

PEDOMAN EVALUASI
PROSES DAN HASIL BELAJAR
(BAHAN DISKUSI DI PUSBIKTEK)

YAYA SUNARYA

PEDOMAN EVALUASI PROSES DAN HASIL BELAJAR

I. KONSEP-KONSEP DASAR TENTANG EVALUASI

Kegiatan belajar mengajar pada intinya bertujuan untuk membantu mahasiswa memperoleh pengalaman belajar seperti yang sudah dirumuskan dalam GBPP atau kurikulum atau satuan acara perkuliahan. Hal ini berarti bahwa seorang dosen yang melakukan KBM, akan mendasarkan kegiatan KBM tersebut untuk mencapai tujuan. Dengan demikian, dalam mengajar dosen akan selalu memikirkan materi apa yang cocok diajarkan, metoda mengajar yang bagaimana yang sesuai, fasilitas dan sumber belajar apa yang dibutuhkan, dan sebagainya.

Dari hal di atas, tugas di akhir KBM yang harus dilakukan dosen adalah merumuskan bagaimana cara untuk mendeteksi apakah tujuan KBM tersebut tercapai atau tidak, apakah materi pelajaran yang telah disajikan dikuasai atau tidak, apakah metoda mengajar yang digunakan sudah cukup tepat ? Semua pertanyaan ini dapat dijawab melalui kegiatan pengukuran dan penilaian.

Pengukuran terhadap kemajuan belajar mahasiswa tidak dapat hanya didasarkan pada kesan umum dosen tentang perilaku mahasiswa, karena hal ini akan terlalu subjektif. Oleh karena itu dosen dituntut untuk mampu menemukan tingkah laku mahasiswa secara komprehensif. Pedoman ini merupakan salah satu petunjuk ke arah itu.

A. Tes, Pengukuran dan Penilaian

Terdapat tiga istilah yang sering dikacaukan penggunaannya dalam bidang evaluasi, yaitu tes, pengukuran, dan penilaian. Sebenarnya ketiga istilah tersebut merupakan istilah yang berjenjang (hierarkhis), dengan tes merupakan tingkatan yang paling dasar (paling awal).

Tes merupakan himpunan pertanyaan yang harus dijawab, atau pernyataan-pernyataan yang harus dipilih/ditanggapi, atau tugas-tugas yang harus dilakukan oleh orang yang di tes dengan tujuan untuk mengukur suatu aspek perilaku atau memperoleh informasi tentang trait atau atribut pendidikan dari orang yang di tes. Dalam setiap pertanyaan, pernyataan, atau tugas yang diberikan tersebut terdapat jawaban atau ketentuan yang dianggap benar. Dengan demikian maka setiap tes akan menuntut respon atau jawaban dari orang yang di tes (subjek/testee) yang dapat disimpulkan sebagai trait dari subjek yang sedang dicari informasinya. Dari uraian ini tersirat bahwa *tes berfungsi sebagai alat (instrumen) atau pun sebagai cara pengungkap informasi atau pengumpul data* tentang sesuatu.

Istilah tes tidak hanya digunakan dalam bidang pendidikan, tetapi dalam berbagai bidang. Misalnya dalam bidang kesehatan ada tes urine, dalam bidang pekerjaan ada tes penempatan, dalam bidang psikologi ada tes inteligensi, tes kepribadian, dan sebagainya. Dalam bidang pendidikan kita mengenal istilah tes hasil belajar, tes diagnostik, ujian tengah semester, ujian akhir semester, dan lain-lain.

Dalam tes hasil belajar, yang hendak diukur atau dicari informasinya ialah tingkat kemampuan seorang siswa/mahasiswa/karyasiswa dalam menguasai bahan pelajaran yang telah diajarkan kepadanya. Dalam hal ini perlu dibedakan antara prestasi

belajar (*achievement*) dan hasil belajar (*learning outcome*). Hasil belajar meliputi aspek pembentukan watak seseorang. Yang akan lebih banyak dibahas dalam pedoman ini adalah prestasi belajar (*achievement*) yang lebih bersifat mengukur pengetahuan (aspek kognitif).

Tes untuk mengukur prestasi belajar (*achievement test*), baik untuk aspek pengetahuan maupun aspek keterampilan, dapat dibedakan menjadi dua jenis, yaitu tes kemampuan (*power test*) dan tes kecepatan (*speed test*).

Power Test. Prinsip utama power test adalah tidak adanya pembatasan waktu yang ketat bagi testee untuk mengerjakan pekerjaan (tes) tersebut. Menurut prinsip ini jika waktu pengerjaan tes tidak dibatasi secara ketat, maka hasil tes akan benar-benar mengungkap kemampuan testee. Pembatasan waktu dalam mengerjakan tes, diperkirakan akan menyebabkan testee menjadi tidak dapat menunjukkan seluruh kemampuan yang dimilikinya. Namun demikian, dalam penyelenggaraan tes diberikan secara klasikal tidaklah efisien jika testee yang paling lambat pun ditunggu semau dia. Untuk mengatasi hal ini, para ahli berpendapat bahwa suatu tes dapat dianggap sebagai power test (tes kemampuan), jika sebagian besar orang yang di tes (testee) dapat menyelesaikannya dalam waktu yang sudah ditetapkan.

Speed Test. Dalam tes ini yang diukur ialah kecepatan di dalam memikirkan atau mengerjakan suatu persoalan. Itulah sebabnya tes ini disebut tes kecepatan (*speed test*). Semakin sedikit waktu yang dibutuhkan untuk mengerjakan suatu tugas (biasanya tugasnya relatif mudah), maka semakin baiklah prestasi seseorang. Biasanya model ini lebih banyak digunakan dalam tes bakat atau tes kecerdasan. Tes prestasi belajar yang disusun oleh dosen/instruktur, pada umumnya berupa tes kemampuan (*power test*) bukan tes kecepatan (*speed test*).

Tes dapat dipilah-pilah ke dalam beberapa kelompok. Dilihat dari konstruksinya maka tes dapat dikelompokkan/diklasifikasikan sebagai berikut:

1. Dilihat dari cara siswa/mahasiswa melakukannya, maka tes dapat diklasifikasikan ke dalam tiga bentuk, yaitu tes/ujian tertulis (*written test*), tes/ujian lisan (*oral test*) dan tes/ujian perbuatan (*performance test*)

a. Tes/ujian tertulis

Tes atau ujian tertulis biasanya dilakukan secara kelompok dengan mengambil tempat di suatu ruangan. Pada tes atau ujian yang kurang bersifat formal, misalnya ujian tengah semester (UTS) atau ujian akhis semester (UAS), maka pelaksanaan tes dilakukan oleh dosen masing-masing bidang studi. Dalam bentuk ini, dosen yang bersangkutan berperan sebagai perencana, pelaksana, pengawas, korektor, pengolah hasil, dan sekaligus sebagai penentu hasil bagi mahasiswa. Pada tes yang lebih formal, seperti UMPTN, seleksi calon karya siswa, pekerjaan-pekerjaan tadi dilakukan oleh panitia yang ditunjuk secara resmi oleh lembaga.

b. Tes/Ujian Lisan

Bentuk ini digunakan untuk mengevaluasi hasil belajar dalam bentuk kemampuan mengemukakan ide, pendapat atau gagasan secara lisan. Bagi lembaga yang menyiapkan siswa untuk berhubungan dengan orang lain (misalnya pendidikan), maka tes atau ujian bentuk ini mempunyai kedudukan yang penting. Untuk pelaksanaan tes bentuk ini alat yang diperlukan adalah panduan yang berisi pokok-pokok persoalan yang akan ditanyakan. Bentuk soal atau pertanyaannya biasa digunakan bentuk esai.

- c. Tes Perbuatan
Bentuk ini lebih dikenal dengan sebutan ujian praktek. Tes bentuk ini dapat digunakan untuk mengevaluasi proses penyelesaian suatu pekerjaan, keterampilan dan ketepatan menyelesaikan pekerjaan, kecepatan dan kemampuan merencanakan pekerjaan, dan mengidentifikasi suatu piranti. Yang dievaluasi dengan tes adalah prosesnya, tetapi produknya juga seringkali ikut dievaluasi sehingga mungkin paduan keduanya. Alat yang dipergunakan adalah Lembar Pengamatan.
2. Menurut Bentuk Soal
Menurut bentuk soalnya, secara umum tes dapat diklasifikasikan ke dalam soal/tes bentuk uraian (*essay test*) dan soal/tes bentuk objektif (*objective test*). Kedua bentuk ini dapat dipilah lagi ke dalam berbagai tipe.
3. Menurut Tipe Soal
Butir soal bentuk uraian dapat diklasifikasikan ke dalam dua tipe yaitu tes uraian terbatas (*restricted essay*) dan tes uraian bebas (*extended essay*). Butir soal bentuk objektif dapat dibagi menjadi tiga tipe, yaitu tipe soal benar-salah (*true-false*), pilihan ganda (*multiple choice*) dan tipe soal menjodohkan (*matching choice*).
4. Menurut ragam soal
Menurut ragamnya, tiap tipe pada butir tiga dapat dipilah ke dalam beberapa ragam soal sebagai berikut:
- a. Tipe tes uraian terbatas
 - 1) ragam soal jawaban singkat
 - 2) ragam soal melengkapi
 - b. Tipe tes uraian bebas
 - 1) Ragam uraian bebas sederhana
 - 2) Ragam uraian bebas ekspresif
 - c. Tipe soal objektif benar-salah
 - 1) ragam benar-salah sederhana
 - 2) ragam benar salah dengan koreksi
 - d. Tipe tes pilihan ganda:
 - 1) ragam pilihan ganda biasa
 - 2) ragam pilihan ganda hubungan antar hal
 - 3) ragam pilihan ganda variasi (kompleks)
 - e. Tipe tes objektif menjodohkan :
 - 1) ragam menjodohkan sederhana
 - 2) ragam menjodohkan sebab akibat

Selain dengan cara tes, alat atau cara pengumpulan data dapat pula dilakukan dengan cara *non tes* yang dilaksanakan dalam bentuk wawancara, observasi, angket, atau skala. Perbedaan utama antara tes dan nontes terletak dalam tiga hal. *Pertama* bahwa pada tes ada jawaban benar dan salah, sedangkan pada nontes jawaban benar dan salah sangat kondisional. Misalnya jawaban atas pertanyaan “Berapa jumlah saudara kamu? Apa pekerjaan orang tua kamu? “ akan sangat bervariasi, dan semuanya bisa betul. *Kedua* adalah hasil pada non tes lebih bersifat kualitatif, sedangkan tes lebih kuantitatif (walau pun akhirnya dapat dikualitatifkan). *Ketiga* pelaksana tes adalah orang profesional, sedangkan nontes tidak selamanya harus orang profesional. Dalam dunia pendidikan cara

non tes sering digunakan untuk mengungkap hasil belajar (learning outcome), yang banyak menyangkut aspek afektif, dan pada prestasi belajar (learning achievement), aspek kognitif dan psikomotor.

Pengukuran (Measurement). Dalam pengertian sehari-hari pengukuran merupakan suatu usaha untuk mengetahui keadaan sesuatu sebagaimana adanya atau pengumpulan data tentang sesuatu. Misalnya usaha untuk mengetahui kedalaman sebuah sumur, disebut pengukuran kedalaman sumur.

Dalam pengertian ilmu evaluasi pengukuran dapat diartikan sebagai pemberian angka kepada suatu atribut atau karakteristik tertentu yang dimiliki oleh orang, hal, atau objek tertentu menurut aturan atau formulasi yang jelas. Ada dua karakteristik utama dari pengukuran, yaitu (1) menggunakan angka atau skala tertentu, dan (2) menurut suatu aturan atau formulasi tertentu. Misalnya untuk mengukur tinggi atau berat badan seseorang dengan mudah kita memahami karena skalanya telah jelas dan aturannya telah diketahui secara umum. Tetapi untuk mengukur pendengaran, penglihatan atau kepekaan seseorang jauh lebih kompleks dari itu dan tidak semua orang dapat memahaminya. Dalam kegiatan seperti ini mungkin saja aturan atau formulasi yang diikuti tidak lagi sederhana (tidak diketahui dan dipahami setiap orang, hanya diketahui oleh orang tertentu). Kegiatan pengukuran ini menjadi lebih kompleks lagi bila kita akan mengukur karakteristik psikologis seseorang, seperti tentang kecerdasan, emosi, daya juang atau kepribadian seseorang.

Demikian halnya dalam dunia pendidikan. Kita tidak mengukur orangnya, tetapi hanya mengukur atribut atau karakteristiknya dengan formulasi tertentu. Seorang dosen dapat mengukur penguasaan mahasiswanya dalam suatu mata kuliah tertentu yang telah dilatihkannya, tetapi tidak mengukur mahasiswanya. Pengukuran pendidikan merupakan pekerjaan profesional guru, instruktur, atau dosen. Karenanya orang yang tidak memahami pengukuran pendidikan tidak akan mampu memahaminya. Jadi intinya dengan pengukuran atau melalui pengukuran kita akan mendapatkan data atau informasi (yang berupa angka) tentang sesuatu.

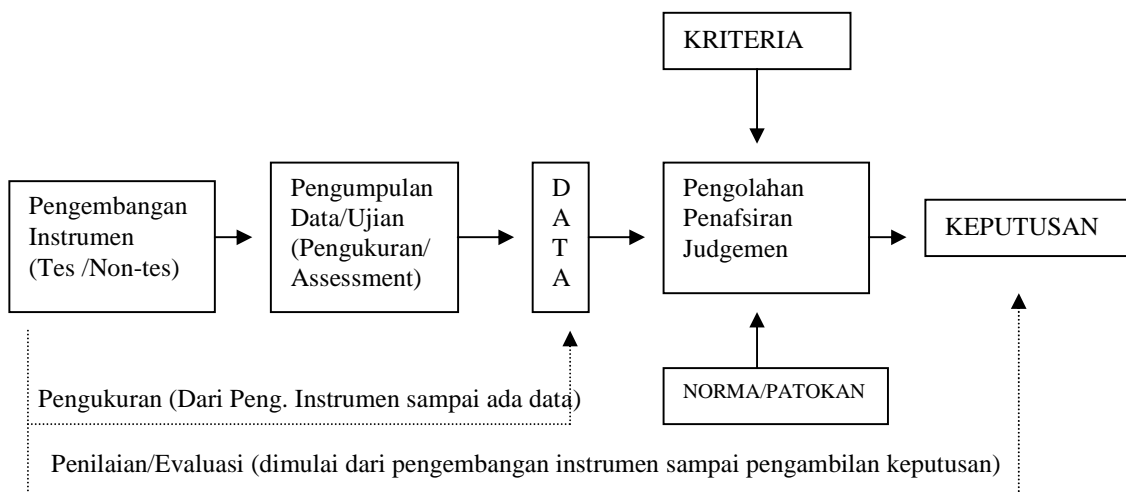
Dalam pengukuran, skala angka itu dapat diklasifikasikan kedalam 4 (empat) kategori yang bersifat hierarkhis, dari yang paling rendah berurutan yaitu *skala nominal*, *skala ordinal*, *skala interval*, dan *skala rasio*.

1. **Skala Nominal**, yaitu skala yang bersifat katagorikal. Dengan skala nominal ini kita hanya dapat memahami apakah seseorang itu berada dalam kategori mana (Jenis kelamin: pria atau wanita; warna kulit: hitam, putih, atau sawo matang; agama: islam, keristen, hidu, budha, atau lainnya). Dalam ulangan, misalnya jika sebutir soal dapat dijawab dengan *benar* oleh seorang mahasiswa, maka ia mendapat skor 1 (satu), sedangkan jika jawaban mahasiswa tersebut *salah*, dia mendapat skor 0 (nol); lulus atau tidak lulus. Dalam beberapa hal kita tidak bisa mengatakan bahwa suatu kategori itu lebih baik dari yang lain.
2. **Skala Ordinal**, yaitu angka yang menunjukkan adanya urutan, tanpa mempersoalkan jarak antar urutan tersebut. Misalnya angka yang menunjukkan urutan ranking atau peringkat mahasiswa dalam suatu mata kuliah. Mahasiswa yang memperoleh ranking 1 tidak berarti dua kali lebih pandai dari mahasiswa ranking 2. Jarak kepandaian antara mahasiswa ranking 1 dengan mahasiswa ranking 2 tidaklah sama dengan jarak kepandaian mahasiswa ranking 2 dengan mahasiswa ranking 3, dan seterusnya. Namun disini selain kita dapat menyatakan bahwa seseorang itu termasuk dalam

kategori tertentu (sifat data pertama), kita juga dapat menyatakan bahwa ranking satu lebih baik dari ranking kedua atau ranking yang ada dibawahnya.

3. **Skala Interval**, yaitu angka yang selain memiliki dua ciri di atas (menunjukkan klasifikasi dan kedudukan seseorang dalam kelompok) juga menunjukkan adanya jarak yang sama dari angka yang satu ke angka yang lain (equality of interval) jika berada dalam urutan. Misalnya angka Kg untuk mengukur berat atau angka Km untuk mengukur jarak. Jarak antara Kg 1 dengan Kg 2 sama dengan jarak antara Kg 3 dengan Kg 4; jarak antara Km 15 dengan Km 16 sama dengan jarak antara Km 21 dengan Km 22, dan seterusnya. Demikian juga dalam angka hasil ujian yang dihasilkan dari suatu tes, jarak skor 30 dengan 35 sama dengan jarak skor 40 dengan 45.
4. **Skala atau angka Rasio**, yaitu angka atau data yang selain memiliki tiga sifat data di atas (menunjukkan klasifikasi, kedudukan dalam suatu kelompok, dan memiliki jarak interval yang sama), juga memiliki nilai 0 (nol) mutlak. Misalnya tinggi badan seseorang, bila ada tinggi badan manusia 0 cm, maka tinggi tersebut tidak mempunyai makna, sama halnya dengan berat badan 0 kg. Kita dapat membuat perbandingan (rasio) terhadap orang yang memiliki berat badan 70 kg, baik laki-laki maupun perempuan (nominal) berarti lebih berat dari 35 kg (ordinal) dan sama dengan dua kali lipat dari 35 kg (interval).

Penilaian. Penilaian adalah suatu proses sistematis untuk mengambil keputusan dengan menggunakan data atau informasi yang diperoleh dari hasil pengukuran, baik dengan menggunakan instrumen yang berupa tes atau pun non-tes. Pengertian ini menjelaskan kepada kita bahwa antara tes, pengukuran dan penilaian merupakan kegiatan yang hierarkhis. Tes merupakan instrumen pengungkap data, pengukuran merupakan prosedur pengumpulan data tentang sesuatu, dan penilaian bermaksud memberikan nilai tentang kualitas sesuatu atas dasar standar tertentu. Penilaian adalah padanan kata evaluasi. Untuk memahami proses ini perhatikan bagan berikut.



Suatu proses penilaian akan selalu didahului oleh pengukuran. Suatu proses pengukuran akan menjadi penilaian jika terhadap hasil pengukuran tersebut dilakukan judgement (penimbangan) atas dasar kriteria/standar/patokan atau norma tertentu. Hasil pertimbangan antara data yang diperoleh dari hasil pengukuran dengan kriteria/standar tertentu, itulah yang disebut penilaian yang hasilnya berupa keputusan. Sebagai ilustrasi, seorang dosen mengadakan ujian akhir semester (kegiatan pengukuran). Selesai ujian diperiksa, maka diperoleh data si A mendapat skor 70 (nilai B/C), si B mendapat skor 60 (nilai C), dan seterusnya. Ini masih berupa pengukuran, belum penilaian; karena apa arti skor 70 (B/C) atau skor 60 (C), ini belum berarti apa-apa. Nah baru dikatakan melakukan penilaian jika misalnya si A dan si B dinyatakan tidak lulus karena nilainya minimal harus B atau skornya minimal 75 (nilai B atau skor 75 adalah kriteria atau norma atau standar). Contoh lain, hasil pengukuran kedalaman sebuah sumur 20 m. Dalamkah sumur itu ? Jika rata-rata kedalaman sumur di tempat itu adalah 15 m, maka sumur itu boleh dikatakan dalam. Tetapi jika rata-rata sumur di daerah itu adalah 30 m, maka tentu sumur yang 20 m itu adalah dangkal. Untuk memahami hal ini perhatikan ilustrasi berikut yang merupakan hubungan antara hasil pengukuran, patokan, dan hasil penilaian dari objek-objek pengukuran/penilaian tersebut.

Objek	Hasil Pengukuran	Patokan (pembanding)	Hasil membandingkan (Penilaian)
Kedalaman Sumur	7 meter	15-16 meter	Amat dangkal
Ujian Akhir	Ahmad = 70	- Angka terendah 40	Jauh diatas angka terendah
		- Angka tertinggi 99	Jauh dibawah angka tertinggi
		- rata-rata : 69	Sedikit di atas rata-rata
		- Batas lulus : 60	Melampaui batas lulus (lulus)
		93 – 99 : A 83 – 92 : A/B 75 – 83 : B 68 – 74 : B/C 60 – 68 : C	Mendapat nilai : B/C
		Prestasi yang sudah jauh di bawah temannya	Prestasi yang cukup bagus (ada kenaikan)
		Kemampuan IQ : 109	Hasil yang sesuai dengan kemampuan
Penguasaan kata kerja	100 buah	Rata-rata = 101	Hampir sama dengan rata-rata

Setiap penyelenggara pendidikan (guru, dosen, widyaiswara) pada saat-saat tertentu harus membuat keputusan pendidikan, yaitu keputusan yang terkait dengan soal-soal pendidikan, baik yang menyangkut diri sendiri maupun orang lain. Keputusan tersebut dapat mempunyai lingkup yang besar, seperti keputusan seorang Mendiknas tentang penerapan sistem baru dalam penyelenggaraan pendidikan, pergantian kurikulum,

seorang Rektor menetapkan batas lulus bagi mahasiswa baru. Dapat pula mempunyai lingkup yang kecil, seperti orang tua menetapkan perlu tidaknya anaknya mengikuti les setiap hari, seorang dosen menetapkan beberapa orang mahasiswa perlu ujian ulang, atau seorang mahasiswa memutuskan untuk mengontrak ulang beberapa mata kuliah yang nilainya jelek, dsb.

Dalam dunia pendidikan (khususnya pengajaran di kelas) penilaian dimaksudkan untuk mendapatkan gambaran atau untuk mengukur besar kecilnya perubahan perilaku siswa (mahasiswa) sebagai hasil belajar secara kuantitatif. Perilaku-perilaku yang dimaksud adalah perilaku yang telah dirumuskan oleh guru/dosen dalam persiapan mengajar, yang berupa rumusan-rumusan tujuan instruksional (tujuan pembelajaran).

Tujuan dan kegunaan penilaian pendidikan dapat berkaitan dengan perencanaan, pengelolaan, proses, dan tindak lanjut pendidikan baik yang menyangkut perorangan, kelompok, maupun kelembagaan. Menurut Thorndike dan Hagen (1991) berbagai keputusan yang dapat diambil berkaitan dengan pendidikan dan pengajaran dapat diarahkan untuk mengambil keputusan yang menyangkut *pengajaran, hasil belajar, diagnosis dan usaha perbaikan, penempatan, seleksi, bimbingan dan penyuluhan, dan penilaian kelembagaan*.

Istilah lain yang sering terkait dengan penilaian dan pengukuran adalah *assessment* dan penelitian. **Assessment** (di Indonesiakan : asesmen), adalah suatu proses pengumpulan data dengan menggunakan alat tertentu namun lebih bersifat kualitatif. Asesmen berbeda dengan evaluasi. Suatu asesmen tidak dirancang untuk mengevaluasi suatu program atau suatu proses tertentu. Fungsi asesmen hanya untuk mengestimasi tingkat prestasi mahasiswa di suatu daerah setelah diberikan perlakuan atau pengaruh tertentu. Dengan asesmen juga tidak dimaksudkan untuk mengambil generalisasi tertentu. Asesmen perhatiannya kepada penentuan kemajuan pelajar dalam mencapai tujuan pendidikan pada suatu waktu tertentu. Asesmen juga berbeda dengan penelitian. Asesmen tidak dimaksudkan untuk mengambil generalisasi yang dianalisis dari hubungan antar variabel (Best, 1977).

Penelitian (*research*) berbeda dengan evaluasi walau pun antara keduanya terdapat kesamaan. Keduanya berusaha untuk mengembangkan pengetahuan, mesti dilakukan oleh mereka yang kompeten, dan sama-sama mempersyaratkan penyelidikan-penyelidikan ilmiah.

Perbedaannya adalah bahwa penelitian dilakukan untuk mengembangkan prinsip-prinsip dan generalisasi-generalisasi dengan jalan menganalisis hubungan antar variabel. Pengetahuan yang diperolehnya bisa menjadi dasar bagi suatu prediksi, generalisasinya berlaku bagi orang, waktu, serta lingkungan lainnya. Evaluasi perhatiannya pada aplikasi tertentu yang lebih mendesak, berusaha menentukan sumbangan atau andil dari suatu produk, proses atau program pendidikan tertentu. Suatu evaluasi berisi *judgment-judgment*, tidak berkepentingan dengan generalisasi yang bisa diberlakukan untuk lingkungan-lingkungan lain.

B. Keterkaitan Penilaian dengan Pengajaran

Dalam konsep yang sederhana, tugas guru/dosen dalam KBM dapat dikategorikan ke dalam empat kegiatan utama, yaitu melakukan persiapan (merencanakan pengajaran), mengorganisasikan pengajaran, melakukan penilaian, dan melakukan analisis hasil penilaian guna membantu siswa yang diduga mengalami kesulitan.

Dari konsep di atas, ternyata bahwa kegiatan penilaian (evaluasi) merupakan bagian yang takterpisahkan (integral) dari kegiatan KBM secara keseluruhan. Stiggins (Furqon, 1999:1) menyatakan bahwa “ ... assessment and teaching can be one and the same”. Keterkaitan dan interdependensi antara efektifitas pengajaran dengan penilaian digambarkan oleh Gronlund (1998: 4) sebagai berikut:

A. Instruction	B. Assessment
<p>Instruction is most effective when:</p> <ol style="list-style-type: none"> 1. Directed toward a clearly defined set of intended learning outcomes 2. The methods and materials of instruction are congruent with the outcome to be assessed 3. The instruction is designed to fit the characteristics and needs of the students. 4. Instructional decisions are based on information that is meaningful, dependable, and relevant. 5. Student are periodically informed concerning their learning progress. 6. Remediation is provided for students not achieving the intended learning. 7. Instructional effectiveness is periodically reviewed and the intended learning outcomes and instruction modified as needed. 	<ol style="list-style-type: none"> 1. Assessment is most effective when Designed to assess a clearly defined set of intended learning outcomes. 2. The nature and function of the assessments are congruent with the outcomes to be assessed. 3. The assessments are designed to fit relevant student characteristics and are fair to everyone. 4. Assessment provide information that is meaningful, dependable, and relevant. 5. Provision is made for giving students early feedback of assessment results. 6. Specific learning weaknesses are revealed by the assessment results. 7. Assessment results provide information useful for evaluating the appropriateness of the objectives, the methods, and materials of instruction.

Tabel di atas memperlihatkan bahwa penilaian (*assessment*) dan pengajaran (*instruction*) merupakan dua sisi yang saling bergantung dalam mendukung efektivitas siswa belajar. Antara keduanya akan efektif jika guru/dosen memiliki seperangkat tujuan yang dipahami, diyakini, dan dirumuskan dengan jelas. Evaluasi harus dijadikan sebagai strategi dan upaya untuk meningkatkan mutu KBM. Dengan demikian maka evaluasi harus menyediakan informasi yang tepat dan relevan baik bagi guru maupun bagi siswa. Sebagai konsekwensinya maka evaluasi harus dipahami, direncanakan, dan dilaksanakan selain sebagai upaya meningkatkan (*improvement*) mutu KBM, tapi juga sebagai jaminan mutu (*quality assurance*) KBM.

Untuk mendukung tercapainya kegiatan penilaian sebagai bagian integral dari KBM dan penilaian sebagai strategi dalam meningkatkan kualitas KBM, maka kemampuan yang dituntut dari dosen/guru adalah:

1. kemampuan memilih dan merumuskan perubahan-perubahan perilaku yang signifikan yang diyakininya sebagai hasil belajar yang penting (*important learning outcomes*), secara kreatif dan mandiri;
2. kemampuan mengidentifikasi informasi evaluatif yang diperlukan dirinya untuk mampu meningkatkan kemampuan dan keefektifan mengajarnya, dan

mengidentifikasi yang diperlukan siswa untuk mampu meningkatkan keberhasilan belajarnya;

3. kemampuan mengembangkan dan menggunakan berbagai alternatif penilaian yang sesuai dengan hasil belajar yang diharapkan dan informasi evaluatif yang diperlukan.

C. Fungsi Penilaian/Evaluasi

Sejalan dengan uraian di atas, maka evaluasi akan memiliki fungsi sebagai berikut.

1. Fungsi Motivasi

Jika evaluasi telah dilakukan secara terpadu, sistematis dan kontinu, (baik dalam bentuk tugas, latihan, kuis, maupun ujian) maka dia akan berfungsi sebagai motivator dan mengkondisikan siswa/mahasiswa untuk tetap giat belajar . Untuk ini maka hasil evaluasi harus dikomunikasikan kepada mahasiswa. Mahasiswa yang mengetahui hasil ujiannya, diharapkan hasil tersebut menjadi kekuatan atau daya juang (adversity) bai dirinya.

2. Fungsi Jaminan mutu (Quality assurance)

Jika evaluasi dilakukan relevan dengan tujuan pembelajaran, maka prestasi yang dicapai oleh mahasiswa/siswa akan menggambarkan kualitas tingkat pemahaman atau tingkat penguasaan (tingkat ketuntasan) siswa/mahasiswa atas materi pelajaran yang diberikan. Pada gilirannya, keadaan ini akan menjamin kualitas/mutu pendidikan. Pemahaman atas suatu materi pelajaran akan mendorong siswa senang belajar lebih jauh. Sebaliknya, ketidaktuntasan atau ketidakpahaman siswa atas suatu materi akan menghambat dan menyebabkan kesulitan pada seseorang. Jika hal ini berakumulasi bukan tidak mustahil akan menyebabkan siswa/mahasiswa frustrasi.

3. Fungsi Balikan (Feed back)

Data yang diperoleh dari hasil ujian/ulangan akan merupakan bahan yang sangat berharga baik bagi guru/dosen maupun bagi siswa/mahasiswa untuk introspeksi, mengevaluasi diri. Dosen/guru dapat menelaah kembali tentang keefektifan metoda atau cara mengajar yang telah ia lakukan. Siswa/mahasiswa juga dapat melihat bahkan memperkirakan kelancaran studi mereka.

4. Fungsi Diagnosis

Dari hasil ujian, baik menggunakan kriteria norma kelompok (*norm reference test*) maupun dengan menggunakan kriteria mutlak (*criterion reference test*), siswa akan terbagi kepada dua kelompok, yaitu yang memenuhi kriteria dan yang tidak, yang dinyatakan lulus dan yang tidak. Khusus berkenaan dengan yang tidak lulus atau tidak memenuhi kriteria, apakah mereka akan dibiarkan? Bagi kelompok kedua ini, dari hasil tes atau ujian, dosen dapat meneliti (mendiagnosis) dimanakah kelemahan mereka itu, apa penyebabnya, mungkinkah dia dibantu, bagaimana cara membantunya, dan sebagainya. Inilah peran bimbingan seorang guru/dosen.

5. Fungsi Seleksi

Pada suatu saat guru/dosen sering diharuskan untuk menentukan mahasiswa mana yang paling unggul secara akademik atau karena jatah beasiswa terbatas maka perlu dipilih mahasiswa berprestasi. Dalam hal ini guru/dosen perlu mengadakan seleksi. Selain itu, misalnya karena tempat untuk pelatihan sangat terbatas, maka perlu dipilih calon peserta yang benar-benar unggul. Suatu lembaga memerlukan karyawan yang kualified, maka perlu diadakan tes seleksi yang ketat.

6. Fungsi Administrasi

Secara profesional, guru/dosen memiliki sejumlah tugas administratif. Diantaranya dosen harus melaporkan hasil ujian ke lembaga (biro administrasi akademik) atau mungkin juga harus melaporkan kepada lembaga pengirim mahasiswa, atau mungkin dosen harus melaporkan nilai ke bagian komputer atau bagian akademik. Disamping itu, tatkala seorang dosen/guru memerlukan bukti fisik untuk kenaikan pangkat, laporan hasil tes dapat dijadikan bahan usulan kumulatif.

D. Prinsip-prinsip Penilaian/Evaluasi

Agar keputusan yang diambil tepat dan akurat serta dapat dipertanggungjawabkan, maka ada sejumlah prinsip penilaian yang harus diperhatikan dalam melakukan penilaian.

1. *Prinsip keterpaduan*

Kegiatan penilaian yang dilakukan bukan merupakan kegiatan tersendiri, tetapi merupakan kegiatan yang harus terpadu dengan kegiatan lain. Dalam melakukan penilaian harus juga terkait dengan kebijakan lembaga, tujuan lembaga (institusi), dan memiliki daya dukung terhadap keberadaan lembaga.

2. *Prinsip Komprehensif*

Penilaian yang dilakukan harus menyeluruh, mencakup semua aspek perilaku, mencakup/mewakili semua materi (representatif), mencakup berbagai teknik penilaian, dan mewakili berbagai tingkatan soal (mencakup soal yang mudah, sedang, sukar).

3. *Prinsip kesinambungan dan keteraturan*

Penilaian tidak cukup hanya satu kali dilakukan (harus berkelanjutan/kontinu) dan harus terprogram. Siswa/mahasiswa adalah makhluk yang terus berkembang dan berubah. Oleh karena itu hasil ujian di tingkat satu tidak dapat digunakan untuk kenaikan di tingkat tiga.

4. *Prinsip Objektivitas*

Penilaian yang akurat tidak bisa dilakukan hanya atas dasar observasi dan pertimbangan subjektif. Penilaian yang tepat harus didasarkan data yang objektif, data apa adanya (data yang nyata diperoleh dari testee), tidak mesti di tambah atau dikurangi.

5. *Prinsip Relevansi*

Prinsip ini mengandung maksud bahwa pengambilan keputusan harus didasarkan pada data yang relevan, sesuai dengan tujuannya. Untuk menetapkan penjurusan kita tidak bisa mengambil data hasil tes masuk (kecuali jika dalam tes tersebut sudah dikonstruksi juga untuk tujuan tersebut). Dalam hal ini perlu adanya kesesuaian antara tujuan penilaian, data yang ada, dan instrumen yang digunakan.

E. Jenis Pelaksanaan Penilaian

Dilihat dari waktu penyelenggaraan serta tujuannya, penilaian dapat dibedakan menjadi penilaian formatif, sumatif, dan diagnostik.

Penilaian formatif adalah penilaian yang bertujuan untuk mengetahui efektivitas proses belajar mengajar, karenanya waktu pelaksanaannya pun dilakukan di akhir setiap pertemuan KBM. Informasi yang diperoleh merupakan balikan untuk melihat apakah KBM yang kita lakukan efektif atau tidak, dan berguna untuk memperbaiki kekurangan

atau kelemahan kegiatan belajar mengajar. Penilaian formatif juga berguna untuk meramalkan keberhasilan seseorang dalam penilaian sumatif. Bagi siswa/mahasiswa sendiri penilaian formatif berguna untuk melihat seberapa jauh dia menguasai tujuan yang telah dirumuskan atau materi mana yang belum dikuasai.

Penilaian sumatif adalah penilaian yang tujuannya menitikberatkan pada mengetahui hasil belajar siswa setelah mengalami kegiatan belajar mengajar selama periode tertentu, sifatnya lebih menyeluruh dibandingkan dengan penilaian formatif. Jadi dilihat dari waktu pelaksanaannya penilaian sumatif dilakukan diakhir suatu program (catur wulan atau semester); identik dengan ujian akhir semester untuk perguruan tinggi. Informasi yang diperoleh dari tes sumatif, digunakan untuk menentukan tingkat keberhasilan akhir atau kelulusan siswa dalam mengikuti suatu program.

Penilaian (tes) diagnostik adalah tes atau penilaian yang bertujuan untuk menelaah atau mengungkap kelemahan atau kesulitan yang dialami oleh siswa/mahasiswa. Waktu pemberiannya tidak terjadwal pasti, namun tatkala seorang guru atau dosen ingin mengetahui kesulitan apa atau materi mana-mana saja yang dirasakan sulit oleh mahasiswa saat itulah tes diberikan. Bentuk tes ini adalah tes biasa namun yang berbeda adalah tujuan dan cara menganalisis atau menafsirkannya.

F. Pendekatan dalam Penilaian

Di muka telah dikemukakan bahwa hasil pengukuran dapat diperbandingkan terhadap berbagai patokan/kriteria/norma. Dalam bagian ini akan diuraikan dua pendekatan penilaian, yaitu (1) pendekatan penilaian yang membandingkan hasil pengukuran seseorang dengan hasil pengukuran yang diperoleh orang-orang lain dalam kelompoknya, yaitu yang dinamakan Penilaian Acuan Norma (*Norm Reference Evaluation*); dan (2) pendekatan penilaian yang membandingkan hasil pengukuran seseorang dengan patokan “batas lulus” yang telah ditetapkan, yaitu yang dinamakan Penilaian Acuan Patokan (*Criterion Reference Evaluation*).

1. Penilaian Acuan Norma (PAN)

Secara singkat dapat dikatakan bahwa PAN ialah penilaian yang membandingkan hasil belajar mahasiswa terhadap hasil mahasiswa lain dalam kelompoknya. Pendekatan ini sering dikatakan sebagai pendekatan “apa adanya”, dalam arti, bahwa patokan pembanding semata-mata diambil dari kenyataan-kenyataan yang diperoleh pada saat pengukuran/penilaian itu berlangsung. Penilaian ini sama sekali tidak dikaitkan dengan ukuran atau patokan apa pun yang terletak diluar hasil pengukuran kelompok mahasiswa tersebut.

PAN pada dasarnya menggunakan kurve normal dan hasil-hasil perhitungannya sebagai dasar penilaian. Kurve ini dibentuk dengan mengikutsertakan semua angka hasil pengukuran yang diperoleh. Dua patokan nyata yang dipergunakan untuk membandingkan atau menafsirkan angka yang diperoleh masing-masing mahasiswa adalah angka rata-rata (mean) dan angka simpangan baku (standard deviation). Mudah dipahami bahwa patokan ini bersifat relatif, karena angka-angka yang diperoleh dapat naik atau dapat turun, sesuai dengan besarnya dua kenyataan yang diperoleh dari kurve tersebut. Dengan demikian patokan ini dapat berubah dari kurve normal yang satu ke kurve normal yang lain. Jika hasil ujian mahasiswa pada umumnya naik (lebih baik) dan akan menghasilkan angka rata-rata yang lebih tinggi,

maka patokan menjadi bergeser ke atas (naik). Sebaliknya jika hasil ujian mahasiswa pada umumnya turun, patokannya akan bergeser ke bawah (turun). Dengan demikian angka yang sama pada dua kurve yang berbeda akan mempunyai arti yang berbeda.

2. Penilaian Acuan Patokan (PAP)

PAP pada dasarnya berarti penilaian yang membandingkan hasil belajar mahasiswa dengan patokan yang telah ditetapkan sebelumnya. Hal ini bermakna bahwa sebelum usaha pengukuran atau penilaian dilakukan terlebih dahulu harus ditetapkan patokan yang akan dipakai untuk membandingkan angka hasil pengukuran agar mempunyai makna tertentu. Jadi berbeda dengan PAN, di sini patokan tidak dicari dari hasil ujian. Patokan yang telah ditetapkan itu biasanya disebut “batas lulus (passing grade) atau tingkat penguasaan minimum (*minimum acceptable*)”. Mahasiswa yang dapat mencapai patokan ini bahkan melampauinya dapat dinyatakan/diputuskan *lulus*, dan sebaliknya yang belum mencapai batas tersebut dinyatakan *tidak lulus*. Mereka yang dinyatakan lulus dapat melanjutkan ke program berikutnya, sedangkan yang belum itulah yang perlu diremedial sampai mencapai batas lulus.

PAP dikenal dengan kriteria mutlak. Patokan ini dapat digunakan untuk mahasiswa mana saja yang menempuh program yang sama. Pendekatan ini terutama digunakan pada sistem atau pendekatan belajar tuntas (*mastery learning*). Pendekatan ini akan lebih menggambarkan atau menjamin kualitas pembelajaran, jika dipergunakan dengan tepat.

II. PENGEMBANGAN ALAT EVALUASI HASIL BELAJAR

A. Syarat-syarat (Ciri-Ciri) Tes yang Baik

Telah dijelaskan di muka bahwa tes merupakan sebuah alat ukur, seperti halnya alat ukur panjang (mistar), alat ukur berat (timbangan), alat ukur suhu (termometer), dan sebagainya. Perbedaannya adalah tes dimaksudkan untuk mengukur aspek perilaku manusia. Oleh sebab itu, segala persyaratan bagi sebuah alat ukur yang baik, berlaku pula bagi sebuah tes (dalam hal ini tes hasil belajar). Persyaratan atau ciri tersebut setidaknya mencakup lima hal, yaitu validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan kepraktisan.

1. Validitas

Validitas menunjukkan tingkat ketepatan suatu alat (tes) dalam mengukur aspek yang hendak diukur, atau mengungkap data yang hendak diungkap. Setiap alat ukur harus hanya mengukur satu dimensi atau aspek saja. Suatu tes hasil belajar dikatakan valid kalau hanya mengungkap hasil belajar tertentu saja. Mistar hanya mengukur panjang atau jarak, timbangan hanya mengukur berat, tes matematika soal-soalnya harus hanya mengukur pengetahuan matematika saja, dan sebagainya. Tidak mustahil, dalam soal matematika terdapat unsur bahasa atau bentuk soal yang belum dikenal oleh siswa, sehingga berhasil tidaknya siswa menjawab soal tersebut tidak semata-mata ditentukan oleh pengetahuan di bidang matematika, melainkan ditentukan pula oleh kemampuan memahami bahasa, bentuk soal, dsb. Tes seperti ini kurang valid. Suatu tes yang hanya mengukur satu dimensi, biasanya soal yang satu dengan soal yang lain memiliki keterkaitan yang erat. Karena itu disyaratkan bahwa setiap aspek/subtes/soal harus berkorelasi tinggi dengan satu sama lain sehingga dapat dijadikan bukti bahwa semua aspek tersebut memang merupakan bagian dari aspek yang lebih luas. Jika tidak, konsekwensinya skor-skor untuk masing-masing soal jangan dijumlahkan begitu saja sebagai skor total.

Berdasarkan konsep di atas, maka validitas itu akan selalu terkait dengan pertanyaan: valid dalam hal apa dan untuk siapa? Tes yang valid untuk mengukur bakat, tidak akan valid jika digunakan untuk mengukur minat. Demikian juga tes yang valid untuk siswa SMU kelas III, tidak akan valid untuk mahasiswa PPKKT.

Dengan demikian, menguji validitas suatu tes berarti kita membandingkan tes yang kita buat dengan suatu kriteria tertentu. Berdasarkan cara/prosedur pengujiannya kita akan mengenal validitas rasional dan validitas empirik. Berdasarkan standar yang digunakan, kita akan mengenal validitas permukaan (*face validity*), validitas isi (*content validity*), dan validitas kriteria (*criterion validity*) yang terdiri dari *construct validity* dan *predictive validity*.

2. Reliabilitas

Reliabilitas tes menunjukkan tingkat keajegan suatu tes, yaitu sejauhmana tes tersebut dapat dipercaya untuk menghasilkan skor yang ajeg/konsisten. Kecermatan hasil pengukuran ditentukan oleh banyaknya informasi yang dihasilkan dan sangat berkaitan dengan satuan ukuran dan jarak rentang (range) dari skala yang digunakan. Dalam

mengukur berat sebuah cincin emas, pengukuran dengan timbangan yang bersatuan miligram dan berjarak rentang antara 0 - 1000 mg, tentu akan menghasilkan ukuran yang lebih teliti dari pada menggunakan timbangan dengan satuan kilogram dengan berjarak rentang 0 – 100 kg. Begitu pula dengan tes prestasi belajar. Sebuah tes dengan jumlah soal yang banyak dan seluruh soalnya bertaraf kesukaran sedang (on-target) bagi orang yang menepuh, tentu akan menghasilkan informasi yang lebih teliti mengenai orang yang diukur, jika dibandingkan dengan tes yang soalnya sedikit dan tingkat kesukarannya rendah (off-target). Dengan kata lain soal-soal sebuah tes jangan terlalu di bawah atau di atas kemampuan tingkat pencapaian belajar siswa, dan tingkat kesukaran butir soalnya harus relatif homogen..

Menurut Balitbangdikbud (1998), mengenai keajegan (consistency) dari skor suatu tes, dapat dibedakan menjadi keajegan internal dan keajegan eksternal. Keajegan internal ialah sejauhmana butir-butir soal sebuah tes itu homogen baik dari segi tingkat kesukaran maupun dari segi bentuk soal/prosedur menjawabnya. Jadi tingkat kesukaran soal harus sesuai dengan tingkat kemampuan siswa. Tingkat keterhandalan skor tes dalam arti (1) homogenitas butir soal dan (2) kehandalan butir-butir soal dalam mengungkap perbedaan kemampuan yang terdapat di kalangan siswa, dapat diukur dengan sebuah indeks yang disebut indeks alfa dari Cronbach yang telah disederhanakan oleh Kuder dan Richardson.

Keajegan eksternal adalah sejauhmana skor yang dihasilkan dari tes tersebut kepada sekelompok orang akan tetap sama sepanjang orang tersebut belum berubah. Hal ini dapat diuji dengan indeks korelasi dari tes re-test atau dengan paralel form (disederhanakan dengan splithalp method).

3. Daya Pembeda (*Discriminating Power/Discriminating Indeks*)

Soal-soal dari suatu tes yang baik akan mampu membedakan antara testee yang benar-benar mampu dengan testee yang kurang mampu, antara testee yang benar-benar belajar dengan testee yang tidak belajar. Secara empirik hal ini akan ditunjukkan dengan adanya perbedaan skor/hasil yang diperoleh oleh orang yang termasuk kelompok unggul dengan skor yang diperoleh oleh orang dari kelompok asor. Jadi orang-orang dari kelompok unggul akan lebih banyak benar dibandingkan dengan orang dari kelompok asor.

4. Tingkat Kesukaran (*Difficulty Indeks*)

Soal-soal suatu tes yang baik akan memiliki tingkat kesukaran yang seimbang. Seimbang disini berarti berkenaan dengan proporsi penyebaran soal mudah, sedang, dan sukar. Proporsinya bisa 20% mudah, 60% sedang, dan 20% sukar, atau komposisi yang lain (1:2:1). Soal yang mudah diperlukan untuk memberikan motivasi kerja, sedangkan soal yang sukar diperlukan untuk seleksi.

5. Kepraktisan

Suatu tes yang baik harus praktis dilihat dari cara menggunakannya, praktis melaksanakannya, serta praktis pula pengadministrasikannya. Semakin mudah ter itu diadministrasikan tes itu akan semakin baik

B. Prosedur Pengembangan Tes Hasil Belajar

Agar persyaratan tes yang baik seperti telah dikemukakan di atas tercapai, maka tes itu harus direncanakan, dilaksanakan dan diadministrasikan dengan cermat. Untuk itu prosedur atau langkah yang harus ditempuh adalah: (1) Perumusan tujuan tes, (2) Penyusunan Kisi-kisi, (3) Penulisan Soal, (4) Penelaahan soal (Review dan revisi soal), (5) Uji coba Soal, (6) Perakitan Soal, (7) Penyajian Tes, (8) Skoring, (9) Pelaporan Hasil tes, dan (10) Pemanfaatan Hasil Tes

Langkah-langkah diatas terutama berkaitan dengan pengembangan tes bentuk tes tertulis (written test). Untuk langkah pengembangan tesl perbuatan (performance test) akan diuraikan pada bagian IV. Berikut ini akan diuraikan langkah-langkah di atas walau pun serba singkat.

1. Penentuan Tujuan Tes

Langkah pertama yang harus dirumuskan adalah tujuan, untuk apa tes itu dilakukan. Secara umum penyelenggaraan tes dalam kaitnya dengan PBM, tes dapat ditunjukkan untuk:

- a. mengukur efektifitas PBM (tes formatif)
- b. mengukur tingkat keberhasilan belajar, dalam arti mengukur penguasaan materi pelajaran oleh siswa (tes sumatif)
- c. mengidentifikasi kesulitan belajar siswa (tes diagnostik)
- d. menyeleksi calon-calon siswa untuk suatu program tertentu (seleksi atau penempatan).

Sebelum pelaksanaan tujuan tersebut harus jelas yang mana, sehingga dapat memberikan arah dan lingkup pengembangan tes selanjutnya.

2. Penyusunan Kisi-kisi

Kisi-kisi dikenal pula dengan nama blue-print, layout, atau table of specification. Fungsi utamanya adalah sebagai pedoman dalam menulis dan merakit butir soal, karena kisi-kisi ini merupakan deskripsi mengenai ruang lingkup dan isi dari apa yang akan diujikan, serta memberikan perincian mengenai soal-soal yang diperlukan. Kisi-kisi dikembangkan berdasarkan tujuan penggunaan tes. Karena itu, penyusunan kisi-kisi merupakan langkah penting yang harus dilakukan sebelum menulis soal. Dengan kisi-kisi yang baik prinsip komprehensivitas dan representativeness akan terpenuhi.

Untuk mengisi format kisi-kisi diperlukan kejelian dalam memilih materi esensial yang ada dalam setiap pokok bahasan. Mungkin karena banyaknya, tidak semua materi akan diujikan (terutama dalam tes yang bersifat sumatif). Dalam memilih atau menentukan materi esensial perlu diperhatikan beberapa indikator berikut:

- a. *Urgensi* materi atau pokok bahasan, yaitu materi yang secara teoritik mutlak harus dikuasai
- b. *Kontinuitas*, yaitu materi atau pokok bahasan lanjutan yang merupakan pendalaman dari materi sebelumnya
- c. *Relevansi*, yaitu materi atau pokok bahasan yang diperlukan untuk mempelajari materi lain (prasyarat/prerequisite bagi materi lain)
- d. *Keterpakaian*, yaitu materi atau pokok bahasan yang memiliki nilai terapan tinggi dalam kehidupan sehari-hari

Contoh matrik kisi-kisi serta beberapa aspeknya

KISI-KISI PENULISAN SOAL UJIAN AKHIR

Nama Lembaga :
Mata Kuliah :
Semester :
Alokasi waktu tes : menit
Jumlah soal : butir

No. Ur.	Tujuan Pembelajaran Umum	Pokok Bahasan	Materi	Indikator	Bentuk soal	Jml. Soal

3. Penulisan Soal

Penulisan soal merupakan langkah penting untuk dapat menghasilkan alat ukur atau tes yang baik. Penulisan soal adalah penjabaran indikator jenis dan tingkat perilaku yang hendak diukur menjadi pernyataan-pernyataan yang karakteristiknya sesuai dengan rincian dalam kisi-kisi. Dalam soal harus jelas apa yang dinyatakan dan jawaban apa yang dituntut. Mutu setiap soal akan menentukan mutu tes secara keseluruhan.

Untuk mendapatkan mutu soal yang baik, berikut ini diberikan beberapa kaidah penulisan soal untuk tiap-tiap tipe/jenis soal.

a. Kaidah Penulisan Soal Pilihan Ganda

Secara umum setiap soal pilihan ganda akan terdiri dari pokok soal (stem) dan pilihan jawaban (option). Pilihan jawaban terdiri atas kunci jawaban (jawaban yang benar atau paling benar), dan pengecoh atau distractor (jawaban tidak benar yang diharapkan/mungkin dapat mengecoh seseorang yang tidak menguasai bahan pelajaran).

Keunggulan soal bentuk pilihan ganda adalah dapat di skor dengan mudah, cepat, objektif, dan dapat mencakup ruang lingkup yang luas. Namun soal ini relatif lebih susah/lama disusun, susah membuat pengecoh yang homogen, dan terdapat peluang untuk menebak jawaban.

Beberapa kaidah penulisan yang harus diperhatikan adalah berkenaan dengan *materi soal, konstruksi dan bahasa*.

Materi

- Soal harus sesuai dengan indikator. Artinya soal harus menanyakan perilaku dan materi yang hendak diukur sesuai dengan tuntutan indikator.

Contoh:

Indikator

Karyasiswa dapat menjelaskan konsep sifat-sifat cahaya yang berkaitan dengan fungsi bagian-bagian mata pada manusia.

Soal yang tidak sesuai dengan indikator misalnya:

Alat optik yang dapat membantu manusia melihat sesuatu yang jauh menjadi lebih jelas adalah

- a. mikroskop
- b. kamera
- c. teleskop
- d. proyektor

Soal yang sesuai dengan indikator adalah:

Bagian mata yang mengatur banyaknya cahaya yang masuk adalah

- a. pupil
- b. retina
- c. kornea
- d. selaput iris

- Pengecoh harus logis dan berfungsi

Contoh soal yang salah:

“Bukan satu, bukan dua, bukan sepuluh, tetapi sudah beratus-ratus tokoh yang tamat dari lembaga ini”.

Kalimat di atas menggunakan majas

- a. klimaks
- b. sastra
- c. ilmiah
- d. anti klimaks

Soal yang lebih baik adalah:

“Bukan satu, bukan dua, bukan sepuluh, tetapi sudah beratus-ratus tokoh yang tamat dari lembaga ini”.

Kalimat di atas menggunakan majas

- a. klimaks
- b. anti klimaks
- c. paralelisme
- d. repetisi

- Setiap soal harus mempunyai satu jawaban yang benar atau paling benar. Artinya satu soal hanya mempunyai satu kunci jawaban. Jika terdapat beberapa jawaban yang benar, maka kunci jawaban adalah jawaban yang paling benar.

Contoh soal yang salah (kunci jawaban lebih dari satu):

Bentuk pengamalan Pancasila yang sesuai dengan sila kelima (keadilan sosial bagi seluruh rakyat Indonesia) adalah

- a. tidak bergaya hidup mewah
- b. saling mencintai sesama manusia
- c. menghormati hak-hak orang lain
- d. mengembangkan sikap tenggang rasa

Contoh soal yang lebih baik (setelah diperbaiki):

Bentuk pengamalan Pancasila yang sesuai dengan sila kelima (keadilan sosial bagi seluruh rakyat Indonesia) adalah

- a. berani membela kebenaran dan keadilan
- b. saling mencintai sesama manusia
- c. menghormati hak-hak orang lain
- d. mengembangkan sikap tenggang rasa

Konstruksi

- Pokok soal harus dirumuskan secara jelas dan tegas. Artinya soal tidak bermakna ganda, dan setiap soal harus hanya mengandung satu persoalan.

Conton pokok soal yang tidak jelas dan tidak tegas:

Presiden Republik Indonesia ialah

- a. Soekarno
- b. Suharto
- c. B.J. Habibie
- d. Abdurachman Wahid

Soal yang lebih baik adalah:

Presiden Republik Indonesia yang kedua adalah ...

- a. Soekarno
- b. Suharto
- c. B.J. Habibie
- d. Abdurachman Wahid

- Rumusan pokok soal dan pilihan jawaban harus merupakan pernyataan yang diperlukan saja.

Contoh soal yang mengandung informasi yang tidak diperlukan:

Majelis Permusyawaratan Rakyat merupakan lembaga tertinggi negara RI. Penulisan singkatan kata *Majelis Permusyawaratan Rakyat* yang benar adalah

...

- a. MPR
- b. M.P.R
- c. MPR.
- d. M.P.R.

Soal tersebut cukup ditulis:

Penulisan singkatan kata *Majelis Permusyawaratan Rakyat* yang benar adalah

...

- a. MPR
- b. M.P.R
- c. MPR.
- d. M.P.R.

- Pokok soal jangan memberikan petunjuk ke arah jawaban yang benar. Artinya pada pokok soal jangan terdapat kata, frase, atau ungkapan yang dapat memberikan petunjuk ke arah jawaban yang benar.

Contoh soal yang memberi petunjuk kepada kunci jawaban yang benar:

Generator listrik di PLTA Saguling digerakan oleh tenaga

- air
- upa panas
- gas bumi
- matahari

Soal yang lebih memadi adalah :

Generator listrik Saguling digerakan oleh tenaga

- air
- upa panas
- gas bumi
- matahari

- Pokok soal jangan mengandung pernyataan yang bersifat negatif ganda. Artinya dalam setiap soal jangan mengandung dua kata atau lebih yang mengandung arti negatif, supaya terhindar dari kesalahan penafsiran oleh testee.

Contoh soal yang mengandung pernyataan negatif ganda:

Hewan-hewan yang hidup di air dibawah ini **tidak** berkembang biak dengan cara melahirkan, **kecuali**

- ikan mujair
- ikan gurame
- ikan paus
- ikan mas

Contoh soal yang lebih memadai adalah

Di bawah ini adalah hewan-hewan yang hidup di air yang berkembang biak dengan cara bertelur, **kecuali**

- ikan mujair
- ikan gurame
- ikan paus
- ikan mas

- Pilihan jawaban harus homogen dan logis ditinjau dari segi materi. Artinya semua pilihan jawaban harus berasal dari materi yang sama seperti yang dinyatakan oleh pokok soal, penulisannya harus setara, dan semua pilihan jawaban harus berfungsi.

Contoh soal yang salah

Diantara nama-nama berikut, yang tidak pernah menjabat sebagai Menteri Pendidikan dan Kebudayaan Republik Indonesia adalah

- Fuad Hasan
- Yahya Muhaimin
- Mahatmagandhi
- Nugroho Notosusanto

Soal yang lebih memadai adalah

Diantara nama-nama berikut, yang tidak pernah menjabat sebagai Menteri Pendidikan dan Kebudayaan Republik Indonesia adalah

- a. Fuad Hasan
- b. Yahya Muhaimin
- c. Fuad Bawazir
- d. Nugroho Notosusanto

- Panjang rumusan pilihan jawaban harus relatif sama. Karena pilihan jawaban yang panjang itu lebih lengkap dan akan merujuk ke pilihan jawaban benar.

Contoh soal yang salah

Orangtua memelihara dan mendidik anak merupakan perwujudan

- a. hak azasi manusia
- b. pelaksanaan cita-cita anak
- c. cinta dan kasih sayang orangtua terhadap anaknya supaya anak menjadi pintar dan terhormat di masa depannya
- d. penghormatan anak kepada orang tua

Soal yang lebih memadai adalah

Orang tua memelihara dan mendidik anak merupakan perwujudan

- a. pelaksanaan cita-cita anak
- b. pelaksanaan cita-cita orang tua
- c. cinta kasih sayang orang tua
- d. penghormatan anak kepada orang tua

- Pilihan jawaban jangan mengandung pernyataan “semua pilihan di atas benar” atau “semua pilihan jawaban di atas salah”.

Contoh

Menurut William Stern, faktor-faktor yang mempengaruhi pertumbuhan dan perkembangan anak adalah

- a. pembawaan
- b. lingkungan
- c. kematangan
- d. semuanya benar

Contoh soal yang lebih baik

Menurut William Stern, faktor-faktor yang mempengaruhi pertumbuhan dan perkembangan anak adalah

- a. pembawaan dan lingkungan
- b. lingkungan dan kematangan
- c. kematangan dan pembawaan
- d. pembawaan, lingkungan, dan kematangan

- Pilihan jawaban yang berbentuk angka, harus disusun berdasarkan urutan besar kecilnya nilai angka tersebut . Boleh dari yang besar ke yang kecil, atau sebaliknya. Demikian juga pilihan jawaban yang menunjukkan waktu harus disusun secara kronologis. Pola pengurutan tersebut dimaksudkan untuk memudahkan siswa memilih pilihan jawabannya.

Contoh soal yang kurang baik

Hari Sumpah Pemuda diperingatu setiap tanggal

- a. 5 Juli
- b. 2 Mei
- c. 28 Oktober
- d. 30 Oktober

Soal yang lebih baik adalah

Hari Sumpah Pemuda diperingatu setiap tanggal

- a. 2 Mei
- b. 5 Juli
- c. 28 Oktober
- d. 30 Oktober

- Gambar, grafik, tabel, diagram, dan sejenisnya yang terdapat pada soal harus jelas dan berfungsi. Artinya apa saja yang menyertai suatu soal yang ditanyakan harus jelas, terbaca, dapat dimengerti oleh siswa.

Contoh soal yang tidak jelas



Gambar ini menunjukkan kedudukan matahari, bulan, dan bumi. Peristiwa yang sedang terjadi adalah

- a. gerhana matahari
- b. herhana bulan
- c. gerhana bumi
- d. revolusi bumi

Contoh soal yang lebih memadai adalah



Gambar ini menunjukkan kedudukan matahari, bulan dan bumi. Peristiwa yang sedang terjadi adalah

- a. gerhana matahari
- b. herhana bulan
- c. gerhana bumi
- d. revolusi bumi

- Butir soal jangan bergantung pada jawaban soal sebelumnya. Soal yang tergantung pada jawaban soal sebelumnya, akan menyebabkan siswa tidak dapat menjawabnya dengan benar pada soal tersebut.

Contoh

Zat yang diperlukan untuk proses pembakaran adalah

- CO₂
- O₂
- N₂
- H₂O

Berapa persen dari zat yang dihasilkan pada jawaban di atas terdapat di sekitar kita ?

- 10 %
- 20 %
- 25 %
- 59 %

Soal di atas lebih baik diubah menjadi

Berapa persenkah zat O₂ yang terdapat di udara sekitar kita ?

- 10 %
- 20 %
- 25 %
- 59 %

Bahasa

- Setiap soal harus menggunakan bahasa yang sesuai dengan dengan kaidah bahasa Indonesia yang benar.

Contoh

Partai Demokrasi Indonesia, kongresnya yang ke dua dilaksanakan pada tahun

- 1990
- 1994
- 1996
- 1997

Soal tersebut dibuat menjadi lebih baik seperti

Kongres Partai Demokrasi Indonesia yang ke dua diselenggarakan pada tahun

- 1990
- 1994
- 1996
- 1997

- Bahasa yang digunakan harus komunikatif, sehingga mudah dimengerti testee.

Contoh bahasa yang kurang komunikatif

Huruf pada kata-kata berikut yang diucapkan berbeda terdapat pada kata

- bebas
- beli
- heran
- merah

Soal yang menggunakan bahasa yang lebih komunikatif adalah

- Bunyi /e/ yang berbeda terdapat pada kata
- bebas
 - beli
 - heran
 - merah
- Jangan menggunakan bahasa yang berlaku setempat, jika soal akan digunakan untuk daerah lain atau nasional.
Contoh soal yang menggunakan istilah bahasa setempat
Enyeng termasuk binatang yang berkembangbiak dengan cara
 - bertelur
 - membelah diri
 - melahirkan
 - generatif
 Soal yang lebih memadai
Kucing termasuk binatang yang berkembang biak dengan cara
 - bertelur
 - membelah diri
 - melahirkan
 - generatif
 - Pilihan jawaban jangan mengulang kata/frase yang bukan merupakan suatu pkesatuan pengertian. Letakkanlah kata atau frase tersebut pada pokok soal.
Contoh:
"Have you ever been to Bali?"
"Yes,"
 - I have been to Bali twice
 - I have been to Bali for two years
 - I have been to Bali in 1994
 - I have been to Bali since January
 Soal tersebut sebaiknya dibuat menjadi
"Have you ever been to Bali?"
"Yes, I have been to Bali"
 - twice
 - for two years
 - in 1994
 - since January

b. Kaidan Penulisan Soal Bentuk Uraian

Soal bentuk uraian adalah suatu soal yang jawabannya menuntut testee untuk mengingat dan mengorganisasikan gagasan-gagasan atau hal-hal yang telah dipelajarinya dengan cara mengemukakan atau mengekspresikan gagasan tersebut dalam bentuk uraian tertulis. Berdasarkan nya soal bentuk uraian dapat dibedakan menjadi uraian objektif dan uraian non-objektif.

Soal uraian objektif adalah suatu soal atau pernyataan yang menuntut sehimpunan jawaban dengan pengertian/konsep tertentu, sehingga penyekorannya dapat dilakukan secara objektif. Soal bentuk uraian non-objektif adalah suatu soal yang menuntut sehimpunan jawaban dengan pengertian/konsep menurut pendapat masing-masing siswa, sehingga penyekorannya mengandung unsur subjektifitas (sukar untuk dilakukan secara objektif).

Jadi perbedaan yang prinsip adalah pada kepastian penyekorannya. Pada soal uraian objektif, kunci jawaban dan pedoman penyekorannya lebih pasti (diuraikan secara jelas hal atau komponen yang diskor dan berapa besarnya skor untuk setiap komponen tersebut. Pada bentuk uraian non-objektif, bentuk pedoman penyekorannya paling banter dalam bentuk rentang skor, dan karena pengaruh subjektifitas penilai walaupun rentang skor dan garis besar jawaban sudah ada, tetap akan ada perbedaan skor jika dilakukan penilaian/koreksi oleh orang yang berbeda.

Dalam menulis soal bentuk uraian, penulis soal harus mempunyai gambaran tentang ruang lingkup materi yang ditanyakan dan lingkup jawaban yang diharapkan, kedalaman dan panjang jawaban, atau rincian jawaban yang mungkin diberikan oleh siswa. Dengan adanya batasan ruang lingkup, kemungkinan terjadinya ketidakjelasan soal dapat dihindari, serta dapat mempermudah pembuatan kriteria atau pedoman penyekoran. Karena itu kaidah umum yang terpenting dalam menulis soal bentuk uraian adalah, segera tulis kunci jawaban atau pokok-pokok jawaban yang mungkin diberikan oleh testee beserta kriteria atau rentang skor yang mungkin diberikan, begitu selesai menulis soal.

Kaidah khusus penulisan soal bentuk uraian adalah sebagai berikut:

Materi

- Soal harus sesuai dengan indikator pada kisi-kisi. Artinya soal harus menyatakan perilaku dan materi yang hendak diukur sesuai dengan tuntutan indikator.
- Batasan pertanyaan dan jawaban yang diharapkan (ruang lingkup) harus jelas.

Contoh yang tidak baik

Apa yang anda ketahui tentang PAP dan PAN ?

Soal yang lebih memadai

Jelaskan sekurang-kurangnya tiga perbedaan pokok antara PAP dan PAN dalam proses penilaian !

- Isi materi sesuai dengan tujuan pengukuran.
- Isi materi yang ditanyakan sesuai dengan jenjang, jenis sekolah dan tingkat kelas.

Konstruksi

- Rumusan kalimat soal atau pertanyaan harus menggunakan kata tanya atau perintah yang menuntut jawaban terurai; seperti : mengapa, uraikan, jelaskan, hubungkan, tafsirkan, buktikan, hitunglah, dsb. Jangan menggunakan kata tanya yang tidak menuntut jawaban uraian, misalnya: siapa, dimana, kapan. Demikian juga kalimat tanya yang menuntut jawaban “ya” atau “tidak”, jangan digunakan.
- Buatlah petunjuk yang jelas tentang cara mengerjakan soal.
- Buatlah pedoman penyekoran segera setelah soal selesai ditulis dengan cara menguraikan komponen yang akan dinilai atau kriteria penyekorannya, besarnya

skor bagi setiap komponen, serta rentang skor yang dapat diperoleh untuk soal yang bersangkutan.

- Hal-hal lain yang menyertai soal (grafik, tabel, gambar, peta, dsj.) harus jelas dan terbaca, sehingga tidak menimbulkan penafsiran yang berbeda.

Bahasa

- Rumusan kalimat soal harus komunikatif, yaitu menggunakan bahasa yang sederhana, dan menggunakan kata-kata yang sudah dikenal testee.
- Butir soal menggunakan bahasa Indonesia yang baik dan benar.
- Rumusan soal tidak menggunakan kata/kalimat yang menimbulkan penafsiran yang berbeda (salah pengertian).
- Jangan menggunakan bahasa yang berlaku setempat, jika soal peserta berasal dari berbagai daerah.
- Rumusan soal tidak menggunakan kata-kata yang dapat menyinggung perasaan testee.

Contoh Penyusunan Pedoman Penyelesaian

Pedoman penyelesaian merupakan panduan atau petunjuk yang menjelaskan tentang:

- Batasan atau kata-kata kunci untuk melakukan penyelesaian terhadap soal-soal bentuk uraian objektif.
- Kriteria-kriteria jawaban yang digunakan untuk melakukan penyelesaian terhadap soal-soal uraian non-objektif.

Uraian Objektif

Indikator : Siswa dapat menghitung isi bangun ruang (balok) dan mengubah satuan ukurannya.

Butir Soal : Sebuah bak penampung air berbentuk balok berukuran panjang 150 cm, lebar 80 cm, dan tinggi 75 cm. Berapa literkah ini bak penampung air tersebut ?

Pedoman Penyelesaian

Langkah	Kunci Jawaban	Skor
1.	Rumus isi balok = panjang x lebar x tinggi	1
2.	= 150 cm x 80 cm x 75 cm	1
3.	= 900.000 cm ³	1
4.	Isi balok dalam liter 900.000 = ----- liter 1000	1
5	= 900 liter	1
	Skor maksimum	5

Uraian Non-objektif

Butir Soal : Jelaskan alasan apa saja yang membuat kita perlu bangga sebagai bangsa Indonesia !

Kriteria Jawaban: Jawaban boleh bermacam-macam, namun pada pokoknya jawaban dapat dikompokkan sebagai berikut:

Kriteria Jawaban	Rentang Skor
Kebanggaan yang berkaitan dengan kekayaan alam Indonesia	0 – 2
Kebanggaan yang berkaitan dengan keindahan tanah air Indonesia (pemandangan alam, geografis, dsb.).	0 – 2
Kebanggaan yang berkaitan dengan keanekaragaman budaya, suku, adat istiadat, tetapi dapat bersatu	0 – 3
Keganggaan yang berkaitan dengan keramahtamahan masyarakat Indonesia	0 – 2
Skor maksimum	9

4. Penelaahan soal (Review dan revisi soal)

Langkah ini merupakan hal yang penting untuk diperhatikan, karena seringkali kekurangan yang terdapat pada suatu soal tidak terlihat oleh penulis soal. Review dan revisi soal ini idealnya dilakukan oleh orang lain (bukan penulis soal, ahli lain, judge) dan terdiri dari satu tim penelaah yang terdiri dari ahli-ahli bidang studi, pengukuran, dan bahasa. Hal ini sulit untuk dilakukan karena berbagai keterbatasan (waktu, dana, dsb). Sebagai jalan keluarnya, si penulis soal dapat melakukannya sendiri dengan mengambil waktu. Misalnya soal ditulis malam hari, besok siang si penulis dapat menelaah/membacanya sendiri dengan menggunakan kartu telaah.

Telaahan mutu soal atau analisis secara kualitatif pada dasarnya adalah penelaahan butir soal ditinjau dari segi kaidah penulisan soal, yaitu menelaah soal dilihat dari segi (1) isi atau materi, (2) konstruksi, dan (3) bahasa.

Telaah isi soal. Suatu soal dapat ditelaah untuk mengetahui apakah isi atau materi soal yang dinyatakan sesuai dengan tujuan pertanyaan yang tersirat dalam indikator. Misalnya menelaah soal-soal IPS, apakah yang ditanyakan memang masalah IPS dan bukan hal lainnya. Atau soal matematika, mungkin bukan hanya menanyakan materi matematika tetapi juga menuntut pemahaman bahasa, karena soal tersebut menggunakan kalimat yang panjang. Telaah mengenai kesesuaian termaksud, dapat didasarkan pada buku acuan kurikulum.

Telaah Konstruksi soal. Soal yang baik harus memenuhi kaidah-kaidah penulisan soal. Pada soal pilihan ganda, misalnya, pokok soal jangan memberikan petunjuk kearah jawaban yang benar, pilihan jawaban harus homogen dan logis, gambar atau grafik harus berfungsi dan jelas keterkaitannya dengan soal, dan sebagainya.

Telaah Bahasa dan Budaya. Telaah dari segi bahasa adalah untuk melihat apakah bahasa dari suatu soal sudah jelas dan komunikatif, sehingga mudah dimengerti testee, dan tidak menimbulkan pengertian atau penafsiran yang berbeda. Dari segi budaya, perludilihat apakah suatu soal dapat menguntungkan kelompok budaya tertentu dan merugikan kelompok budaya lain (aspek bias).

Untuk melakukan telaahan terhadap soal (baik objektif maupun uraian) berikut ini diberikan contoh kartu telaah. Kartu ini digunakan untuk tiap soal (soal demi soal).

KARTU TELAAH SOAL PILIHAN GANDA

Nomor Soal :		Perangkat :	
No.	KRITERIA PENELAAHAN	YA	TIDAK
A. MATERI			
1	Soal sesuai dengan indikator		
2	Pengecoh berfungsi		
3	Mempunyai satu jawaban yang benar atau paling benar		
B. KONSTRUKSI			
4	Pokok soal dirumuskan secara tegas dan jelas		
5	Rumusan soal dan rumusan jawaban merupakan pernyataan yang diperlukan saja		
6	Pokok soal tidak memberi petunjuk ke arah jawaban benar		
7	Pokok soal tidak mengandung pernyataan yang bersifat negatif ganda		
8	Pilihan jawaban homogen dan logis ditinjau dari segi materi		
9	Panjang rumusan jawaban relatif sama		
10	Pilihan jawaban tidak mengandung pernyataan “Semua pilihan jawaban di atas benar” atau “semua pilihan di atas salah”.		
11	Pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan urutan besarkecilnya nilai angka tersebut atau kronologis waktunya		
12	Gambar, Grafik, tabel, diagram dan sejenisnya jelas dan berfungsi		
13	Butir soal tidak bergantung pada jawaban soal sebelumnya		
C. BAHASA			
14	Soal menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia		
15	Bahasa yang digunakan komunikatif		
16	Tidak menggunakan bahasa yang berlaku setempat		
17	Pilihan jawaban tidak mengandung kata/frase yang bukan merupakan satu kesatuan pengertian		
Catatan:			

Jika terdapat tanda cek pada kolom “tidak”(satu atau lebih), soal tersebut harus diperbaiki sehingga merupakan soal benar.

KARTU TELAAH SOAL URAIAN

Nomor 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 pengukuran		
4	Isi materi yang ditanyakan sudah sesuai dengan jenjang, jenis sekolah, atau tingkat kelas		
B. KONSTRUKSI			
5	Rumusan kalimat soal atau pertanyaan harus menggunakan kata 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 yang berlaku setempat		
13	Rumusan soal tidak mengandung kata-kata yang dapat menyinggung perasaan testee		
Catatan:			

5. Uji coba Soal

Uji coba pada prinsipnya bertujuan untuk mendapatkan data/informasi secara empirik mengenai sejauhmana sebuah soal dapat mengukur apa yang hendak diukur. Informasi empirik tersebut pada umumnya menyangkut segala hal yang dapat mempengaruhi validitas soal, seperti aspek keterbacaan soal, tingkat kesukaran soal, pola jawaban, daya pembeda, pengaruh budaya, dan sebagainya. Uji coba dilakukan terhadap testee yang memiliki kesamaan karakteristik dengan testee yang akan di tes sesungguhnya dengan tes tersebut. Data hasil uji coba dianalisis/digunakan untuk memperbaiki kualitas soal yang dianggap jelek.

6. Perakitan Soal

Agar skor hasil tes dapat dipercaya, diperlukan soal yang banyak. Oleh karena itu dalam penyajiannya butir-butir soal yang banyak itu dirakit menjadi suatu alat ukur yang terpadu. Urutan nomor soal, pengelompokkan butir atau bentuk-bentuk soal (kalau terdapat beberapa bentuk soal), tata lay-out soal, akan mempengaruhi validitas skor-skor tes.

Hal yang harus diperhatikan dalam merakit soal adalah:

- Memberi petunjuk pengerjaan untuk setiap bagian
- Pengurutan nomor soal harus sesuai dengan nomor yang ada pada kisi-kisi.
- Setiap soal tidak boleh memberi petunjuk jawaban terhadap soal yang lain.
- Penyebaran kunci jawaban harus acak, tidak berpola.
- Masing-masing pilihan jawaban yang merupakan kunci jawaban harus mengikuti rumus:

$$\text{- Jumlah maksimum} = \frac{\text{Jumlah soal}}{\text{Jumlah pilihan jawaban}} + 3$$

$$\text{- Jumlah maksimum} = \frac{\text{Jumlah soal}}{\text{Jumlah pilihan jawaban}} - 3$$

7. Penggunaan (Penyajian Tes)

Setelah disusun (dirakit), naskah tes diperbanyak sejumlah peserta, tes sudah siap digunakan. Hal yang perlu diperhatikan dalam pelaksanaan tes adalah waktu, petunjuk

cara mengerjakan atau menjawab tes (langsung pada lembar soal atau pada lembaran khusus), dan ruangan/ tempat duduk

8. Skoring

Pada prinsipnya skoring harus dilaksanakan secara objektif. Ini dapat dilihat dari hasil yang diperoleh, jika diperiksa ulang atau diperiksa oleh dua orang akan menghasilkan data/informasi yang berupa angka yang sama. Hal yang harus diperhatikan dalam skoring adalah kunci jawaban dan bobot soal. Setelah diperoleh informasi kuantitatif dari setiap peserta langkah berikut adalah mengolah skor kedalam angka baku tertentu (skala 5, skala 10 atau skala 100). Mengenai pengolahan skor ini akan dibahas lebih lanjut pada bagian khusus.

Skoring dalam tes objektif adalah lebih mudah, karena setiap jawaban benar diberi skor 1 dan jika salah adalah 0. Dengan demikian skor yang diperoleh akan sama dengan jumlah jawaban yang benar. Untuk mempermudah pemeriksaan (skoring), dapat digunakan lembar jawaban yang dilubangi atau menggunakan plastik transparan.

Berbeda dengan tes uraian (lebih lagi uraian yang non-objektif), skoring lebih sulit. Karena rentang skor untuk setiap soal dapat berbeda, bahkan jawaban yang benar juga dapat mempunyai rentang yang berbeda, subjektivitas akan mempengaruhi skor yang diperoleh. Untuk mengurangi subjektivitas dan meningkatkan reliabilitas (keajegan) penskoran, dalam memeriksa soal uraian perlu diperhatikan sara-saran berikut.

- a. Siapkan garis-garis besar jawaban yang dikehendaki terlebih dahulu sebelum penskoran dimulai.
- b. Tetapkan, bagaimana menangani faktor-faktor yang tidak relevan dengan hasil belajar yang sedang diukur. Yang menonjol misalnya tulisan, ejaan, struktur kalimat, tanda baca, dan kerapian. Penilai harus berusaha untuk tidak terpengaruh oleh faktor-faktor tersebut. Evaluasilah isinya. Namun bila hal itu akan diperhitungkan, pisahkanlah skornya dari isi jawaban.
- c. Gunakanlah metoda penskoran yang paling tepat. Ada dua metoda yang dapat digunakan, yaitu point method dan rating method. Dengan point method setiap jawaban dibandingkan dengan jawaban ideal yang telah ditetapkan dalam kunci, dan skor yang diberikan tergantung pada kualitas jawaban yang diberikan.. Sedangkan dengan rating method, setiap jawaban testi ditetapkan dalam salah satu kelompok yang sudah dipilah-pilah sesuai dengan kualitas jawaban. Jadi kalau suatu soal akan diberi skor maksimum 4 maka buatlah jawaban kedalam 5 kelompok. Kelompok yang terbaik mendapat skor 4 dan yang terjelek 0.
- d. Evaluasilah semua jawaban terhadap satu pertanyaan sebelum pindah ke jawaban berikutnya. Gunakan whole method daripada sparated method.
- e. Evaluasilah setiap jawaban testi tanpa melihat nama atau identitasnya.
- f. Lakukanlah pemeriksaan oleh lebih dari satu orang.

9. Pelaporan Hasil tes

Setelah selesai diolah langkah berikutnya adalah melaporkan hasil kepada peserta, kepada orangtua, ke lembaga (bagian akademik) atau ke pihak lain yang berkepentingan. Masing-masing pihak yang berkepentingan akan menggunakan hasil tes itu sesuai tujuannya.

10. Pemanfaatan Hasil Tes

Hasil pengukuran yang diperoleh melalui hasil pengetesan, dapat dimanfaatkan untuk berbagai kepentingan. Dengan menggunakan kriteria tertentu, hasil tes dapat ditafsirkan. Penggunaan hasil dapat ditujukan untuk memperbaiki atau menyempurnakan KBM, menentukan kelulusan, atau untuk mengambil keputusan dan kebijakan lain yang relevan.

C. Standardisasi Alat Evaluasi

Seerti sudah dijelaskan pada bagian sebelumnya, bahwa suatu alat ukur yang baik memiliki beberapa ciri, seperti harus valid, harus reliable, dan butir-butir soalnya harus memiliki daya pembeda dan tingkat kesukaran yang seimbang. Tes yang baik tersebut dinamakan tes standar (standardized test). Untuk mendapatkan tes yang baik, idealnya tes itu harus diujicobakan. Namun demikian sebelum diujicobakan pun tes tersebut harus menunjukkan indikator tes yang baik. Maka idealnya, tes itu harus di analisis baik sebelum ujicoba maupun sesudahnya. Analisis sebelum ujicoba dilakukan didasarkan pada karakteristik yang tampak pada soal sebelum diujicobakan dinamakan analisis rasional. Sedangkan analisis yang dilakukan terhadap data hasil uji coba dinamakan analisis empiris.

Prosedur analisis rasional lebih keanalisis validitas item, yaitu menganalisis soal dari segi isi, konstruksi dan bahasa (sudah kita tempuh pada bagian sebelumnya). Berikut ini kita akan mencoba menganalisis soal secara empiris.

1. Analisis Tingkat Kesukaran

Tingkat kesukaran soal adalah proporsi jumlah peserta tes yang menjawab benar, yang dapat dicari dengan cara membandingkan jumlah peserta yang menjawab betul untuk suatu soal dengan jumlah peserta tes seluruhnya. Bila diwujudkan dalam bentuk rumus :

$$P = \frac{B}{T}$$

P = tingkat kesukaran soal

B = jumlah subjek yang menjawab betul untuk soal ke 1

T = jumlah seluruh subjek yang menjawab soal

Tingkat kesukaran ini akan berkisar antara 0 sampai 1. Semakin tinggi tingkat kesukaran (semakin mendekati 1) berarti semakin mudalah soal tersebut. Semakin rendah angka yang diperoleh (semakin dekat dengan 0) berarti soal tersebut semakin sukar. Sebenarnya pendekatan seperti ini lebih tepat kalau disebut tingkat kemudahan, dan bila kita ingin menghitung tingkat kesukaran bisa dihitung sebaliknya, yaitu yang dihitung jumlah yang menjawab salah. Kriterianya adalah

0,00 - 0,15 = Sangat sukar

0,16 - 0,30 = Sukar

0,31 - 0,70 = Sedang

0,71 - 0,85 = Mudah

0,86 - 1,00 = Sangat Mudah

Berikut ini adalah contoh hasil perhitungan tingkat kesukaran dalam bentuk matrik (tabel)

Nomor Soal	Jumlah yang betul (B)	Jumlah pengikut (T)	Tingkat Kesukaran (P)	Kualifikasi
1	0	40	0,00	Sanga sukar
2	10	40	0,25	Sukar
3	20	40	0,50	Sedang
4	30	40	0,75	Mudah
5	40	40	1,00	Sangat Mudah
Dst				

Fungsi tingkat kesukaran ini biasanya dikaitkan dengan tujuan tes, misalnya untuk seleksi calon siswa (UMPTN), maka soal-soal yang sulit yang digunakan; untuk keperluan sertifikasi seperti EBANAS, soal yang sedang dan sebagian yang mudah dapat digunakan. Untuk keperluan sehari-hari soal yang sangat mudah atau sangat sukar sebaiknya dibuang saja.

2. Analisis Daya Pembeda

Yang dimaksud dengan daya pembeda adalah kemampuan suatu soal untuk membedakan antara siswa yang mampu dengan yang kurang mampu, antara siswa yang termasuk kelompok unggul dengan kelompok asor. Dengan kata lain daya pembeda merupakan kemampuan suatu soal untuk membedakan antara siswa yang menguasai dengan siswa yang tidak menguasai materi pelajaran yang ditanyakan dalam soal.

Langkah yang harus ditempuh untuk melakukan analisis ini adalah:

- Susun lembar jawaban yang telah di skor mulai dari yang terbesar (di atas) sampai yang terkecil (di bawah).---→ diranking.
- Tentukan kelompok siswa yang termasuk kelompok unggul (uper group) sebanyak 27% sampai 33% dari peserta, dan kelompok siswa yang termasuk kelompok asor (lower group) dengan jumlah yang sama dengan kelompok unggul.
- Hitung jumlah siswa yang menjawab betul untuk setiap soal, demikian juga untuk kelompok asor
- Hasil perhitungan itu masukan kedalam rumus:

$$DP = \frac{\Sigma Bu - \Sigma Ba}{N}$$

DP = daya pembeda soal yang dianalisis

ΣBu = jumlah siswa yang menjawab betul pada kelompok unggul

ΣBa = jumlah siswa yang menjawab betul pada kelompok asor

N = jumlah siswa kelompok unggul atau kelompok asor

Daya pembeda soal akan berada pada interval -1 sampai dengan 1. Semakin tinggi daya pembeda soal, semakin kuat soal itu membedakan kelompok unggul dengan kelompok asor, dengan demikian soal tersebut semakin baik mutunya. Sebaliknya, semakin kecil daya pembeda maka semakin lemah soal itu membedakan kelompok unggul dengan kelompok asor. Daya pembeda yang negatif (< 0), berarti kelompok asor lebih banyak yang betul. Ini merupakan soal

yang jelek. Soal yang dianggap baik adalah soal yang memiliki daya pembeda 0,25 ke atas.

Kriteria untuk daya pembeda secara lengkap adalah sebagai berikut:

- 0,00 - 0,19 = Jelek
- 0,20 - 0,29 = Cukup
- 0,30 - 0,39 = Baik
- 0,40 ke atas = Baik sekali

Contoh

Suatu tes diikuti oleh 50 orang siswa. Maka jumlah siswa yang termasuk kelompok unggul (antara 27% sd. 33%) adalah antara 14 sampai 17 orang. Setelah diurutkan skornya ambil saja 15 orang (ranking 15 terbaik). Demikian juga untuk kelompok asor (jumlahnya sama), ambil 15 orang ranking terjelek. Setelah dikelompokkan masukan data yang diperoleh kedalam tabel berikut.

Nomor Soal	ΣBu	ΣBa	$(\Sigma Bu - \Sigma Ba)/N$	Tafsiran	Rekomendasi
1	15	3	$12/15 = 0,80$	Baik sekali	Digunakan
2	8	8	$0/15 = 0,00$	Jelek	Dibuang
3	7	13	$-6/15 = -0,4$	Jelek/terbalik	Dibuang
4	7	3	$4/15 = 0,27$	Cukup	Direvisi
5	12	9	$3/15 = 0,2$	Jelek	Revisi
Dst.					

Catatan: untuk menguji tingkat kesukaran juga bisa menggunakan pendekatan ini, dengan rumus:

$$DP = \frac{\Sigma Bu + \Sigma Ba}{2N}$$

3. Penyebaran (distribusi jawaban)

Analisis distribusi jawaban bertujuan untuk menganalisis berfungsi tidaknya jawaban yang disediakan (baik kunci jawaban maupun distractor/pengecoh). Suatu pilihan jawaban (pengecoh) dapat dikatakan berfungsi apabila:

- a. Paling tidak dipilih oleh 2,5% peserta tes.
- b. Pengecoh lebih banyak dipilih oleh kelompok bawah.

4. Analisis Keباikan Perangkat Tes

Selain analisis soal seperti diatas, hal lain yang harus diuji adalah analisis perangkat tes secara keseluruhan. Di sini ada dua hal yang harus diuji, yaitu validitas dan reliabilitas tes.

- a. Analisis validitas

Seperti suda dikatakan bahwa validitas merujuk kepada ketepatan suatu tes dalam mengukur aspek yang hendak diukur. Ada beberapa jenis validitas.

1) Validitas permukaan (*face validity*)

Validitas ini dapat diketahui dengan cara analisis rasional. Hal yang dianalisis dalam validitas ini adalah:

- a) Apakah bahasa dan susunan kalimat pada tiap butir soal cukup jelas dan sesuai dengan kemampuan siswa ?
- b) Apakah isi jawaban yang diminta tidak membingungkan ?
- c) Apakah cara menjawab sudah dipahami siswa ?
- d) Apakah tes (soal-soalnya) itu telah disusun berdasarkan kaidah penulisan soal ?

2) Validitas Isi

Validitas ini juga dapat diketahui dengan analisis rasional, yaitu mencocokkan setiap butir soal dengan kisi yang telah dikembangkan berdasarkan GBPP. Yang diuji adalah

- a) Apakah keseluruhan tes telah sesuai dengan kisi-kisi ?
- b) Apakah tiap butir soal tidak menuntut siswa untuk menguasai bahan diluar mata pelajaran yang bersangkutan ?

3) Validitas Kriteria (Criterion validity)

Validitas ini diketahui dengan cara empirik, yaitu menghitung koefisien korelasi antara tes bersangkutan dengan tes lain sebagai kriterianya. Tes lain yang dapat dijadikan kriteria adalah tes yang sudah dianggap valid. Misalnya skor tes bahasa Inggris dikorelasikan dengan skor tes bahasa Inggris yang sudah dibakukan. Kesulitannya adalah mencari skor tes yang akan dijadikan kriteria.

Contoh menghitung koefisien validitas kriteria. Misalnya kita akan mengkorelasikan skor tes biologi (hasil dari tes yang baru kita buat) sebagai variabel X dengan skor tes biologi (hasil dari tes lain yang sudah dianggap baku) sebagai variabel Y sebagai berikut:

Siswa	Skor var. X	Skor Var. Y	X	Y	XY
A	65	76	4225	5776	4940
B	43	56	1849	3136	2408
C	78	70	6084	4900	5460
D	87	89	7569	7921	7743
E	43	59	1849	3481	2537
F	65	70	4225	4900	4550
G	38	34	1444	1156	1292
H	70	81	4900	6561	5670
I	56	67	3136	4489	3752
J	85	84	7225	7056	7140
K	64	70	4096	4900	4480
N = 11	$\Sigma X = 694$	$\Sigma Y = 756$	$\Sigma X^2 = 46602$	$\Sigma Y^2 = 54276$	49972

Harga-harga di atas dimasukkan kedalam rumus Pearsons Product Moment

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

$$r_{xy} = \frac{11 (49972) - (694) (756)}{\sqrt{\{ 11 (46602) - (694)^2\} \{ 11 (54276) - (756)^2\}}}$$

$$= 0,89 \rightarrow (0,70 \text{ s.d. } 0,90 \text{ korelasi tinggi})$$

Angka 0,89 yang dihasilkan menunjukkan bahwa antara variabel X dengan variabel Y di atas memiliki korelasi (keterkaitan) yang tinggi. Hal ini dapat ditafsirkan pula bahwa tes yang dianalisis di atas memiliki validitas yang tinggi. Masih ada jenis validitas lain yang masih harus dihitung (biasanya pada tes psikologis). Jenis validitas tersebut adalah:

- Validitas ramalan (predictive validity) untuk mengetahui sejauhmana skor tes yang bersangkutan dapat digunakan untuk meramalkan keberhasilan seseorang di masa yang akan datang dalam bidang tertentu.
- Validitas konstruk (Construct validity), yaitu mengetahui apakah tes yang bersangkutan benar-benar mengukur konstruk tertentu.

b. Analisis Reliabilitas

Dimuka sudah dikatakan bahwa reliabilitas merujuk kepada keajegan suatu tes dalam menghasilkan skor yang relatif konsisten. Ini berarti bahwa tes yang reliabel akan mampu memberikan skor yang relatif konstan walau pun diberikan pada situasi yang berbeda-beda. Ada tiga cara untuk mengetahui reliabilitas, yang prinsipnya adalah menghitung indeks korelasi. Tiga cara tersebut adalah;

1) Metoda Tes Ulang (Test-retest Method)

Metoda ini dilakukan dengan memberikan tes dua kali terhadap subjek dengan jangka waktu tertentu. Skor hasil tes pertama dikorelasikan dengan skor tes kedua. Misalnya hasil dua kali tes terhadap 12 orang siswa sebagai berikut:

Nama Siswa	Skor hasil tes pertama (X)	Skor hasil tes kedua (Y)
A	45	49
B	78	75
C	54	63
D	65	68
E	76	72
F	89	72
G	33	43

H	65	67
I	71	74
J	67	62
K	56	63
L	80	80

Dari data di atas, diperoleh $N = 12$ $X = 779$ $Y = 788$ $X^2 = 53327$ $Y^2 = 53014$ dan $XY = 52867$. Dengan menggunakan rumus perhitungan seperti pada contoh perhitungan validitas (Pearsons Product Moment) maka diperoleh angka koefisien reliabilitas = 0,91

2) *Metoda tes paralel (Parallel Test Method)*

Cara ini menuntut dua tes yang paralel, yaitu tes yang disusun dengan tujuan sama (mengukur hal yang sama tetapi berbeda dalam redaksi atau kalimatnya). Kedua tes ini diberikan kepada siswa yang sama dengan waktu yang berbeda beberapa hari saja. Kemudian skor yang diperoleh dari masing-masing tes itu dihitung indeks korelasinya (seperti di atas). Koefisien korelasi yang diperoleh menunjukkan tingkat reliabilitasnya.

3) *Teknik Belah Dua (Split-half Method)*

Cara ini paling mudah, yaitu dengan cara membagi dua skor total yang diperoleh dari satu kali tes. Pembagian skor itu bisa dengan cara memisahkan skor dari soal-soal yang bernomor ganjil dengan skor dari soal-soal nomor genap, atau dengan cara membelah dua (setengah-setengah). Masing-masing bagian dijadikan satu variabel. Kemudian angka dari masing-masing variabel itu dihitung korelasinya seperti pada contoh di atas, hasilnya adalah koefisien korelasi r_{hh} atau koefisien korelasi ganjil-genap. Karena tes dibelah menjadi dua, maka untuk mendapatkan korelasi seluruh tes, koefisien korelasinya ganjil genap itu dihitung lagi, menjadi r_{tt} dengan rumus:

$$r_{tt} = \frac{2 \times r_{hh}}{1 + r_{hh}}$$

Contoh penggunaan rumus tersebut adalah:

Nama siswa	Skor kelompok soal gasal (X)	Skor kelompok soal genap (Y)	Skor total (Ganjil + Genap)
A	20	25	45
B	38	40	78
C	28	26	54
D	33	32	65
E	33	43	76
F	44	45	89
G	17	16	33
H	33	32	65

I	36	35	71
J	35	32	67
K	30	26	56
L	42	38	80
Σ	389	390	779

Dengan menggunakan rumus korelasi Pearson, maka diperoleh harga r_{xy} atau $r_{hh} = 0,88$. Maka koefisien korelasi atau indeks reliabilitas seluruh tes (r_{tt}) dengan rumus di atas adalah;

$$r_{tt} = \frac{2 \times 0,88}{1 + 0,88} = \frac{1,72}{1,88} = 0,94$$

III. MENGOLAH SKOR

Pada bagian ini akan diuraikan bagaimana mengolah skor yang diperoleh (skor mentah) menjadi skor matang. Skor matang ini sering disebut sebagai nilai akhir.

Ada beberapa jenis skor matang (skala nilai), yaitu nilai skala 100, skala 10, skor Z, skor T dan skor/nilai skala abjad (skala 0 – 4). Di perguruan tinggi jenis nilai akhir (skor matang) yang digunakan adalah bentuk yang terakhir, yaitu nilai skala 0 – 4 atau skala abjad (A – E). Di sini akan dibahas/dicontohkan cara mengubah skor mentah kedalam skor matang skala abjad saja (0 – 4).

Untuk mengubah skor mentah kedalam skor matang dalam skala 0 – 4 dapat digunakan tabel konversi sebagai berikut:

SKALA SKOR MENTAH	SKALA NILAI MATANG	
	0 – 4	Dengan huruf
$X + 1,50 \text{ SD}$	4	A
$X + 0,50 \text{ SD}$	3	B
$X - 0,50 \text{ SD}$	2	C
$X - 1,50 \text{ SD}$	1	D

Catatan: X = rata-rata

SD = standar deviasi

Rata-rata ada dua macam, yaitu rata-rata ideal dan rata-rata aktual:

$X \text{ ideal} = \frac{1}{2}$ dari skor ideal

SD ideal = $\frac{1}{3}$ dari rata-rata ideal

Skor ideal adalah skor maksimum yang bisa dicapai oleh mahasiswa, jika semua soal dijawab dengan betul.

Misalnya skor ideal untuk 50 butir soal pilihan ganda adalah 50. Maka dapat dihitung:

$$X \text{ idealnya} = \frac{1}{2} \times 50 \\ = 25$$

$$SD \text{ idealnya} = \frac{1}{3} \times 25 \\ = 8,33$$

Dari hasil perhitungan di atas, maka tabel konversi di atas dapat diubah menjadi :

SKALA SKOR MENTAH	RENTANGAN SKOR	SKALA NILAI MATANG	
		0 – 4	Dengan huruf
$25 + 1,5 (8,33) = 37,49$	≥ 37	4	A
$25 + 0,5 (8,33) = 29,17$	29 – 36	3	B
$25 - 0,5 (8,33) = 20,84$	21 – 28	2	C
$25 - 1,5 (8,33) = 12,50$	12 – 20	1	D
	≤ 11	0	E/G

Dengan menggunakan tabel di atas, maka seorang mahasiswa yang mendapat skor 37 ke atas akan mendapat nilai akhir A; mahasiswa yang mendapat skor 35 akan mendapat nilai

akhir B; dan mahasiswa yang mendapat skor 11 ke bawah berarti mendapat nilai E atau G, yang berarti tidak lulus.

Jika menggunakan pendekatan skor aktual, maka rata-rata dan simpangan baku harus dihitung dari skor aktual.

Rumus untuk menghitung rata-rata aktual (untuk data yang tidak dikelompokkan adalah:

$$X \text{ aktual} = \frac{\Sigma X}{N}$$

$$S \text{ aktual} = \sqrt{\frac{\Sigma (X - X)^2}{N - 1}}$$

Misalnya, suatu ujian yang terdiri dari 50 soal pilihan ganda yang diikuti oleh 50 orang mahasiswa mendapatkan skor mentah sebagai berikut;

30	28	27	27	26	26	26	25	25	25
25	25	24	24	24	24	24	23	23	23
23	23	23	23	23	23	23	23	23	22
22	22	22	22	21	21	21	21	21	21
20	20	20	19	19	19	18	17	16	15

Dari data di atas setelah dihitung diperoleh :

Rata-rata = 22,6

Simpangan baku = 2,96

Dengan angka-angka hasil perhitungan di atas, maka tabel konversi di atas dapat berubah menjadi berikut:

SKALA SKOR MENTAH	RENTANGAN SKOR	SKALA NILAI MATANG	
		0 - 4	Dengan huruf
$22,6 + 1,5 (2,96) = 27,04$	≥ 27	4	A
$22,6 + 0,5 (2,96) = 24,08$	24 - 26	3	B
$22,6 - 0,5 (2,96) = 21,12$	21 - 23	2	C
$22,6 - 1,5 (2,96) = 18,16$	18 - 20	1	D
	≤ 17	0	E/G

Dengan tabel di atas, maka mahasiswa yang mendapat skor 27 ke atas akan mendapat nilai akhir A, yang mendapat skor 22 mendapat nilai akhir C, dan yang mendapat skor 17 atau kurang akan mendapat nilai E atau G.

Jika dibandingkan dari dua tabel di atas akan tampak bahwa pada pendekatan ideal skor mahasiswa yang mendapat nilai E lebih kecil dibandingkan dengan dengan pendekatan aktual (idealnya 11 dan aktualnya 17). Jadi pada pendekatan ideal mungkin bisa lulus semua, tetapi pada pendekatan aktual dapat dipastikan akan ada mahasiswa yang tergolong tidak lulus.

Cara mengolah atau mengubah angka mentah kedalam skala 0 – 4 sangat ditentukan oleh pendekatan mana yang akan digunakan (PAP atau PAN, atau pendekatan skor aktual atau skor ideal). Namun pada kedua pendekatan ini ada kesamaannya , yaitu menentukan atau menghitung batas lulus. Batas lulus tersebut dapat dihitung dengan rumus:

$$\text{Batas Lulus (Passing Grade)} = X + 0,25 \text{ SD}$$

Skor batas lulus dengan rumus di atas akan ekuivalen dengan nilai 6 pada standar nilai 0 – 10.

Jika kita sudah menghitung batas lulus (PG) baik yang ideal maupun aktual, maka tabel konversi di atas tidak perlu digunakan lagi. Mahasiswa yang skor mentahnya berada di bawah batas lulus, langsung diberi nilai E atau G. Untuk menentukan kelompok nilai yang lain (A sampai D), maka rentangan yang diperoleh dibagi empat.

Misal, dari data aktual di atas, maka batas lulus dapat dihitung yaitu

$$\begin{aligned} \text{PG} &= 22,6 + 0,25 (2,96) \\ &= 23,34 \end{aligned}$$

Batas lulus di atas menyatakan bahwa mahasiswa yang mendapat skor kurang dari 23 itu dinyatakan tidak lulus. Maka rentangan skor yang dapat menduduki nilai A – D (jika nilai D dianggap lulus) adalah 23 sampai skor tertinggi (dalam hal ini 30). Maka pembagiannya:

$$\begin{aligned} \text{Rentang} &= 30 - 23 \\ &= 8 \end{aligned}$$

$$\text{Rentang untuk masing-masing nilai} = 8 : 4 = 2$$

Pembagian skor selengkapnya adalah:

NILAI	RENTANG SKOR
A	29 - 30
B	27 - 28
C	25 - 26
D	23 - 24

Dalam pendekatan Belajar Tuntas, maka batas lulus akan sama dengan batas minimal tingkat penguasaan, mungkin 60 persen dari soal, atau mungkin 70 persen dari soal, atau yang lainnya.

Penafsiran Hasil Ujian.

Prosedur pengolahan seperti di atas, biasanya dilakukan dalam penilaian ujian akhir semester (lebih bersifat sumatif). Sedangkan dalam penilaian formatif (UTS), yang tujuannya lebih terarah kepada perbaikan proses belajar mengajar, maka batas lulus biasanya ditetapkan secara purposif. Misalnya menggunakan ukuran persentase tingkat penguasaan materi. Kriterianya adalah sebagai berikut (Dirjen Dikti, Depdikbud, 1980)

Derajat Penguasaan	Nilai Akhir
--------------------	-------------

90 % - 100 %	A
80% - 89%	B
65% - 79%	C
55% - 64%	D
≤ 55%	E

Dengan menggunakan kriteria di atas, maka untuk mengkonversikan skor mentah mahasiswa menjadi skor matang, dilakukan dengan cara membagi skor mentah dengan skor maksimal ideal kali 100%. Batas lulusnya adalah 55%.

$$\text{Penguasaan} = \frac{\text{Skor mentah}}{\text{Skor ideal}} \times 100 \%$$

Batas lulus untuk 50 soal pilihan ganda yang masing-masing dibobot 1 dapat dihitung dengan rumus :

$$\frac{55}{100} \times 50 = 28 \text{ soal}$$

Jadi minimal seorang mahasiswa dinyatakan lulus jika dia paling sedikit dapat menyelesaikan 28 soal (Untuk contoh di atas).

Misalnya mahasiswa yang mendapat skor mentah 28 dari 50 soal maka tingkat penguasaannya adalah :

$$\frac{28}{50} \times 100\% = 56\% \text{ (Nilai D).}$$

Analisis Ketuntasan

Hal lain yang perlu dianalisis adalah tingkat ketuntasan mahasiswa dalam menguasai materi perkuliahan. Indikatornya adalah pencapaian penguasaan atas materi yang dituangkan dalam soal. Biasanya satu soal mengungkap satu tujuan pembelajaran yang notabene mengungkap satu satuan materi perkuliahan. Ketuntasan ini dapat bersifat ketuntasan individual dan ketuntasan kelompok

Misalnya pada setengah semester (UTS) kita mempunyai 4 sub pokok bahasan. Masing-masing sub pokok bahasan memiliki empat kehususan materi. Sehingga dalam setengah semester tersebut kita mempunyai 16 tujuan khusus. Untuk setiap tujuan dibuat satu soal-satu soal. Untuk UTS tersebut dibuat kisi-kisi soal sebagai berikut:

Sub Pokok Bahasan	Materi Pokok	Tujuan Pembelajaran	No. Soal
A	A-1	A-1-1	1
	A-2	A-2-2	2
	A-3	A-3-3	3
	A-4	A-4-4	4
B	B-1	B-1-1	5
	B-2	B-2-2	6
	B-3	B-3-3	7
	B-4	B-4-4	8

C	C-1	C-1-1	9
	C-2	C-2-2	10
	C-3	C-3-3	11
	C-4	C-4-4	12
D	D-1	D-1-1	13
	D-2	D-2-2	14
	D-3	D-3-3	15
	D-4	D-4-4	16

Dari hasil ulangan diperoleh jawaban siswa sebagai berikut:

Siswa	Nomor Soal																Σ	%
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		
Aa	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	6
Bb	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	19
Cc	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	4	25
Dd	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	5	31
Ee	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	6	38
Ff	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	8	50
Gg	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	9	56
Hh	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	10	62
Ii	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	13	81
Jj	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100
Σ	10	9	8	7	6	5	4	3	2	1	2	2	1	1	5	10		
%	100	90	80	70	60	50	40	30	20	10	20	20	10	10	50	100		

Analisis Individual

Tujuan analisis ini adalah untuk melihat siswa-siswa mana saja yang sudah mencapai tingkat ketuntasan dan siswa mana yang belum. Dari data di atas, siswa Aa hanya betul satu soal saja (no. 1) dari 16 soal yang disajikan, ini berarti tingkat ketuntasan dia hanya 6% saja. Siswa Ee hanya betul 6 soal dari 16, berarti tingkat ketuntasan dia 38%. Ini berarti siswa Aa mengalami kesulitan yang paling banyak.

Jika kita menggunakan kriteria ketuntasan individual minimal 75% (sama dengan 12 soal), maka siswa yang dinyatakan menguasai materi di atas hanya 2 orang, yaitu siswa Ii (betul 13 soal = 81%) dan siswa Jj (betul semua atau 100%). Dengan kriteria ini berarti kepada siswa yang belum mencapai tingkat ketuntasan (yaitu 8 orang), harus di berikan remedial, dan siswa-siswa inilah yang diduga sebagai siswa yang mengalami *kesulitan belajar*. Siswa-siswa inilah yang harus diberikan bimbingan belajar.

Suatu pengajaran dikatakan tuntas jika terdapat minimal 65% siswa dapat mencapai tingkat ketuntasan. Data diatas menunjukkan hanya 20% saja siswa yang mencapai tingkat ketuntasan, dan ini berarti menunjukkan kasus kesulitan belajar kelompok. Ini perlu dicari, apa penyebabnya ?

Jika dianalisis dari sisi efektivitas PBM secara menyeluruh, bila sebagian besar siswa mengalami kesulitan (kasus kelompok) maka penyebab kesulitan itu dapat dipastikan besumber pada kesalahan metodologis. Sedangkan jika kasusnya individual

(hanya sebagian kecil saja siswa yang mengalami kesulitan) sumber kelemahan atau kesalahan akan terdapat pada individu (misalnya kelemahan intelektual, atau individu punya masalah pribadi).

Analisis Ketuntasan Materi

Tujuan analisis ini adalah untuk melihat apakah materi yang diajarkan dikuasai oleh siswa, dan bagian materi mana yang masih belum dikuasai. Jika pada analisis individual, pada tabel di atas kita melihatnya ke arah kiri/horizontal (siswa), maka pada analisis ini cara melihatnya secara vertikal (soal demi soal).

Misalnya, pada soal nomor 1 dan soal nomor 16 semua siswa dapat menjawab dengan betul. Ini berarti materi yang terkait dengan soal nomor 1 dan 16 sudah dikuasai oleh semua siswa. Bila kita menggunakan standar 70% (artinya pada setiap soal/materi/tujuan harus dikuasai oleh minimal 70% siswa), maka dari data di atas hanya ada 5 materi/soal/ tujuan saja (31%) yang sudah dikuasai oleh siswa, yaitu soal/materi/tujuan nomor 1, 2, 3, 4, dan 16. Sedangkan soal-soal yang lain (11 soal/materi/tujuan) belum sepenuhnya dikuasai oleh siswa.

Bila kita melihat kisi-kisi, soal nomor 1 sampai 4 mengukur sub pokok bahasan A atau pokok materi A-1 sampai A-4 atau mengukur tujuan A-1-1 sampai A-1- 4, dan tingkat penguasaan atas soal ini sudah di atas 70%, ini berarti bahwa semua materi dan tujuan belajar yang terumuskan dari sub pokok bahasan A sudah dikuasai. Sedangkan soal nomor 5 sampai 15 yang mengungkap materi B-1 sampai D-4 atau tujuan pembelajaran B-1-1 sampai D-4-4, belum dikuasai sepenuhnya oleh siswa (tingkat ketuntasannya kurang dari 70%). Ini berarti materi yang terangkum dalam sub pokok bahasan B-1 sampai D-4 harus diajarkan ulang (remedial teaching) kepada semua siswa.

IV. PELAKSANAAN EVALUASI

A. Mengevaluasi Hasil Belajar

Seperti sudah diungkapkan pada bagian sebelumnya, bahwa perilaku hasil belajar itu (Benyamin Bloom, dkk.) dapat berwujud pengetahuan (kognitif), sikap (afektif), dan keterampilan (psikomotor). Masing-masing aspek tersebut menuntut alat tes dan bentuk pelaksanaan yang berbeda. Hasil belajar yang berupa pengetahuan (kognitif) lebih menuntut cara penilaian yang tertulis, hasil belajar yang berupa sikap (afektif) lebih tepat dengan observasi atau pun tertulis dengan menggunakan alat pedoman skala penilaian (skala sikap, dsb.), dan hasil belajar yang berupa keterampilan (psikomotor) lebih menuntut cara penilaian perbuatan atau praktek.

1. Segi-segi yang Dievaluasi

Segi-segi hasil belajar yang dapat dievaluasi mencakup hal-hal di atas. Pada bagian ini akan diuraikan serba singkat dua hal saja, yaitu aspek kognitif dan afektif, sedangkan aspek psikomotor akan banyak disinggung pada bagian lain.

Ranah Kognitif

Menurut taksonomi Bloom, dalam *aspek kognitif* terdapat enam tahapan tataran perilaku. Keenam jenis perilaku ini bersifat bertahap atau berjenjang. Beturut-turut dari yang paling rendah sampai kepada yang paling tinggi adalah: *pengetahuan (knowledge)*, *pemahaman (comprehension)*, *aplikasi*, *analisis*, *sintesis*, dan *evaluasi*.

a. *Pengetahuan (knowledge)*, adalah kemampuan mengingat kembali hal-hal yang telah diperoleh atau dialami yang khusus maupun yang umum. Mengingat kembali fakta, konsep, generalisasi, prinsip, dalil, metoda, proses, pola, struktur atau susunan, dsb. Untuk keperluan penilaian maka situasi mengingat kembali tersebut tidak sekedar membawa hal-hal yang diinginkan itu ke dalam ingatan, akan tetapi lebih dari itu. Dalam hal ini lebih ditekankan pada proses psikologis dari mengingat tersebut. Aspek pengetahuan ini meliputi pengetahuan akan hal yang khusus (simbol, dsb.), istilah, fakta-fakta, cara-cara bekerja sesuatu, kecenderungan dan urutan, klasifikasi dan kategori, prinsip-prinsip, dan sebagainya. Secara operasional kata-kata yang dapat mengungkap aspek ini diantaranya: menyebutkan, menyatakan kembali, mendefinisikan, menunjukkan, menjelaskan istilah, dan menyebutkan urutan.

b. *Pemahaman*, adalah kemampuan memberi arti atau pengertian yang derajatnya paling rendah. Pemahaman merupakan kemampuan pengetahuan, ide atau gagasan seseorang akan apa yang sedang dikomunikasikan tanpa melihat materi yang lain, atau kemampuan berpikir seseorang untuk memahami bahan-bahan atau materi yang sedang dipelajari. Kemampuan ini mencakup penterjemahan, tafsiran, dan eksplorasi seseorang tentang sesuatu. Secara operasional kata-kata yang dapat mengungkap hal ini diantaranya: menjelaskan, merumuskan dengan kata-kata sendiri, memberi contoh.

c. *Penerapan (aplikasi)*, adalah kemampuan menggunakan abstraksi-abstraksi dalam situasi khusus dan konkrit, merupakan kemampuan untuk menggunakan teori-

teori, prinsip-prinsip, dan rumus-rumus dalam situasi konkrit. Misalnya mengaplikasi fenomena-fenomena yang dibahas di dalam sebuah makalah untuk membuat makalah-makalah yang lain. Secara operasional kata-kata yang menunjukkan kemampuan aplikasi diantaranya: menggunakan, mengoperasikan, menghitung, menghasilkan.

d. *Analisis*, adalah kemampuan memecahkan suatu kesatuan kedalam unsur-unsur atau bagian-bagian sedemikian rupa sehingga hierarkhi ide-idenya menjadi jelas. Jadi kemampuan ini merupakan kemampuan menguraikan sesuatu keseluruhan atau suatu sistem hubungan kedalam unsur-unsur yang membentuknya. Analisis ini mencakup analisis unsur-unsur, analisis hubungan-hubungan, dan analisis terhadap prinsip-prinsip yang terorganisasikan. Misalnya kemampuan mengenali aspek-aspek pembangunan pemukiman baru yang memenuhi syarat kesehatan, sosial, budaya, dan seni. Kata kerja operasional yang menunjukkan kemampuan ini diantaranya: menguraikan, memerinci, memisah-misahkan, mengidentifikasi, memilih.

e. *Sintesis*, adalah kemampuan memadukan unsur-unsur dan bagian-bagian sedemikian rupa sehingga membentuk suatu keseluruhan. Sintesis ini meliputi proses bekerja dengan potongan-potongan, bagian-bagian atau unsur-unsur sehingga menjadi suatu kesatuan yang sebelumnya belum ada. Sintesis meliputi hal: membuat komunikasi yang unik (kemampuan menulis dengan menggunakan ide-ide yang jitu), membuat suatu rencana atau serentetan kegiatan (merencanakan pembangunan pemukiman yang baru), menurunkan seperangkat hubungan-hubungan abstrak (kemampuan merumuskan hipotesis berdasarkan analisis terhadap faktor-faktor yang terlibat). Kata kerja operasional yang mengungkap hal ini diantaranya: menciptakan, menyusun, mengorganisasikan.

f. *Evaluasi*, yaitu kemampuan memberikan penimbangan kepada nilai materi atau metoda tertentu untuk maksud yang tertentu pula, kemampuan untuk mempertimbangkan suatu ide, situasi, nilai, metoda atas dasar kriteria tertentu. Penimbangan ini dapat bersifat kualitatif maupun kuantitatif. Misalnya seorang karya siswa memutuskan untuk membangun pondasi di suatu tempat cukup digunakan batu kali ukuran 40 cm x 60 cm dan tidak mesti dengan komposisi beton cakar ayam. Kata kerja operasional yang merujuk kepada hal ini diantaranya : menilai, mempertimbangkan.

Penentuan aspek mana atau aspek apa yang perlu dievaluasi, dapat ditentukan berdasarkan rumusan tujuan pembelajaran khusus (TPK) yang telah dirumuskan dalam satuan acara perkuliahan (SAP) dan sifat materi yang diajarkan.

Ranah Afektif

Ranah atau *aspek afektif* merupakan perilaku yang berhubungan dengan penghayatan individu terhadap sesuatu (lingkungan) dengan jalan menerima, merespon, menilai, mengorganisasikan, dan menginternalisasikan sesuatu.

a. *Penerimaan atau menaruh perhatian*. Dalam proses belajar seseorang menyadari adanya suatu fenomena. Kemudian setelah melalui proses belajar, seseorang dapat menjadi peka terhadap fenomena dan rangsangan tertentu, yakni mau menerima atau mau memperhatikannya. Tataran ini mencakup aspek: kesadaran, keihlasan menerima, dan perhatian yang terpimpin atau yang terpilih. Kata kerja operasional yang mencerminkan hal ini seperti mengikuti, menjawab, memilih, menunjukkan.

b. *Merespon*. Pada tahap ini seseorang tidak hanya mau memperhatikan fenomena, akan tetapi berpartisipasi aktif terhadap sesuatu yang terjadi. Ia berkeinginan

dan memiliki kepuasan untuk merespon. Tataran ini mencakup hal yang lebih kecil seperti mengizinkan merespon (tunduk, menurut), keihlasan merespon (berbuat secara sukarela), dan kepuasan dalam merespon (melibatkan emosional). Kegiatan yang termasuk tataran ini diantaranya menyambut, memperbincangkan, menyesuaikan, menyetujui.

c. *Menghargai*. Dalam tataran ini seseorang sudah memberikan nilai tertentu kepada sesuatu yang diterimanya. Ia tidak hanya menerima atau menyetujui, akan tetapi sudah memberikan penghargaan dan makna terhadap sesuatu, ada unsur internalisasi seperangkat nilai-nilai yang khusus dan ideal. Tataran ini mencakup: penerimaan terhadap nilai-nilai sesuatu (keyakin, ada penerimaan emosional), preferensi nilai (merasa terlibat untuk mencapai dan memburunya), dan pelibatan terhadap nilai (mengabdikan diri pada suatu paham). Indikator perilaku pada tataran ini diantaranya: mengusulkan, memprakarsai, mengidentifikasi diri, mengakui dengan tulus.

d. *Pengorganisasian*. Pada tahap ini orang yang belajar berhasil menginternalisasikan nilai-nilai dengan nilai yang sudah ada pada dirinya (nilai menjadi bagian dari dirinya). Bagian dari tataran ini adalah: konseptualisasi nilai (menghubungkan nilai lama dengan nilai baru yang akan ia pegang secara abstrak dan simbolik) dan kemudian mengorganisasikan sistem nilai (menata dan mengatur sistem nilai yang baru dengan yang sudah ada).

e. *Karakterisasi sistem nilai*. Pada tahap ini internalisasi sistem nilai telah mendapat tempat dihierarkhi nilai individu dan akhirnya nilai baru yang diterima telah mengendalikan perilaku individu dalam jangka waktu tertentu. Kata-kata kerja yang merupakan indikator tahap ini diantaranya: mengintegrasikan, mempertahankan, menyelaraskan.

Pendek kata aspek afektif ini mencakup penerimaan nilai, pemikiran nilai dan pemasukan nilai secara selektif oleh individu kedalam dirinya. Oleh karena itu aspek afektif sering disejajarkan dengan sikap yang merupakan kecenderungan seseorang untuk melakukan tindakan atas dasar nilai-nilai tertentu.

2. Cara Mengevaluasi Hasil Belajar

Penilaian hasil belajar yang telah diuraikan di atas (aspek kognitif dan afektif) secara umum dapat dievaluasi baik dengan menggunakan tes tertulis maupun tes lisan. Bentuk tes tertulis ini dapat dilakukan pada setiap akhir KBM, setiap akhir satu satuan acara perkuliahan (formatif), setelah selesai beberapa satuan acara perkuliahan (Ujian tengah semester), atau pun di akhir program semester (ujian akhir semester). Bentuk soal yang digunakan dapat berupa soal-soal tulisan objektif atau pun uraian (essay), sesuai dengan tujuan pembelajaran yang sudah dirumuskan. Soal-soal yang diberikan itu dapat berupa kuis atau pun tes yang terencana seperti sudah dijelaskan pada bagian sebelumnya.

Secara khusus, untuk mengevaluasi aspek atau tataran afektif lebih disarankan (lebih tepat) jika menggunakan tes tertulis dengan menggunakan model skala (scaling techniques model), seperti rating scale atau skala sikap.

Skala sikap biasanya berupa sejumlah pernyataan yang mengungkap tentang objek atau situasi tertentu. Terhadap setiap pernyataan tersebut para mahasiswa diminta untuk memberikan persetujuan dalam skala kontinum, yang dimulai skala sangat tidak setuju sampai kepada skala sangat setuju.

B. Mengevaluasi Proses Belajar di Kelas

1. Evaluasi Penyajian Laporan/Tugas

Dalam proses belajar-mengajar dosen sering memberikan tugas kepada para mahasiswa. Mungkin tugas tersebut berupa laporan hasil kerja atau hasil kunjungan (survey), mungkin juga berupa tugas lain (membuat makalah, misalnya). Mungkin tugas yang diberikan tersebut, harus disajikan (diseminarkan) oleh mahasiswa di depan kelas atau laporan tersebut tidak mesti disajikan di kelas tetapi hanya diperiksa oleh dosen.

Dalam hal yang pertama (tugas disajikan/diseminarkan di kelas), maka ada dua aspek yang dapat dievaluasi. Pertama adalah kemampuan/keterampilan mahasiswa dalam menyajikan laporan tersebut dihadapan teman-temannya, dan kedua adalah format/bentuk atau isi laporannya itu sendiri. Jika yang kedua yang dilakukan (hanya memberikan laporan saja atau berupa makalahnya saja), maka yang dinilai itu hanya bentuk atau isi laporan yang diberikan.

a. Menilai Penyajian

Langkah pertama yang harus ditempun adalah menentukan aspek yang dinilai dan kedua menentukan bagaimana cara menilainya.

Aspek yang perlu dinilai dalam penyajian laporan di depan kelas di antaranya meliputi: kemampuan menyajikan laporan dan kemampuan mempertahankan isi atau ide yang tertuang dalam laporan (makalah) tersebut. Aspek-aspek tersebut selengkapnya dapat dilihat pada tabel berikut ini.

FORMAT PENILAIAN PENYAJIAN LAPORAN/MAKALAH

Nama Karyasiswa :
NIM :
Jurusan/Program :
Semester/Tingkat :

No	Aspek yang dinilai	Rentang nilai			
		1	2	3	4
1	Kemampuan menyajikan makalah/laporan				
	a. Penggunaan Bahasa				
	b. Sistematika berfikir/berbicara				
	c. Sikap dalam menyajikan				
	d. Aspek lainnya				
2	Kemampuan mempertahankan isi laporan/makalah				
	a. Orisinalitas isi laporan/makalah				
	b.				
3	Kesan tampilan secara keseluruhan				
	Jumlah nilai				

$$\text{Nilai Penyajian} = \frac{\text{Jumlah Nilai}}{\text{Jumlah aspek}} \quad \text{Dosen/Penilai}$$

$$= \dots\dots\dots$$

b. Menilai Isi Laporan/Makalah

Aspek yang perlu diperhatikan dari sebuah makalah atau laporan di antaranya adalah : format laporan/makalah (sistematika laporan/makalah, apakah sesuai dengan yang ditentukan), kelengkapan laporan/makalah (apakah laporan/makalah itu lengkap atau tidak), isi laporan/makalah (apa ada unsur duplikasi atau tidak, kemampuan mengungkapkan ide/gagasan runtut atau loncat-loncat), tampilan laporan/makalah (termasuk pengetikan, kerapihan), serta aspek lain yang dianggap perlu. Aspek itu semua dapat dimuat dalam tabel berikut.

FORMAT PENILAIAN LAPORAN/MAKALAH

Nama Karyasiswa :
 NIM :
 Jurusan/Program :
 Semester/Tingkat :

No	Aspek yang dinilai	Rentang nilai			
		1	2	3	4
A	Isi Laporan/makalah				
	1. Sistematika pemikiran				
	2. Keluasan dan kedalaman isi				
	3. Kebenaran dan ketepatan fakta, konsep, teori				
	4. Ketepatan dan kedalaman analisis dan Pembahasan				
	5. Ketepatan dan kebermaknaan kesimpulan dan Rekomendasi				
B	Tampilan laporan/makalah				
	6. Kelengkapan dan sistematika isi				
	7. Bahasa				
	8. Perwajahan dan pengetikan				
	9. lain-lain				
	Jumlah nilai				

$$\text{Nilai Makalah} = \frac{\text{Jumlah Nilai}}{\text{Jumlah aspek}} \quad \text{Dosen/Penilai}$$

$$= \dots\dots\dots$$

Tips menilai laporan/makalah

Sebenarnya menilai laporan atau makalah memiliki prinsip yang sama dengan memeriksa ujian bentuk soal essay. Untuk meningkatkan objektivitas dan keajegan dalam memeriksa mahalah/laporan sebaiknya dilakukan langkah sebagai berikut:

1. Abaikan dulu (Jangan dulu melihat) identitas si pembuat laporan/makalah.
2. Lakukan pemeriksaan berdasarkan kesan umum (jangan dulu memeriksa laporan/makalah aspek demi aspek). Dari hasil penilaian kesan umum itu kelompokkan laporan/makalah berdasarkan judgement (dari yang terbaik sampai kepada yang paling kurang)
3. Setelah dikelompokkan, lakukan penilaian terhadap aspek demi aspek.
4. Isikan nilai setiap aspek pada format penilaian.

2. Evaluasi Diskusi/Seminar Kelas

Diskusi sering dipilih sebagai metoda yang cukup komprehensif untuk menilai proses belajar mengajar atau untuk menilai kemampuan mahasiswa dalam menanggapi/memecahkan suatu masalah. Diskusi dilakukan jika dosen bertujuan untuk melihat dinamika kelompok yang terjadi dalam suatu kelas, melihat kemampuan memimpin seseorang, melihat daya nalar seseorang, atau untuk melihat aktivitas seseorang. Dalam suatu diskusi baik aktivitas pemimpin diskusi maupun aktivitas peserta dapat dinilai. Dari seorang pemimpin diskusi dapat diprediksi kemampuan memimpinnnya, sedangkan dari peserta diskusi/seminar hal yang dapat dinilai adalah keaktifannya.

Format berikut adalah salah satu model untuk melihat aktivitas serta kecenderungan seseorang dalam melakukan atau berpartisipasi dalam diskusi.

FORMAT OBSERVASI PERILAKU DISKUSI

No	Jenis Aktivitas	Nama Peserta						Jml
		A	B	C	D	E	F	
1	Menunjukkan solidaritas							
2	Menunjukkan sikap lunak							
3	Menunjukkan persetujuan							
4	Memberi saran							
5	Memberi pendapat							
6	Memberi informasi							
7	Meminta informasi							
8	Meminta pendapat							
9	Meminta saran							
10	Menunjukkan ketidaksetujuan							
11	Menunjukkan sikap keras							
12	Menunjukkan antagonis							
	Jumlah							

Pada tabel di atas, aspek 1 sampai 3 menunjukkan fungsi sosial emosional(Social Emotional Functions) yang aktif, aspek 4 sampai 6 menunjukkan fungsi tugas (Task

function) yang aktif, aspek 7 sampai 9 menunjukkan fungsi tugas yang pasif, dan 10 sampai 12 menunjukkan fungsi sosial emosional yang pasif. Dari tabel di atas dapat ditafsirkan bahwa seseorang itu lebih cenderung task function yang aktif atau social emotional function-nya yang aktif atau sebaliknya.

C. Evaluasi Proses Belajar di Laboratorium atau lapangan.

Evaluasi yang dilakukan terhadap kegiatan belajar di laboratorium atau lapangan, merupakan jenis evaluasi tindakan (performance test). Tujuannya lebih diarahkan kepada sejauhmana karyasiswa mampu/terampil menggunakan suatu alat tertentu atau untuk menilai karyasiswa menghasilkan sesuatu produk. Dengan demikian maka aspek yang dinilai dari karyasiswa dalam berpraktek ditempat-tempat itu ada dua hal, yaitu (1) kemampuan karyasiswa menggunakan suatu alat, dan (2) hasil dari penggunaan alat tersebut. Untuk aspek kedua, biasanya dituangkan dalam laporan

Dengan melihat konsep di atas, maka langkah yang harus ditempuh adalah menentukan aspek yang perlu dinilai dan bagaimana menilainya. Setidaknya aspek yang perlu dinilai dalam praktek tersebut adalah sebagai berikut:

FORMAT PENILAIAN PRAKTEK DI LABORATORIUM ATAU LAPANGAN

Nama Karyasiswa :
 NIM :
 Jurusan/Program :
 Semester/Tingkat :

No	Aspek yang dinilai	Rentang nilai			
		1	2	3	4
A	Proses kerja				
	1. Penyiapan alat dan bahan				
	2. Pemahaman petunjuk/lembar kerja				
	3. Ketepatan prosedur (langkah)				
	4. Ketelitian, kesungguhan bekerja				
	5. Kerjasama dan tanggung jawab				
	6. Pengamanan dan perawatan alat				
B	Hasil kerja				
	7. Ketuntasan hasil kerja				
	8. Mutu hasil kerja				
	9. Jumlah hasil kerja				
C	Lain-lain				
	Jumlah nilai				

Jumlah Nilai

Nilai Makalah = $\frac{\text{Jumlah Nilai}}{\text{Jumlah aspek}}$ Dosen/Penilai

=

D. Evaluasi Hasil Penugasan

Tugas juga merupakan salah satu metoda mengajar. Tugas dapat berupa tugas lapangan, atau tugas penelaahan buku/referensi. Realisasi dari tugas biasanya diwujudkan dalam bentuk laporan tugas atau makalah. Keduanya hampir mirip, karena itu aspek yang dinilai dan cara penilaiannya juga mirip (memiliki kesamaan).

Seperti sudah disinggung di muka bahwa tugas memiliki aspek keharusan (sesuatu yang diminta/ditentukan) serta ada aspek bersifat asesoris. Aspek yang perlu di nilai terutama berkenaan dengan sesuatu yang diminta tersebut (apa yang harus ada dalam laporan tersebut).

Secara rinci, mungkin hasil laporan di lab dengan hasil laporan di tempat lain (bengkel, studio, lapangan) memiliki perbedaan. Untuk hal ini para dosen diharapkan dapat mengembangkan format tersendiri sesuai dengan kebutuhan.

Berikut ini adalah contoh tugas lapangan, seorang dosen gambar teknik yang menugaskan seorang mahasiswa atau sekelompok mahasiswa untuk merencanakan pembangunan sebuah rumah (gedung) di suatu lokasi. Setidaknya dia akan meminta mahasiswa untuk menyusun dan melaporkan hal-hal sebagai berikut:

- A. Keadaan Tanah/Lahan
 - Struktur/kekuatan tanah
 - Kemiringan (Kondisi tanah)
 - Jenis tanah
- B. Struktur Fondasi
 - Dimensi
 - Bahan
 - Jenis struktur
- C. Struktur Gedung
 - Struktur Beton
 - Struktur rangka
(Untuk fondasi, dinding, atap)
- D. Dinding
 - Jenis
 - Struktur
 - Dimensi
- E. Kusen
 - Jendela
 - Pintu
- F. Atap
 - Struktur
 - Konstruksi
 - Atap
- G. Gambar Teknik
 - 2. Denah
 - Organisasi Ruang

- Dimensi
 - Ukuran-ukuran detail
3. Tampak
 - Representasi bentuk
 - Citra bahan/kesan
 4. Potongan
 - Dimensi ukuran
 - Jenis material
 - Konstruksi
 5. Gambar Tiga Dimensi
 6. Maket

Berdasarkan struktur (sistematika) laporan yang diminta di atas, maka seorang dosen dapat melihat atau menilai baik tidaknya laporan mahasiswa darisegi-segi berikut:

1. Kelengkapan aspek yang diminta. Apakah aspek-aspek yang ditentukan di atas ada atau tidak. Berapa banyakkah tidak adanya ? Semakin banyak aspek yang hilang makin nilainya semakin berkurang.
2. Fisibilitas masing-masing isi atau materi yang dilaporkan. Apakah materi yang dilaporkan fisibel untuk dilakukan sesuai dengan kondisinya ? Semakin banyak isi atau materi yang dilaporkan yang bertentangan atau tidak fisibel untuk dilakukan, maka semakin kurang juga nilai yang diperoleh.
3. Kebenaran model/isi/materi yang direkomendasikan. Apakah model atau materi yang direkomendasikan sesuai dengan teori yang ada ? Atau apakah model yang disarankan (yang dibuat) benar atau salah ?
4. Keabsahan data yang diberikan. Apakah data yang diberikan benar sesuai dengan kenyataan di lapangan (ukuran, dsb.)?
5. Dan sebagainya, sebagaimana layaknya suatu ajuan proyek pembangunan.

E. Penentuan Nilai Akhir

Nilai akhir yang diperoleh oleh seorang karyasiswa dapat dilihat dari:

1. Kelengkapan komponen nilai yang ditentukan. Misalnya komponen tersebut meliputi:
 - a. Kehadiran mengikuti kuliah. Kehadiran kuliah dapat menentukan apakah nilai seseorang dapat diproses atau tidak. Biasanya kehadiran minimal adalah 80 % dari jumlah tatap muka.
 - b. Aktivitas dalam kegiatan belajar di kelas (diskusi atau penyajian laporan).
 - c. Ujian Tengah Semester (UTS)
 - d. Ujian Akhir Semester (UAS)
 - e. Tugas-tugas (Tugas lapangan, laboratorium).

Di sini dipertimbangkan apakah setiap komponen itu ada nilainya, jika komponen tersebut tidak lengkap sampai batas yang ditentukan, maka seorang mahasiswa dapat dinyatakan gugur atau mengulang (remedial).
2. Kualitas dari komponen-komponen tersebut. Baik tidaknya nilai yang diperoleh setiap mahasiswa dalam setiap komponen akan menentukan berapa besar nilai akhir yang diperolehnya.

3. Untuk menentukan nilai akhir dari setiap mahasiswa, setiap komponen penilaian tersebut di kalikan dengan bobot.

Umpamanya:

- a. Kehadiran dipertimbangkan sebagai penentu apakah nilai seseorang dapat diproses atau tidak.
- b. Nilai Aktivitas di kelas diberi bobot 1
- c. Nilai Ujian Tengah Semester diberi bobot 2
- d. Nilai Ujian Akhir Semester diberi bobot 3
- e. Nilai Tugas-tugas *) diberi bobot 1

$$\text{Jumlah Bobot} = 7$$

Misalnya seorang Karyasiswa mendapat perolehan nilai sebagai berikut

- a. Kehadiran Diatas 80 % (memenuhi)
- b. Nilai Aktivitas di kelas $3,25 \times 1 = 3,25$
- c. Nilai Ujian Tengah Semester $2,75 \times 2 = 5,50$
- d. Nilai Ujian Akhir Semester $3,00 \times 3 = 9,00$
- e. Nilai Tugas-tugas $3,5 \times 1 = 3,50$

$$\text{Jumlah nilai} \times \text{Bobot} = 21,25$$

$$\text{Nilai akhir} = \frac{21,25}{7}$$

$$= 3,03 \text{ (B)}$$

*) Tergantung berat/ringannya dan banyak /sedikit tugas

DAFTAR PUSTAKA

- Asmawi Zainul dan Noehi Nasoetion, (1997), *Penilaian Hasil Belajar*, Jakarta: Dirjen Dikti
- Best, John W. (1977), *Research in Education*, Third Edition, Prentice-Hall of Indiana
- Depdikbud, (1981), *Materi Dasar Pendidikan Program Akta Mengajar V: Buku IIIID, Penilaian dalam Pendidikan*, Jakarta: Dirjen Dikti
- Depdikbud, (1992), *Pedoman Penelaahan, Perbaikan, dan Perakitan Soal*, Jakarta: Sisjian-Balitbangdikbud.
- Depdikbud, (1992), *Penyusunan, Penskoran, dan Penggunaan Tes Prestasi Belajar Bentuk Uraian: Buku Panduan Guru*, Jakarta: Sisjian-Balitbangdikbud
- Furon, (1997), *Statistika Terapan untuk Penelitian*, Bandung: CV. Alfabeta
- Furqon, (2000), *Sistem Penilaian Kelas Terpadu (Strategi untuk Meningkatkan Mutu KBM)*, Bandung: FIP FIP UPI
- Gronlund, N.E. (1998), *Assessment of Student Achievement*, Needham Heights, MA: A Viacom Company
- Stiggins, R. J. (1994), *Stugent-Centered Classroom Assessment*, New York: Macmillan College Publishing
- Subino, (1987), *Konstruksi dan Analisis Tes: Suatu Pengantar kepada Tes dan Pengukuran*, Jakarta: P2LPTK
- Sunaryo Kartadinata, dan Karno To, (1991), *Evaluasi Hasil Belajar*, Bandung: IKIP
- Umar, Jahja, dkk. (1997), *Bahan Penataran Pengujian Pendidikan*, Jakarta: Sisjian-Balitbangdikbud

PEDOMAN

EVALUASI PROSES DAN HASIL BELAJAR

YAYA SUNARYA

JURUSAN PSIKOLOGI PENSISIKAN DAN BIMBINGAN
UNIVERSITAS PENDIDIKAN INDONESIA

- Tes, Pengukuran, Penilaian
- a. Jenis evaluasi (formatif, sumatif, diagnostik, dsb)
 - b. Kriteria Evaluasi (PAN dan PAP)
 - c. Macam-macam bentuk, Teknik tipe Evaluasi
- I. Pengembangan Alat Evaluasi
- a. Penyusunan Kisi-kisi

- b. Penyusunan butir-butir soal
- c. Standarisasi alat evaluasi
- II. Pengubahan angka mentah ke angka 0 – 4 (A – E)
- III. Pelaksanaan Evaluasi
 - a. Mengevaluasi Hasil Belajar (KUSIS, UTS, UAS)
 - 1. Segisegi yang dievaluasi
 - 2. Cara Mengevaluasi
 - b. Evaluasi Proses Belajar di Kelas
 - 1. Evaluasi Penyajian Laporan/Tugas (Segi dan cara evaluasi)
 - 2. Evaluasi Diskusi/Seminar Kelas (Segi dan Cara evaluasi)
 - c. Evaluasi PBM di LAB, Bengkel, Studio, lapangan
 - 1. Segi-segi yang dievaluasi
 - 2. Cara mengevaluasi
 - d. Evaluasi Hasil Penugasan
 - 1. Segi yang dievaluasi
 - 2. Cara Mengevaluasi
 - e. Penentuan Nilai Akhir
 - 1. Pembobotan angka
 - 2. Nilai Akhir