah konsep yang dapat diobservasi (*observable*) dan dapat diukur (*measurable*). Validitas konstruk sering juga disebut validitas logis (*logical validity*). Validitas konstruk berkenaan dengan pertanyaan hinggamana suatu tes betul-betul dapat mengobservasi dan mengukur fungsi psikologis yang merupakan deskripsi perilaku peserta didik yang akan diukur oleh tes tersebut. Validitas konstruk banyak dikenal dan digunakan dalam tes-tes psikologis untuk mengukur gejala

perilaku yang abstrak, seperti kesetiakawanan, kematangan emosi, sikap, motivasi, minat, dan sebagainya.

Untuk menguji validitas konstruk dapat dilakukan dengan berbagai sumber, antara lain validitas isi, validitas prediktif, dan validitas konkuren. N.E. Gronlund (1985) mengemukakan *"It is a matter of accumulating evidence from many different sources. We may use content validity, predictive validity, and concurrent validity as partial evidence to support construct validity, but none of them alone is sufficient"*. Analisis statistika yang digunakan dalam validitas konstruk antara lain dengan analisis faktor (*factor analysis*), sehingga dapat diketahui :

- b. Aspek-aspek apa saja yang diukur oleh setiap butir soal.
- c. Berapa besar suatu butir soal berisi faktor-faktor tertentu.
- d. Faktor-faktor apa yang diukur oleh suatu butir soal.

Produk analisis faktor ini dapat menganalisis dan mempertimbangkan apakah suatu tes betul-betul dapat mengukur fungsi psikologis yang merupakan deskripsi perilaku peserta didik yang hendak diukur oleh tes yang bersangkutan.

5. Validitas faktor

Dalam evaluasi atau penilaian sering digunakan skala pengukuran tentang suatu variabel yang terdiri atas beberapa faktor. Faktor-faktor tersebut diperoleh berdasarkan dimensi/indikator dari variabel yang diukur sesuai dengan apa yang terungkap dalam konstruksi teoritisnya. Meskipun variabel terdiri atas beberapa faktor, tetapi prinsip homogenitas untuk keseluruhan faktor harus tetap dipertahankan, sehingga tidak terjadi tumpang tindih antara satu faktor dengan faktor yang lain. Dengan demikian, kriterium yang digunakan dalam validitas faktor ini dapat diketahui dengan menghitung homogenitas skor setiap faktor dengan total skor, dan antara skor dari faktor yang satu dengan skor dari faktor yang lain.

B. RELIABILITAS

Reliabilitas suatu tes adalah tingkat atau derajat konsistensi tes yang bersangkutan. Reliabilitas berkenaan 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 atau kesempatan yang berbeda. Anastasi (1976: 103) mengemukakan *"reliability refers to the consistency of scores obtained by the same persons when reexamined the same test on different occasion, or with different sets of equivalent items or under other variable examining conditions"*. Hal senada dikemukakan

Gronlund (1985: 87) bahwa “*reliability refers to the results obtained with an evaluation instrument and not to the instrument it self*”.

Sementara itu, Kerlinger (1986: 443) mengemukakan “*reliabilitas dapat diukur dari tiga kriteria, yaitu stability, dependability, dan predictability*”. *Stability* menunjukkan keajegan suatu tes dalam mengukur gejala yang sama pada waktu yang berbeda. *Dependability* menunjukkan kemantapan suatu tes atau seberapa jauh tes dapat diandalkan. *Predictability* menunjukkan kemampuan tes untuk meramalkan hasil pada pengukuran gejala selanjutnya. Untuk meningkatkan reliabilitas suatu tes, antara lain dapat dilakukan dengan memperbanyak butir soal.

Selanjutnya, Gronlund (1985: 100) mengemukakan ada empat faktor yang dapat mempengaruhi reliabilitas, yaitu “*panjang tes, sebaran skor, tingkat kesukaran, dan objektivitas*”.

1. Panjang tes (*length of test*)

Panjang tes berarti 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 banyak, sehingga faktor tebakan (*guessing*) akan semakin rendah.

2. Sebaran skor (*spread of scores*)

Besarnya sebaran skor akan membuat tingkat reliabilitas menjadi lebih tinggi, karena koefisien reliabilitas yang lebih besar diperoleh ketika peserta didik tetap pada posisi yang relatif 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 (*difficulty indeks*)

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 antar peserta 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 (*obyektivity*)

Objektivitas disini 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 (*error of measurement*). 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 (*standard error of measurement*) rendah.

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

3. Koefisien stabilitas

Koefisien stabilitas (*coefficient of stability*) adalah 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 kedua cukup lama, kemudian diadakan latihan-latihan tambahan, maka bisa jadi nilai tes kedua akan lebih besar daripada tes 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 demikian, 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*.

4. Koefisien ekuivalen

Koefisien ekuivalen (*coefficient of equivalence*) adalah jika mengorelasikan dua buah tes yang paralel pada kelompok dan waktu yang sama. Metode yang digunakan untuk memperoleh koefisien ekuivalen adalah metode dengan menggunakan dua buah bentuk tes yang paralel (*equivalen*) atau *equivalence forms method* atau disebut juga *parallel or alternate-forms method*. Syarat-syarat yang harus dipenuhi kedua tes paralel adalah kriteria 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.

5. Koefisien konsistensi internal

Koefisien konsistensi internal (*coefficient of internal consistency*) adalah 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 juga 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. Di samping 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 ekuivalens, 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:

$$r_n = \frac{2r_{1,2}}{1+(n-1)r_{1,2}}$$

Keterangan:

n = panjang tes yang selalu sama dengan 2 karena seluruh tes = $2 \times \frac{1}{2}$

Contoh: 10 orang peserta didik dites dalam mata pelajaran Ekonomi dan Akuntansi. Jumlah soal masing-masing lima buah. Dua buah nomor genap diambil dari tes Ekonomi dan tiga buah nomor ganjil diambil dari tes Akuntansi. Data diperoleh sebagai berikut:

Nilai 10 Orang Peserta Didik Dalam Mata Pelajaran Ekonomi dan Akuntansi

Nama	Skor Ekonomi No. Genap (2 dan 4)		Skor Akuntansi No. Ganjil (1,3 dan 5)		
	A	8	6	8	7
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 Koefisien Konsistensi Internal

X	Y	x	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	+1	4	1	2
14	22	+1	+3	1	9	3
130	190			22	90	29
$\bar{X} = 13$	$\bar{Y} = 19$					

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{29}{\sqrt{(22)(90)}} = 0,65$$

Untuk menghitung seluruh tes itu, kita gunakan rumus *Spearman Brown* sebagai berikut:

$$r_n = \frac{2r_{12}}{1+(n-1)r_{12}} = \frac{(2)(0,65)}{1+(2-1)(0,65)} = 0,787$$

Di samping itu, Anda dapat pula menggunakan teknik *Kuder-Richardson* (dua orang ahli psikometri yang merumuskan persamaan untuk mencari reliabilitas) yang lebih populer dengan istilah KR20.

Teknik *Kuder-Richardson* biasanya digunakan jika instrumen digunakan untuk mengukur satu gejala psikologis atau perilaku yang sama. Artinya, tes tersebut dapat dikatakan reliabel bila terbukti ada konsistensi jawaban antara soal yang satu dengan soal yang lain. Jika sifat dan tingkatan homogenitas antar soal tidak terpenuhi, maka tes tersebut dianggap mengukur lebih dari satu variabel. Jika dalam suatu tes terdapat lebih dari satu skala pengukuran atau mengukur lebih dari satu variabel dan setiap variabel memiliki beberapa aspek, maka pengecekan reliabilitas dilakukan terhadap masing-masing skala pengukuran. Teknik ini lebih cocok untuk tes yang menggunakan soal dua pilihan dengan salah satu jawaban benar.

Teknik lain yang biasa digunakan untuk menguji konsistensi internal dari suatu tes adalah *Cronbach's Alpha* atau Koefisien Alpha. Perbedaannya dengan teknik *Kuder-Richardson* adalah teknik ini tidak hanya digunakan untuk tes dengan dua pilihan saja, tetapi penerapannya lebih luas, seperti menguji reliabilitas skala pengukuran sikap dengan tiga, lima atau tujuh pilihan.

C. KEPRAKTISAN

Dalam kenyataannya, banyak tes yang dibuat orang tidak menunjukkan kepraktisan. Padahal, kepraktisan merupakan syarat suatu tes standar. Kebanyakan orang membuat tes hanya untuk kepentingan dirinya sendiri, tidak berpikir untuk orang lain. Akibatnya, ketika tes tersebut digunakan orang lain, maka orang tersebut akan merasa kesulitan. Kepraktisan bukan hanya dipertimbangkan ketika memilih tes yang sudah dipublikasikan, tetapi siapapun yang mengembangkan tes harus memenuhi syarat ini. Kepraktisan mengandung arti kemudahan suatu tes, baik dalam mempersiapkan, menggunakan, mengolah dan menafsirkan, maupun mengadministrasikannya. Dimiyati dan Mudjiono (1994: 184) mengemukakan faktor-faktor yang mempengaruhi kepraktisan instrumen evaluasi meliputi: "kemudahan administrasi, waktu yang disediakan untuk melancarkan evaluasi, kemudahan menskor, kemudahan interpretasi dan aplikasi, tersedianya bentuk instrumen evaluasi yang ekuivalen atau sebanding".

4. Kemudahan mengadministrasi.

Jika instrumen evaluasi diadministrasikan oleh guru atau orang lain dengan kemampuan yang terbatas, kemudahan pengadministrasian adalah suatu kualitas penting yang diminta dalam instrumen evaluasi. Untuk memberikan kemudahan pengadministrasian instrumen evaluasi dapat dilakukan dengan jalan memberi petunjuk yang sederhana dan jelas, subtes sebaiknya relatif sedikit, dan pengaturan tempo tes sebaiknya tidak menimbulkan kesulitan. Kesalahan-kesalahan dalam mengadministrasikan alat ukur atau instrumen evaluasi akan menurunkan kepraktisannya, sehingga dapat menyebabkan berkurangnya validitas dan reliabilitas suatu alat ukur.

5. Waktu yang disediakan untuk melancarkan evaluasi.

Kepraktisan dipengaruhi pula oleh faktor waktu yang disediakan untuk melancarkan evaluasi. Waktu antara 20 menit sampai 60 menit yang disediakan untuk melancarkan evaluasi merupakan waktu yang cukup memberikan kepraktisan.

6. Kemudahan penskor.

Hal yang membosankan dan mengganggu dalam melancarkan kegiatan evaluasi adalah penskoran. Guru seringkali bekerja keras untuk melaksanakan tugas ini. Hal ini tentu mengurangi kepraktisan alat ukur. Untuk memberikan kemudahan penskoran diperlukan upaya berupa perbaikan petunjuk penskoran dan lebih memudahkan kunci penskoran, pemisahan lembar jawaban dari lembar soal, dan penskoran menggunakan mesin.

7. Kemudahan interpretasi dan aplikasi.

Dalam analisis terakhir, keberhasilan atau kegagalan evaluasi ditentukan oleh penggunaan hasil evaluasi. Jika hasil evaluasi ditafsirkan secara tepat dan diterapkan secara efektif, maka hasil evaluasi akan mendukung terhadap keputusan-keputusan pembelajaran yang lebih tepat. Untuk memudahkan interpretasi dan aplikasi hasil evaluasi diperlukan petunjuk yang jelas. Semakin mudah interpretasi dan aplikasi hasil evaluasi, semakin meningkatkan kepraktisan evaluasi.

8. Tersedianya bentuk instrumen evaluasi yang ekuivalen atau sebanding.

Untuk berbagai kegunaan pendidikan, bentuk-bentuk ekuivalen untuk tes yang sama sering kali diperlukan. Bentuk-bentuk ekuivalen dari sebuah tes mengukur aspek-aspek perilaku melalui butir-butir tes yang memiliki kesamaan dalam isi, tingkat kesulitan, dan karakteristik lainnya. Dengan demikian, satu bentuk tes dapat menggantikan yang lain. Sedangkan alat ukur atau instrumen evaluasi yang sebanding

adalah instrumen evaluasi yang memiliki kemungkinan dibandingkan makna dari skala skor umum yang dimiliki, sehingga untuk tes berseri cukup menggunakan satu skala skor. Adanya bentuk-bentuk yang ekuivalen atau sebanding dari instrumen evaluasi akan mempraktikkan kegiatan evaluasi.

BAB IV

MENGANALISIS KUALITAS BUTIR SOAL

A. TINGKAT KESUKARAN SOAL (*DIFFICULTY INDEX*)

Berkualitas atau tidaknya butir-butir item tes hasil belajar pertama-tama dapat diketahui dari derajat kesukaran atau taraf kesulitan yang dimiliki oleh masing-masing butir item tersebut. Butir-butir item tes hasil belajar dapat dinyatakan sebagai butir-butir item yang baik, apabila butir-butir item tersebut tidak terlalu sukar dan tidak pula terlalu mudah. Dengan kata lain derajat kesukaran item tersebut adalah sedang atau cukup.

Bertitik tolak dari pernyataan tersebut, maka butir-butir item tes hasil belajar dimana seluruh testee tidak dapat menjawab dengan betul (karena terlalu sukar) tidak dapat disebut sebagai item yang baik. Demikian juga sebaliknya, butir-butir item tes hasil belajar dimana seluruh testee dapat menjawab dengan betul (karena terlalu mudah) juga tidak dapat digolongkan kedalam kategori item yang baik.

Perhitungan tingkat kesukaran soal adalah pengukuran seberapa besar derajat kesukaran suatu soal. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik. Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah.

1. Menghitung tingkat kesukaran soal bentuk objektif

Untuk menghitung tingkat kesukaran soal bentuk objektif dapat digunakan dengan dua cara, yaitu menggunakan rumus tingkat kesukaran (TK):

$$TK = \frac{(WL+WH)}{(nL+nH)} \times 100\%$$

Keterangan:

WL = jumlah peserta didik yang menjawab salah dari kelompok bawah

WH = jumlah peserta didik yang menjawab salah dari kelompok atas

nL = jumlah kelompok bawah

nH = jumlah kelompok atas

Sebelum menggunakan rumus di atas, maka Anda harus menempuh terlebih dahulu langkah-langkah sebagai berikut:

- a) Menyusun lembar jawaban peserta didik dari skor tertinggi sampai dengan skor terendah.

- b) Mengambil 27 % lembar jawaban dari atas yang selanjutnya disebut kelompok atas (*higher group*), dan 27 % lembar jawaban dari bawah yang selanjutnya disebut kelompok bawah (*lower group*). Sisa sebanyak 46% disisihkan.
- c) Membuat tabel untuk mengetahui jawaban (benar atau salah) dari setiap peserta didik, baik untuk kelompok atas maupun kelompok bawah. Jika jawaban peserta didik benar, diberi tanda + (plus), sebaliknya jika jawaban peserta didik salah, diberi tanda - (minus).

Sebenarnya, dalam analisis butir soal secara klasikal, tingkat kesukaran dapat diperoleh dengan beberapa cara, antara lain: skala kesukaran linier, skala bivariat, indeks davis, dan proporsi menjawab benar. Cara yang terakhir, yaitu proporsi menjawab benar (*proportion correct*) sangat banyak digunakan karena dianggap lebih mudah. Caranya adalah jumlah peserta didik yang menjawab benar pada soal yang dianalisis dibagi dengan jumlah peserta didik. Persamaan yang digunakan untuk menentukan *proportion correct* (p) adalah:

$$p = \frac{\sum B}{N}$$

Keterangan:

p = tingkat kesukaran

$\sum B$ = jumlah peserta didik yang menjawab benar

N = jumlah peserta didik

Contoh: 40 orang siswa SMA dites dalam mata pelajaran Ekonomi. Dari seluruh peserta didik tersebut, ada 25 orang yang dapat menjawab dengan benar pada soal nomor 1. Dengan demikian, tingkat kesukaran soal nomor 1 itu adalah: $p = \frac{25}{40} = 0,625$

Untuk menafsirkan tingkat kesukaran tersebut, dapat digunakan kriteria sebagai berikut:

$p > 0,70$ = mudah

$0,30 \leq p \leq 0,70$ = sedang

$p < 0,30$ = sukar

Dengan demikian, soal nomor 1 dalam contoh di atas termasuk "sedang".

Tingkat kesukaran model ini banyak mengandung kelemahan, karena tingkat kesukaran ini sebenarnya merupakan "ukuran kemudahan" soal. Semakin tinggi indeks tingkat kesukaran (p), maka semakin mudah soalnya. Sebaliknya, semakin rendah tingkat kesukaran, maka semakin sulit soalnya. Artinya, model tingkat kesukaran seperti ini lebih tepat disebut tingkat kemudahan (*easiness*). Dalam praktiknya, ada soal yang dikategorikan ekstrim sukar (jika p mendekati nol) dan ada soal yang termasuk ekstrim mudah (jika p mendekati satu).

Sehubungan dengan tingkat kesukaran ini, ada beberapa hal yang harus diperhatikan dalam menyusun soal di bank soal, yaitu:

- a) Soal yang termasuk ekstrim sukar atau ekstrim mudah tidak memberikan informasi yang berguna bagi sebagian besar peserta didik. Oleh sebab itu, soal seperti ini kemungkinan distribusi jawaban pada alternatif jawaban ada yang tidak memenuhi syarat.
 - b) Jika ada soal ekstrim sukar atau ekstrim mudah, tetapi setiap pengecoh (distribusi jawaban) pada soal tersebut menunjukkan jawaban yang merata, logis, dan daya bedanya negatif (kecuali kunci), maka soal-soal tersebut masih memenuhi syarat untuk diterima.
 - c) Jika ada soal ekstrim sukar dan ekstrim mudah, tetapi memiliki daya pembeda dan statistik pengecoh memenuhi kriteria, maka soal tersebut dapat dipilih dan diterima sebagai salah satu alternatif untuk disimpan dalam bank soal.
 - d) Jika ada soal ekstrim sukar dan ekstrim mudah, daya pembeda dan statistik pengecohnya belum memenuhi kriteria, maka soal tersebut perlu direvisi dan diujicoba lagi.
2. Menghitung tingkat kesukaran untuk soal bentuk uraian

Cara menghitung tingkat kesukaran untuk soal bentuk uraian adalah menghitung berapa persen peserta didik yang gagal menjawab benar atau ada di bawah batas lulus (*passing grade*) untuk tiap-tiap soal. Untuk menafsirkan tingkat kesukaran soalnya dapat digunakan kriteria sebagai berikut:

- a) Jika jumlah peserta didik yang gagal mencapai 27 %, termasuk mudah.
- b) Jika jumlah peserta didik yang gagal antara 28 % sampai dengan 72 %, termasuk sedang.
- c) Jika jumlah peserta didik yang gagal 72 % ke atas, termasuk sukar.

Contoh: 33 orang peserta didik dites dengan lima soal bentuk uraian. Skor maksimum ditentukan 10 dan skor minimum 0. Jumlah peserta didik yang memperoleh nilai 0 – 5 = 10 orang (berarti gagal), nilai 6 = 12 orang, dan nilai 7 – 10 = 11 orang.

Jadi, tingkat kesukaran (TK) = $\frac{10}{33} \times 100\% = 30,3\%$

Tingkat kesukaran 30,3 berada di antara 28 dan 72, berarti soal tersebut termasuk sedang. Catatan: Batas lulus ideal = 6 (skala 0 - 10).

Setelah berhasil diidentifikasi butir-butir item mana yang derajat kesukarannya termasuk dalam kategori cukup, terlalu sukar dan terlalu mudah maka yang menjadi pokok permasalahan sekarang adalah, bagaimana menindaklanjuti hasil analisis item tersebut? Dalam kaitannya

dengan hasil analisis item dari segi derajat kesukaran, maka tindak lanjut yang perlu dilakukan oleh tester adalah sebagai berikut.

Pertama, untuk butir-butir item yang berdasarkan hasil analisis termasuk dalam kategori baik, seyogyanya butir item tersebut segera dicata dalam buku bank soal. Selanjutnya butir-butir soal tersebut dapat dikeluarkan lagi dalam tes-tes hasil belajar pada waktu-waktu yang akan datang.

Kedua, untuk butir-butir item yang termasuk dalam kategori terlalu sukar, ada tiga kemungkinan tidak lanjut, yaitu: (1) Butir item tersebut dibuang atau didrop dan tidak akan dikeluarkan lagi dalam tes-tes hasil belajar yang akan datang. (2) diteliti ulang, dilacak dan ditelusuri sehingga dapat diketahui factor yang menyebabkan butir item yang bersangkutan sulit dijawab oleh testee; apakah kalimat soalnya kurang jelas, apakah petunjuk petunjuk cara mengerjakan atau menjawab soalnya sulit dipahami, atau dalam soal terdapat istilah-istilah yang tidak jelas. Setelah dilakukan perbaikan kembali, butir-butir item tersebut dikeluarkan lagi dalam tes hasil belajar yang akan datang. (3) haruslah dipahami bahwa tidak setiap butir item yang termasuk dalam kategori terlalu sukar itu sama sekali tidak memiliki kegunaan. Butir-butir item yang terlalu sukar itu sewaktu-waktu masih dapat diambil manfaatnya, yaitu dapat digunakan dalam tes-tes (terutama tes seleksi) yang sifatnya sangat ketat. Dalam kondisi seperti itu sangat tepat apabila butir-butir item yang dikeluarkan adalah butir-butir item yang termasuk kategori terlalu sukar dengan asumsi bahwa testee dengan kemampuan yang rendah dengan mudah akan tersisihkan dari seleksi, sedangkan testee yang memiliki kemampuan tinggi tidak akan terlalu sukar untuk lolos dalam seleksi tersebut.

Ketiga, untuk butir-butir item yang termasuk dalam kategori terlalu mudah, juga ada tiga kemungkinan tindak lanjut, yaitu: (1) butir item tersebut dibuang atau didrop dan tidak akan dikeluarkan lagi dalam tes-tes hasil belajar yang akan datang. (2) diteliti ulang, dilacak dan ditelusuri secara cermat guna mengetahui factor yang menyebabkan butir item tersebut dapat dijawab betul oleh hamper seluruh testee; ada kemungkinan option atau alternative yang dipasangkan pada butir-butir item yang bersangkutan terlalu mudah diketahui oleh testee, mana item yang merupakan kunci jawaban item dan mana option yang berfungsi sebagai pengecoh atau distractor. Disini tester harus berusaha memperbaiki atau menggantinya dengan option yang lain sedemikian rupa sehingga antara kunci jawaban dengan pengecoh sulit dibedakan oleh testee. Setelah dilakukan perbaikan, item yang bersangkutan dicoba untuk dikeluarkan lagi pada tes hasil belajar selanjutnya. Tujuannya

adalah untuk mengetahui apakah derajat kesukaran item itu menjadi lebih baik atau tidak daripada yang sebelumnya. (3) seperti halnya butir-butir item yang terlalu sukar, butir-butir item yang terlalu mudah juga masih mengandung manfaat, yaitu bahwa butir-butir item yang termasuk dalam kategori ini dapat dimanfaatkan pada tes-tes yang sifatnya longgar, dalam arti bahwa sebagian besar dari testee akan dinyatakan lulus dalam tes seleksi tersebut. Dalam kondisi seperti ini sangat bijaksana apabila butir-butir item yang dikeluarkan dalam tes seleksi tersebut adalah butir-butir item yang termasuk dalam kategori terlalu mudah, sehingga tes seleksi itu boleh dikatakan hanya sebagai formalitas saja.

Dari uraian di atas, maka tidak ada jeleknya untuk memasukkan butir-butir item yang termasuk kategori terlalu sukar dan terlalu mudah di dalam buku bank soal, sebab sewaktu-waktu butir-butir semacam itu diperlukan, tester tidak perlu membuat atau menyusun butir-butir item dengan derajat kesukaran dan derajat kemudahan yang tinggi.

B. DAYA PEMBEDA (*DISCRIMINATING POWER*)

Daya pembeda item adalah kemampuan suatu butir item tes hasil belajar untuk dapat membedakan antara testee yang berkembang tinggi (pandai) dengan testee yang berkembang rendah (bodoh) sedemikian rupa sehingga sebagian besar testee dapat menjawab soal dengan betul sementara testee yang memiliki kemampuan rendah tersebut sebagian besar tidak dapat menjawab item dengan betul.

Mengetahui daya pembeda itu penting sekali, sebab salah satu dasar yang dipegangi untuk menyusun butir-butir item tes hasil belajar adalah adanya anggapan, bahwa kemampuan antar testee yang satu dengan testee yang lain itu berbeda-beda, dan bahwa butir-butir item tes hasil belajar itu haruslah mampu memberikan hasil tes yang mencerminkan adanya perbedaan-perbedaan kemampuan yang terdapat dikalangan testee tersebut.

Daya pembeda dapat diketahui dengan melihat besar kecilnya angka indeks deskriminasi item. Angka indeks diskriminasi item adalah sebuah angka atau bilangan yang menunjukkan besar kecilnya daya pembeda yang dimiliki oleh sebutir item. Daya pembeda pada dasarnya dihitung atas dasar pembagiantestee ke dalam dua kelompok, yaitu kelompok atas (kelompok yang tergolong pandai) dan kelompok bawah (kelompok yang tergolong bodoh).

Perhitungan daya pembeda adalah pengukuran sejauhmana suatu butir soal mampu membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum/kurang menguasai kompetensi berdasarkan kriteria tertentu. Semakin tinggi koefisien daya

pembeda suatu butir soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi. Untuk menghitung daya pembeda setiap butir soal dapat digunakan rumus sebagai berikut:

$$DP = \frac{(WL - WH)}{n}$$

Keterangan:

DP = daya pembeda

WL = jumlah peserta didik yang gagal dari kelompok bawah

WH = jumlah peserta yang gagal dari kelompok atas

n = 27 % x N

Contoh:

Jumlah peserta didik (N) = 40

Jumlah sampel (n) = 27 % x 40 = 10,8 = 11 (dibulatkan)

WL = 10

WH = 2

Jadi, Daya Pembedanya (DP) = $\frac{10-2}{11} = 0,73$

Untuk menginterpretasikan koefisien daya pembeda tersebut dapat digunakan kriteria yang dikembangkan oleh Ebel sebagai berikut:

<i>Besarnya Angka Indeks Diskriminasi Item (D)</i>	<i>Klaifikasi</i>	<i>Item evaluation</i>
Kurang dari 0,20	Poor	Butir item yang bersangkutan daya pembedanya lemah sekali (jelek), dianggap tidak memiliki daya pembeda yang baik.
0,20 – 0,40	Satisfactory	Butir item yang bersangkutan telah memiliki daya pembeda yang cukup (sedang).
00,40 – 0,70	Good	Butir item yang bersangkutan telah memiliki daya pembeda yang baik.
0,70 – 1,00	Excellent	Butir item yang bersangkutan telah memiliki daya pembeda yang baik sekali.
Bertanda negatif	-	Butir item yang bersangkutan daya pembedanya negatif (jelek sekali).

Berdasarkan kriteria di atas, nilai DP = 0,73 termasuk soal yang sangat baik. Ini berarti bahwa soal tersebut dapat membedakan antara peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum/kurang menguasai kompetensi.

C. ANALISIS PENGECOH

Pada soal bentuk pilihan-ganda ada alternatif jawaban (opsi) yang merupakan pengecoh. Butir soal yang baik, pengecohnya akan dipilih secara merata oleh peserta didik yang menjawab salah. Sebaliknya, butir soal yang kurang, pengecohnya akan dipilih secara tidak merata.

Pengecoh dianggap baik bila jumlah peserta didik yang memilih pengecoh itu sama atau mendekati jumlah ideal. Indeks pengecoh dihitung dengan rumus.

$$IP = \frac{P}{(N-B)(n-1)} \times 100\%$$

Keterangan :

- IP = indeks pengecoh
- P = jumlah peserta didik yang memilih pengecoh
- N = jumlah peserta didik yang ikut tes
- B = jumlah peserta didik yang menjawab benar pada setiap soal
- n = jumlah alternatif jawaban (opsi)
- 1 = bilangan tetap

Catatan: Jika semua peserta didik menjawab benar pada butir soal tertentu (sesuai kunci jawaban), maka IP = 0 yang berarti soal tersebut jelek. Dengan demikian, pengecoh tidak berfungsi.

Contoh: 50 orang peserta didik di tes dengan 10 soal bentuk pilihan-ganda. Tiap soal memiliki 5 alternatif jawaban (a, b, c, d dan e). Kunci jawaban (jawaban yang benar) soal nomor 8 adalah c. Setelah soal nomor 8 diperiksa untuk semua peserta didik, ternyata dari 50 orang peserta didik, 20 peserta didik menjawab benar dan 30 peserta didik menjawab salah. Idealnya, pengecoh dipilih secara merata, artinya semua pengecoh secara merata ikut menyesatkan peserta didik. Perhatikan contoh soal nomor 8 berikut ini:

Alternatif jawaban	a b c d e
Distribusi jawaban peserta didik	7 8 20 7 8
IP	93% 107% ** 93% 107%
Kualitas pengecoh	++ ++ ** ++ ++

Keterangan:

- ** : kunci jawaban
- ++ : sangat baik
- + : baik

- : kurang baik
- _ : jelek
- : sangat jelek

Pada contoh di atas, IP butir a, b, d, dan e adalah 93%, 107%, 93% dan 107%. Semuanya dekat dengan angka 100%, sehingga digolongkan sangat baik, sebab semua pengecoh itu berfungsi. Jika pilihan jawaban peserta didik menumpuk pada satu alternatif jawaban, misalnya seperti berikut:

Alternatif jawaban	a	b	c	d	e	
Distribusi jawaban peserta didik	20	2	20	8	0	
IP	267%	27%	**	107%		0%
Kualitas pengecoh	_	-**	++	_		

Dengan demikian, dapat ditafsirkan pengecoh (d) yang terbaik, pengecoh (e) dan (b) tidak berfungsi, pengecoh (a) menyesatkan, maka pengecoh (a) dan (e) perlu diganti karena termasuk jelek, dan pengecoh (b) perlu direvisi karena kurang baik. Adapun kualitas pengecoh berdasar indeks pengecoh adalah:

Sangat baik IP	= 76% - 125%
Baik IP	= 51% - 75% atau 126% - 150%
Kurang baik IP	= 26% - 50% atau 151% - 175%
Jelek IP	= 0% - 25% atau 176% - 200%
Sangat jelek IP	= lebih dari 200%

Untuk analisis pengecoh perlu dibuat tabel khusus agar setiap butir soal diketahui berapa banyak peserta didik yang menjawab a, b, c dan seterusnya. Hal ini tentu saja sangat memakan waktu dan tenaga. Jika diolah dengan komputer dan data sudah dimasukkan dalam disket, pengolahan ini hanya memerlukan waktu beberapa detik saja.

D. EFEKTIFITAS FUNGSI OPSI

Setelah tingkat kesukaran soal, daya pembeda, homogenitas dan analisis pengecoh dihitung, selanjutnya kita perlu mengetahui pula apakah suatu opsi (alternatif jawaban) dari setiap soal berfungsi secara efektif atau tidak. Untuk itu, kita dapat menggunakan langkah-langkah sebagai berikut:

1. Menentukan jumlah peserta didik (N).
2. Menentukan jumlah sampel (n), baik untuk kelompok atas maupun kelompok bawah, yaitu $27\% \times N$
3. Membuat tabel pengujian efektifitas opsi seperti berikut:

Tabel 9.1. Pengujian Efektifitas

Ops Kelompok	A	B	C	d	e
Atas					
Bawah					

4. Menghitung jumlah alternatif jawaban yang dipilih peserta didik, baik untuk kelompok atas maupun kelompok bawah.

5. Menentukan efektifitas fungsi opsi dengan kriteria:

1) Untuk opsi kunci:

a) Jumlah pemilih kelompok atas dan kelompok bawah berada diantara 25% - 75%.

Rumusnya adalah $\frac{\Sigma PKA + \Sigma PKB}{n1 + n2} \times 100\%$

Keterangan :

ΣPKA = jumlah pemilih kelompok atas

ΣPKB = jumlah pemilih kelompok bawah

n1 = jumlah sampel kelompok atas (27 %)

n2 = jumlah sampel kelompok bawah (27 %)

b) Jumlah pemilih kelompok atas harus lebih besar daripada jumlah pemilih kelompok bawah.

2) Untuk opsi pengecoh

a) Jumlah pemilih kelompok atas dan kelompok bawah tidak kurang dari:

$25\% \times \frac{1}{2(\Sigma d)} \times (Ka + Kb)$

Keterangan:

d = jumlah opsi pengecoh

Ka = kelompok atas

Kb = kelompok bawah

b) Jumlah pemilih kelompok bawah harus lebih besar daripada jumlah pemilih kelompok atas.

Contoh: Diketahui.

Jumlah peserta didik (N) = 40 orang

n (27% x 40) = 10,80 = 11 (dibulatkan)

Jumlah soal = 10.

Bentuk soal = pilihan-ganda.

Jumlah opsi = 5 (a, b, c, d, e)

Kunci jawaban (opsi kunci) soal nomor 1 (misalnya) adalah (c) dan opsi pengecohnya adalah (a), (b), (d), dan (e).

Distribusi pilihan peserta didik terhadap opsi untuk kelompok atas adalah: opsi (a) = 0; opsi (b) = 1; opsi (c) = 7; opsi (d) = 3; opsi (e) = 0.

Distribusi pilihan peserta didik terhadap opsi untuk kelompok bawah adalah: opsi (a) = 2; opsi (b) = 6; opsi (c) = 2; opsi (d) = 1; opsi (e) = 0.

Dengan demikian, kita dapat membuat tabel distribusi seperti berikut:

Opsi Kelompok	a	B	C	d	e
Atas	0	1	7	3	0
Bawah	2	6	2	1	0

Berdasarkan tabel di atas, kita dapat menentukan efektif tidaknya fungsi opsi sebagai berikut:

- ✓ Untuk opsi (c) sebagai opsi kunci berfungsi efektif, karena jumlah pemilih kelompok atas dan kelompok bawah adalah $7 + 2 = 9$. $9 \times 100\% = 40,91\%$. Angka ini berada diantara $25\% - 75\%$. Di samping itu, jumlah pemilih kelompok atas (7 orang) lebih besar daripada jumlah pemilih kelompok bawah (2 orang).
- ✓ Untuk opsi (a) sebagai opsi pengecoh berfungsi efektif, karena jumlah pemilih kelompok atas dan kelompok bawah 2 orang. Jumlah ini di atas minimal dari: $24\% \times \frac{1}{2 \times 4} \times (11 + 11) = 25\% \times 2,75 = 0,69$. Di samping itu, jumlah pemilih kelompok bawah (2 orang) lebih besar daripada jumlah pemilih kelompok atas (tidak ada pemilih).
- ✓ Untuk opsi (b) sebagai opsi pengecoh berfungsi efektif, karena jumlah pemilih kelompok atas dan kelompok bawah 7 orang. Jumlah ini di atas minimal dari: $25\% \times \frac{1}{2 \times 4} \times 22 = 0,69$. Di samping itu, jumlah pemilih kelompok bawah (6 orang) lebih besar daripada jumlah pemilih kelompok atas (1 orang).
- ✓ Untuk opsi (d) sebagai opsi pengecoh tidak berfungsi secara efektif, karena jumlah pemilih kelompok atas (3 orang) lebih besar daripada jumlah pemilih kelompok bawah (1 orang).
- ✓ Untuk opsi (e) sebagai opsi pengecoh tidak berfungsi secara efektif, karena jumlah pemilih kelompok atas dan kelompok bawah kurang dari 0,69.

BAB X

TEKNIK PENENTUAN NILAI AKHIR DAN PENYUSUNAN RANKING

A. PENTINGNYA MEMANFAATKAN HASIL EVALUASI

Salah satu manfaat hasil evaluasi adalah untuk memberikan umpan balik (*feed-back*) kepada semua pihak yang terlibat dalam pembelajaran, baik secara langsung maupun tidak langsung. Menurut QCA (2003) "*feedback is the mean by which teachers enable children to close the gap in order to take learning forward and improve children's performance*". Umpan balik dapat dijadikan sebagai alat bagi guru untuk membantu peserta didik agar kegiatan belajarnya menjadi lebih baik dan meningkatkan kinerjanya. Peserta didik akan dapat mengukur sejauhmana tingkat penguasaannya terhadap materi, jika hasil pekerjaan mereka mendapat umpan balik dari gurunya. Umpan balik tersebut dapat dilakukan secara langsung, tertulis atau demonstrasi. Dalam memberikan umpan balik, guru hendaknya memperhatikan kualitas pekerjaan peserta didik dan tidak membandingkannya dengan hasil pekerjaan peserta didik yang lain. Hal ini dapat membuat perasaan minder bagi peserta didik yang memiliki kemampuan kurang. Umpan balik sifatnya memberikan saran dan perbaikan, sehingga peserta didik termotivasi untuk memperbaiki dan meningkatkan kualitas proses belajar serta hasil pekerjaannya.

Crooks (2001) menyimpulkan agar umpan balik dapat bermanfaat untuk memotivasi peserta didik, maka harus difokuskan pada:

1. Kualitas pekerjaan peserta didik dan bukan membandingkannya dengan hasil pekerjaan peserta didik yang lain.
2. Cara-cara yang spesifik dimana pekerjaan peserta didik dapat ditingkatkan.
3. Peningkatan pekerjaan peserta didik yang harus dibandingkan dengan pekerjaan sebelumnya.

Apa yang dikemukakan Crooks lebih menekankan kepada kualitas pekerjaan peserta didik sebagai faktor utama dalam menentukan jenis umpan balik yang diberikan, bukan membandingkan hasil pekerjaan peserta didik yang satu dengan lainnya, apalagi membandingkan dengan pekerjaan peserta didik yang mayoritas jawabannya benar atau sesuai dengan kriteria yang telah ditetapkan sebelumnya. Sehubungan dengan hal tersebut, Clarke (2003) menyarankan enam prinsip yang harus diperhatikan, yaitu (1) umpan balik harus fokus pada tugas-tugas yang sesuai dengan tujuan pembelajaran, dan bukan membandingkan dengan anak yang lain (2) guru menggunakan bahasa verbal dan non-verbal serta

memberikan pesan yang baik pada peserta didik tentang kemampuan yang mereka peroleh (3) penilaian setiap bagian pekerjaan dapat mengarah pada penurunan moral bagi peserta didik yang mencapai prestasi rendah dan kepuasan bagi peserta didik yang mencapai prestasi tinggi (4) penghargaan eksternal sama seperti *grades* (5) perlu memberikan umpan balik spesifik yang terfokus kepada kesuksesan dan peningkatan daripada mengoreksi, dan (6) peserta didik perlu diberi kesempatan untuk meningkatkan kualitas pekerjaan mereka.

Selanjutnya, Remmer (1967) mengemukakan *“we discuss here the use of test results to help students understand them selves better, explain pupil growth and development to parents and assist the teacher in planning instruction”*. Pendapat Remmers ini menunjukkan, paling tidak ada tiga manfaat penting dari hasil evaluasi, yaitu untuk membantu pemahaman peserta didik menjadi lebih baik, untuk menjelaskan pertumbuhan dan perkembangan peserta didik kepada orang tua, dan membantu guru dalam membuat perencanaan pembelajaran. Berikut akan dijelaskan beberapa manfaat hasil evaluasi dalam hubungannya dengan pembelajaran.

1. Untuk Perbaikan Rencana Pelaksanaan Pembelajaran

Perencanaan merupakan bagian penting sekaligus menjadi pedoman dan panduan bagi guru dalam melaksanakan kegiatan pembelajaran. Hasil belajar akan dipengaruhi oleh berbagai faktor diantaranya perencanaan pembelajaran. Perencanaan pembelajaran yang baik tidaklah cukup karena memerlukan kesungguhan dalam mengorganisasi rencana tersebut, melaksanakan kegiatan pembelajaran, dan melakukan evaluasi pembelajaran. Sebagaimana Anda ketahui bahwa rencana pelaksanaan pembelajaran merupakan bagian integral dari tugas guru sebagai pelaksana dan pengelola pembelajaran. Oleh sebab itu, seorang guru harus mampu menganalisis bagian mana dari rencana pelaksanaan pembelajaran yang masih perlu dilakukan perbaikan. Harapannya adalah melalui perbaikan maka kualitas proses pembelajaran akan menjadi lebih baik.

Rencana pelaksanaan pembelajaran berperan sebagai skenario proses pembelajaran, karena itu harus disusun secara fleksibel dan membuka kemungkinan bagi guru untuk menyesuaikan diri dengan kondisi yang ada dalam proses pembelajaran. Dalam penyusunannya, sebaiknya dilakukan sendiri oleh guru agar guru senantiasa sadar dan paham apa yang harus disampaikan kepada peserta didik, bagaimana cara menyampaikan materi pelajaran, bahan dan alat apa yang diperlukan, dan kemana peserta didik akan diarahkan. Untuk itu, guru perlu memahami komponen-komponen rencana pembelajaran dan

bagaimana langkah-langkah pembelajaran yang akan ditempuh. Komponen rencana pelaksanaan pembelajaran biasanya terdiri atas identitas mata pelajaran, standar kompetensi dan kompetensi dasar, indikator hasil belajar, materi pembelajaran, langkah-langkah pembelajaran, media pembelajaran, penilaian, dan sumber bacaan.

Kesalahan dalam pembelajaran dapat dilihat dari perencanaan, pengorganisasian, pelaksanaan, sampai dengan evaluasi pembelajaran. Kekeliruan pada tahap perencanaan akan mengakibatkan kekeliruan dalam melaksanakan pembelajaran. Jika kesalahan atau kekeliruan sudah dapat ditemukan, maka guru dapat mencari solusi untuk memperbaikinya. Memperbaiki rencana pelaksanaan pembelajaran dapat dilakukan dengan jalan memeriksa kelengkapan komponen-komponennya, kesesuaian antar komponen, kemungkinan proses pelaksanaan, operasional tidaknya indikator, kesesuaian indikator dengan kompetensi dasar, ketepatan dalam memilih metoda, media dan sumber belajar, serta ketepatan teknik penilaian yang digunakan. Merancang perbaikan rencana pelaksanaan pembelajaran berarti melakukan perbaikan dari rencana sebelumnya.

2. Untuk Mengoptimalkan Proses Pembelajaran

Optimalisasi proses pembelajaran adalah upaya memperbaiki proses pembelajaran sehingga peserta didik mencapai keberhasilan proses dan hasil belajar. Tujuannya adalah untuk memperbaiki aspek-aspek pembelajaran yang dianggap masih kurang optimal. Semua guru tentu berharap bahwa proses pembelajaran dapat berjalan secara optimal. Persoalannya adalah bagaimana caranya mengoptimalkan proses pembelajaran? Hal ini dapat dilakukan melalui (a) evaluasi diri secara jujur dan teliti terhadap semua aspek pembelajaran (b) identifikasi faktor-faktor penyebab kegagalan dan keberhasilan pembelajaran. Melalui cara ini, guru dapat menindaklanjuti upaya-upaya memantapkan keberhasilan (pengayaan) dan upaya-upaya memperbaiki kegagalan (remidi). Artinya, melalui pengayaan dan remedial diharapkan proses pembelajaran menjadi optimal.

Secara umum, dapat dikatakan bahwa pemanfaatan hasil evaluasi berkaitan erat dengan tujuan menyelenggarakan evaluasi itu sendiri. Sebagaimana ditegaskan Julian C. Stanley (1964: 321) "*just what is to be done, of course, depends on the purpose of the program*". Artinya, pemanfaatan hasil evaluasi sangat bergantung kepada tujuan evaluasi. Misalnya, tujuan evaluasi formatif adalah untuk memperbaiki kegiatan pembelajaran. Perbaikan tersebut dilakukan atas dasar hasil monitoring kemajuan belajar peserta didik. Hasil monitoring ini seharusnya dapat dimanfaatkan sebagai *feedback*, baik bagi guru maupun bagi peserta

didik untuk mengetahui materi mana yang belum dikuasai. Hasil evaluasi formatif dapat dimanfaatkan untuk mengulangi pelajaran, memperbaiki strategi pembelajaran, atau melanjutkan pelajaran.

Begitu juga dengan evaluasi sumatif yang bertujuan untuk memberikan nilai sebagai dasar menentukan kenaikan kelas atau kelulusan peserta didik dan pemberian sertifikat bagi peserta didik yang telah menyelesaikan pelajaran dengan baik. Hasil evaluasi sumatif dapat dimanfaatkan untuk kenaikan kelas atau kelulusan peserta didik. Menurut Oemar Hamalik (1989: 177) “penggunaan hasil evaluasi dapat mengacu kepada fungsi evaluasi itu sendiri, yaitu fungsi instruksional, fungsi administratif, dan fungsi bimbingan”. Dalam konteks fungsi instruksional, guru dapat menggunakan hasil evaluasi untuk memperbaiki sistem pembelajaran. Begitu juga dalam fungsi administratif, guru dapat membuat laporan dan menetapkan kedudukan peserta didik dalam kelas. Sedangkan dalam fungsi bimbingan, guru dapat memberikan bimbingan kepada peserta didik yang mengalami kesulitan belajar.

B. MANFAAT HASIL EVALUASI

Dalam praktiknya, masih banyak guru yang tidak atau kurang memahami pemanfaatan hasil evaluasi, sehingga hasil evaluasi formatif atau sumatif (misalnya) banyak dimanfaatkan hanya untuk menentukan kenaikan kelas dan mengisi buku rapor. Namun demikian, untuk melihat pemanfaatan hasil evaluasi ini secara komprehensif, kita dapat meninjaunya dari berbagai pihak yang berkepentingan, yaitu:

1. Bagi peserta didik, hasil evaluasi dapat dimanfaatkan untuk:
 - a. Membangkitkan minat dan motivasi belajar.
 - b. Membentuk sikap yang positif terhadap belajar dan pembelajaran.
 - c. Membantu pemahaman peserta didik menjadi lebih baik.
 - d. Membantu peserta didik dalam memilih metoda belajar yang baik dan benar.
 - e. Mengetahui kedudukan peserta didik dalam kelas
2. Bagi guru, hasil evaluasi dapat dimanfaatkan untuk:
 - a. Promosi peserta didik, seperti kenaikan kelas atau kelulusan.
 - b. Mendiagnosis peserta didik yang memiliki kelemahan atau kekurangan, baik secara perorangan maupun kelompok.
 - c. Menentukan pengelompokan dan penempatan peserta didik berdasarkan prestasi masing-masing.
 - d. *Feedback* dalam melakukan perbaikan terhadap sistem pembelajaran.
 - e. Menyusun laporan kepada orang tua guna menjelaskan pertumbuhan dan perkembangan peserta didik.

- f. Dijadikan dasar pertimbangan dalam membuat perencanaan pembelajaran.
 - g. Menentukan perlu tidaknya pembelajaran remedial.
3. Bagi orang tua, hasil evaluasi dapat dimanfaatkan untuk:
- a. Mengetahui kemajuan belajar peserta didik.
 - b. Membimbing kegiatan belajar peserta didik di rumah.
 - c. Menentukan tindak lanjut pendidikan yang sesuai dengan kemampuan anaknya.
 - d. Memprakirakan kemungkinan berhasil tidaknya anak tersebut dalam bidang pekerjaannya.
4. Bagi administrator madrasah, hasil evaluasi dapat dimanfaatkan untuk:
- a. Menentukan penempatan peserta didik.
 - b. Menentukan kenaikan kelas.
 - c. Pengelompokan peserta didik di madrasah mengingat terbatasnya fasilitas pendidikan yang tersedia serta indikasi kemajuan peserta didik pada waktu mendatang.

Hasil evaluasi dapat dimanfaatkan oleh peserta didik untuk membangkitkan minat dan motivasi belajar. Hal ini dapat dilakukan jika peserta didik mengetahui hasil evaluasi yang dicapainya, mengetahui kesalahan-kesalahannya dan bagaimana solusinya. Di samping itu, hasil evaluasi dapat membentuk sikap positif peserta didik terhadap mata pelajaran, termasuk juga terhadap guru, proses pembelajaran, lingkungan dan evaluasi pembelajaran, bahkan dapat membantu pemahaman peserta didik menjadi lebih baik. Banyak hasil penelitian yang menunjukkan adanya hubungan yang positif dan signifikan antara minat, motivasi dan sikap peserta didik terhadap hasil belajarnya. Oleh sebab itu, hasil evaluasi perlu dimanfaatkan oleh peserta didik untuk mengembangkan sikap, minat, dan motivasi belajar.

Hasil evaluasi bisa juga dimanfaatkan peserta didik untuk memilih teknik belajar yang tepat dan benar. Seperti kita ketahui, banyak sekali teknik atau cara belajar peserta didik yang digunakan selama ini kurang baik. Misalnya, melakukan kegiatan belajar jika besok mau ujian, mengerjakan tugas atau latihan dengan "copy-paste", dan sebagainya. Teknik belajar yang salah ini dapat merugikan peserta didik, baik fisik maupun mental yang pada gilirannya dapat mempengaruhi prestasi belajar. Teknik belajar ini tentu tidak bisa dibiarkan begitu saja. Guru dan orang tua harus ikut bertanggung jawab setiap tindakan belajar peserta didik, karena itu harus cepat mengambil tindakan pencegahan, seperti memberitahukan hasil evaluasi kepada peserta didik dan orang tua membimbing teknik belajar di rumah. Dengan kata lain, jika hasil evaluasi

peserta didik kurang baik dan mereka mengetahuinya, tentu diharapkan mereka dapat memperbaiki teknik belajarnya.

Manfaat hasil evaluasi yang lain adalah untuk menentukan kedudukan belajar dalam kelas. Kedudukan belajar ini dapat dilihat secara kelompok maupun perorangan. Secara kelompok, maksudnya guru melihat kedudukan peserta didik secara kelompok melalui perhitungan rata-rata kelompok dan membandingkan antara kelompok satu dengan kelompok yang lain. Secara perorangan, maksudnya guru melihat kedudukan belajar melalui perhitungan prestasi belajar peserta didik secara perorangan. Guru juga dapat membandingkan antara prestasi belajar seorang peserta didik dengan peserta didik lainnya dalam satu kelompok.

Dalam rangka promosi peserta didik, baik untuk menentukan kenaikan kelas atau kelulusan, guru dapat memanfaatkan hasil evaluasi terutama hasil evaluasi sumatif. Guru dapat menafsirkan dan memutuskan sejauhmana taraf kesiapan peserta didik dapat melanjutkan ke kelas atau ke jenjang pendidikan berikutnya sesuai dengan kemampuan peserta didik masing-masing. Artinya, jika penafsiran guru peserta didik dianggap sudah siap, maka peserta didik dapat melanjutkan ke kelas atau jenjang pendidikan berikutnya. Sebaliknya, jika penafsiran guru ternyata peserta didik belum siap, maka peserta didik harus mengulang lagi di kelas semula. Hasil evaluasi dapat juga digunakan guru untuk mendiagnosis peserta didik yang memiliki kelemahan atau kekurangan, baik secara perorangan maupun kelompok. Berdasarkan kelemahan-kelemahan ini, maka guru harus mencari faktor-faktor penyebabnya, antara lain: dari sistem evaluasi itu sendiri, materi pelajaran, kemampuan guru, kemampuan peserta didik, perencanaan dan pelaksanaan pembelajaran serta lingkungan sekolah.

Hasil evaluasi dapat juga dimanfaatkan guru untuk menentukan pengelompokan dan penempatan peserta didik berdasarkan prestasi masing-masing. Pengelompokan tersebut didasarkan atas penafsiran guru terhadap susunan kelompok. Jika kelompok atau kelas yang dihadapi mempunyai susunan yang normal dan homogen, maka kelas tersebut tidak perlu dibagi-bagi dalam kelompok. Sebaliknya, jika kelas tersebut heterogen maka kelas itu perlu dibagi dalam beberapa kelompok berdasarkan prestasi masing-masing. Selanjutnya, hasil evaluasi dapat dijadikan *feedback* bagi guru dalam melakukan perbaikan terhadap sistem pembelajaran. Jika prestasi belajar peserta didik kurang baik, pada umumnya guru "menyalahkan" peserta didiknya, tetapi jika prestasi belajar peserta didik baik atau memuaskan maka guru akan mengatakan itu merupakan hasil dari perbuatan mengajarnya. Pernyataan ini tentu

tidak dapat disalahkan atau dibenarkan karena banyak faktor yang mempengaruhi prestasi belajar peserta didik. Faktor tersebut dapat timbul dari guru atau peserta didik itu sendiri.

Hasil evaluasi dapat dimanfaatkan guru untuk menyusun laporan kepada orang tua guna menjelaskan pertumbuhan dan perkembangan peserta didik. Hal ini dimaksudkan agar orang tua mengetahui kemajuan dan prestasi yang dicapai oleh siswa. Secara tegas, Oemar Hamalik (1989: 179) menjelaskan “tujuan dari pelaporan adalah untuk mengikhtisarkan, mengorganisasi dan menafsirkan hasil tes sehingga dapat memberikan gambaran tentang status dan kemajuan perorangan siswa, kelas dan sekolah”. Bentuk laporan dapat dilakukan secara tertulis maupun lisan. Laporan lisan dimaksudkan agar terjadi komunikasi secara efektif antara sekolah dengan pihak yang menerima laporan, dan juga membentuk hubungan emosional yang lebih kental antara kedua belah pihak. Sedangkan laporan tertulis dimaksudkan agar dapat memberikan petunjuk yang permanen. Laporan tertulis dapat didokumentasikan dan pada waktunya dapat dijadikan data untuk dianalisis lebih lanjut.

Dalam perencanaan pembelajaran, kita dapat memanfaatkan hasil evaluasi, seperti merumuskan kompetensi dasar dan indikator, menyusun tingkat kesulitan materi, menentukan strategi pembelajaran yang tepat, dan mengembangkan alat evaluasi yang akurat. Jika kompetensi dasar sudah dirumuskan dalam silabus, berarti guru tinggal merumuskan indikator. Sebagaimana kita ketahui bahwa indikator harus dirumuskan dengan mengacu kepada kompetensi dasar dan harus sesuai dengan sub pokok bahasan atau sub topik. Seringkali guru merumuskan indikator yang salah, karena tidak menggunakan kata kerja operasional, yaitu kata kerja yang spesifik, dapat diukur dan dapat diamati. Jika tidak maka guru akan kesulitan menentukan langkah-langkah pembelajaran berikutnya termasuk menyusun alat evaluasi itu sendiri.

Manfaat hasil evaluasi berikutnya adalah menentukan perlu tidaknya pembelajaran remedial. Sehubungan dengan hal tersebut, maka guru harus melakukan penafsiran terhadap prestasi kelompok. Misalnya, materi pelajaran dapat dilanjutkan jika seluruh peserta didik menguasai minimal 80 % materi yang telah disampaikan. Sebaliknya, jika kurang dari standar minimal tersebut, maka materi pelajaran harus diulang. Pengulangan suatu pelajaran dapat juga dilihat dari hasil penafsiran prestasi kelompok. Jika prestasi kelompok dianggap sudah mencapai prestasi yang baik, maka materi pelajaran tidak perlu diulang, tetapi bila prestasi kelompok dianggap masih kurang, maka materi pelajaran perlu diulang.

Bagi orang tua, hasil evaluasi dapat dimanfaatkan untuk mengetahui kemajuan belajar peserta didik. Sebagai orang tua tentu berharap agar putra-putrinya berhasil. Untuk itu, orang tua harus mengetahui perkembangan kemajuan belajar anaknya, baik fisik maupun mental terutama berkaitan dengan prestasi belajar. Hal ini penting terutama apabila ada diantara peserta didik yang memperoleh prestasi belajar kurang memuaskan. Orang tua dapat menentukan langkah-langkah apa yang harus ditempuh untuk memajukan prestasi belajar anaknya. Orang tua juga dapat membimbing kegiatan belajar peserta didik di rumah. Jika tidak mampu, orang tua dapat menyuruh anaknya mengikuti bimbingan di luar atau juga mendatangkan guru ahli ke rumah. Berdasarkan hasil evaluasi, orang tua dapat menentukan tindak lanjut pendidikan yang sesuai dengan kemampuan anaknya. Jangan sampai orang tua memaksakan anaknya masuk ke sekolah favorit sedangkan kemampuan anaknya tidak mendukung. Akibatnya, anak akan menjadi malas, minder, dan acuh terhadap pelajaran. Orang tua juga dapat memprakirakan kemungkinan berhasil tidaknya anak dalam bidang pekerjaannya.

Bagi administrator madrasah, hasil evaluasi dapat dimanfaatkan untuk menentukan penempatan peserta didik sesuai dengan kemampuannya, untuk menentukan kenaikan kelas, dan pengelompokan peserta didik di madrasah mengingat terbatasnya fasilitas pendidikan yang tersedia serta indikasi kemajuan peserta didik pada waktu mendatang. Hasil evaluasi tidak hanya dimanfaatkan untuk laporan ke berbagai pihak, tetapi juga untuk memotivasi dan menghargai peserta didik itu sendiri, baik dalam rangka promosi maupun melihat kelebihan dan kekurangannya, bukan sebaliknya untuk menakut-nakuti peserta didik atau untuk menjatuhkan mentalnya. Diadakannya evaluasi dalam proses pengembangan sistem pembelajaran dimaksudkan untuk keperluan perbaikan sistem, pertanggungjawaban kepada pemerintah dan masyarakat, serta penentuan tindak lanjut hasil pengembangan.

Hasil evaluasi dapat dimanfaatkan guru untuk membuat kelompok sesuai dengan prestasi yang diperoleh peserta didik. Bahkan sekarang ini sudah banyak sekolah yang membentuk "kelas unggulan", yaitu kelompok peserta didik yang mempunyai prestasi belajar di atas rata-rata. Biasanya diambil dari peserta didik yang memperoleh peringkat 10 terbesar. Padahal dalam kenyataannya, kelompok ini mendapat perlakuan biasa saja, bahkan cenderung sama dengan kelas lainnya, baik yang menyangkut materi pelajaran, pemberian tugas, latihan, pengembangan kreatifitas atau diskusi kelompok. Sebaliknya, peserta didik yang memperoleh prestasi di bawah rata-rata kelompok justru kurang

mendapat perhatian lebih. Seharusnya, peserta didik kelompok ini yang mendapat perhatian serius, seperti memberi bimbingan belajar, latihan atau tugas-tugas khusus, bahkan bagi peserta didik yang tergolong “sangat kurang” harus dilakukan diagnosis atau identifikasi tentang faktor-faktor penyebabnya. Hasil diagnosis atau identifikasi tersebut dapat dijadikan bahan pertimbangan bagi guru untuk memperbaiki kegiatan pembelajaran berikutnya.

Di lain pihak, memang guru-guru tidak dapat disalahkan sepenuhnya, karena mungkin saja ada guru yang mau bahkan sudah menawarkan bimbingan belajar kepada peserta didik tersebut, tetapi kurang mendapat respon yang baik dari orang tua, mengingat ekonomi orang tua peserta didik termasuk “pas-pasan”, sehingga permasalahan tersebut tidak pernah memperoleh solusi yang terbaik. Padahal, untuk memberikan bimbingan belajar secara khusus, waktu bagi guru sangat sulit, karena begitu padatnya materi yang harus disampaikan sesuai dengan target dan tugas-tugas rutin yang harus juga diselesaikan. Jika guru ingin memberikan bimbingan, maka ia harus mencari waktu di luar jam pelajaran, baik sebelum atau sesudah jam pelajaran.

Bagi guru, faktor waktu sangat penting. Artinya, apakah waktu yang tersedia masih memungkinkan guru untuk memberikan bimbingan atau tidak. Jika hanya mengandalkan jam dinas, maka guru sulit mencari waktu untuk memberikan bimbingan apalagi melakukan diagnosis. Hasil evaluasi dapat juga digunakan untuk kepentingan yang lain. Maksudnya, andaikan diagnosis dan bimbingan ini memang sulit dilaksanakan, maka tentu guru harus mencari alternatif penggunaan hasil evaluasi yang lain, seperti: perlu tidaknya materi pelajaran diulang, perbaikan proses pembelajaran dan sebagainya.

C. FAKTOR-FAKTOR YANG TURUT DIPERHITUNGGAN DALAM PENILAIAN

Walaupun hal yang dinilai tidak sama bagi setiap sekolah, namun secara garis besar dapat ditentukan unsur umum dalam penilaian yang menyangkut faktor-faktor yang harus dipertimbangkan. Unsur umum tersebut adalah sebagai berikut:

1. Prestasi/pencapaian (*achievement*)

Nilai prestasi harus mencerminkan tingkatan-tingkatan siswa sejauh mana telah dapat mencapai tujuan yang ditetapkan di setiap bidang studi. Simbol yang digunakan untuk menyatakan nilai, baik huruf maupun angka, hendaknya hanya merupakan gambaran tentang prestasi saja. Unsur pertimbangan atau kebijaksanaan guru tentang

usaha dan tingkah laku siswa tidak boleh ikut berbicara pada nilai tersebut.

2. Usaha (*effort*)

Terpisah dan nilai prestasi, guru dapat menyampaikan laporannya kepada orang tua siswa. Laporan atau nilai tidak boleh dicampuri dengan nilai prestasi sama sekali. Yang sering terjadi adalah kecenderungan dari guru untuk menilai unsur usaha ini lebih rendah bagi anak yang restasinya rendah dan sebaliknya.

3. Aspek pribadi dan sosial (*personal and social characteristics*)

Unsur ini juga perlu dilaporkan terutama yang berhubungan dengan berlangsungnya proses belajar mengajar, misalnya, mentaati tata tertib sekolah. Dalam memberikan nilai pribadi ini harus hati-hati sekali. Rentangan nilai sebaiknya tidak usah lebar-lebar (lebih baik 6-10). Lebih baik lagi jika diterangkan dengan khusus dan jelas sehingga mudah dimengerti oleh guru pembimbing dan siapa saja.

4. Kebiasaan bekerja (*working habits*)

Yang dimaksud di sini adalah hal-hal yang berhubungan dengan kebiasaan melakukan tugas. Misalnya: segera mengerjakan PR, keuletan dalam usaha, beekerja teliti, kerapihan kerja, dan sebagainya.

D. TEKNIK PENENTUAN NILAI AKHIR

Sebelum dibicarakan lebih lanjut mengenai cara-cara yang dapat ditempuh dalam rangka menentukan nilai akhir perlu kiranya diingatkan lagi tentang adanya dua bentuk penilaian, yaitu: penilaian dalam bentuk tes formatif dan penilaiann dalam bentuk tes sumatif.

Penilaian yang diberikan oleh pendidik dalam bentuk tes-tes formatif sebenarnya dimaksudkan untuk memperbaiki proses belajar mengajar dan untuk mengetahui sampai sejauh mana tingkat pencapaian peserta didik terhadap tujuan instruksional yang telah dirumuskan dalam setiap satuan pelajaran. Adapun tes sumatif bertujuan untuk menilai prestasi peserta didik terhadap penguasaan bahan pelajaran yang telah diberikan kepada mereka selama jangka waktu tertentu. Akan tetapi oleh karena tes sumatif itu pada umumnya tidak sering dilakukan, maka untuk dapat menjaga kesinambungan penilaian dan hasil penilaian yang dipandang lebih mantap bagi setiap peserta didik, maka penentuan nilai akhir pada umumnya dilaksanakan dengan jalan menggabungkan nilai-niali hasil tes formatif dengan nilai hasil tes sumatif.

Dlam pelaksanaanya, dicarillah nilai-nilai rata-rata hitung dari nilai-nilai hasil tes formatif dan nilai-nilai hasil tes sumatif; nilai-nilai mana sebelum dicari rata-rata hitungnya terlebih dahulu diubah atau dikonversikan ke dalam nilai standar berskala sepuluh.

Penentuan nilai akhir pada umumnya dilakukan pada saat guru akan mengisi buku laporan pendidikan atau mengisi ijazah. Dalam praktek mereka telah dibimbing oleh suatu peraturan atau pedoman yang ditetapkan oleh pihak yang berwenang. Karena itu, dalam praktek kita jumpai berbagai macam cara yang biasanya digunakan oleh guru dalam menentukan nilai akhir tersebut.

Berikut ini dikemukakan tiga macam contoh cara yang sering dipergunakan dalam penentuan nilai akhir.

1. Nilai akhir diperoleh dengan jalan memperhitungkan nilai hasil tes formatif, yaitu nilai rata-rata hasil ulangan harian, dengan nilai hasil tes sumatif, yaitu nilai hasil ulangan semester, dengan menggunakan rumus sebagai berikut:

$$N_A = \frac{\frac{(F_1 + F_2 + F_3 + \dots + F_n)}{n} + 2S}{3}$$

Keterangan:

- NA = Nilai Akhir
- F1 = Nilai hasil tes formatif ke-1
- F2 = Nilai hasil tes formatif ke-2
- F3 = Nilai hasil tes formatif ke-3
- F4 = Nilai hasil tes formatif ke-n
- n = Banyaknya kali tes formatif dilaksanakan
- 2 & 3 = Bilangan konstan (2=bobot tes formatif, 3= bobot tes secara keseluruhan)

Contoh: Tes formatif (ulangan harian) mata pelajaran pendidikan agama dilaksanakan 4 kali dalam satu catur wulan dan ulangan umum bersama (tes sumatif) dilaksanakan 1 kali. A murid Sekolah Dasar kelas V berhasil memperoleh nilai-nilai sebagai berikut:

- Nilai tes formatif I = 8
- Nilai tes formatif II = 7,5
- Nilai hasil tes formatif = 6,5
- Nilai hasil tes formatif IV = 7
- Nili hasil tes sumatif = 8

Dengan demikian nilai akhir yang dapat diberikan kepada A:

$$N_A = \frac{\frac{(8+7,5+6,5+7)}{4} + (2 \times 8)}{3} = 7,75 = 8 \text{ (dibulatkan ke atas)}$$

2. Cara kedua ini dipergunakan untuk keperluan pengisian nilai dalam ijazah atau Surat Tanda Tamat Belajar (STTB). Disini nilai akhir diperoleh dari nilai: nilai rata-rata hasil ulangan harian (H), diberi bobot 1, ditambah dengan nilai Evaluasi Tahap Akhir (EBTA), diberi bobot 2. Jika dituangkan dalam bentuk rumus:

$$N_A = \frac{\sum H}{N} \times 2E$$

Contoh: A, siswa kelas XI SMA, untuk ulangan harian I mendapat nilai 7, ulangan harian II, mendapat nilai 8, ulangan harian II mendapat nilai 9. Sedangkan nilai EBTA = 6. Dengan demikian nilai yang diberikan kepada A adalah:

$$N_A = \frac{(7+8+9) + (2 \times 6)}{3} = 6,666 = 7 \text{ (dibulatkan ke atas)}$$

Dalam pembulatan nilai-nilai yang akan dicantumkan dalam buku rapor atau STTB, umumnya dipergunakan pedoman sebagai berikut:

- Jika dibelakang tanda desimal terdapat bilangan yang lebih kecil dari 50, dianggap = 0 (dibulatkan ke bawah). contoh: nilai 5,43 dibulatkan ke bawah menjadi 5.
- Jika dibelakang tanda desimal terdapat bilangan yang besarnya = 50, maka nilai akhir tidak dibulatkan. Jadi dituliskan apa adanya. contoh: 6,50 tetap dicantumkan 6,5.
- Jika di belakang tanda desimal terdapat bilangan yang lebih besar atau di atas 0,50 dibulatkan ke atas. contoh: nilai 5,75 dibulatkan ke atas menjadi 6.

E. TEKNIK PENYUSUNAN RANKING

Yang dimaksud dengan ranking adalah letak seorang siswa di dalam urutan tingkatan. Untuk dapat diketahui ranking dari siswa-siswa di suatu kelas, maka harus diadakan pengurutan nilai siswa-siswa tersebut dari yang paling atas sampai ke nilai yang paling bawah. Dengan mengurutkan nilai-nilai maka dengan mudah dapat ditentukan nomor yang menunjukkan kedudukan siswa dalam tingkatannya.

Ada bermacam-macam untuk menentukan ranking atau kedudukan siswa dalam kelompoknya. Akan tetapi, didalam uraian ini hanya akan diberikan 4 (empat) cara saja, yaitu:

1. Dengan ranking sederhana (*simple rank*)

Simple Rank adalah urutan yang menunjukkan letak/kedudukan seseorang dalam kelompoknya dan dinyatakan dengan nomor atau angka biasa.

contoh: Skor dari ulangan Bahasa Inggris bagi 20 orang siswa adalah sebagai berikut:

A = 45	F = 70	K = 75	P = 78
B = 50	G = 81	L = 75	Q = 74
C = 39	H = 75	M = 69	R = 65
D = 61	I = 68	N = 60	S = 49
E = 63	J = 46	O = 73	T = 60

Hanya dengan melihat deretan skor yang masih berserakan ini, kita belum dapat menentukan ranking atau kedudukan seseorang dalam kelompoknya. Untuk itu maka skor-skor tersebut terlebih dahulu harus kita susun,urut dari skor yang paling tinggi sampai ke skor yang paling rendah, dengan urutan ke bawah. Setelah itu tentukan urutan nomor dari atas, yaitu nomor 1,2,3,4,5, dan seterusnya sampai seluruh siswa memperoleh nomor. Perlu diingat di sini bahwa apabila ada dua atau tiga orang yang kebetulan memiliki skor yang sama, harus diberi nomor urut atau ranking yang sama pula, yaitu rata-rata dari urutan orang-orang yang memiliki skor sama tersebut.

Untuk memahami bagaimana menentukan simple rank atau ranking sederhana, marilah kita urutkan dahulu skor-skor A sampai dengan T. seperti pada table 10.1.

Tabel 10.1. Simpel Rank Dari 20 Orang Siswa

Nama Siswa	Skor	Rangking
G	81	1
P	78	2
H	75	3
K	75	4
L	75	5
Q	74	6
O	73	7
F	70	8
M	69	9
I	68	10
R	65	11
E	62	12
D	61	13
N	60	14
T	60	15
B	50	16
S	49	17
J	46	18
A	45	19
C	39	20

- a) Siswa yang mempunyai skor sama, juga mempunyai ranking sama, sehingga ada nomor-nomor yang tidak digunakan sebagai nomor urut.
 - b) Rank terakhir selalu sam dengan nomor urut siswa atau banyaknya siswa dalam kelompok, kecuali ada beberapa siswa yang mempunyai persamaan skor.
2. Dengan ranking persentase (*percentile rank*)

Percentile Rank atau ranking persentase adalah kedudukan seseorang dalam kelompok, yang menunjukkan banyaknya persentase yang berada di bawahnya. Jadi dalam hal ini siswa dibandingkan dengann siswa lain yang mempunyai skor atau lebih kecil daripadanya.

Contoh: Jika seorang siswa memiliki PR (Percentile Rank) 85 ini menunjukkan bahwa kecakapan siswa tersebut sama atau melebihi 85% dari seluruh kelompok.

Dengan ranking persentase, lebih diketahui gambaran kecakapan siswa, karena angka ranking menunjukkan besarnya persentase siswa dalam kelompok itu yang berhasil dilampaui.

Cara menentukan Percentile Rank adalah sebagai berikut:

- a) Menentukan dahulu Percentile Rank.
- b) Mencari banyaknya siswa dalam kelompok itu, yang ada di bawahnya.
- c) Mengalikan dengan 100, setelah dibagi dengan kelompok.

Contoh: Dengan kelompok yang terdapat pada “tabel simple rank untuk 20 orang”, siswa F mendudukan ranking 8 dalam simple rank. Maka banyaknya siswa yang ada di bawahnya adalah (20-8) orang atau 12 orang.

$$PR \text{ untuk F adalah } \frac{12}{20} \times 100 = 60$$

Ini berarti bahwa siswa F itu letaknya dalam kelompok mengalahkan sebanyak 60% untuk persentasi yang bersangkutan. Untuk menentukan PR kita tidak boleh menentukan SR terlebih dahulu. Rumus untuk menentukan PR adalah:

$$PR = \frac{N-SR}{N} \times 100$$

3. Standar deviasi.

Yang dimaksud dengan penentuan kedudukan siswa dengan standar deviasi adalah penentuan kedudukan dengan membagi kelas atas kelompok-kelompok. Tiap kelompok di batasi oleh suatu standar deviasi tertentu. Penentuan kedudukan dengan standar deviasi dapat dilakukan dengan 2 cara, yaitu:

- a) Pengelompokan atas 3 ranking

b) Pengelompokan atas 11 ranking

c) Pengelompokan atas 3 ranking

Langkah-langkah dalam menentukan kedudukan siswa dalam 3 ranking

- 1) Menjumlahkan skor semua siswa.
- 2) Mencari nilai rata-rata (Mean) dan simpangan baku (deviasi satandar atau standar deviasi).
- 3) Menentukan batas-batas kelompok.
 - Kelompok atas
Semua siswa yang mempunya skor sebanyak skor rata-rata plus satu standar deviasi ke atas.
 - Kelompok sedang
Semua siswa yang mempunyai skor antara -1 SD dan +SD.
 - Kelompok kurang
Semua siswa yang mempunyai skor -1 SD dan yang kurang dari itu.

Di bawah ini akan diberikan rumus mencari Mean dan standar deviasi, langsung dari angka (skor) kasar.

$$\text{Mean } (\bar{X}) = \frac{\sum X}{N}$$

Jadi untuk mencari nilai rata-rata, tinggal menjumlahkan semua skor, kemudian dibagi dengan banyaknya siswa memiliki skor itu. Rumus standar deviasi:

$$SD = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2}$$

Apabila dilalui tabel ini atau data kelompok, maka digunakan rumus-rumus yang lain:

$$\text{Mean} = \frac{\sum fX}{N}$$

Sedangkan rumus Standar deviasi adalah:

$$SD = \sqrt{\frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2}$$

4. Dengan menggunakan z-score.

Standar Score atau z-Score adalah angka yang menunjukkan perbandingan perbedaan skor seseorang dan mean, dengan standar deviasinya. Standar Score ini lebih mempunyai arti dibandingkan dengan skor itu sendiri karena telah dibandingkan dengan suatu standar yang sama.

a) Rata-rata skor dari kelompok.

b) Standar Deviasi dari skor-skor tersebut

$$\text{Rumus: } z = \frac{X-M}{SD}$$

z = Nilai baku

DAFTAR PUSTAKA

- Anas Sudjiono. (2013) *Pengantar Evalausi Pendidikan*. Jakarta: PT. Raja Grafindo Persada.
- Arifin, Zainal. (2011) *Evaluasi Pembelajaran: Prinsip-Teknik-Prosedur*, Bandung: PT. Remaja Rosdakarya.
- Arifin, Zainal. (2011) *Konsep dan Model Pengembangan Kurikulum*, Bandung: PT. Remaja Rosdakarya.
- Arikunto. Suharsimi. (2012) *Dasar-Dasar Evaluasi Pendidikan*, Jakarta: Bumi Aksara.
- Arikunto, S., dan Jabar, C.S.A, (2007) *Evaluasi Program Pendidikan*, Cetakan ke-2, Jakarta: Bumi Aksara.
- Cronbach, L.E. (1963) *Course Improvement Through Evaluation dalam Educational Evaluation: Theory and Practice* (ed.Worthen, B.R., dan Sanders, J.R.,) California, Belmont: Wadworth Publishing Co.
- Davis, R. H., L.T.Alexander, and S.L.Yelon. (1974) *Learning System Design: An Approach to The Improvement of Instruction*, Michigan State University: McGraw-Hill Book Company.
- Depdiknas. (2003) *Materi Pelatihan Peningkatan Kemampuan Guru Dalam Penyusunan dan Penggunaan Alat Evaluasi Serta Pengembangan Sistem Penghargaan Terhadap Siswa*, Jakarta: Direktorat PLP – Ditjen Dikdasmen.
- Depdiknas. (2002) *Kurikulum Berbasis Kompetensi*, Jakarta: Pusat Kurikulum – Balitbang Depdiknas.
- Dimiyati dan Mudjiono. (1994) *Belajar dan Pembelajaran*, Jakarta: P3MTK-Ditjen Dikti-Depdikbud.
- Ebel, R.L. (1972) *Essentials of Educational Measurement*, Engelwood Cliffs, New Guba, E.G. and Lincoln, Y.S. (1985) *Effective Evaluation*, San Francisco: Jossey – Bass Pub.

- Gronlund, N.E. (1985) *Measurement and Evaluation in Teaching*, Fifth Edition, New York: Mc Millan Publishing Co., Inc.
- Guba, E.G. and Lincoln, Y.S. (1985) *Effective Evaluation*, San Francisco: Jossey – Bass Pub.
- Hamalik, Oemar. (1989) *Teknik Pengukuran dan Evaluasi Pendidikan*. Bandung: Mandar Maju.
- Hasan, S. Hasan, (1988) *Evaluasi Kurikulum*, Jakarta: P2LPTK-Ditjen Dikti-Depdikbud.
- Koran, J.J., E.J.Montague, and G.E.Hall (1969) *Use Behavioral Objectives in Science Instruction*, Austin (Texas): National Science Teachers Association, The University of Texas.
- Mursell, J., dan Nasution, S., (tanpa tahun) *Mengajar dengan Sukses*, Bandung: Jemmars.
- Nitko, A. J., (1996) *Educational Assessment of Students*, Second Edition, New Jersey: Englewood Cliffs.
- Nurkancana, Wayan dan Sunartana (1990) *Evaluasi Hasil Belajar*, Surabaya: Usaha Nasional.
- Popham, W.J., (1974) *Evaluation in Education: Current Application*, Los Angeles: University of California.
- Purwanto, Ngalim (1995) *Prinsip-Prinsip dan Teknik Evalausi Pengajaran*, Bandung: Rosdakarya.
- Remmer, H.H., at.al., (1967) *A Practical Introduction to Measurement and Evaluation*, Apleton-Century Crafts, Inc.
- Stiggins, R.J., (2004) *New Assessment Beliefs for A New School Mission*, Phi Delta Kappan.
- Sudjana, N., dan Ibrahim, R., (2007) *Penelitian dan Penilaian Pendidikan*, Cetakan ke-4, Bandung: Sinar Baru Algensindo.
- Surapranata, S., dan Hatta, M., (2004) *Penilaian Portofolio*, Bandung: PT.Remaja Rosdakarya.

Suwandi Sarwiji. (2011), *Model-Model Asesmen Dalam Pembelajaran*,
Surakarta: Yuma Pustaka.