

PEMBELAJARAN EVALUASI PENGUKURAN



**Disusun oleh :
Dameria Sinaga**

**PROGRAM PASCA SARJANA
UNIVERSITAS KRISTEN INDONESIA
JAKARTA 2019**

Daftar Isi

Daftar Isi	2
Pendahuluan	3
Evaluasi Pengukuran	6
Distribusi Probabilitas Normal & Logistik	10
Penyelenggara Uji Tes	10
Penilaian Hasil Belajar	12
Penilaian Acuan Patokan (PAP)	13
Penilaian Acuan Norma (PAN)	17
Kriteria Pemberian Penilaian	22
Afeksi	25
Tes Essay	28
Paper and pensil	31
Analisis butir soal	32
Konseptualisasi Definisi nominal Definisi operasional Pengukuran	38
Daftar Pustaka	43

Pendahuluan

Istilah evaluasi pembelajaran sering disamaartikan dengan ujian. Meskipun saling berkaitan, akan tetapi tidak mencakup keseluruhan makna yang sebenarnya. Ujian ulangan harian yang dilakukan guru di kelas atau bahkan ujian akhir sekolah sekalipun, belum dapat menggambarkan esensi evaluasi pembelajaran, terutama bila dikaitkan dengan penerapan kurikulum 2013. Sebab, evaluasi pembelajaran pada dasarnya bukan hanya menilai hasil belajar, tetapi juga proses-proses yang dilalui pendidik dan peserta didik dalam keseluruhan proses pembelajaran.

Istilah tes, pengukuran (measurement), penilaian (assesment) dan evaluasi sering disalahartikan dan disalahgunakan dalam praktik evaluasi. Secara konsepsional istilah-istilah tersebut sebenarnya berbeda satu sama lain, meskipun mempunyai keterkaitan yang sangat erat. Tes adalah pemberian suatu tugas atau rangkaian tugas dalam bentuk soal atau perintah/suruhan lain yang harus dikerjakan oleh peserta didik. Hasil pelaksanaan tugas tersebut digunakan untuk menarik kesimpulan-kesimpulan tertentu terhadap peserta didik.

Pengukuran (measurement) adalah suatu proses untuk menentukan kuantitas daripada sesuatu. Sesuatu itu bisa berarti peserta didik, starategi pembelajaran, sarana prasana sekolah dan sebagainya. Untuk melakukan pengukuran tentu dibutuhkan alat ukur. Dalam bidang pendidikan, psikologi, maupun variabel-variabel sosial lainnya, kegiatan pengukuran biasanya menggunakan tes sebagai alat ukur.

Sedangkan penilaian (assesment) adalah suatu proses atau kegiatan yang sistematis dan berkesinambungan untuk mengumpulkan informasi tentang proses dan hasil belajar peserta didik dalam rangka membuat keputusan-keputusan berdasarkan kriteria dan pertimbangan tertentu (Arifin, 2013:4). Jika dilihat dalam konteks yang lebih luas, keputusan tersebut dapat menyangkut keputusan tentang peserta didik (seperti nilai yang akan diberikan), keputusan tentang kurikulum dan program atau juga keputusan tentang kebijakan pendidikan.

Selanjutnya, istilah evaluasi telah diartikan para ahli dengan cara berbeda meskipun maknanya relatif sama. Guba dan Lincoln (1985:35), misalnya, mengemukakan definisi evaluasi sebagai “a process for describing an evaluand and judging its merit and worth”. Sedangkan Gilbert Sax (1980:18) berpendapat bahwa “evaluation is a process through which a value judgement or decision is made from a variety of observations and from the background and training of the evaluator”.

Dalam buku *Measurement and Evaluation in Education and Psychology* ditulis William A. Mohrens (1984:10) istilah tes, measurement, evaluation dan assesment dijelaskan sebagai berikut:

1. Tes, adalah istilah yang paling sempit pengertiannya dari keempat istilah lainnya, yaitu membuat dan mengajukan sejumlah pertanyaan yang harus dijawab. Sebagai hasil jawabannya diperoleh sebuah ukuran (nilai angka) dari seseorang.
2. Measurement, pengertiannya menjadi lebih luas, yakni dengan menggunakan observasi skala rating atau alat lain yang membuat kita dapat memperoleh informasi dalam bentuk kuantitas. Juga berarti pengukuran dengan berdasarkan pada skor yang diperoleh.
3. Evaluasi, adalah proses penggambaran dan penyempurnaan informasi yang berguna untuk menetapkan alternatif. Evaluasi bisa mencakup arti tes dan measurement dan bisa juga berarti di luar keduanya. Hasil Evaluasi bisa memberi keputusan yang professional. Seseorang dapat mengevaluasi baik dengan data kuantitatif maupun kualitatif.
4. Assesment, bisa digunakan untuk memberikan diagnosa terhadap problema seseorang. Dalam pengertian ia adalah sinonim dengan evaluasi. Namun yang perlu ditekankan disini bahwa yang dapat dinilai atau dievaluasi adalah karakter dari seseorang, termasuk kemampuan akademik, kejujuran, kemampuan untuk mengejar dan sebagainya.

Kita juga sebenarnya hampir setiap hari melakukan pengukuran, yakni membandingkan benda-benda yang ada dengan ukuran tertentu, setelah itu kita menilai, menentukan pilihan mana benda yang paling memenuhi ukuran itulah yang kita ambil. Dua langkah kegiatannya dilalui sebelum mengambil barang untuk kita, itulah yang disebut mengadakan evaluasi yakni mengukur dan menilai. Kita tidak dapat mengadakan penilaian sebelum kita mengadakan pengukuran.

- Mengukur adalah membandingkan sesuatu dengan satu ukuran. Pengukuran bersifat kuantitatif.
- Menilai adalah mengambil suatu keputusan terhadap suatu dengan ukuran baik buruk. Penilaian bersifat Kualitatif.
- Mengadakan evaluasi meliputi kedua langkah di atas. Yakni mengukur dan menilai. (Suharsimi:2002:2-3)

Sejalan dengan pengertian evaluasi yang disebutkan di atas, Arifin (2013:5) mengemukakan bahwa pada hakikatnya evaluasi adalah suatu proses yang sistematis dan berkelanjutan untuk menentukan kualitas (nilai dan arti) daripada sesuatu, berdasarkan pertimbangan dan kriteria tertentu dalam rangka mengambil suatu keputusan.

Berdasarkan pengertian tersebut, Arifin selanjutnya menjelaskan beberapa hal tentang evaluasi, bahwa:

- Evaluasi adalah suatu proses bukan suatu hasil (produk). Hasil yang diperoleh dari kegiatan evaluasi adalah gambaran kualitas daripada sesuatu, baik yang menyangkut tentang nilai atau arti. Sedangkan kegiatan untuk sampai kepada pemberian nilai dan arti

itu adalah evaluasi. Gambaran kualitas yang dimaksud merupakan konsekuensi logis dari proses evaluasi yang dilakukan. Proses tersebut tentu dilakukan secara sistematis dan berkelanjutan, dalam arti terencana, sesuai dengan prosedur dan aturan, dan terus menerus.

- Tujuan evaluasi adalah untuk menentukan kualitas daripada sesuatu, terutama yang berkenaan dengan nilai dan arti.
- Dalam proses evaluasi harus ada pemberian pertimbangan (judgement). Pemberian pertimbangan ini pada dasarnya merupakan konsep dasar evaluasi. Melalui pertimbangan inilah ditentukan nilai dan arti (worth and merit) dari sesuatu yang sedang dievaluasi. Tanpa pemberian pertimbangan, suatu kegiatan bukanlah termasuk kategori kegiatan evaluasi.
- Pemberian pertimbangan tentang nilai dan arti haruslah berdasarkan kriteria tertentu. Tanpa kriteria yang jelas, pertimbangan nilai dan arti yang diberikan bukanlah suatu proses yang dapat diklasifikasikan sebagai evaluasi. Kriteria ini penting dibuat oleh evaluator dengan pertimbangan (a) hasil evaluasi dapat dipertanggungjawabkan secara ilmiah (b) evaluator lebih percaya diri (c) menghindari adanya unsur subjektifitas (d) memungkinkan hasil evaluasi akan sama sekalipun dilakukan pada waktu dan orang yang berbeda, dan (e) memberikan kemudahan bagi evaluator dalam melakukan penafsiran hasil evaluasi.

I. Evaluasi Pengukuran

Pengukuran pendidikan terdiri dari beberapa bidang ilmu, bidang ilmu itu terdiri dari bidang kognitif; bidang afektif; dan bidang psikomotor. Bidang kognitif diukur melalui uji test; bidang afektif diukur melalui kuisioner, wawancara, bisa juga melalui suatu pengamatan; bidang psikomotor dapat diukur melalui perbuatan dan pengamatan.

Berdasarkan perspektif uji test dan kuisioner dapat dianggap menjadi 3 bagian kegiatan. Bagian pertama terdiri dari konstruksi, uji test, dan kuisioner. Bagian kedua adalah dilakukannya penyelenggaraan uji test dan pengisian kuisioner oleh responden. Nabnagian ketiga terdiri dari pensekoran dan penganalisisan butir dari uji tes dan kuisioner.

Pensekoran terdapat di dalam pengukuran pendidikan, digunakan dalam mengukur kemampuan, keberhasilan belajar, sikap, minat, atau ciri terpendam lainnya yang dimiliki oleh para peserta pengukuran tersebut. Karena sifat pengukuran ini terpendam yang tidak kelihatan yang dimiliki oleh peserta maka kita perlu memberikan sejumlah stimulus baik dalam uji test, atau dalam bentuk kuisioner. Bila stimulus itu mengenai sasaran maka akan terlihat kemampuan keberhasilan belajar, sikap, minat, atau ciri-ciri lain yang dimiliki para peserta itu yang kita dapat ukur. Respon yang kelihatan itu dapat kita tafsirkan dengan memberikan nilai yang sesuai.

Pengukuran pendidikan terdiri dari beberapa hal, yaitu: 1. Untuk mengukur ciri terpendam yang tidak kelihatan yang dimiliki para peserta; 2. Mengukur ciri terpendam yang tidak terlihat yang kita berikan pada para peserta, stimulus yang terdiri dari uji test dan kuisioner yang tepat; 3. Para peserta memberikan responsi terhadap stimulus itu dengan berharap bahwa responsi tersebut mencerminkan dengan benar ciri dari terpendam yang ingin kita ukur; 4. Responsi yang terlihat diberikan skor yang dapat ditafsirkan dengan memadai.

Bila urutan kegiatannya tepat maka skor yang kita dapat dari uji test atau kuisioner maka skor yang kita peroleh dari uji test dan kuisioner dapat dianggap sebagai skor dari ciri terpendam yang dimiliki peserta. Pengukuran ciri terpendam peserta memiliki beberapa masalah, yaitu sejauh mana responsi dari para peserta dapat mengukur secara tepat ciri terpendam yang dimiliki peserta untuk diukur; uji test dan kuisioner sudah menjadi stimulus, namun dipertanyakan kemampuan para peserta untuk mengungkapkan secara benar ciri terpendam yang dimiliki oleh para peserta. Maka sebaiknya uji test dan kuisioner harus divalidasi terlebih dahulu lalu kemudian di reliabilitas. Setelah dilakukan validitas dan realibilitas uji test dan kuisioner, maka dapat kita terapkan kepada responsi untuk mengukur sifat terpendam yang dimiliki oleh para peserta.

Skor klasik tidak sama dengan skor modern dalam skala pengukuran pendidikan.

Pengukuran modern belum banyak digunakan orang hanya baru diterapkan di negara-negara maju, sedangkan skor klasik masih digunakan sampai sekarang baik dalam uji test dan kuisioner. Ciri dari pengukuran secara klasik ditunjukkan oleh kenyataan bahwa kelompok butir pada uji test dan kuisioner tidak dapat dipisahkan dari kelompok peserta yang mengikuti uji test dan mengisi kuisioner. Tidak dapat dipisahkan itu maksudnya adalah kelompok butir soal uji test dan pengisian kuisioner diisi dengan kelompok yang sama, kalau kelompok butir soal uji test dan kuisioner diisi kelompok peserta yang berbeda maka hasilnya merupakan ciri atau karakteristik umum dari kelompok butir itu berubah, berarti taraf kesukarannya berubah karena kelompok pesertanya berubah dan berbeda. Dengan demikian pada pengukuran skor klasik uji test dan pengisian kuisioner sangat bergantung kepada butir-butir dari uji test dan pengisian kuisioner dan pesertanya. Begitu juga skor selalu disertai dengan butir dan peserta terlibat dalam penskoran.

Hal-hal yang perlu diperhatikan dalam skala pengukuran klasik yang dimiliki para peserta terdiri dari;

1. Semua peserta yang cirinya akan di skor atau dinilai perlu menempuh uji test atau mengisi butir kuisioner yang sama pada saat yang sama dan bersamaan. Hindari kebocoran butir uji test sebelum uji test dilakukan pada para peserta.
2. Interpretasi atau tafsiran skor menjadi bersifat khusus karena butir soal terkait dengan para peserta, maka tafsiran skor mengacu kepada kelompok para peserta yang menanggapi butir soal tersebut.
3. Bila uji test dan pengisian kuisioner sangat mudah atau sangat sukar bagi para peserta yang mengikuti uji test dan pengisian kuisioner maka hasilnya akan memberikan gambaran skor yang tidak mencerminkan dengan benar kemampuan peserta itu sesungguhnya.
4. Bila uji test dan pengisian kuisioner sangat mudah atau sangat sukar bagi dua atau lebih peserta maka skor tidak dapat lagi membedakan kemampuan bagi peserta karena mereka memiliki perbedaan kemampuan.

Skala Modern digunakan untuk menghilangkan kelemahan pada pengukuran skor klasik. Kegunaan pengukuran modern adalah melepaskan keterpisahan diantara butir uji test dengan peserta uji test. Pada pengukuran skor modern ciri butir soal akan tetap sama tidak menjadi masalah peserta mana yang menempuhnya. Pada skala pengukuran skor modern teori yang cukup terkenal adalah teori responsi butir (item response theory). Pada teori responsi ada 3 hal, yaitu:

1. Proporsi jawaban benar yang dilakukan oleh peserta uji test terhadap butir uji test. Dalam

hal sikap ini merupakan proporsi jawaban yang positif yang dilaksanakan oleh responden kuisioner terhadap butir kuisioner.

2. Ciri butir soal tersebut yang menyebabkan para peserta memberi tanggapan seperti itu terhadap soal tersebut.
3. Mempunyai ciri terpendam bagi peserta sehingga peserta memberikan tanggap demikian terhadap butir soal tersebut.

Skala Pengukuran terdiri dari nominal, ordinal, interval, ratio. Nominal dari yang rendah sampai yang tertinggi yaitu Ratio. Makin tinggi skala pengukuran maka makin tinggi teori yang dapat dibangun. Sifat ini berlaku untuk umum dan dapat untuk diprediksi.

- Skala nominal
Klasifikasi sederhana berdasarkan karakteristik tertentu. Angka yang diberikan disini hanya sebagai alat identifikasi, tidak menunjukkan magnitude.
- Skala Ordinal
Objek diurut berdasarkan dimensi tertentu, juga tidak dapat menunjukkan magnetude karakteristik, sebab tidak ada jarak baku antara unit- unit tersebut.
- Skala Interval
Adalah perbedaan antara objek dinyatakan dengan suatu magnetude yang selalu ekuivalen besarnya meliputi: klarifikasai, magnetude kunut pengukuran yang selalu sama besarnya.
- Skala Ratio
Adalah skala pengukuran yang tertinggi mempunyai angka 0 yang bersifat absolut, dapat dipakai untuk perhitungan-perhitungan dan prediksi.

Skala respon:

1. Nominal : gender: -Laki-laki

-Perempuan

Pekerjaan :Pns, ABRI, pegawai swasta, wiraswasta, pengangguran

2. Ordinal : slalu-sring-jarang-tidk pernah. Slalu-netral-ts-sts (sangat tidak setuju)
3. Ordinal bahaya karna tidak ada kadang-kadang.
4. Ordinal jarang berupa kali sering : o, absolut.

Macam-macam skala ada dua yaitu diskontiniu dan kontiniu. Diskontiniu adalah kategori diurut tetapi bersifat diskrit atau terpisah satu dengan yang lain misalnya besar keluarga. Kontiniu adalah terdapat suatu kontiniunitas yang menunjukkan gradasi jumlah karakteristik misalnya tinggi rendah.

Variansi atau simpangan baku adalah penyebaran nilai disekitar nilai rata. Simpangan baku adalah akar dua positif dan variansi. Dengan demikian, simpangan baku populasi adalah

$$\sigma x = \sqrt{\sigma^2 x}$$

dan simpangan baku sampel adalah

$$s_x = \sqrt{s_x^2}$$

Hasil belajar terdiri Kognitif % ? ; Psikomotor %? Afektif %?.

Distribusi Probabilitas Normal

Probabilitas normal merupakan distribusi yang paling banyak digunakan orang dalam bidang ilmu pendidikan khususnya di bidang uji test juga digunakan dalam penyebaran uji skor. Kalaupun terjadi penyimpangan pada distribusi probabilitas normal, biasanya penyimpangan itu terjadi bila dibanding kan atau di rujukan dengan distribusi, probabilitas normal. Biasanya distribusi probabilitas normal membentang dari $-\infty$ sampai dengan ∞ . Distribusi probabilitas atau karakteristik butir diduga mencerminkan keadaan sesungguhnya adalah model ojaif normal. Sebenarnya ojail normal berasal dari distribusi probabilitas normal. Distribusi probabilitas normal memiliki dua parameter, kedua para meter itu adalah rerata dan simpangan baku. Setiap pasang parameter yakni satu rerata dan satu simpangan baku membentuk seperangkat distribusi probabilitas normal. Dengan μ sebagai rerata dan σ sebagai simpangan baku, fungsi densitas distribusi probabilitas normal adalah

$$N(X; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Rerata hitung/ Arithmetic Mean (μ, x) merupakan hasil bagi antara jumlah data dengan banyak data untuk data tunggal maupun data kelompok.

Distribusi Probabilitas Logistik

Distribusi Probabilitas Logistik adalah model model lain yang menyerupai probabilitas normal, tetapi analisis matematikanya lebih sederhana dan tidak rumit biasanya memiliki lengkungan logistik. Lengkung logistik ini biasanya digunakan untuk pertumbuhan penduduk. Karena bentuknya yang menyerupai model ojaif normal, teori responsi butir juga menggunakan lengkungan fungsi distribusi logistik. Untuk itu, teori responsi butir menggunakan rumus tertentu dari lengkungan logistik itu.

Penyelenggaraan Uji Tes

Sebelum dilakukan uji tes, soal-soal yang telah dibuat oleh para pakar dan setiap butir tes itu sudah dilengkapi dengan kunci jawabannya. Setelah itu harus diuji cobakan kepada sekelompok oran yang jumlahnya tidak terlalu kecil. Setelah selesai uji coba, sebelum dimasukan ke dalam bank butir soal harus di evaluasi dahulu melalui patokan yang sudah ditetapkan. Bank butir soal harus disediakan wadah dan pengolahan yang cukup. Karena terlihat rumit dan memerlukan pengolahan yang besar, dengan demikian wadah yang tepat

adalah komputer. Karena pada komputer terdapat perangkat keras dengan daya tampung serta kecepatan kelola yang memadai yang dimiliki komputer. Selain itu komputer juga mempunyai perangkat lunak, perangkat lunak itu digunakan sebagai pengadministrasian dan melayani permintaan penguji berupa butir yang memadai spesifikasi tertentu. Lama kelamaan bank butir terus mengalami pengembangan dan perubahan-perubahan sehingga kian lama menjadi sangat rumit dalam pengelolaan dan perawatan bank butir tersebut. Maka menjadilah suatu kegiatan tersendiri yang terpisah dari kegiatan lain dengan para spesialisasi bidang penguji tes.

II. Penilaian Hasil Belajar

Hal-hal yang perlu diperhatikan dalam penilaian hasil belajar meliputi mahasiswa, dosen (mudah ditemui atau tidak), dosen (memperhatikan pengajaran dengan baik atau tidak), tugas-tugas banyak, namun dirasakan relevan.

Dalam melakukan pemberian nilai perlu memperhatikan hal berikut ini, yaitu: Perbandingan antara siswa berdasarkan prestasi untuk membedakan jumlah pengetahuan yang dikuasai siswa. Pada umumnya nilai rata-rata C. Baik untuk kelompok yang heterogen; Perbandingan berdasarkan kemampuan/ability; Hati-hati terhadap halo effect; Instrumen apa yang dipakai untuk mengukur kemampuan; Ada kemungkinan regresi statistic.

Pemberian nilai juga memiliki kriteria yang ada seperti, perbandingan antara siswa di berdasarkan prestasi, perbandingan antara individu berdasarkan tingkat kemampuan, Usaha siswa, meningkatkan penambahan prestasi, tingkat prestasi yang telah ditentukan sebelumnya (PAP=penilaian acuan patokan), distribusi normal (PAN=penilaian acuan normatif), lulus atau tidak lulus, nilai komposit, dan skor presentase.

Kinerja diukur dari tugas yang telah dapat diselesaikan atau berapa % tujuan belajar yang telah dicapai. Syarat skor presentase: tujuan belajar harus jelas, butir-butir soal relevan dengan tujuan, tersedia dengan banyak soal , kondisi ujian optimal, rangking.

Kedudukan siswa di dalam kelompoknya, belum menggantungkan arti skor siswa sebab rangking tergantung pada jumlah siswa.

Nilai komposit

Pemberian bobot berdasarkan sifat tujuan komponen ujian kerja laboratorim/bengkel.

Bobot	Skor/baku	Skor yang dibobot
2	9	18
1	7	7
1	8	8
4		33 (skor komposit)

Rata-rata komposit

Skor komposit = $33 = 8,25$

Bobot 4

Skor mentah yang diperlukan dari sebagai hasil pengukuran tidak berpartisipasi kecuali dibandingkan dengan sesuatu kadang-kadang ditransformasikan kedalam bentuk lain.

Skor mentah mahasiswa (raw score) terdiri dari:

- a) Tes objektif yaitu jumlah jawaban yang benar dengan korelasi maka skor mentah :
$$\frac{\text{jumlah jawaban yang benar} - \text{jumlah jawaban yang salah}}{(n-1)}$$
$$N:\text{jumlah jawaban}$$
- b) Tes essay: jumlah jawaban yang benar x bobot tiap aspek yang dijawab (yang khas) ditentukan sebelumnya.

Penilaian Acuan Patokan (PAP)

Penilaian acuan patokan (PAP) atau dikenal dengan istilah Criterion Referenced Test adalah penilaian acuan patokan adalah penilaian yang mengacu kepada kriteria pencapaian tujuan pembelajaran yang telah dirumuskan sebelumnya (Slameto, 1988). Nilai-nilai yang diperoleh peserta didik dikaitkan dengan tingkat pencapaian penguasaan (mastery) peserta didik tentang materi pengajaran sesuai dengan tujuan pembelajaran yang telah ditetapkan. Hal senada diungkapkan Shirran (2008) menjelaskan PAP menfokuskan pada apa yang mampu dikerjakan peserta didik dan apakah peserta didik tersebut menguasai mata pelajaran. Noeng Muhadjir (1994) menjelaskan bahwa PAP ini lebih tepat digunakan untuk mata pelajaran yang bersifat teknologik atau keterampilan tertentu yang di dalamnya dituntut kemampuan peserta didik secara tepat sesuai dengan rumusan ilmu pengetahuan, yang apabila salah bisa berakibat fatal. Misalnya, mata pelajaran statistik yang apabila keliru dalam penghitungan mean, median, modus, korelasi dan lainlain akan berakibat pada kesalahan interpretasi data dan penarikan kesimpulan. Demikian juga misalnya dalam mata pelajaran agama, seperti Fikih, Tauhid dan lain-lain apabila salah bisa berakibat fatal. Namun tidak menutup kemungkinan untuk menggunakan PAP pada mata pelajaran lain asalknya kriterianya dapat dibuat secara teliti.

Lembaga pendidikan yang membuat kriteria atau patokan penilaian berdasarkan persentase dengan skala nilai 0 – 100, maka peserta didik yang memperoleh nilai atau skor 75 dipandang telah memiliki 75% kemampuan atau penguasaan pengetahuan dan keterampilan mengenai mata pelajaran yang bersangkutan. Demikian pula peserta didik yang memperoleh nilai 50% saja dari mata pelajaran tersebut. Kemudian nilai-nilai ini ditransformasikan ke dalam nilai huruf dengan kriteria tertentu pula. Nilai-Nilai 80 – 100 ditransformasikan menjadi nilai A, nilai 70 – 79 ditransformasikan nilai B dan seterusnya. Selanjutnya ditetapkan pula ketentuan batas lulus (passing grade) misalnya 60.

Ketika diperoleh data tentang nilai peserta didik 70 maka peserta didik tersebut telah melewati batas passing grade yang ditetapkan sebesar 60. Demikian juga ketika terdapat peserta didik yang memperoleh nilai 50 maka peserta didik tersebut belum menguasai pengetahuan dan keterampilan yang dituntut dalam mata pelajaran tersebut karena memperoleh nilai dibawah passing grade. Bagi peserta didik yang telah memperoleh nilai yang berada di atas nilai passing grade dapat diberikan materi pengayaan yang dapat memperluas wawasan pengetahuannya, sedangkan bagi peserta didik yang belum memperoleh nilai pada batas nilai passing grade dapat diberikan kegiatan remedial.

Demikian pula misalnya dalam kegiatan belajar mandiri yang mempergunakan modul sebagai bahan ajar, maka di dalam modul tersebut dinyatakan bahwa untuk dapat dinyatakan lulus maka peserta didik harus memperoleh nilai minimal 80% dari tes akhir yang terdapat dalam modul. Dalam hal ini apabila terdapat peserta didik setelah mempelajari modul tersebut dan mengerjakan tes akhir modul memperoleh nilai 70, maka hal ini dapat dinyatakan peserta didik tersebut hanya menguasai bahan sebesar 70%, sehingga peserta didik tersebut masih harus mempelajari kembali bagian-bagian dari modul yang belum dikuasainya. Untuk selanjutnya dites lagi sampai akhirnya peserta didik tersebut dapat memperoleh nilai 80 atau lebih.

Dari contoh di atas dapat dimaknai bahwa lembaga pendidikan menggunakan kriteria penilaian tertentu. yaitu berdasarkan kriteria tingkat kemampuan penggunaan pengetahuan dan keterampilan sesuai dengan tujuan kurikulum sehingga nilai yang diperoleh peserta didik sekaligus mencerminkan tingkat kemampuan atau penguasaan peserta didik terhadap materi ajar yang diteskan. Kriteria atau patokan yang digunakan dalam PAP bersifat mutlak, artinya kriteria itu bersifat tetap, setidak-tidaknya untuk jangka waktu tertentu. Ketentuan tersebut berlaku bagi seluruh peserta didik yang mengikuti kegiatan pembelajaran di lembaga pendidikan tersebut.

Tujuan PAP adalah untuk mengukur secara pasti tujuan atau kompetensi yang ditetapkan sebagai kriteria keberhasilannya. Penilaian acuan patokan bermanfaat dalam upaya meningkatkan kualitas hasil belajar, sebab peserta didik diusahakan untuk mencapai standar yang telah

ditentukan dan hasil belajar peserta didik dapat diketahui derajat pencapaiannya (Arifin, 2009). Untuk mencapai tujuan PAP tersebut maka dalam hal ini Davies (1991) menjelaskan tiga syarat yang harus dipenuhi:

- a) Tepat. Tes PAP harus sesuai dengan tujuan-tujuannya, dengan bahan pelajaran, dengan strategi pembelajaran yang digunakan serta dengan peserta didik yang akan menjawabnya.
- b) Efektif. Tes PAP harus dapat melakukan tugasnya dengan baik. Ini berarti bahwa hal itu harus dapat diandalkan (reliabel) dan sah.
- c) Praktis. Dalam pengertian ini, tes PAP harus dapat diterima baik oleh guru maupun peserta didik. Hal itu harus realistis dalam pembiayaan dan waktu yang digunakan dalam pelaksanaan serta mudah digunakan dan digunakan kembali.

Selanjutnya untuk menentukan batas lulus (passing grade) dengan pendekatan PAP maka setiap skor peserta didik dibandingkan dengan skor ideal yang mungkin dicapai oleh peserta didik. Misalnya dalam suatu tes ditetapkan skor idealnya adalah 100, maka peserta didik yang memperoleh skor 65 sama dengan memperoleh nilai 6,5 dalam skala 0 – 10. Demikian seterusnya.

Penilaian Acuan Patokan (PAP) adalah model pendekatan penilaian yang mengacu kepada suatu kriteria pencapaian tujuan (TKP) yang telah ditetapkan sebelumnya. PAP merupakan suatu cara menentukan kelulusan siswa dengan menggunakan sejumlah patokan. Bilamana siswa telah memenuhi patokan tersebut maka dinyatakan berhasil. Tetapi bila siswa belum memenuhi patokan maka dikatakan gagal atau belum menguasai bahan pembelajaran tersebut. Nilai-nilai yang diperoleh siswa dihubungkan dengan tingkat pencapaian penguasaan siswa tentang materi pembelajaran sesuai dengan tujuan yang telah ditetapkan.

Siswa yang telah melampaui atau sama dengan kriteria atau patokan keberhasilan dinyatakan lulus atau memenuhi persyaratan. Guru tidak melakukan penilaian apa adanya melainkan berdasarkan kriteria keberhasilan yang telah ditetapkan sejak pembelajaran dimulai. Guru yang menggunakan model pendekatan PAP ini dituntut untuk selalu mengarahkan, membantu dan membimbing siswa kearah penguasaan minimal sejak pembelajaran dimulai, sedang berlangsung dan sampai berakhirnya pembelajaran. Kompetensi yang dirumuskan dalam TKP merupakan arah, petunjuk, dan pusat kegiatan dalam pembelajaran. Penggunaan tes formatif dalam penilaian ini sangat mendukung untuk mengetahui keberhasilan belajar siswa. Pelaksanaan PAP tidak memerlukan perhitungan statistik melainkan hanya tingkat penguasaan kompetensi minimal.

Sebagai contoh misalnya: untuk dapat diterima sebagai calon tenaga pengajar di perguruan tinggi adalah IP minimal 3,00 dan setiap calon harus lulus tes potensi akademik yang diadakan oleh lembaga yang bersangkutan. Berdasarkan kriteria di atas siapapun calon yang tidak memenuhi persyaratan di atas maka dinyatakan gagal dalam tes atau tidak diterima sebagai calon tenaga pengajar.

Seperti uraian di atas tingkat kemampuan atau kelulusan seseorang ditentukan oleh tercapai tidaknya kriteria. Misalnya seseorang dikatakan telah menguasai satu pokok bahasan / kompetensi bilamana ia telah menjawab dengan benar 75% dari butir soal dalam pokok bahasan / kompetensi tersebut. Jawaban yang benar 75% atau lebih dinyatakan lulus, sedang jawaban yang kurang dari 75% dinyatakan belum berhasil dan harus mengulang kembali.

Muncul pertanyaan bahwa apakah siswa yang dapat menjawab benar 75% ke atas juga akan memperoleh nilai yang sama? Hal ini tergantung pada sistem penilaian yang digunakan. Jika hanya menggunakan kriteria lulus dan tidak lulus, berarti siswa yang menjawab benar 75% ke atas adalah lulus, demikian juga sebaliknya siswa yang menjawab benar kurang dari 75% tidak lulus. Apabila sistem penilaian yang digunakan menggunakan model A, B, C, D atau standar yang lain, kriteria ditetapkan berdasarkan rentangan skor atau skala interval. Perlu dijelaskan bahwa kriteria atau patokan yang digunakan dalam PAP bersifat mutlak. Artinya kriteria itu bersifat tetap, setidaknya untuk jangka waktu tertentu dan berlaku bagi semua siswa yang mengikuti tes di lembaga yang bersangkutan.

Kelebihan atau keunggulan Penilaian Acuan Patokan adalah dapat mengukur penguasaan siswa terhadap pengetahuan TIK Dikawasan kognitif, psikomotor, afektif; dapat langsung dipakainuntuk menginterpretasi kinerja siswa; mengukur kemampuan yang harus diketahui/dikuasai siswa; untuk pembelajaran yang bersifat individual. Patokan TIK(tujuan intruksional khusus) terdiri dari patokan kurikulum kelulusan perlu ditentukan terlebih dahulu.

Kelemahan Penilaian Acuan Patokan yaitu tidak dapat menunjukan kedudukan siswa terhadap kelompoknya; karena sukar untuk menyatakan TIK secara eksplisit. PAP tidak dapat dipakai untuk mengukur kemampuan siswa dalam aturan yang luas; adanya TIK membuat penggunaan sangat terbatas demikian juga proses belajar siswa; patokan harus ditentukan dengan baik dan benar; tidak dapat dipakai untuk membandingkan kinerja siswa dengan kelompok diluar kelompoknya.

Bagaimana membuat distribusi kelas juling kekiri Dosen harus: menentukan patokan/batas lulus, memberikan pengarahan/bimbingan selama proses belajar mengajar, memberikan formatif

testing, memberikan umpan balik secepat mungkin.

PAP (penilaian Acuan Patokan)

1. Skor dinyatakan dalam %
2. Tingkat kelulusan tetap berdasarkan patokan yang telah ditentukan dalam TIL.
3. Untuk mata kuliah profesi serta prerequisite harus dikuasai benar-benar.

Penilaian berdasarkan acuan patokan menurut departemen pendidikan dan kebudayaan tahun 1983

Derajat Penggunaan	Nilai
90%-100%	A
80%-89%	B
70%-79%	C
60%-69%	D
<60%	E

Batas lulus ditentukan terlebih dahulu dan tergantung dari tujuan belajar yang harus dicapai.

Penilaian Acuan Norma (PAN)

Penilaian acuan norma (PAN) atau dikenal dengan istilah Norm Referenced Test adalah penilaian yang dilakukan dengan mengacu pada norma kelompok. Nilai-nilai yang diperoleh peserta didik diperbandingkan dengan nilai-nilai peserta didik lainnya yang termasuk di dalam kelompoknya (Slameto, 1988). Istilah “norma” menunjukkan kapasitas atau prestasi kelompok, sedangkan yang dimaksudkan dengan “kelompok” adalah semua peserta didik yang mengikuti tes tersebut. Jadi pengertian “kelompok” yang dimaksudkan dapat berarti sejumlah peserta didik dalam satu kelas, sekolah, rayon, propinsi atau wilayah.

Penilaian acuan norma dapat diilustrasikan sebagai berikut: hasil ujian nasional (UN) dikenal adanya nilai UN murni yang berasal dari penilaian yang dilakukan secara komputerisasi dengan menggunakan persentase yang menunjukkan tingkat kemampuan atau

penguasaan peserta didik tentang materi ajar yang diujikan. Dengan kata lain nilai UN murni merupakan penilaian dengan cara PAP. Akan tetapi, setelah diketahui bahwa nilai-nilai UN murni ini pada umumnya rendah bahkan sangat rendah sehingga tidak memenuhi syarat untuk dapat dinyatakan lulus, maka nilai UN murni itu kemudian diolah ke dalam PAN dengan menggunakan rumus-rumus tertentu dengan maksud agar nilai-nilai tersebut dapat diperbesar.

Penilaian Acuan Norma (PAN) adalah penilaian yang dilakukan dengan mengacu pada norma kelompok atau nilai-nilai yang diperoleh siswa dibandingkan dengan nilai-nilai siswa lain dalam kelompok tersebut. Dengan kata lain PAN merupakan sistem penilaian yang didasarkan pada nilai sekelompok siswa dalam satu proses pembelajaran sesuai dengan tingkat penguasaan pada kelompok tersebut. Artinya pemberian nilai mengacu pada perolehan skor pada kelompok itu.

Dalam hal ini “norma” berarti kapasitas atau prestasi kelompok, sedangkan “kelompok” adalah semua siswa yang mengikuti tes tersebut dapat kelompok siswa dalam satu kelas, sekolah, rayon, provinsi, dan lain-lain. Pan juga dapat dikatakan penilaian “apa adanya” dengan pengertian bahwa acuan pembandingnya semata-mata diambil dari kenyataan yang diperoleh (rata-rata dan simpangan baku) pada saat penilaian dilakukan dan tidak dikaitkan dengan hasil pengukuran lain.

PAN menggunakan prinsip-prinsip yang berlaku pada kurva normal. Hasil-hasil perhitungannya dipakai sebagai acuan penilaian dan memiliki sifat relatif sesuai dengan naik turunnya nilai rata-rata dan simpangan baku yang dihasilkan pada saat itu. Penggunaan sistem PAN membiarkan siswa berkembang seperti apa adanya. Namun demikian guru tetap merumuskan Tujuan Khusus Pembelajaran (TKP) sesuai dengan tuntutan kompetensi. TKP yang berorientasi pada kompetensi tetap dipakai sebagai tumpuan dalam penyusunan evaluasi akan tetapi pada saat pemberian skor yang diperoleh siswa maka TKP tidak dipergunakan sebagai pedoman. Batas kelulusan tidak ditentukan oleh penguasaan minimal siswa terhadap kompetensi yang ditetapkan dalam TKP, melainkan didasarkan pada nilai rata-rata dan simpangan baku yang dihasilkan kelompoknya.

Dengan demikian kelemahan sistem PAN dapat terlihat jelas bahwa tes apapun, dalam kelompok apapun, dengan kadar prestasi yang bagaimanapun pemberian nilai dengan model pendekatan PAN selalu dapat dilakukan. Oleh karena itu penggunaan model pendekatan ini dapat dilakukan dengan baik apabila memenuhi syarat antara lain: a). skor nilai terpengar atau dapat dianggap terpengar sesuai dengan pengar kurva normal; b). jumlah yang dinilai minimal 50 orang atau lebih dari 100 orang dalam arti sampel yang digunakan besar.

Dalam penerapan sistem PAN ada dua hal pokok yang harus ditetapkan yaitu: banyaknya

siswa yang akan lulus dan penetapan batas lulus. Terdapat dua cara di dalam menentukan batas kelulusan antara lain: menetapkan terlebih dahulu jumlah yang diluluskan, misalnya 75% dari seluruh peserta tes, kemudian skor tiap siswa disusun dan diranking sehingga akan diketemukan skor terendah. Cara kedua dengan menggunakan data statistik yang terdapat dalam kurva normal dengan menggunakan nilai rata-rata dan simpangan baku, sehingga akan diketemukan luas daerah kurva normal atau jumlah anak yang diluluskan.

Keunggulan PAN (penilaian acuan norma), yaitu:

- Dapat dipakai mengukur kemampuan termaksimal mungkin.
- Untuk mengukur/interpretasi kinerja siswa tingkat tinggi dikawasan afektif/psikomotor
- Dipakai dalam pengukuran kemampuan siswa di mana memang diharapkan adanya perbedaan-perbedaan
- Afektif untuk pengujian
- Seleksi calon-calon

Kelemahan PAN (penilaian acuan norma)

- Tidak memadai untuk mengukur penguasaan materi/ketrampilan
- Tidak dapat mengukur apakah siswa sudah siap/mampu untuk menerima pengajaran berikutnya
- Hasil pengukuran tidak dapat langsung diinterpretasikan.

Penilaian berdasarkan acuan normal (tahun 83)

Batas daerah dalam kurva	Nilai	Jumlah(%)
$X+1,05$ s.d.atau lebih	A	6,68%
$X+1,05$ s.d- $X+0,50$.sd	B	24,17%
$X+0,50$ s.d- $X-0,50$ s.d	C	38,30%
$X-0,50$ s.d- $X-1,05$ s.d	D	24,17%
$<X-1,05$ sd	E	6,68%

Aspek Pengukuran	PAN	PAP
------------------	-----	-----

1.Pengajaran	Bahan materi yang luas	Penguasaan TIK
2.Sifat/kemajuan yang diukur	Mencakup rintangan	Cukup atau tidak
3.Kawasan	Kognitif, afektif, psikomotor	Kognitif, afektif, psikomotor
4.Butir-butir soal	Harus dapat membedakan antara siswa	Mengukur penguasaan patokan
5.Tahun kelulusan	medium	Tergantung patokan TIK
6.Validitas	Validitas isi	Validitas isi
7.Rentangan skor	Rendah-tinggi	Diharapkan meningkat
8.Interpretasi skor	Berdasarkan norma	Berdasarkan patokan

Interpretasi skor:

A.Memperoleh skor 75

1.Berdasarkan PAP

70-80 (C)

81-90 (B)

91-100 (A)

Jadi A memperoleh nilai (C)

2.Berdasarkan PAN

Dengan $X=62$

s.d =5

maka $A > X + 2 \cdot s.d. \dots\dots\dots > (A)$

Penilaian berdasarkan Acuan Norma (tahun 83)

Batas daerah dalam kurva nilai jumlah %

Batas daerah dalam kurva	Nilai	Jumlah %
$X + 2 \cdot S.d.$ atau lebih	A	2,28 %
$X + 2 \cdot S.d - X + 1 \cdot s.d$	B	13,59 %

$X + 1.s.d - X - 1.s.d$	C	68,26 %
$X - 1.s.d - X - 2.S.d$	D	13,59 %
$< X - 2.s.d$	E	2,28 %X

III. Kriteria Pemberian Penilaian

1. Berdasarkan antar siswa berdasarkan prestasi untuk membedakan jumlah pengetahuan yang dikuasai siswa. Beda umumnya nilai rata-rata C baik untuk kelompok yang heterogen
2. Perbandingan berdasarkan kemampuan/ability
3. Usaha siswa
4. Meningkatkan Prestasi (game score)

Kelemahan

1. Tidak reliabel
2. Jumlah meningkat yang sama belum tentu mempunyai arti yang sama

Skor presentase

Kinerja diukur dari tugas yang telah dapat diselesaikan atau % tujuan belajar yang tidak dicapai syarat-syarat tujuan belajar

Cara menginterpretasi hasil evaluasi

1. Berdasarkan acuan patokan
 - Apabila sebagai patokan dipakai tujuan instruksional
 - Selalu tetap standartnya
2. Berdasarkan acuan norma apabila sebagai patokan dipakai distribusi normal PAP: untuk mahasiswa:
 - prerequisite
 - profesi

PAN: untuk seleksi Syarat tes harus:

1. Mengukur proses/hasil belajar yang perlu dicapai mahasiswa
2. Mencakup semua tugas belajar dari yang mudah sampai yang sulit
3. Mengukur suatu sampel yang representatif dari tugas-tugas tersebut
4. Mengandung bermacam-macam bentuk soal yang memadai untuk mengukur tugas/hasil belajar mahasiswa
5. Didasarkan atas rencana untuk apa hasil tes tersebut dipakai (seleksi, diagnosa, kesukaran belajar, pengukuran penguasaan materi yang diajarkan)
6. Dapat menghasilkan skor yang bebas dari kesalahan pengukuran dan memberikan hasil yang konsisten

Faktor-faktor yang mempengaruhi jumlah butir soal yang dikembangkan:

1. Bentuk soal yang dipakai

2. Umur/tingkat pendek yang diuji
3. Tingkat kemampuan mahasiswa
4. Panjang kekompleksan butir soal
5. Macam tujuan belajar/proses belajar yang akan diukur
6. Jumlah hitungan atau berpikir kuantitatif yang diperlukan untuk menjawab soal
7. Waktu yang tersedia

Penyusunan kisi-kisi

Pedoman penyusunan instrumen yang menunjukkan:

1. Macam tujuan belajar, yang harus diukur
2. Ruang lingkup materi yang diujikan
3. Bentuk pertanyaan yang dikembangkan
4. Jumlah pertanyaan yang diperlukan
5. Jumlah pertanyaan yang luas
6. Tingkat kesukaran
7. Distribusi pertanyaan
8. Waktu yang tersedia

Yang perlu diperhatikan berapa bobot tiap-tiap topik/tujuan Penyusunan Kisi

1. Identifikasi tujuan/hasil belajar dan topik yang harus diukur
2. Pemberian bobot kepada tujuan belajar dan topik-topik tersebut
3. Distribusi butir-butir soal secara proporsional kedalam tiap-tiap sel

Tujuan memperoleh tes yang representatif Kisi-kisi Blue Print Ujian

1. Berapa bobot yang diberikan untuk tiap-tiap tujuan belajar?
2. Bentuk instrumen apakah yang sesuai untuk mengukur tujuan-tujuan tersebut
3. Berapa banyak soal yang diperlukan? (d disesuaikan dengan waktu, tingkat kesukaran soal/pertanyaan)

Referensi untuk pembuatan kisi-kisi

1. Tujuan belajar
2. Kurikulum/materi yang diajarkan Kisi-kisi Instrumen

Topik	Kognitif	Psikomotor	Afektif	Total
	C1 C2 C3 –C5			
I				%
II				
III				
Dst				100%

Pembobotan ditentukan oleh:

1. Banyaknya waktu yang dipakainuntuk mempelajari topik/tujuan
2. Pentingnya topik/tujuan belajar untuk pengalaman belajar mahasiswa secara keseluruhan
3. Pentingnya hasil belajar untuk retensi transformasi
4. Penilaian para ahli

Referensi penyusunan kisi-kisi

1. Tujuan belajar
2. Kurikulum/materi yang diajarkan

IV. Afeksi

Clasification of tes item type

- Selection
- True-false
- Always-sometimes-never
- Multiple choice Increasing
- Objektiviti Supply Completion Short answer
- Restricted essay Extended essay
- Dimensi format tes

1. Alternatif-bebas (respons)

- Ada alternatif yang harus dipilih siswa
- Siswa bebas dalam memberikan respons

2. Kecepatan=Kemampuan

- Ada batas waktu butir-butir soal mudah
- Bermacam-macam tingkat kesulitan

3. Kertas-kertas potlot-penampilan

- Biasa mengukur kawasan kognitif
- Biasa mengukur tingkah laku/manipulasi sesuatu

4. Kelompok individual

- Biasanya kertas/potlot untuk beberapa orang sekaligus
- Hanya seorang setiap kali tes lisan atau penampilan

5. Terstruktur-Proyektif

- Stimulasi + tugas subjek jelas
- Stimulasi + tugas subjek tidak jelas

Bentuk Soal

1. Selection type

- Pilihan ganda (M.C)
- Benar-salah (true-false)
- Penjodohan (matching)

2. Supply Type

Jawaban pendek (short answer)

- Essay dengan respons terbatas (structured/restricted respons)
- Essay dengan respons bebas

3. Soal benar-Salah

- .-Mudah dalam penyusunan
- Materi yang tercakup dapat luas
- Hanya mengukur kemampuan berpikir yang sederhana
- Tidak semua bidang studi dapat dinyatakan dalam bentuk seperti absolut
- Dapat menebak jawaban

Sasaran Penyusunan soal benar-salah

- 1.Pernyataan jangan bersifat umum
- 2.Pernyataan jangan terlalu sederhana
- 3.Pernyataan jangan negatif
- 4.Kalimat jangan panjang dan kompleks
- 5.Jangan ada dua gagasan dalam satu pernyataan
- 6.Apabila mengukur suatu pendapat beri sumbernya
- 7.Pernyataan yang benar sama jumlahnya/panjangnya dengan yang salah

Tes Penjodohan

- 1-Dapat mengukur materi yang luas
 - Mudah menyusunnya
- 2.-Hanya mengukur tingkat kemampuan berpikir rendah
 - Ada kemungkinan petunjuk-petunjuk kearah jawaban yang benar

Penyusunan soal penjodohan

- 1.Pilih bahan yang homogen
- 2.Jumlah pernyataan/kata dalam kolom satu tak sama dengan kolom lainnya
- 3.Kolom kiri pernyataan, kolom kanan respon
- 4.10 butir soal/kolom
5. Urutan respon harus logis
- 6.Letakkan dalam/halaman yang sama

Soal dengan jawaban pendek

- 1.-Sangat mudah disusun
 - Kemungkinan menebak kecil

- Pengetahuan sepotong tak dapat dipakai untuk menjawab dengan benar
- 2.-Tak dapat dipakai untuk mengukur hasil belajar yang kompleks
- Sulit memberikan apabila formulasi tak jelas/benar

Penyusunan butir soal

- 1.Jawaban harus pendek + jelas
- 2.Jangan ambil pernyataan-pernyataan dari buku teks begitu saja
- 3.Nyatakan dalam bentuk pernyataan, bagian yang kosong letakan dibagian akhir kalimat

Tes Pilihan ganda

- 1.-Dapat dipakai untuk mengukur hasil belajar disemua jenjang berpikir
 - Mempunyai reliabilitas tinggi
 - Makin banyak alternatif jawaban, makin kecil kemungkinan menebak
- 2.-Hanya mengukur hasil belajar yang bersifat verbal
 - Karena hanya memilih jawaban yang benar tak dapat dipakai untuk mengukur kemampuan mengorganisasi gagasan
 - Sukar menyusun alternatif jawaban yang setara

Penyusunan tes objektif M.C.Q

- 1.Pernyataan pokok harus mencerminkan dengan nyata dan jelas masalah yang ditanyakan harus lebih panjang dari pilihan
- 2.Makin banyak jumlah pilihan makin baik
- 3.Pilihan harus secara gramatikal benar dan menyatu dengan pokok
4. Pilihan harus mempunyai kemungkinan untuk dipilih, meskipun harus ada
5. Panjang pilihan harus kira-kira sama
- 6.Tanyakan hal-hal yang sifatnya baru sehingga tidak hanya mengingat kembali
7. Butir yang mengukur pemahaman harus merupakan sesuatu yang baru
8. Semua distraktor harus mungkin dipilih
9. Panjang alternatif lebih kurang sama
10. Asosiasi verbal antara batang-jawaban yang benar harus dihindari

V. Tes Essay

Dipakai apabila ingin mengukur:

1. Keberhasilan belajar mahasiswa tingkat tinggi/kompleks
2. Kemampuan mahasiswa menulis/mengutamakan gagasan (Hopkins, Stanley dan Hopkins, 1990)

Contoh: - Mengapa media pembelajaran perlu digunakan dalam proses pembelajaran
 - Ada key word

Macam-macam tes essay

Tergantung dari macam respons yang diinginkan dari mahasiswa

1. Jawaban singkat dan jelas

Penyusunan tes essay

1. Pertanyaan harus terfokus
2. Isi buat terstruktur, ditentukan panjang pertanyaan, jumlah pertanyaan dapat lebih banyak, jumlah diskusi kurang
3. Minta teman mereview tes/pertanyaan
4. Hindari pemilihan pertanyaan, semua mahasiswa harus dapat menjawab semua pertanyaan
5. Pemakaian butir essay hanya untuk tujuan-tujuan tertentu
6. Pertanyaan jangan panjang-panjang
7. Alokasi waktu cukup untuk memberi jawaban

Tujuan	1	2	3	4	5
Topik					
I					
II					
III					

Penyusunan soal essay:

1. Hanya dipakai untuk pengukuran hasil belajar tertentu

2. Penyusunan pertanyaan harus hati-hati
3. Mahasiswa harus tau:
 - Apa yang harus dikerjakan
 - Berapa bobot tiap-tiap aspek
4. Ada perkiraan waktu yang memadai
5. Jangan minta mahasiswa memilih soal yang akan dijawab
6. Susun soal-soal yang meminta jawaban pendek-pendek

Penyusunan tes essay

1. Apa yang ditanyakan harus jelas + diberi pembatasan
2. Sedapat mungkin dibuat dalam bentuk setengah tertutup
3. Bobot penilaian diberitahukan kepada mahasiswa

Tes essay dengan jawaban singkat (restricted respon) Keunggulan.

1. Mudah dibuat
2. Lebih mengukur suatu hasil belajar
3. Mudah pemberian skor

Kelemahan

1. Mahasiswa tidak dapat menunjukkan kreatifitasnya

Tes essay dengan jawaban bebas

1. Perlu hati-hati dalam pemakaian
2. Perlu diberi batasan-batasan juga
3. Sulit dalam memberi skor, kadang-kadang tergantung dari bahasa, tulisan tangan mahasiswa
4. Mahasiswa dapat dengan bebas mengutarakan
5. Tidak reliable, mahasiswa dapat kreatif
6. Halo effect, dipengaruhi sifat-sifat

Kelemahan tes essay dengan jawaban bebas didalam penilaian:

1. Tidak reliable/konsisten
2. Halo effect: dipengaruhi sifat-sifat lain, mahasiswa
3. Item to item carry over effect dipengaruhi

Pemberian skor tes essay:

1. Tidak reliable

- 2.Halo effect
- 3.Kemungkinan adanya item-item carry over effect
- 4.Tes to test carry over effect
- 5.Order effect

VI. Paper and pencil tes yang digunakan untuk guru menilai pada laboratorium (pengukuran kognitif)

Evaluasi proses dan hasil belajar dikawasan Psikomotor:

1. Pengukuran kognitif melalui paper and pencil tes

Mengukur pengetahuan siswa tentang ketrampilan yang akan dilaksanakan (diteliti). Dapat dilaksanakan secara bersama

2. Pengukuran kegiatan (psikomotor) melalui observasi

Kegiatan Psikomotor melalui observasi maksimum 6-7 orang (satu kali observasi). Satu kelas tidak mungkin. Untuk itu diperlukan lembar observasi yang disusun sebelumnya

3. Pengukuran afektif melalui kebersihan, kesetiaan

Afektif mencakup ketelitian, kebersihan, dan tepat waktu Kawasan Afektif. Sukar dilihat dan sukar diukur. Kita mempunyai beberapa variabel. Variabel yang disusun dari beberapa pengertian yang tidak dapat dilihat

Misal: minat (afektif), cinta (afektif), motivasi (afektif).

Variasi konstruk ini harus didefinisi operasionalkan artinya: dapat diukur yang dapat dilihat dari indikatornya. Misal:

1. Minat dapat dilihat dari waktu, perhatian, uang
2. Sikap indikatornya: positif dan negatif

Yang penting dalam penjalaran kawasan afektif ialah bagaimana kita mengukur variabel tersebut, bagaimana kita mengetahui orang itu punya minat

- atau melalui kelompok yang berbeda (diferensial groups)

Contoh: matematika adalah ilmu yang sangat penting untuk kehidupan oleh orang teknik, sementara bagi jurusan lain belum tentu perlu di lakukan uji T harus ada perbedaan yang nyata antara kedua kelompok tersebut

-Melalui sebelum dan setelah pada kerjaan apakah ada perbedaan kalau ada berarti memang mengukur variabel afeksi tersebut

VII. Analisis butir soal

Prosedur siswa dari yang tertinggi-terendah ambil lebih kurang $\frac{1}{4}$ tertinggi (KA)
kelompok A, $\frac{1}{4}$ terendah (KB)

Indeks kesukaran:

Jumlah benar KA + jumlah benar KB

KA + KB

Indeks/kemampuan diskriminasi

Jumlah benar KA – Jumlah benar KB

$\frac{1}{2} (KA + KB)$

Kekuatan distraktor/pilihan sedapat mungkin harus ada yang memiliki (minimal 10 %)

	a	B +	c	d	
A	2	5	1	7	N=15
B	3	7	1	4	
	A +	b	c	d	
B	1	4	2	3	A=10
B	3	2	2	3	

Item analisis

Diskriminating power (daya pembeda)

$$DP = \frac{4 - L}{\frac{1}{2} N}$$

DP : Indeks

U : No of upper gruve answering

L :No of lower

N : No of local grove

$$\text{Difficulty indeks : } P = \frac{4 + L}{N} \times 100$$

$$\text{Efectivelesst of distractors } D = \frac{4 - L}{\frac{1}{2} N}$$

Motor Skills

1.Kognitif phase

3.Intermediate phase organization

3.Protecting

Komponen Psikomotor

1.Kognisi

2.Fiksasi

3.Otomasi

Pengembangan standart suatu pekerjaandidasarkan pada:

1.Filsafat

2.Pengetahuan

3.Seni : -Factor-factor eksternal

- internal

Kompetensi (kemampuan)

1. Generik

2. Khusus

Kompetensi adalah diperoleh melalui pendidikan dan pengalaman Kualifikasi adalah diperoleh melalui pendidikan. Pengukuran keterampilan terdiri dari :

1. Proses

2. Prodak

Pendidikan profesi :

1. Apakah kriteria yang harus diperhatikan didalam penilaian
2. apakah standar / patokan yang akan dipakai.

Kriteria penilaian:

Didasarkan atas nilai tugas secara:

1. teoritik: studi literatur
2. lapangan : - observasi
 - wawancara dengan eksperts kinerja kerja performance teks
3. measure : - proses
 - produk
 - Proses dan produk

Instrumen pengukuran ketrampilan;

1. tes penampilan dengan kertas- pensil
 - (paper – pensil performance test)
2. test identifikasi
3. penampilan dalam situasi simulasi
4. test sampel kerja (work sampel)

Tingkat realitas tergantung pada :

1. macam tujuan
2. urutan pengajaran
3. kendala praktis
4. jenis tugas

Penyusunan tes ketrampilan:

1. spesifikasi hasil belajar ketrampilan yang akan diukur
2. pemilihan tingkat realitas yang memadai tergantung dari
 - macam tujuan
 - urutan pengajaran
 - kendala praktis
 - jenis tugas
3. intruksi / petunjuk yang perlu.
4. penentuan format penilaian, penamilan, siswa
5. penentuan tingkat keberhasilan

1	Komponen yang dilihat	Komponen yang tidak dilihat
2		
3		
4		
5		

Rating skile

Komponen yang dilihat	Diagnosa nilai	baik	sedang	kurang	Kurang sekali
1					
2					
3					
4	Yang penting tentukan kompetensi kritis		Tentukan batas lulus		
5	Yang penting tentukan kompetensi kritis				

Obserfasi:

- 1.apa yang diobserfasi
- 2.kapan dan bagaimana mencatat data
- 3.sejauh mana akan diambil inferensi dari data yang diperoleh.

Obserfasi (terstruktur)

Pengukuran ;

1. Tentukan tingkah laku yang akan diobserfasi
2. batasi jumlah tingkah laku sampai yang paling penting-penting saja
3. batasi jumlah orang yang akan diobserfasi
4. tentukan alat yang akan dipakai (ceklis, rating acak)
5. susun alat yang udah dipakai
6. obserfasi dilakukan secara objektif mungkin
7. Pencatatan dilakukan secepatnya
8. sedapat mungkin obserfasi dilakukan secara alamiah mungkin
9. hindari pengaruh hawtone, joh dan hery prakmation

Instrumen obserfasi:

1. ceklis : ya/tidak
 2. rating sklae : dengan deskripsi kualitatif tentang tingkah laku yang diukur. hasil pengukuran menunjukkan kualitas yang dinyatakan dalam bentuk ranking (urutan)
- slalu
 - bagus sekali
 - tak pernah
 - Kurang sekali

Lembaran obserfasi

Ceklis : ya / tidak kurang teliti

Objektif

Rating skale 1. bagus jelek 2. subjektif

Dapat dipakai untuk mengadakan diferinsiasi (perbedaan) Kesalahan-esalahan pada pengukuran dengan rating skale 1. penyipamngan individual pada pemberian skore / observasi 2. pengaruh halo 3. kesalahan logis

Ceklis: Untuk menilai tes essay

1. apakah bentuk butir semacam ini memadai untuk mengukur hasil belajar
2. apakah butir sesuai dengan tugas belajar yang dikukur
3. apakah soal / pertanyaan memang mengukur hasil belajar yag kompleks
4. apakah pertanyaan-pertanyaan jelas menunjukan apa yang harus diukur
5. apakah pertanyaan-pertanyaan jelas menunjukan apa yang harus dikukur
6. apakah istila-istilah jelas

- 7.apakah semua mahasiswa harus menjawab pertanyaan yang sama
- 8.apakah tersedia cukup waktu untuk menjawab semua pertanyaan
- 9.apakah ada rambu-rambu pemberian skore?

VIII. Konsep konstruk Konseptualisasi Definisi nominal Definisi operasional Pengukuran

Konseptualisasi = mendeskripsikan dengan jelas apa yang dimaksud dengan istilah tersebut. Sesuatu yang mempunyai nilai banyak variabel

Konsep = Abstrak mengenai suatu realitas

Konstruk = konsep psikologis yang tidak dapat dilihat/bersifat abstrak dibangun dari beberapa indikator

Konstruk Definisi nominal Indikator

Pernyataan-pernyataan

Konsep = abstraksi dari sesuatu yang dapat dilihat dan mempunyai suatu arti tertentu

Konstruk yaitu konsep yang tidak dapat terlihat dengan mudah (abstrak)/konsep psikologis.

Variabel konsep/konstruk yang mempunyai bermacam-macam nilai.

Minat menurut definisi operasional = sesuatu yang dapat membangkitkan gairah seseorang dan yang menyebabkannya menggunakan waktu yang energi untuk itu (Rath, Harmin dan Sowor)

Hati-hati pengukuran sikap, minat dan sebagainya merupakan self-inventory > hasil yang diperoleh perlu ditafsirkan secara hati-hati. Instruksi yang biasa dipakai kuesioner dengan skala likert (setuju – tidak setuju) atau SS – S ? Ts – STS

Selalu sering – kadang-kadang – tak pernah

Minat menurut definisi nominal tingkah laku yang ditunjukkan oleh seseorang apabila ia mempunyai kebebasan untuk memilih.

Self Inventeri: Quesioner bila mengukur sesuatu kita pakai chek list

Observasi

1.Check list: ya/tidak

2.Rating scale: -Selalu – Kadang-kadang – Tidak bebas

-Bagus – Sedang – Kurang – Jelek sekali

Komponen yang dinilai	Ya	Tidak
A		
B		

C		
D		

Observasi					
Bagus Sekali	Bagus	Baik	Sedang	Kurang	Kurang Sekali

Tetapi ada suatu kelompok yang perlu anda ketahui yaitu apa yang disebut Kompetensi kritis (yang harus ya/tidak)

, maka ada komponen ini yang harus tidak, kita mesti harus melihat itu.

Validitas Instrumen

Sejahu mana informasi yang diperoleh pada pengukuran, dapat diinterpretasikan sebagai tingkah laku atau karakteristik yang diukur.

Macam-macam Validitas

1. Validitas Isi

2. Validitas Sehubungan dengan suatu kriterium

- Validitas konkuren

Dua macam instrumen : - Diketahui mengukur

-Instrumen> bisa

-Validitas Prediktif>memerlukan waktu

3. Validitas Konstruk : paling sulit

Validitas butir: instruksional indistensy apakah tiap-tiap butir mengukur yang sama.

Internal consistensy method:

1. Split-half method

2. Test divided into 2 parts/holveseq odd-even items

3. Determine corr coef of the two halves = r

4. Reliability of total test

$r^1 = 2 \times \text{reliability for } \frac{1}{2} \text{ tes}$

$1 + \text{reliability for } \frac{1}{2} \text{ tes}$

KR – 20 rumusnya ada (cari) Kuder Richardson Yang diukur reliability

Reliability dapat dipercaya

.....> hasil testing dapat kira-kira sama

Mengukur keefektifan siswa instruksional mengukur keefektifan sistem intruksional

Prinsip-prinsip umum pengukuran

1. Harus ditentukan/dijelaskan apa yang diukur

2. Sekali pengukuran harus dipilih berdasarkan kegunaan pengukuran

3. Penilaian yang komprehensif memerlukan sistem gabungan berbagai teknik pengukuran

4. Masing-masing teknik pengukuran mempunyai keterbatasan

Macam-macam Evaluasi

1. Penempatan (placement)

Dilaksanakan pada sipengajar untuk melihat kemampuan anak

2. Formatif

Dilaksanakan selama pengujian untuk melihat kemajuan belajar

3. Diagnostik

Dilaksanakan selama pengajaran untuk melihat kesulitan belajar

4. Sumatif

Dilaksanakan pada akhir pengajaran untuk melihat prestasi atau keberhasilan belajar mahasiswa

Pengukuran: ganda multiple measures mengapa:

1. Untuk menghilangkan kelemahan

2. Keterbatasan pengukuran

3. Pengukuran yang tidak- reliabel-valid (sahi)

4. Perhatian kepada hasil yang tidak dikumpulkan sebelumnya

5. Perbedaan nilai pada kelompok-kelompok yang berbeda

Pengukuran (kuantitatif) dan pertimbangan (kualitatif)

Assesment proses pengumpulan informasi

Interprestasi dan sintesis lebih dari sekedar tes tertulis dan pemberian skor

Assesment = untuk

Dosen ----- → mahasiswa

Prinsip-prinsip umum pengukuran

1. Harus jelas/ditentukan apa yang akan diukur
2. Teknik pengukuran harus dipilih berdasarkan kegunaan pengukuran tersebut
3. Pengukuran yang komprehensif memerlukan adanya sistem gabungan berbagai teknik pengukuran
4. Masing-masing teknik pengukuran mempunyai keterbatasan/kekuatannya sendiri

Fungsi pengukuran

1. Mendeskripsikan gejala sosial/psikologis
2. Memperoleh data yang dapat dipakai untuk manipulasi statistik
3. Memungkinkan pengujian hipotesis/tesis
4. Memungkinkan peneliti membedakan objek-objek penelitian berdasarkan tingkah laku masing-masing

Kegunaan Pengukuran

1. Membantu dosen dalam mengambil keputusan
2. Meningkatkan motivasi mahasiswa
3. Meningkatkan retensi dan transfer of learning pada mahasiswa
4. Menentukan rangking mahasiswa
5. Melihat mana tugas yang sudah atau belum dikuasai mahasiswa

Evaluasi (Penilaian)

Pengukuran

1. Proses untuk memperoleh data yang bersifat kuantitatif
2. Dengan bantuan suatu alat instrumen
3. Hasil data numerik

Penilaian

1. Proses untuk mengumpulkan informasi yang dipakai sebagai dasar pengambilan keputusan
- Informasi..... > pertimbangan

Keputusan

Data bersifat kualitatif dan kuantitatif

Pengukuran

1. Proses pemberian angka untuk dapat menyatakan sifat-sifat atau karakteristik suatu objek, kejadian, orang
2. Mencakup pemakaian prosedur tertentu sesuai dengan aturan-aturan yang tidak ditentukan
3. Dinyatakan dalam suatu skala yang jelas

Pengujian

1. Suatu prosedur, sistematis yang dipakai untuk pengukuran tingkah laku/karakteristik seseorang
2. Hanya mengukur respons
3. Penguji mengadakan inferens (kesimpulan) berdasarkan respons yang diberikan tersebut

Daftar Pustaka

- Aunudin. 1989. Analisis Data. PAU Ilmu Hayat, IPB. Bogor
- Cochran, W.G. 1977. Sampling Techniques. 4th ed. Wiley. New York.
- Departemen Pendidikan Nasional, Pengembangan Sistem Penilaian, Jakarta: Direktorat Pendidikan Menengah Umum, 2004
- Departemen Pendidikan Nasional, Bahan Persentasi Penetapan KKM, 2006
- Green, P.E., D. S.Tull & G. Albaum. 1988. Research for Marketing Decission. 5th ed. Prentice Hall. New Jersey
- Greenacre, M. J. 1984. Theory and Application Correspondence Analysis. Academic Press, London.
- Kerlinger, F. N. 1973. Foundation of Behavioral Research. Holt, Rinehart and Winston, Inc. New York.
- Naga, Dali. S. 1992. Pengantar Teori Skor Pada Pengukuran Pendidikan. Gunadarma, Jakarta.
- Oppenheim, A. N. 1992. Questionnaire Design, Interviewing and Attitude Measurement. St. Martin Press. New York.
- Walpole, R.E. & R.H. Myers. 1978. Probability and Statistics for Engineers and Scientist. 2nd ed. Collier Macmillan Publisher. London.
- Zainal Arifin, Evaluasi Pembelajaran: Prinsip-Teknik-Prosedur, Bandung: PT. Remaja Rosdakarya, 2013.