

**PENGEMBANGAN SISTEM EVALUASI
PENDIDIKAN AGAMA ISLAM**

OLEH:

Dr. Andi Warisno, M.MPd

Kata Pengantar

Puji syukur kehadiran Allah SWT atas limpahan rahmat dan karunia-Nya kami dapat menyelesaikan buku “Pengembangan Sistem Evaluasi Pendidikan Agama Islam” ini. Sholawat dan salam semoga senantiasa tercurahkan kepada junjungan Nabi Agung kita, yaitu Nabi Muhammad SAW yang telah menunjukkan kepada kita jalan yang lurus berupa ajaran Agama Islam yang sempurna dan menjadi anugerah serta rahmat bagi seluruh alam semesta.

Penulis sangat bersyukur karena telah menyelesaikan buku ini. Disamping itu, kami mengucapkan banyak terima kasih kepada semua pihak yang telah membantu kami selama pembuatan buku ini berlangsung sehingga terealisasikan buku ini.

Besar harapan kami agar buku ini dapat diterima dan menjadi referensi dalam pembelajaran guna perkembangan ilmu pengetahuan.

Demikian yang dapat kami sampaikan, semoga buku ini bisa bermanfaat sebagai bahan bacaan dan bahan ajar serta rujukan dan kritik dan saran terhadap buku ini agar kedepannya bisa diperbaiki.

Lampung Selatan, Februari 2017

Penulis

DAFTAR ISI

HALAMAN COVER	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iii
BAB I PENDAHULUAN.....	1
BAB II MATERI POKOK SISTEM EVALUASI PENDIDIKAN.....	6
A. Pengetahuan Dasar tentang Evaluasi Pendidikan	7
1. Pengertian Evaluasi Pendidikan.....	7
2. Tujuan dan Fungsi Evaluasi.....	8
3. Dasar al-Qur'an tentang evaluasi.....	9
4. Ciri-ciri dan prinsip-prinsip evaluasi pendidikan	10
5. Jenis Evaluasi Pendidikan.....	12
B. Teknik dan Bentuk Evaluasi Hasil Belajar	23
1. Teknik Evaluasi non tes.....	23
2. Teknik Evaluasi tes	23
C. Teknik pembuatan soal	31
1. Soal tes tertulis.....	31
2. Soal tes perbuatan	32
3. Soal instrumen non tes	32
D. Perencanaan penulisan tes.....	45
1. Langkah – langkah menyusun tes.....	45
2. Membuat kisi – kisi soal.....	45
3. Tingkat kesukaran (TK) dan Daya Pembeda (DP).....	47
E. Pemberian Skor dan Sistem Penilaian	59
F. Validitas dan Reliabilitas Tes	75
G. Tes standar dan tes non standar.....	85
H. Analisis hasil evaluasi dan pengembangannya	95

RANGKUMAN
DAFTAR PUSTAKA

BAB I PENDAHULUAN

Pendidikan karakter ditempatkan sebagai landasan untuk mewujudkan visi pembangunan nasional, yaitu terwujudnya masyarakat yang berakhlak mulia, bermoral, beretika, berbudaya, dan beradab berdasarkan falsafah Pancasila. Hal ini sekaligus menjadi upaya untuk mendukung perwujudan cita-cita sebagaimana diamanatkan Pancasila dan pembukaan UUD 1945.

Upaya pembentukan karakter sesuai dengan budaya bangsa dapat dilakukan dengan adanya pelaksanaan evaluasi yang berbasis karakter. Penilaian melalui pendidikan karakter diharapkan mampu mewujudkan cita-cita bangsa. Selain itu penilaian berbasis karakter dapat melatih siswanya untuk belajar secara aktif, mandiri dan integratif.

Penilaian merupakan suatu proses, yang menurut Cronbach adalah proses pengumpulan data dan penggunaan informasi yang dipergunakan sebagai dasar pembuatan keputusan tentang program pendidikan. Keputusan adalah pilihan diantara berbagai arah tindakan. Jadi, penilaian menurut Cronbach memiliki komponen pengumpulan informasi dan pembuatan keputusan. Penilaian adalah kegiatan untuk menentukan pencapaian hasil pembelajaran.

Pembentukan karakter juga tidak lepas dari peran guru, karena segala sesuatu yang dilakukan oleh guru mampu mempengaruhi karakter peserta didik. Karakter terbentuk dari tiga macam bagian yang saling berkaitan yakni pengetahuan moral, perasaan moral, dan perilaku moral.

Menurut Kamus Besar Bahasa Indonesia, karakter adalah sifat-sifat kejiwaan, akhlak atau budi pekerti yang membedakan seseorang dengan yang lain.

Jadi, penilaian berbasis karakter adalah pencapaian hasil belajar yang berorientasi pada dua aspek yaitu aspek afektif dan psikomotor.

Tujuan penilaian berbasis karakter hampir sama dengan tujuan evaluasi, karena keduanya sama-sama bertujuan untuk mengetahui hasil belajar siswa. Adapun tujuan penilaian tersebut yaitu

1. Untuk mengetahui tingkat kemajuan yang telah dicapai oleh siswa dalam suatu kurun waktu proses belajar tertentu.
2. Untuk mengetahui posisi atau kedudukan seorang siswa dalam kelompok kelasnya.
3. Untuk mengetahui tingkat usaha yang dilakukan siswa dalam belajar.
4. Untuk mengetahui hingga sejauh mana siswa telah mendayagunakan kapasitas afektifnya.
5. Dan untuk mengetahui tingkat daya guna dan hasil guna metode mengajar yang telah di gunakan guru dalam proses mengajar-belajar (PMB).

Disamping memiliki tujuan, penilaian berbasis karakter juga memiliki fungsi yaitu sebagai bahan pertimbangan pengembangan hasil belajar, untuk membentuk tiga macam bagian yang saling berkaitan yakni pengetahuan moral, perasaan moral, dan perilaku moral. Sehubungan dengan tujuannya penilaiannya ini maka yang menjadi sasaran penilaian kawasan afektif adalah perilaku anak didik, bukan pengetahuannya.

Karakter merupakan bagian dari ranah afektif dan ranah psikomotor. Karakter berorientasi pada keduanya dan saling berkaitan. Adapun ranah afektif dan ranah psikomotor sebagai berikut:

1. Ranah afektif

Afektif selalu berhubungan dengan minat dan sikap, seperti komitmen, tanggung jawab, disiplin, percaya diri, jujur, menghargai pendapat orang lain, pengendalian diri dan sebagainya. Oleh karena itu, ketika strategi pembelajaran afektif di terapkan dalam proses pembelajaran, secara otomatis akan berorientasi pada penanaman nilai-nilai karakter tersebut.

Dalam merencanakan penyusunan instrumen tes presentasi siswa yang berdimensi afektif (ranah rasa) jenis-jenis presentasi internalisasi dan karakterisasi

seyogyanya mendapat perhatian khusus. Alasannya, karena kedua jenis prestasi ranah rasa itulah yang lebih mengendalikan sikap dan perbuatan siswa.

Pengukuran ranah afektif tidaklah semudah mengukur ranah kognitif. Pengukuran ranah afektif tidak dapat dilakukan setiap saat (dalam arti pengukuran formal) karena perubahan tingkah laku siswa tidak dapat berubah sewaktu-waktu. Perubahan sikap seseorang memerlukan waktu yang relatif lama. Demikian juga pengembangan minat dan penghargaan serta nilai-nilai. Ranah afektif biasa dilihat dari sisi sikap. Aspek sikap dapat dinilai dengan cara berikut :

a. Observasi

Merupakan teknik penilaian yang dilakukan secara berkesinambungan dengan menggunakan indera, baik secara langsung maupun tidak langsung dengan menggunakan format observasi yang berisi sejumlah indikator perilaku yang diamati. Hal ini dilakukan saat pembelajaran maupun diluar pembelajaran.

Observasi adalah suatu pengamatan langsung terhadap siswa dengan memperhatikan tingkah lakunya. Sehingga dalam observasi guru/dosen tidak perlu mengadakan komunikasi langsung dengan siswa.

Observasi dapat dilakukan di berbagai tempat misalnya di kelas pada waktu pelajaran, di halaman sekolah ketika murid bermain-main, dilapangan ketika murid olahraga dll.

Sebagai alat evaluasi, observasi dapat dipakai untuk :

- 1) Menilai minat, sikap, dan nilai-nilai yang terkandung dalam diri siswa.
- 2) Melihat proses kegiatan yang dilakukan oleh seorang siswa atau kelompok.
- 3) Memperoleh informasi balikan guru idalam kegiatan belajar mengajar. Untuk melihat apakah seorang guru efektif pengajarannya, dilakukan observasi terhaap tingkah laku siswanya. Seperti :
 - a) Apakah siswanya dapat mendengar dengan baik?
 - b) Apakah mereka merespon dengan baik terhadap tugas yang di berikan?
 - c) Apakah jawaban-jawaban siswa memberi petunjuk-petunjuk yang diberikan telah mereka pahami?

Observasi digunakan oleh guru dengan cara mengamati kegiatan peserta didik baik secara langsung maupun tidak langsung.

Langkah-langkah dalam merancang pedoman observasi antara lain:

- (1) Merumuskan tujuan observasi
- (2) Membuat kisi-kisi observasi
- (3) Menyusun pedoman observasi
- (4) Menyusun aspek-aspek yang akan diobservasi
- (5) Melakukan uji coba pedoman observasi
- (6) Memperbaiki pedoman observasi berdasarkan hasil uji coba
- (7) Melaksanakan observasi pada saat kegiatan berlangsung
- (8) Mengolah dan menafsirkan hasil observasi.

b. Penilaian diri

Merupakan teknik penilaian dengan cara meminta peserta didik untuk mengemukakan kelebihan dan kekurangan dirinya dalam konteks pencapaian kompetensi. Instrument yang digunakan berupa lembar penilaian diri.

c. Penilaian antar teman

Merupakan teknik penilaian dengan cara meminta peserta didik untuk saling menilai terkait dengan sikap dan perilaku keseharian peserta didik. Instrument penilaian berupa lembar penilaian antarpeserta didik.

d. Skala sikap

Dalam skala sikap ini perilaku peserta didik dievaluasi melalui kegiatan pengukuran sikap. Salah satu model skala sikap yang sering digunakan adalah skala Likert.

Dalam penggunaannya, peserta didik tidak hanya dapat memilih pernyataan-pernyataan yang positif saja, tetapi juga pernyataan-pernyataan yang negatif. Biasanya pernyataan positif diberi skor 5,4,3,2,1 sedang pernyataan negative 1,2,3,4,5.

Untuk dapat merancang skala Likert, guru dapat mengikuti langkah-langkah berikut ini:

- 1) Memilih berbagai variabel afektif yang akan diukur
- 2) Menetapkan berbagai variabel afektif
- 3) Membuat beberapa pernyataan tentang variabel afektif
- 4) Mengklasifikasikan pernyataan positif dan negative

- 5) Menentukan jumlah gradual dan frase atau angka yang dapat menjadi alternatif pilihan
- 6) Menyusun pernyataan dan pilihan jawaban menjadi sebuah alat penilaian
- 7) Melakukan uji coba
- 8) Membuang butir-butir pernyataan yang kurang baik
- 9) Melaksanakan penilaian

BAB II

MATERI POKOK SISTEM EVALUASI PENDIDIKAN

A. Pengetahuan Dasar tentang Evaluasi Pendidikan

1. Pengertian Evaluasi Pendidikan

Evaluasi adalah proses penilaian. Dalam perusahaan, dapat didefinisikan sebagai proses evaluasi akan mengukur Efektivitas strategi yang digunakan untuk mencapai tujuan perusahaan. Data yang diperoleh dari hasil pengukuran akan digunakan sebagai analisis situasi program berikutnya.

Ada hal yang mendasari atau proses sebelum melakukan evaluasi, yaitu :

- Mengembangkan konsep dan melakukan penelitian awal. Konsep harus direncanakan dengan baik sebelum eksekusi terorganisir dan pesan harus tes untuk memeriksa kompatibilitas antara draft yang disiapkan oleh pesan eksekusi.
- Dengan uji coba yang dilakukan, evaluator mencoba untuk mencari respon dari penonton. Respon penonton penting untuk mengukur efektivitas pesan.
- Dalam melakukan proses evaluasi, ada beberapa hal yang akan dibahas adalah apa bahan evaluasi, proses evolusi, ketika evaluasi diadakan, mengapa harus ada evaluasi, di mana proses evaluasi diadakan, dan pihak evaluasi.
- Hal yang harus dilakukan evaluasi ini sumber yang ada, efektivitas menyebarkan pesan, pemilihan media dan keputusan yang tepat keputusan dalam anggaran mengadakan sejumlah promosi dan iklan.

Evaluasi tersebut perlu dilakukan untuk menghindari kesalahan perhitungan pembiayaan, memilih strategi terbaik dari berbagai alternatif strategi yang tersedia, meningkatkan efisiensi iklan pada umumnya, dan melihat apakah tujuan telah tercapai

2. Tujuan dan Fungsi Evaluasi

Ada beberapa tujuan dan atau fungsi penilaian dalam pengajaran di sekolah, yaitu:

1. Untuk mengetahui apakah tujuan-tujuan pengajaran yang telah ditetapkan telah tercapai dalam kegiatan pembelajaran.
2. Untuk memberikan objektivitas pengamatan kita terhadap perilaku hasil belajar siswa.
3. Untuk mengetahui kemampuan siswa dalam bidang/topik tertentu.
4. Untuk menentukan kelayakan siswa, misalnya naik kelas, lulus.
5. Untuk memberikan umpan balik bagi kegiatan pembelajaran yang dilakukan.

Evaluasi memiliki beberapa fungsi yaitu:

1. Fungsi normatif, yaitu berfungsi untuk perbaikan sistem pembelajaran
2. Fungsi diagnostik, yaitu untuk mengetahui faktor kesulitan siswa dalam proses pembelajaran
3. Fungsi sumatif, yaitu berfungsi untuk mengetahui tingkat kemampuan peserta didik.
4. Fungsi penempatan

Penilaian hasil belajar pada dasarnya adalah mempermasalahkan, bagaimana pengajar (guru) dapat mengetahui hasil pembelajaran yang telah dilakukan. Pengajar harus mengetahui sejauh mana pembelajar (learner) telah mengerti bahan yang telah diajarkan atau sejauh mana tujuan/kompetensi dari kegiatan pembelajaran yang dikelola dapat dicapai. Tingkat pencapaian kompetensi atau tujuan instruksional dari kegiatan pembelajaran yang telah dilaksanakan itu dapat dinyatakan dengan nilai.

Dalam konteks pelaksanaan pendidikan, evaluasi memiliki beberapa tujuan, antara lain sebagai berikut:

1. Untuk mengetahui kemajuan belajar siswa setelah mengikuti kegiatan pembelajaran dalam jangka waktu tertentu.

2. Untuk mengetahui efektivitas metode pembelajaran.
3. Untuk mengetahui kedudukan siswa dalam kelompoknya.
4. Untuk memperoleh masukan atau umpan balik bagi guru dan siswa dalam rangka perbaikan.

Selain fungsi di atas, penilaian juga dapat berfungsi sebagai alat seleksi, penempatan, dan diagnostik, guna mengetahui keberhasilan suatu proses dan hasil pembelajaran. Penjelasan dari setiap fungsi tersebut adalah:

1. Fungsi seleksi. Evaluasi berfungsi atau dilaksanakan untuk keperluan seleksi, yaitu menyeleksi calon peserta suatu lembaga pendidikan/kursus berdasarkan kriteria tertentu.
2. Fungsi Penempatan. Evaluasi berfungsi atau dilaksanakan untuk keperluan penempatan agar setiap orang (peserta pendidikan) mengikuti pendidikan pada jenis dan/atau jenjang pendidikan yang sesuai dengan bakat dan kemampuannya masing-masing.
3. Fungsi Diagnostik. Evaluasi diagnostik berfungsi atau dilaksanakan untuk mengidentifikasi kesulitan belajar yang dialami peserta didik, menentukan faktor-faktor yang menyebabkan terjadinya kesulitan belajar, dan menetapkan cara mengatasi kesulitan belajar tersebut.

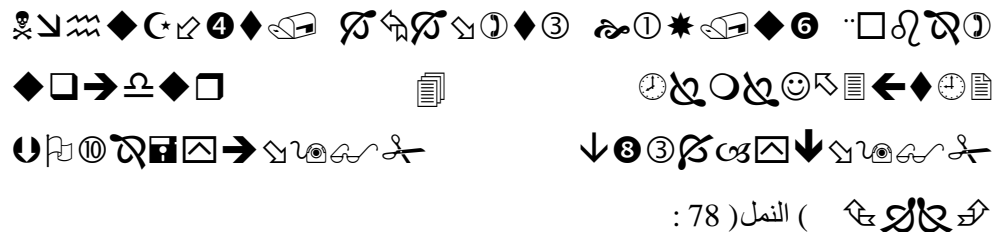
3. Dasar al-Qur'an tentang evaluasi

Istilah qur'ani tentang evaluasi memang tidak ditemukan di dalam ayat-ayat al-Qur'an, namun menurut analisis Prof. Ramayulis, ada beberapa terma dalam al-Qur'an yang maknanya dekat dengan evaluasi, di antaranya adalah :

1. Al-Hisāb/al-Muhāsabah

Artinya : (Dialah Allah) yang menjadikan mati dan hidup, supaya Dia menguji kamu, siapa di antara kamu yang lebih baik amalnya. dan Dia Maha Perkasa lagi Maha Pengampun,(Q.S. al-Mulk: 2).

3. Al-Hukm



Artinya : Sesungguhnya Tuhanmu akan menyelesaikan perkara antara mereka dengan keputusan-Nya, dan Dia Maha Perkasa lagi Maha mengetahui.(Q.S. AL-Naml : 78)

Berdasarkan dari ayat dan hadits di atas, pengertian evaluasi dapat dipahami bahwa evaluasi merupakan suatu usaha untuk memikirkan, memperkirakan, menimbang, mengukur, dan menghitung aktifitas yang telah dikerjakan, dikaitkan dengan tujuan yang dicanangkan untuk meningkat usaha dan aktifitas menuju tujuan yang lebih baik diwaktu mendatang, segi-segi yang mendukung dikembangkan dan segi-segi yang menghambat ditinggalkan.

4. Ciri–ciri dan prinsip–prinsip evaluasi pendidikan

Evaluasi hasil belajar dapat dikatakan terlaksana dengan baik apabila dalam pelaksanaannya senantiasa berpegang pada tiga prinsip berikut ini:

1. Prinsip Keseluruhan

Prinsip keseluruhan atau prinsip menyeluruh juga dikenal dengan istilah prinsip komprehensif. Dengan prinsip komprehensif dimaksudkan disini bahwa evaluasi hasil belajar dapat dikatakan terlaksana dengan baik apabila evaluasi tersebut dilaksanakan secara bulat, utuh atau menyeluruh.

2. Prinsip kesinambungan

Prinsip kesinambungan juga dikenal dengan istilah prinsip kontinuitas. Dengan prinsip kesinambungan dimaksudkan disini bahwa evaluasi hasil belajar adalah evaluasi hasil belajar yang dilaksanakan secara teratur dan sambung menyambung dari waktu ke waktu.

3. Prinsip objektivitas

Prinsip objektivitas mengandung makna, bahwa evaluasi hasil belajar dapat dinyatakan sebagai evaluasi yang baik apabila dapat terlepas dari faktor-faktor yang sifatnya subjektif.

Sebagai suatu bidang kegiatan, evaluasi hasil belajar memiliki ciri-ciri yang khas yang membedakannya dari bidang kegiatan yang lain. Diantara ciri-ciri yang dimiliki oleh evaluasi hasil belajar ada lima ciri, yaitu:

1. Bahwa evaluasi yang dilaksanakan dalam rangka mengukur keberhasilan belajar peserta didik itu , pengukurannya dilakukan secara tidak langsung.
2. Bahwa pengukuran dalam rangka menilai keberhasilan belajar peserta didik pada umumnya menggunakan ukuran-ukuran yang sifatnya kuantitatif, atau lebih sering menggunakan symbol-simbol angka.
3. Bahwa dalam kegiatan evaluasi pada umumnya digunakan unit-unit atau satuan-satuan yang tetap.
4. Bahwa prestasi belajar yang dicapai oleh para peserta didik dari waktu ke waktu adalah bersifat relative, dalam arti : bahwa hasil-hasil evaluasi terhadap keberhasilan belajar peserta didik itu pada umumnya tidak selalu menunjukkan kesamaan.
5. Bahwa dalam kegiatan evaluasi hasil belajar, sulit untuk dihindari terjadinya kekeliruan pengukuran.

5. Jenis Evaluasi Pendidikan

Pada prinsipnya, evaluasi hasil belajar merupakan kegiatan berencana dan berkesinambungan. Oleh karena itu, ragamnyapun banyak, mulai yang paling sederhana sampai yang paling kompleks.

Penilaian ada beberapa jenis, yaitu:

1. Penilaian Formatif; yaitu penilaian untuk mengetahui hasil belajar yang di capai oleh para peserta didik setelah menyelesaikan program dalam satuan materi pokok pada suatu bidang studi tertentu:
 - a. Fungsi
Untuk memperbaiki proses pembelajaran kearah yang lebih baik dan efisien atau memperbaiki satuan atau rencana pembelajaran.
 - b. Tujuan
Untuk mengetahui hingga dimana penguasaan peserta didik tentang materi yang diajarkan dalam satu rencana atau satuan pembelajaran.
 - c. Aspek penilaian
Aspek yang dinilai pada penilaian normative ialah hasil kemajuan belajar peserta didik yang meliputi pengetahuan, keterampilan, sikap terhadap materi ajar agama yang di sajikan.
2. Penilaian Sumatife; yaitu penilaian yang di lakukan terhadap hasil belajar peserta didik yang telah selesai mengikuti pembelajaran dalam satu catur wulan semester atau akhir tahun.
 - a. Fungsi, untuk mengetahui angka atau nilai murid setelah mengikuti program pembelajaran dalam satu catur wulan/ semester.
 - b. Tujuan, untuk mengetahui taraf hasil belajar yang dicapai oleh peserta didik setelah melakukan program pembelajaran dalam satu catur

wulan, semester, akhir tahun atau akhir suatu program pembelajaran pada suatu unit pendidikan tertentu.

c. Aspek Penilaian, aspek yang di nilai ialah kemajuan hasil belajar meliputi pengetahuan, keterampilan, sikap dan penguasaan murid tentang materi pembelajaran yang di berikan. Di waktu pelaksanaan, penilaian ini di laksanakan sebelum peserta didik mengikuti proses pembelajaran permulaan atau peserta didik tersebut baru akan mengikuti pendidikan di suatu tingkat tertentu.

3. Penilaian Penempatan (placement) yaitu penilaian tentang pribadi peserta didik untuk kepentingan penempatan di dalam situasi belajar yang sesuai dengan kondisi peserta didik.

a. Fungsi; untuk mengetahui keadaan peserta didik sepintas lalu termasuk keadaan seluruh pribadinya, peserta didik tersebut dapat di tempatkan pada posisinya.

b. Tujuan; untuk menempatkan peserta didik pada tempatnya yang sebenarnya, berdasarkan bakat, minat, kemampuan, kesanggupan, serta keadaan diri peserta didik sehingga peserta didik tidak mengalami hambatan dalam mengikuti pembelajaran atau setiap program bahan yang di sajikan guru.

c. Aspek Penilaian; aspek yang di nilai meliputi keadaan fisik dan psikis, bakat, kemampuan, pengetahuan, pengalamn, keterampilan, sikap dan aspek lain yang di anggap perlu bagi kepentingan pendidikan peserta didik selanjutnya, kemungkinan penilaian ini dapat juga di lakukan setelah peserta didik mengikuti pelajaran selama satu catur wulan, satu semester, satu tahun, sesuai denagn maksud lembaga pendidikan yang bersangkutan. Di waktu pelaksanaan, penilaian ini sebaiknya di laksanakan sebelum peserta didik menduduki kelas tertentu sewaktu penerimaan murid baru atau setelah naik kelas.

4. Penilaian Diagnostik; yaitu penilaian yang di lakukan terhadap hasil penganalisaan tentang keadaan belajar peserta didik baik merupakan kesulitan atau hambatan yang di temui dalam proses pembelajaran.

- a. Fungsi; untuk mengetahui masalah-masalah yang di derita atau mengganggu peserta didik, sehingga peserta didik mengalami kesulitan, hambatan atau gangguan ketika mengikuti program pembelajaran dalam suatu bidang study. Kesulitan peserta didik tersebut di usahakan pemecahannya.
- b. Tujuan; untuk membantu kesulitan atau mengetahui hambatan yang di alami peserta didik waktu mengikuti kegiatan pembelajaran pada suatu bidang study atau keseluruhan program pembelajaran.
- c. Aspek Penilaian; aspek yang di nilai, termasuk hasil belajar yang di peroleh murid, latar belakang kehidupannya, serta semua aspek yang berkaitan dengan kegiatan pembelajaran.
- d. Waktu Pelaksanaan; pelaksanaan tes diagnostik ini, sesuai dengan keperluan pembinaan dari suatu lembaga pendidikan, dalam rangka meningkatkan mutu pendidikan para peserta didiknya.

B. Teknik dan Bentuk Evaluasi Hasil Belajar

1. Teknik Evaluasi non tes

Teknik penilaian non-tes dapat digunakan untuk mengetahui proses dan produk dari hasil belajar peserta didik, misalnya berkaitan dengan kompetensi sikap. Berikut ini beberapa jenis teknik penilaian non-tes:

- a. Observasi Observasi atau pengamatan adalah teknik penilaian yang dilakukan dengan menggunakan indera secara langsung. Observasi dilakukan dengan menggunakan pedoman observasi yang berisi sejumlah indikator perilaku yang akan diamati.

Berdasarkan kerangka kerjanya, observasi dibedakan menjadi dua jenis yaitu:

1. Observasi berstruktur, yaitu semua kegiatan guru sebagai observer telah ditetapkan terlebih dahulu berdasarkan kerangka kerja yang berisi faktor – faktor

yang telah diatur skenarionya. Isi dan luas materi observasi telah ditetapkan dan dibatasi dengan jelas dan tegas.

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

Berdasarkan teknis pelaksanaannya, observasi dapat ditempuh melalui tiga cara yaitu.

1. Observasi langsung, yaitu observasi yang dilakukan secara langsung terhadap objek yang diselidiki.

2. Observasi tak langsung, yaitu observasi yang dilakukan melalui perantara, baik teknik maupun alat instrumen.

3. Observasi partisipasi, yaitu observasi yang dilakukan dengan cara ikut ambil bagian atau melibatkan diri dalam situasi objek yang diteliti.

b. Wawancara

Teknik wawancara pada satu segi mempunyai kesamaan arti dengan tes lisan yang telah diuraikan di atas. Teknik wawancara ini diperlukan pendidik untuk tujuan mengungkapkan atau menanyakan lebih lanjut hal-hal yang kurang jelas informasinya. Teknik wawancara ini dapat pula digunakan sebagai alat untuk menelusuri kesukaran yang dialami peserta didik tanpa ada maksud untuk menilai. Wawancara terdiri dari dua jenis yaitu wawancara langsung dan tidak langsung.

1. Wawancara langsung adalah wawancara yang dilakukan secara langsung antara pewawancara atau guru dengan orang yang diwawancara atau peserta didik tanpa melalui perantara.

2. Wawancara tidak langsung adalah pewawancara atau guru menanyakan sesuatu kepada peserta didik melalui perantara orang lain atau media. Berdasarkan bentuknya, pertanyaan wawancara dibagi menjadi tiga jenis yaitu:

1. Bentuk pertanyaan berstruktur, yaitu pertanyaan yang menuntut jawaban agar sesuai dengan apa yang terkandung dalam pertanyaan tersebut. pertanyaan semacam ini biasanya digunakan jika masalahnya tidak terlalu kompleks dan jawabannya sudah kongrit.

2. Bentuk pertanyaan tak berstruktur, yaitu pertanyaan yang bersifat terbuka, peserta didik secara bebas menjawab pertanyaan tersebut. pertanyaan semacam ini tidak memberi struktur jawaban kepada peserta didik karena jawaban dalam pertanyaan itu bebas.

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

c. Penugasan

Penugasan adalah suatu teknik penilaian yang menuntut peserta didik melakukan kegiatan tertentu di luar kegiatan pembelajaran di kelas. Penugasan dapat diberikan dalam bentuk individual atau kelompok. Penugasan ada yang berupa pekerjaan rumah atau berupa proyek. Pekerjaan rumah adalah tugas yang harus diselesaikan peserta didik di luar kegiatan kelas, misalnya menyelesaikan soal-soal dan melakukan latihan.

Proyek adalah suatu tugas yang melibatkan kegiatan perancangan, pelaksanaan, dan pelaporan secara tertulis maupun lisan dalam waktu tertentu dan umumnya menggunakan data lapangan.

d. Penilaian portofolio

Penilaian portofolio adalah suatu model penilaian yang bertujuan untuk mengukur kemampuan peserta didik dalam membangun dan merefleksi suatu tugas atau karya melalui pengumpulan bahan – bahan yang relevan dengan tujuan

dan keinginan yang dibangun oleh peserta didik, sehingga hasil pekerjaan tersebut dapat dinilai dan dikontari oleh guru dalam periode tertentu.

Jadi penilaian portofolio merupakan suatu pendekatan dalam penilaian kinerja peserta didik. Penilaian portofolio dibagi menjadi dua jenis yaitu portofolio proses dan produk.

1) Portofolio proses Portofolio proses yaitu menunjukkan kegiatan pembelajaran untuk mencapai standar kompetensi, kompetensi dasar, dan sekumpulan indikator yang telah ditetapkan dalam kurikulum serta menunjukkan semua hasil dari awal sampai dengan akhir selama kurun waktu tertentu.

Contoh portofolio proses yaitu Portofolio kerja (working portofolio) yang digunakan untuk memantau kemajuan atau perkembangan dan menilai peserta didik dalam mengelola kegiatan belajar mereka sendiri. Pada portofolio kerja ini yang dinilai adalah cara kerja dan hasil kerja siswa.

2) Portofolio produk Penilaian portofolio ini hanya menekankan pada penugasan dari tugas yang dituntut dalam standar kompetensi, kompetensi dasar dan sekumpulan indikator hasil belajar serta hanya menunjukkan evidence paling baik tanpa memperhatikan bagaimana dan kapan evidence tersebut diperoleh. Contoh portofolio produk adalah portofolio penampilan (show portofolio) dan portofolio dokumentasi (documentary portofolio).

a) Portofolio penampilan (show portofolio) Portofolio penampilan digunakan untuk memilih hal-hal yang paling baik yang menunjukkan bahan atau pekerjaan terbaik yang dihasilkan oleh siswa.

b) Portofolio dokumentasi (documentary portofolio) Portofolio dokumentasi adalah seleksi hasil kerja terbaik siswa yang akan diajukan dalam penilaian. Portofolio dokumentasi tidak hanya berisi hasil kerja siswa, tetapi semua proses yang digunakan oleh siswa untuk menghasilkan karya tertentu.

e. Penilaian diri

Penilaian diri merupakan teknik penilaian dengan cara meminta peserta didik untuk mengemukakan kelebihan dan kekurangan dirinya berkaitan dengan kompetensi yang menjadi tujuan pembelajaran.

f. Penilaian antarteman

Penilaian antar teman merupakan teknik penilaian dengan cara meminta peserta didik untuk mengemukakan kelebihan dan kekurangan temannya dalam berbagai hal. Oleh karena itu, perlu ada pedomanan penilaian antarteman yang memuat indikator perilaku yang dinilai.

g. Jurnal

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

2. Teknik Evaluasi tes

Tes merupakan suatu alat untuk melakukan pengukuran, yaitu untuk mengumpulkan informasi karakteristik suatu objek. Menurut Djemari tes merupakan salah satu cara untuk menaksir besarnya kemampuan seseorang secara tidak langsung, yaitu melalui respon seseorang terhadap stimulus atau pertanyaan. Tes dapat diartikan sebagai sejumlah pertanyaan yang harus diberikan tanggapan dengan tujuan untuk mengukur tingkat kemampuan seseorang atau mengungkap aspek tertentu dari orang yang dikenai tes.

Dilihat dari bentuk jawaban peserta didik maka tes dapat dibagi menjadi 3 jenis, yaitu tes tertulis, tes lisan dan tes perbuatan.

a. Tes tertulis

Sering disebut pencil test atau paper test. Adalah tes yang menuntut jawaban dari peserta didik dalam bentuk tertulis. Tes tertulis ada dua bentuk yaitu bentuk uraian (essay) atau subjektif dan bentuk objektif (objektive).

Tes tertulis pada umumnya tidak bisa digunakan secara efektif untuk mengevaluasi keterampilan psikomotorik siswa. Akan tetapi tes tertulis dapat mengevaluasi prinsip-prinsip yang menyertai keterampilan termasuk keterampilan kognitif, afektif dan psikomotorik.

1) Tes subjektif

Pada umumnya berbentuk esai (uraian). Tes bentuk essay adalah sejenis tes kemajuan belajar yang memerlukan jawaban yang bersifat pembahasan atau uraian kata-kata. Ciri-ciri pertanyaannya didahului dengan kata-kata seperti; uraikan, jelaskan, mengapa, bagaimana, bandingkan, simpulkan, dan sebagainya.

Jumlah butir soal dalam tes uraian biasanya tidak banyak, hanya sekitar 5-10 butir soal dalam waktu kira-kira 90-120 menit. Soal-soal bentuk uraian ini menuntut kemampuan peserta tes untuk dapat mengorganisir, menginterpretasi, menghubungkan pengertian-pengertian yang dimiliki. Secara singkat dapat dikatakan bahwa tes uraian menuntut peserta untuk dapat mengingat-ingat dan mengenal kembali, dan terutama harus mempunyai daya kreativitas yang tinggi.

Berdasarkan tingkat kebebasan tingkat peserta tes untuk menjawab soal tes uraian, secara umum tes uraian dapat dibagi menjadi dua bentuk yaitu tes uraian bebas atau uraian terbuka (extended response) dan tes uraian terbatas (restricted response).

a) Tes uraian bebas (extended response test)

Merupakan bentuk tes yang memberikan kebebasan kepada peserta tes untuk mengorganisasikan dan mengekspresikan pikiran dan gagasannya dalam menjawab soal tes. Jawaban peserta tes bersifat terbuka, fleksibel dan tidak terstruktur. Contoh ; jelaskan alasan mengapa sistem ekonomi yang dianut suatu negara berbeda-beda.

Peserta ujian diberi kebebasan untuk menjawab menurut gaya bahasa dan gaya kognitifnya masing-masing, sesuai dengan kemampuan mengingat mereka. Dengan demikian maka keterampilan mengekspresikan pikiran dalam bentuk

tertulis akan besar sekali kontribusinya dalam soal ujian tipe seperti ini. Bentuk soal seperti ini baik untuk mengukur hasil belajar pada tingkatan aplikasi, analisis, evaluasi dan kreativitas.

b) Tes uraian terbatas (restricted response test)

Merupakan bentuk tes yang memberikan batasan-batasan atau rambu-rambu tertentu kepada para peserta tes dalam menjawab soal tes. Batasan tersebut mencakup format, isi dan ruang lingkup jawaban.

Ada beberapa ragam tes uraian terbatas antara lain ragam tes melengkapi dan tes jawaban singkat.

(1) Tipe jawaban melengkapi

Yaitu butir soal yang memerintahkan kepada peserta tes untuk melengkapi kalimat dengan suatu frasa, angka atau satu formula.

(2) Tipe jawaban singkat

Yaitu bentuk soal yang berbentuk pertanyaan yang dapat dijawab dengan satu kata, satu frase, satu angka dan satu formula.

2) Tes objektif

Yaitu bentuk tes yang mengandung kemungkinan jawaban atau respon yang harus dipilih oleh peserta didik. Jadi kemungkinan jawaban atau respon telah disediakan oleh penyusun butir soal.

Tes objektif sering juga disebut tes dikotomi karena jawabannya antara benar atau salah dan skornya antara 1 dan 0. Disebut objektif karena penilaiannya objektif. Siapaun yang mengoreksi tes objektif hasilnya akan sama karena kunci jawabannya sudah jelas dan pasti.

Secara umum ada tiga bentuk tes objektif, yaitu

a) Tipe benar salah (True-false test)

Adalah tes yang butir soalnya terdiri dari pernyataan yang disertai dengan alternatif jawaban atau pernyataan yang benar dan yang salah.

- (1) Kelebihan tipe benar salah
 - (a) Dapat mewakili pokok bahasan atau materi pelajaran lebih luas
 - (b) Mudah penyusunannya
 - (c) Mudah diskor
 - (d) Merupakan instrumen yang baik untuk mengukur fakta dan hasil belajar secara langsung, terutama yang berkaitan dengan ingatan.
- (2) Kekurangan
 - (a) Hanya dapat mengungkap daya ingatan dan penghafalan kembali
 - (b) Mendorong peserta tes untuk menebak jawaban
 - (c) Tipe menjodohkan (matching)

Ada beberapa istilah yang digunakan untuk menunjuk tes menjodohkan (matching test) seperti memasangkan, atau mencocokkan. Butir soal menjodohkan ditulis dalam dua kelompok yaitu pernyataan atau stem dan kelompok jawaban.

- (1) Kelebihan tes menjodohkan
 - (a) Baik untuk menguji hasil yang berhubungan dengan pengetahuan istilah, definisi, dan peristiwa.
 - (b) Dapat menguji kemampuan menghubungkan dua hal yang berhubungan
 - (c) Mudah dalam penyusunan.
- (2) Kelemahan
 - (a) Ada kecenderungan untuk menekan ingatan saja

(b) Kurang baik untuk menilai pengertian atau tafsiran.

c) Tipe pilihan ganda (multiple choice)

Adalah tes dimana setiap butir soalnya memiliki jumlah alternatif jawaban lebih dari satu. Jumlah alternatif jawaban berkisar antara dua sampai lima. Setiap tes pilihan ganda terdiri dari dua bagian yaitu (1) pernyataan atau disebut juga stem dan (2) alternatif pilihan jawaban atau disebut option.

Terdapat beberapa variasi atau modifikasi dari tes pilihan ganda, yaitu:

(a) Pilihan ganda analisis hubungan antar hal

Yaitu terdiri dari dua pernyataan yang dihubungkan oleh kata “sebab”. Jadi ada dua kemungkinan hubungan antara kedua pernyataan tersebut, yaitu ada hubungan sebab akibat atau tidak ada hubungan sebab akibat.

(b) Pilihan ganda analisis kasus

Yaitu peserta tes dihadapkan pada suatu kasus yang disajikan dalam bentuk cerita, peristiwa atau sejenisnya. kemudian diajukan pertanyaan dalam bentuk melengkapi pilihan.

(c) Pilihan ganda asosiasi

Struktur soalnya sama dengan melengkapi pilihan. Perbedaannya adalah kalau pada melengkapi pilihan hanya ada satu jawaban yang paling benar atau paling benar tapi pada melengkapi berganda justru jawaban yang benar lebih dari satu, bisa 2,3,4.

(d) Pilihan ganda dengan diagram, grafik, tabel dan sebagainya

Bentuk soal ini mirip dengan analisis kasus., baik struktur maupun pola pertanyaannya. Bedanya dalam tes bentuk ini tidak disajikan kasus dalam bentuk cerita atau peristiwa tetapi dalam diagram, gambar, grafik maupun tabel.

d) Jawaban singkat (short answer) dan melengkapi (completion)

Tes ini masing-masing menghendaki jawaban dengan kalimat dan atau angka yang hanya dapat dinilai benar atau salah.

(1) Kebaikan tes ini

(a) Sangat baik untuk menilai kemampuan peserta didik berkenaan dengan fakta

(b) Relatif mudah disusun

(c) Menuntut peserta didik untuk mengemukakan pendapat dengan singkat

(2) Kelemahan

(a) Hanya berkenaan dengan kemampuan mengingat saja

(b) Dibutuhkan waktu yang lama dalam mengoreksi

b. Tes lisan

Adalah tes yang menuntut jawaban dari peserta didik dalam bentuk lisan. Peserta didik akan mengucapkan jawaban dengan kata-katanya sendiri sesuai dengan pertanyaan atau perintah yang diberikan.

1) Kebaikan tes ini

a) Dapat mengetahui langsung kemampuan peserta didik dalam mengemukakan pendapatnya secara lisan

b) Tidak perlu menyusun soal-soal secara terurai, tetapi cukup mencatat pokok-pokok permasalahannya saja

c) Kemungkinan peserta didik akan menerka jawaban dan spekulasi dapat dihindari

2) Kelemahan

a) Membutuhkan waktu yang cukup lama

b) Seringnya muncul insur subjektifitas

c. Tes perbuatan (performance test)

Tes perbuatan atau tes praktik adalah tes yang menuntut jawaban peserta didik dalam bentuk perilaku, tindakan atau perbuatan. Lebih jauh Stignis (1994) mengemukakan “ tes tindakan adalah suatu bentuk tes yang peserta didiknya diminta untuk melakukan kegiatan khusus dibawah pengawasan penguji yang akan mengobservasi penampilannya dan membuat keputusan tentang kualitas hasil belajar yang didemonstrasikan.

Misalnya untuk melihat bagaimana cara menggunakan komputer dengan baik dan benar, guru harus menyuruh peserta didik untuk mempraktikkan atau mendemonstrasikan penggunaan komputer yang sesungguhnya sesuai dengan prosedur yang baik dan benar.

Sebagaimana jenis tes lain, tes tindakanpun mempunyai kelebihan dan kekurangan. Kelebihan tes tindakan adalah (1) satu-satunya teknik tes yang dapat digunakan untuk mengetahui hasil belajar dalam bidang keterampilan (2) sangat baik digunakan untuk mencocokkan antara pengetahuan teori dan keterampilan praktik (3) dalam pemggunaannya tidak mungkin peserta didik akan mencontek (4) guru dapat lebih mengenal masing-masing karakter peserta didik.

Adapun kelemahannya adalah (1) memakan waktu yang lama (2) dalam hal tertentu membutuhkan biaya yang besar (3) cepat membosankan (4) membutuhkan syarat pendukung yang lengkap baik waktu tenaga maupun biaya.

Ditinjau dari segi penyusunannya tes hasil belajar dapat dibedakan menjadi dua, yaitu:

a. Tes buatan guru (teacher made-test)

Yaitu tes yang telah disusun sendiri oleh guru yang akan mempergunakan tes tersebut. Tes ini biasanya digunakan untuk ulangan harian, formatif, dan

ulangan umum. Tes ini dimaksudkan untuk mengukur tingkat penguasaan peserta didik terhadap materi yang sudah disampaikan guru. Untuk itu guru harus membuat soal secara logis dan rasional mengenai pokok-pokok materi.

b. Tes yang telah distandarkan (standardised test)

Yaitu tes yang telah mengalami proses standarisasi yakni proses validasi dan keadaaan (reliability) sehingga tes tersebut benar-benar valid dan andal untuk suatu tujuan dan bagi suatu kelompok tertentu.

Suatu tes dikatakan valid jika tes tersebut benar-benar mampu menilai apa yang harus dinilai. Tes tersebut jika digunakan dapat mencapai sasaran sesuai dengan yang telah direncanakan sebelumnya. Dengan kata lain merupakan alat yang jitu karena telah mengalami try-out dan perbaikan. dan suatu tes disebut andal atau dapat dipercaya jika tes tersebut menunjukkan ketelitian pengukuran. Ketelitian itu berlaku untuk setiap orang yang sama. Jika tes itu andal maka skor hasil tes yang dibuat murid itu tetap sama.

Berdasarkan jumlah peserta didik tes hasil belajar ada dua macam, yaitu:

a. Tes perseorangan, yaitu tes yang dilakukan secara perorangan. Guru akan berhadapan dengan seorang peserta didik.

b. Tes kelompok, yaitu tes yang diadakan secara kelompok. Guru akan dihadapkan pada sekelompok peserta didik.

Ditinjau dari segi kegunaan untuk mengukur siswa terdapat tiga macam tes, yaitu:

a. Tes diagnostik

Adalah tes yang digunakan untuk mengetahui kelemahan-kelemahan siswa sehingga berdasarkan kelemahan tersebut dapat dilakukan pemberian perlakuan yang tepat. Secara umum tes ini disebut peninjauan masuk atau dalam istilah Inggris entering behaviour test. Ini dilakukan untuk mengukur tingkat penguasaan pengetahuan dasar untuk dapat menerima pengetahuan lanjutannya. Oleh karena

itu tes ini juga disebut prasyarat tes atau pre request test. Tes ini juga berfungsi sebagai tes penempatan (placement test).

b. Tes formatif

Dari arti kata form yang merupakan dasar istilah formatif maka evaluasi formatif dimaksudkan untuk mengetahui sejauh mana siswa telah terbentuk setelah mengikuti sesuatu program tertentu. Evaluasi formatif atau tes formatif diberikan pada akhir setiap program. Tes ini merupakan post-test atau tes akhir program.

c. Tes sumatif

Evaluasi sumatif atau tes sumatif dilaksanakan setelah berakhirnya pemberian sekelompok program atau sekelompok program yang lebih besar. Dalam pengalaman disekolah tes formatif dapat disamakan dengan ulangan harian sedangkan tes sumatif dapat disamakan dengan ulangan umum yang biasanya dilaksanakan pada akhir semester.

Berdasarkan aspek pengetahuan dan keterampilan maka tes dapat dibedakan menjadi 2 jenis, yaitu:

a. Tes kemampuan (power test)

Prinsip tes kemampuan adalah tidak adanya batasan waktu dalam pengerjaan tes. Jika waktu tes tidak dibatasi maka hasil tes dapat mengungkapkan kemampuan peserta didik yang sebenarnya.

b. Tes kecepatan (speed test)

Aspek yang diukur dalam tes kecepatan adalah kecepatan peserta didik dalam mengerjakan sesuatu pada waktu atau periode tertentu. Pekerjaan tersebut biasanya relatif mudah karena aspek yang diukur benar-benar kecepatan bukan aspek lain.

C. Teknik Pembuatan Soal

1. Soal Tes Tertulis

Penulisan soal merupakan salah satu langkah penting untuk dapat menghasilkan tes yang baik. Penulisan soal adalah karakteristik yang diuraikan dalam kisi-kisi. Soal yang digunakan dalam wilayah kelas, umumnya dibedakan menjadi dua, yaitu: (1) tes objektif yang umumnya sangat terstruktur dan mengharuskan peserta didik mengisi kata atau memilih jawaban yang benar dari sejumlah alternatif yang disajikan. (2) tes subjektif, seperti tes uraian, yang umumnya kurang terstruktur dan mengharuskan peserta didik memilih, mengorganisasikan, dan menyajikan jawaban atas pertanyaan atau pernyataan dengan kalimat sendiri.

Untuk berbagai macam kepentingan, seperti ujian kenaikan kelas, ujian sekolah dasar, atau ujian akhir nasional, tes objektif lebih efisien digunakan dibanding tes uraian. Penggunaan kedua bentuk tes ini harus tepat, sesuai dengan kompetensi dasar dan indikator yang telah ditetapkan. Dalam kurikulum sesuai dengan kelebihan dan kelemahan kedua bentuk tes ini.

1. Tes objektif

Tes objektif mencakup beberapa bentuk, tetapi pada umumnya dapat didiversifikasikan menjadi dua bagian yaitu peserta didik harus menuliskan kata atau kalimat sederhana dan tes yang mengharuskan peserta didik memilih beberapa kemungkinan jawaban yang tersedia. Disebut tes objektif, karena penilaian objektif, yaitu apabila benar diberi skor 1, salah diberi skor 0. Tes objektif seringkali disebut tes dikotomi, yaitu penilaian nol sampai satu (*dichotomously scored item*).

Berbagai bentuk tes objektif diatas, dapat dimodifikasi sesuai dengan kebutuhan seperti kompetensi dasar dan indikator yang hendak diukur. Beberapa bentuk soalh objektif digunakan untuk mengukur berbagai macam kemampuan, seperti pemahaman, kemampuan berpikir, dan kemampuan yang lebih tinggi.

Bentuk tes objektif memiliki ciri-ciri yang khusus dibandingkan tes subjektif, yaitu menghendaki peserta didik untuk merespon yang sangat terbatas, hanya dengan mengisi (dengan kalimat yang terbatas) ataupun memilih alternatif jawaban yang telah disediakan. Untuk sampai pada jawaban yang benar, peserta didik harus bisa menunjukkan kemampuan pengetahuan khusus, pemahaman, dan kemampuan yang berkaitan dengan soal. Peserta didik tidak dapat mengembangkan kemampuannya dalam hal mengorganisasi dan menyajikannya dalam kalimat sendiri.

Peserta didik hanya memilih jawaban dari berbagai alternatif jawaban yang disediakan. Kelemahan pilihan objektif adalah kurang dapat mengukur kemampuan yang berkaitan dengan mengorganisasi dan mengintegrasikan gagasan. Kemampuan tersebut hanya dapat diukur melalui tes subjektif seperti tes uraian.

Kelebihan Tes tulis Tes obyektif) yaitu :

- a. Dapat mencakup ruang lingkup materi yang luas
- b. Lebih representatif mewakili isi dan luas bahan, lebih objektif, dapat dihindari campur tangan unsur-unsur subjektif baik dari segi siswa maupun segi guru yang memeriksa
- c. Lebih mudah dan cepat cara pemeriksaannya karena dapat menggunakan kunci tes bahkan alat-alat hasil kemajuan teknologi.
- d. Pemeriksaannya dapat diserahkan kepada orang lain.
- e. Dalam pemeriksaannya tidak ada unsur subjektif yang mempengaruhi.

Kekurangan tes tulis (tes obyektif) yaitu :

- a. Persiapan untuk menyusunnya jauh lebih sulit daripada tes esay karena soalnya banyak dan harus teliti untuk menghindari kelemahan-kelemahan yang lain (yang diukur cenderung aspek kognitif tingkat rendah)

- b. Soal-soalnya cenderung untuk mengungkapkan ingatan dan daya pengenalan kembali saja, dan sukar untuk mengukur proses mental yang tinggi.
- c. Banyak kesempatan untuk main untung-untungan.
- d. Kerjasama antarsiswa pada waktu mengerjakan soal tes lebih terbuka.
- e. Tidak menuntut penalaran siswa.
- f. Tidak membutuhkan pemikiran analitis maupun sistematis.

2. Tes Subjektif

Salah satu contoh tes subjektif adalah tes uraian. Disebut subjektif karena penilaiannya tidak objektif yaitu tidak nol untuk jawaban salah dan 1 untuk jawaban benar. Tes subjektif sering pula dinamakan tes politomous (polytomously scored item).

Tes uraian dibedakan menjadi dua macam, yaitu extended response essay type, pada bentuk itu peserta didik menjawab hampir tak terbatas. Pada restricted response essay type, peserta didik dituntut untuk menjawab terbatas. Baik ditinjau dari sifat, panjang, organisasi jawaban sangat terbatas.

Kelebihan Tes Subjektif yaitu :

- a. Penyusunan soalnya mudah disiapkan dan disusun.
- b. Tidak memberi banyak kesempatan untuk berspekulasi atau untung-untungan (menebak jawaban).
- c. Mendorong siswa untuk berani mengemukakan pendapat serta menyusun dalam bentuk kalimat yang bagus
- d. Memberikan kesempatan kepada siswa untuk mengutarakan maksudnya dengan gaya bahasa dan caranya sendiri.
- e. Dapat diketahui sejauh mana siswa mendalami suatu masalah yang diteskan.

- f. Dapat melatih siswa berfikir logis, analitis, dan sistematis.

Kekurangan Tes Subjektif yaitu :

- a. Kadar validitas dan realibilitas rendah karena sukar diketahui segi-segi mana dari pengetahuan siswa yang betul-betul telah dikuasai.
- b. Kurang representatif dalam hal mewakili seluruh scope bahan pelajaran yang akan dites karena soalnya hanya beberapa saja (terbatas).
- c. Cara memeriksanya banyak dipengaruhi oleh unsur-unsur subjektif.
- d. Pemeriksanya lebih sulit sebab membutuhkan pertimbangan individual lebih banyak dari penilai.
- e. Waktu untuk koreksinya lama dan tidak dapat diwakilkan kepada orang lain.
- f. Cakupan materi terbatas atau sempit.
- g. Yang diukur cenderung tingkat kecerdasan kognitif tinggi.

2. Soal Tes Perbuatan

Tes perbuatan atau tes praktik merupakan suatu tes yang penilaiannya didasarkan pada perbuatan/praktik peserta didik. Sebelum menulis butir soal untuk tes perbuatan, guru dapat mengecek dengan pertanyaan berikut. Tepatkah kompetensi (yang akan diujikan) diukur dengan tes tertulis? Jika jawabannya tepat, kompetensi yang bersangkutan tidak tepat diujikan dengan tes perbuatan/praktik.

Dalam menilai perbuatan/kegiatan/praktik peserta didik dapat digunakan beberapa jenis penilaian perbuatan di antaranya adalah penilaian kinerja (performance), penugasan (project), dan hasil karya (product).

Dalam menulis butir soal untuk tes perbuatan, penulis soal harus mengetahui konsep dasar penilaian perbuatan/praktik. Maksudnya pernyataan dalam soal harus disusun dengan pernyataan yang betul-betul menilai perbuatan/praktik, bukan menilai yang lainnya.

Penilaian kinerja merupakan penilaian yang meminta peserta didik untuk mendemonstrasikan dan mengaplikasikan pengetahuan ke dalam konteks yang sesuai dengan kriteria yang ditetapkan. Dalam menulis butir soal, perhatikan terlebih dahulu kompetensi dari materi yang akan ditanyakan.

Penilaian penugasan merupakan penilaian tugas (meliputi: pengumpulan, pengorganisasian, pengevaluasian, dan penyajian data) yang harus diselesaikan peserta didik (individu/kelompok) dalam waktu tertentu. Aspek yang dinilai di antaranya meliputi kemampuan (1) pengelolaan, (2) relevansi, dan (3) keaslian.

Penilaian hasil karya merupakan penilaian keterampilan peserta didik dalam membuat suatu produk benda tertentu seperti hasil karya seni, misal lukisan, gambar, patung, dll. Aspek yang dinilai di antaranya meliputi: (1) tahap persiapan: pemilihan dan cara penggunaan alat, (2) tahap proses/produksi: prosedur kerja, dan (3) tahap akhir/hasil: kualitas serta estetika hasil karya. Di samping itu, guru dapat memberikan penilaian pada pembuatan produk rancang bangun/perekayasaan teknologi tepat guna misalnya melalui: (1) adopsi, (2) modifikasi, atau (3) difusi.

Kaidah penulisan soal tes perbuatan adalah seperti berikut.

A. Materi

Soal harus sesuai dengan indikator (menuntut tes perbuatan: kinerja, hasil karya, atau penugasan).

Pertanyaan dan jawaban yang diharapkan harus sesuai

Materi sesuai dengan kompetensi (urgensi, relevansi, kontinuitas, keterampilan sehari-hari tinggi).

Isi materi yang ditanyakan sesuai dengan jenjang jenis sekolah atau tingkat kelas.

B. Konstruksi

Menggunakan kata tanya atau perintah yang menuntut jawaban perbuatan/praktik. Ada petunjuk yang jelas tentang cara mengerjakan soal. Disusun pedoman penskorannya. Tabel, gambar, grafik, peta, atau yang sejenisnya disajikan dengan jelas dan terbaca

C. Bahasa/Budaya

Rumusan kalimat soal komunikatif Butir soal menggunakan bahasa Indonesia yang baku. Tidak menggunakan kata/ungkapan yang menimbulkan penafsiran ganda atau salah pengertian. Tidak menggunakan bahasa yang berlaku setempat/tabu. Rumusan soal tidak mengandung kata/ungkapan yang dapat menyinggung perasaan peserta didik.

3. Soal Instrumen Non Tes

a. Pengertian

Instrumen non-tes adalah instrumen selain tes prestasi belajar. Alat penilaian yang dapat digunakan antara lain adalah: lembar pengamatan/observasi (seperti catatan harian, portofolio, life skill) dan instrumen tes sikap, minat, dsb.

Pada prinsipnya, prosedur penulisan butir soal untuk instrumen non-tes adalah sama dengan prosedur penulisan tes pada tes prestasi belajar, yaitu menyusun kisi-kisi tes, menuliskan butir soal berdasarkan kisi--kisinya, telaah, validasi butir, uji coba butir, perbaikan butir berdasarkan hasil uji coba. Namun,

dalam proses awalnya, sebelum menyusun kisi-kisi tes terdapat perbedaan dalam menentukan validitas isi/konstruksinya.

Dalam tes prestasi belajar, validitas isi diperoleh melalui kurikulum dan buku pelajaran, tetapi untuk non-tes validitas isi/konstruksinya diperoleh melalui "teori". Teori adalah pendapat yang dikemukakan sebagai keterangan mengenai suatu peristiwa atau kejadian.

b. Pengamatan

Pengamatan merupakan suatu alat penilaian yang pengisiannya dilakukan oleh guru atas dasar pengamatan terhadap perilaku peserta didik yang sesuai dengan kompetensi yang hendak diukur. Pengamatan dapat dilakukan dengan menggunakan antara lain lembar pengamatan, penilaian portofolio dan penilaian kecakapan hidup.

Pelaksanaan pengamatan sikap dapat dilakukan guru pada sebelum mengajar, saat mengajar, dan sesudah mengajar. Perilaku minimal yang dapat dinilai dengan pengamatan untuk perilaku/budi pekerti peserta didik, misalnya: ketaatan pada ajaran agama, toleransi, disiplin, tanggung jawab, kasih sayang, gotong royong, kesetiakawanan, hormat-menghormati, sopan santun, dan jujur.

Portofolio merupakan deskripsi peta perkembangan kemampuan individu peserta didik. Jadi portofolio merupakan "kartu sehat" individu peserta didik. Bila ada peserta didik yang "sakit", tugas guru adalah (1) menentukan penyakitnya apa, kemudian (2) memberi obat yang tepat agar peserta didik cepat sembuh dari penyakitnya.

c. Kaidah Penulisan Soal

Dalam penulisan soal pada instrumen non-tes, penulis butir soal harus memperhatikan ketentuan/kaidah penulisannya. Kaidahnya adalah seperti berikut ini.

1. Materi

- a. Pernyataan harus sesuai dengan rumusan indikator dalam kisi-kisi.

b. Aspek yang diukur pada setiap pernyataan sudah sesuai dengan tuntutan dalam kisi-kisi (misal untuk tes sikap: aspek kognisi, afeksi atau konasinya dan pernyataan positif atau negatifnya).

2. Konstruksi

a. Pernyataan dirumuskan dengan singkat (tidak melebihi 20 kata) dan jelas.

b. Kalimatnya bebas dari pernyataan yang tidak relevan objek yang dipersoalkan atau kalimatnya merupakan pernyataan yang diperlukan saja.

c. Kalimatnya bebas dari pernyataan yang bersifat negatif ganda.

d. Kalimatnya bebas dari pernyataan yang mengacu pada masa lalu.

e. Kalimatnya bebas dari pernyataan yang faktual atau dapat diinterpretasikan sebagai fakta.

f. Kalimatnya bebas dari pernyataan yang dapat diinterpretasikan lebih dari satu cara.

g. Kalimatnya bebas dari pernyataan yang mungkin disetujui atau dikosongkan oleh hampir semua responden.

h. Setiap pernyataan hanya berisi satu gagasan secara lengkap.

i. Kalimatnya bebas dari pernyataan yang tidak pasti seperti semua, selalu, kadang-kadang, tidak satupun, tidak pernah.

j. Jangan banyak mempergunakan kata hanya, sekedar, semata-mata. Gunakanlah seperlunya.

3. Bahasa/Budaya

a. Bahasa soal harus komunikatif dan sesuai dengan jenjang pendidikan peserta didik atau responden.

b. Soal harus menggunakan bahasa Indonesia baku.

c. Soal tidak menggunakan bahasa yang berlaku setempat/tabu.

d. Contoh Penulisan Kisi-kisi Non-Tes dan Butir soal

1. Tes Skala Sikap

Berbagai definisi tentang sikap yang telah dikemukakan oleh para ahli, di antaranya adalah Mueller yang menyampaikan 5 definisi dari 5 ahli, adalah seperti berikut ini. (1) Sikap adalah afeksi untuk atau melawan, penilaian tentang, suka atau tidak suka, tanggapan positif/negatif terhadap suatu objek psikologis (Thurstone). (2) Sikap adalah kecenderungan untuk bertindak ke arah atau melawan suatu faktor lingkungan (Emory Bogardus). (3) Sikap adalah kesiapsiagaan mental atau saraf (Goldon Allport). (4) Sikap adalah konsistensi dalam tanggapan terhadap objek-objek sosial (Donald Cambell). (5) Sikap merupakan tanggapan tersembunyi yang ditimbulkan oleh suatu nilai (Ralp Linton, ahli antropologi kebudayaan).

Berdasarkan beberapa definisi di atas, para ahli menyimpulkan bahwa sikap memiliki 3 komponen penting, yaitu komponen: (1) kognisi yang berhubungan dengan kepercayaan, ide, dan konsep; (2) afeksi yang mencakup perasaan seseorang; dan (3) konasi yang merupakan kecenderungan bertingkah laku atau yang akan dilakukan.

Oleh karena itu, ketiga komponen ini dimasukkan di dalam format kisi-kisi "sikap belajar peserta didik" seperti contoh berikut. Adapun definisi operasional sikap belajar adalah kecenderungan bertindak dalam perubahan tingkah laku melalui latihan dan pengalaman dari keadaan tidak tahu menjadi tahu yang dapat diukur melalui: toleransi, kebersamaan dan gotong-royong, rasa kesetiakawanan, dan kejujuran.

2. Tes Minat belajar

Minat adalah kesadaran yang timbul bahwa objek tertentu sangat disenangi dan melahirkan perhatian yang tinggi bagi individu terhadap objek tersebut . Di samping itu, minat juga merupakan kemampuan untuk memberikan stimulus yang mendorong seseorang untuk memperhatikan aktivitas yang dilakukan berdasarkan pengalaman yang sebenarnya.

Berdasarkan kedua pengertian tersebut, minat merupakan kemampuan seseorang untuk memberikan perhatian terhadap suatu objek yang disertai dengan rasa senang dan dilakukan penuh kesadaran.

Peserta didik yang menaruh minat pada suatu mata pelajaran, perhatiannya akan tinggi dan minatnya berfungsi sebagai pendorong kuat untuk terlibat secara aktif dalam kegiatan belajar mengajar pada pelajaran tersebut. Oleh karena itu, definisi operasional minat belajar adalah pilihan kesenangan dalam melakukan kegiatan dan dapat membangkitkan gairah seseorang untuk memenuhi kesediaannya yang dapat diukur melalui kesukacitaan, ketertarikan, perhatian dan keterlibatan.

3. Tes Kreativitas

Kreativitas merupakan proses berpikir yang dapat digunakan untuk memecahkan masalah atau menjawab pertanyaan secara benar dan bermanfaat. Disamping itu, kreativitas juga merupakan kemampuan berpikir divergen yang mencerminkan kelancaran, keluwesan dan orisinal dalam proses.

Ciri-ciri kreativitas berkaitan dengan imajinasi, orisinalitas, berpikir divergen, penemuan hal-hal yang bersifat baru, intuisi, hal-hal yang menyangkut perubahan dan eksplorasi. Desain tes kreativitas terdiri dari dua subtes yaitu dalam bentuk gambar dan verbal yang masing-masing bentuk memiliki ciri kelancaran (fluency), keluwesan (flexibility), keaslian (originality), dan elaborasi (elaboration).

Berdasarkan beberapa pendapat para ahli, definisi konseptual kreativitas adalah kemampuan berpikir divergen. Adapun definisi operasionalnya adalah kemampuan berpikir divergen yang memiliki sifat (dapat diukur melalui) kelancaran, keluwesan, keaslian, elaborasi, dan hasilnya dapat berguna untuk keperluan tertentu.

D. Perencanaan Penulisan Tes

1. Langkah – langkah menyusun tes

Dalam menyusun tes perlu memperhatikan tipe hasil belajar atau tingkat kemampuan berpikir mana saja yang akan diukur atau dinilai. Untuk menentukan hal tersebut, penyusun tes dapat berpedoman kepada tujuan intruksional yang akan dinilai atau kepada tujuan evaluasi itu sendiri. Selain itu, dalam mengembangkan atau menyusun sebuah tes hasil belajar, supaya tes tersebut memiliki karakteristik tes yang baik, berikut langkah-langkah yang harus ditempuh.

1. Menetapkan tujuan penilaian atau tujuan tes. Setiap orang yang akan melakukan kegiatan penilaian harus sadar tujuan akan penilaian tersebut.

2. Menguraikan materi tes dan kompetensi.

Dalam menguraikan isi tes harus menjaga agar tes yang ditulis itu tidak keluar lingkup materi yang telah ditentukan oleh batasan kawasan ukur tetapi juga menjaga agar tidak ada bagian isi yang penting yang terlewatkan dan tidak tertuang dalam tes. Materi tes haruslah komprehensif dan berisi butir-butir yang relevan. Dalam hal ini yang perlu dilakukan antara lain:

- a. Penguraian materi berdasarkan bagian-bagiannya, yakni penguraian disandarkan pada topic-topik dalam kurikulum atau bab-bab buku acuan pengajaran atau berdasarkan bahasan selama proses pembelajaran.

- b. Pemberian bobot tes sesuai dengan kepentingannya. Semakin tinggi bobot bagian suatu materi semakin banyak ia harus dituangkan dalam bentuk item dan semakin rendah bobot maka semakin sedikit ia harus dituangkan dalam bentuk item.

3. Mengembangkan kisi-kisi. Kisi-kisi adalah matrik atau format yang memuat informasi yang dapat dijadikan pedoman oleh penulis soal untuk

menulis soal menjadi tes. Dalam kisi-kisi terdapat 2 komponen utama, yaitu:

a. Identitas, yakni mencakup aspek jenis sekolah atau jenjang sekolah, mata pelajaran, kurikulum yang diacu, tingkat kelas, alokasi waktu, dan jumlah soal.

b. Matriks, yakni mencakup komponen yang ingin di ungkap, indikator hasil belajar, tema/konsep/pokok bahasan/sub pokok bahasan, pokok materi soal, bentuk soal, dan nomor soal.

Adapun langkah-langkah penyusunan kisi-kisi untuk menentukan proporsi materi dan kompetensi adalah sebagai berikut:

a. Mengidentifikasi pokok-pokok materi yang akan diujikan dengan memberikan imbangan bobot untuk masing-masing bahasan.

b. Mengidentifikasi tingkatan ranah kognitif yang termuat dalam rumusan indikator dan memberikan imbangan bobot untuk masing-masing tingkatan ranah.

c. Memasukkan ranah dan pokok-pokok materi yang telah teridentifikasi ke dalam table spesifikasi.

d. Memperinci banyaknya butir soal dalam setiap pokok materi dan ranah yang akan di capai.

4. Pemilihan bentuk tes. Pemilihan ini didasarkan pada beberapa faktor seperti: tujuan tes, jumlah peserta tes, waktu yang tersedia untuk memeriksa lembar jawaban tes, cakupan materi tes, dan karakteristik mata pelajaran yang diujikan.

5. Panjang tes.

Yang dimaksudkan dalam hal ini adalah jumlah soal yang akan diujikan dalam suatu ujian. Ada 3 hal utama yang perlu diperhatikan dalam menentukan jumlah soal yang diujikan, yaitu: bobot masing-masing bagian yang telah ditentukan dalam kisi-kisi, keandalan yang diinginkan dan waktu yang tersedia. Analisis rasional adalah menganalisis kembali soal yang telah dirumuskan, ditimbang, baik oleh sendiri maupun orang lain dengan berpedoman pada kisi-kisi dan aturan penulisan soal.

2. Membuat kisi – kisi soal

Kisi-kisi adalah Suatu format berupa matriks yang memuat pedoman untuk menulis soal atau merakit soal menjadi suatu tes. Kisi-kisi berfungsi sebagai pedoman dalam penulisan soal dan atau dalam melakukan perakitan tes.

Syarat-syarat kisi-kisi yang baik :

- a. Mewakili isi kurikulum/kemampuan yang akan diujikan;
- b. Komponen-komponennya rinci, jelas, dan mudah dipahami;
- c. Soal-soalnya dapat dibuat sesuai dengan indikator dan bentuk soal yang ditetapkan

Komponen kisi-kisi terdiri dari :

A. Kelompok Identitas :

1. Jenis institusi
2. Program/Jurusan
3. Bidang studi/matapelajaran
4. Tahun pelajaran
5. Kurikulum yang diacu/dipergunakan
6. Jumlah soal
7. Bentuk soal

B. Kelompok Matriks

1. Kompetensi Dasar
2. Materi yang akan diberikan/dijadikan soal
3. Indikator
4. Nomor urut soal (jika diperlukan)

KOMPETENSI DASAR

Kompetensi Dasar:Kemampuan minimal yang harus dikuasai siswa setelah mempelajari materi pelajaran tertentu. Kompetensi ini diambil dari kurikulum.

MATERI

Materi merupakan:bahan ajar yang harus dikuasai siswa berdasarkan kompetensi yang akan diukur. Penentuan materi (bahan ajar) yang akan diambil disesuaikan dengan indikator yang akan disusun.

Untuk pembuatan soal, kita harus bisa memilih materi esensial yang akan dikeluarkan dalam tes. Untuk memilih materi esensial kita dapat berpatokan pada kriteria-kriteria berikut ini :

- a. Merupakan materi lanjutan
- b. Pendalaman dari satu materi yang sudah dipelajari sebelumnya
- c. Merupakan materi penting yang harus dikuasai oleh siswa
- d. Merupakan materi yang sering diperlukan
- e. Untuk mempelajari bidang studi lain
- f. Merupakan materi yang berkesinambungan yang terdapat pada semua jenjang kelas
- g. Merupakan materi yang memiliki nilai terapan dalam kehidupan sehari-hari

INDIKATOR

Indikator: berisi ciri-ciri perilaku yang dapat diukur sebagai petunjuk untuk membuat soal. Indikator dikembangkan sesuai dg karakteristik siswa, mata pelajaran, satuan pendidikan, potensi daerah dan dirumuskan dg kata kerja operasional yang terukur dan/atau dapat diobservasi.

Syarat-syarat indikator yang baik adalah :

Memuat ciri-ciri kompetensi dasar yang akan diukur.

Memuat kata kerja operasional yang dapat diukur.

Berkaitan dengan materi (bahan ajar) yang dipilih.

Dapat dibuatkan soalnya.

3. Tingkat kesukaran (TK) dan Daya Pembeda (DP)

a. Taraf kesukaran tes

Asumsi yang digunakan untuk memperoleh kualitas yang baik, disamping memenuhi validitas dan reliabilitas adalah daya keseimbangan dari tingkat kesulitan soal tersebut. Keseimbangan yang dimaksudkan adalah adanya soal-soal yang termasuk mudah sedang dan sukar secara proporsional. Tingkat kesukaran soal dipandang dari kesanggupan atau kemampuan siswa dalam menjawabnya, bukan dilihat dari segi guru dalam melakukan analisis pembuat soal.

Ada beberapa dasar pertimbangan dalam menentukan proporsi jumlah soal kategori mudah sedang dan sukar. Pertimbangan pertama adalah adanya keseimbangan, yakni jumlah soal sama untuk ke tiga kategori tersebut. dan ke dua proporsi jumlah soal untuk ke tiga kategori tersebut artinya sebagian besar soal berada dalam kategori sedang sebagian lagi termasuk kategori mudah dan sukar dengan proporsi yang seimbang.

Perbandingan antara soal mudah sedang sukar bisa di buat 3-4-3. Artinya, 30% soal kategori mudah 40% soal kategori sedang dan 30% lagi soal kategori sukar.

Di samping itu oleh karena suatu tes dimaksudkan untuk memisahkan antara murid-murid yang betul-betul mempelajari suatu pelajaran dengan murid-murid yang tidak mempelajari pelajaran itu, maka tes atau item yang baik adalah tes atau item yang betul-betul dapat memisahkan ke dua golongan murid tadi. Jadi setiap item disamping harus mempunyai derajat kesukaran tertentu, juga harus mampu membedakan antara murid yang pandai dengan murid yang kurang pandai.

Setelah judgment dilakukan oleh guru kemudian soal tersebut di uji cobakan dan dianalisis apakah judgment tersebut sesuai atau tidak. Cara melakukan analisis untuk menentukan tingkat kesukaran soal adalah dengan menggunakan rumus sebagai berikut:

$$I = \frac{B}{N}$$

Keterangan:

I = Indeks kesulitan untuk setiap butir soal

B = Banyaknya siswa yang menjawab benar setiap butir soal

N = Banyaknya yang memberikan jawaban pada soal yang di maksudkan.

Kriteria yang digunakan makin kecil indeks yang di peroleh makin sulit soal tersebut. Sebaliknya makin besar indeks yang diperoleh makin mudah soal tersebut.

Menurut keiteria yang sering di ikuti indeks kesukaran sering di klasifikasikan sebagai berikut :

- Soal dengan $P \leq 0,30$ adalah soal kategori sukar.
- Soal dengan $P \geq 0,31 - 0,70$ adalah soal kategori sedang.
- Soal dengan $P \geq 0,71 - 1,00$ adakah soal kategori mudah.

b. Analisis Daya Pembeda

Salah satu tujuan analisis kuantitatif soal adalah untuk menentukan dapat tidaknya suatu soal membedakan kelompok dalam aspek yang di ukur sesuai dengan perbedaan yang ada dalam kelompok itu.

Indeks yang di gunakan dalam membedakan peserta tes yang berkemampuan tinggi dengan peserta tes yang berkemampuan rendah adalah indeks daya pembeda. Indeks ini menunjukkan kesesuaian antara fungsi soal dengan fungsi tes secara keseluruhan. Dengan demikian validitas soal ini sama dengan daya pembeda soal yaitu daya yang membedakan antara peserta tes yang berkemampuan tinggi dengan peserta tes yang berkemampuan rendah.

c. Hubungan antara tingkat kesukaran dan daya pembeda.

Tingkat kesukaran berpengaruh langsung pada daya pembeda soal. Jika setiap orang memilih benar jawaban ($P = 1$), atau jika setiap orang memiliki benar jawaban ($P = 0$) maka soal tidak dapat digunakan untuk membedakan kemampuan peserta tes. Oleh karena itu soal yang baik adalah soal yang memiliki daya pembeda antara peserta tes kelompok atas dan kelompok rendah. Kelompok rendah memiliki tingkat kemampuan 0.50 dan akan diperoleh daya pembeda kelompok atas maksimal 1.00.

E. Pemberian Skor dan Sistem Penilaian

Adapun teknik-teknik pemeriksaan atau koreksi adalah sebagai berikut :

1. Teknik Pemeriksaan Hasil Tes Tertulis

Tes tertulis dapat dibedakan menjadi 2 golongan yaitu tes hasil belajar (tertulis) bentuk uraian (Subjective test=test essay) dan tes hasil belajar

bentuk objektif (Objective test). Kedua kelompok tersebut mempunyai karakter yang berbeda sehingga teknik pemeriksaannya dan hasil-hasilnya pun pasti berbeda pula.

a. Teknik Pemeriksaan Hasil Tes Hasil Belajar Bentuk Uraian

Dalam bentuk teknik pemeriksaan hasil tes belajar dalam bentuk uraian, evaluator (tester/ pemberi test) sebaiknya membuat kunci jawaban setelah soal tes selesai disusun sebagai dasar pedoman dalam mengoreksi hasil jawaban dari peserta tes dengan cara membandingkan antara jawaban peserta tes dengan pedoman jawaban yang betul.

Hal-hal yang perlu diperhatikan dalam pemeriksaan dengan tes uraian ini adalah :

- 1) Apakah nantinya pengolahan dan penentuan hasil tes uraian itu akan didasarkan pada standar mutlak, atau
- 2) Apakah nantinya pengolahan dan penentuan hasil tes subyektif akan didasarkan pada standar relatif.

Apabila pengolahan dan penentuan hasil tes uraian akan didasarkan pada standar mutlak maka prosedur pemeriksaannya adalah sebagai berikut :

- 1) Membaca setiap jawaban yang diberikan oleh peserta tes untuk setiap butir soal tes uraian dan membandingkannya dengan pedoman jawaban betul yang sudah disiapkan.
- 2) Atas dasar perbandingan antara jawaban peserta tes dengan pedoman/ ancar-ancar jawaban betul yang telah disiapkan, tester lalu memberikan skor untuk setiap butir soal dan menuliskannya di bagian kiri dari jawaban testee tersebut.
- 3) Menjumlahkan skor-skor yang telah diberikan kepada testee (digunakan sebagai bahan dalam pengolahan dan penentuan nilai).

Apabila pengolahan dan penentuan nilai akan didasarkan pada standar relatif (dimana penentuan nilai didasarkan pada prestasi kelompok) maka prosedur pemeriksaannya adalah sebagai berikut :

- 1) Memeriksa jawaban atas butir soal nomor 1 yang diberikan oleh seluruh testee, sehingga diperoleh gambaran secara umum mengenai keseluruhan

jawaban yang ada. Setelah pemeriksaan terhadap seluruh jawaban item nomor 1 dapat diselesaikan, maka tester akan menjadi tahu, testee manakah yang memberikan jawabannya termasuk lengkap, kurang lengkap, menyimpang dan tidak memberikan jawaban sama sekali.

2) Memberikan skor terhadap jawaban soal nomor 1 untuk seluruh testee, misalnya untuk jawaban lengkap diberi skor 2, kurang lengkap diberikan skor 1, dan yang menyimpang atau tidak memberikan jawaban sama sekali diberikan skor 0.

3) Setelah pemeriksaan atas jawaban butir soal nomor 1 untuk seluruh testee dapat diselesaikan, lalu dilanjutkan dengan pemeriksaan terhadap jawaban butir soal nomor 2, dengan cara yang sama.

4) Memberikan skor terhadap jawaban butir soal nomor 2 dari seluruh testee, dengan cara yang sama, dan seterusnya sampai dengan selesai dengan langkah yang sama.

5) Setelah jawaban seluruh butir soal yang diberikan oleh seluruh testee dapat diselesaikan, akhirnya dilakukan penjumlahan skor (yang nantinya akan dijadikan bahan dalam pengolahan dan penentuan nilai).

b. Teknik Pemeriksaan Hasil Tes Hasil Belajar Bentuk Obyektif

Memeriksa atau mengoreksi jawaban soal tes obyektif pada umumnya dilakukan dengan jalan menggunakan kunci jawaban.

Ada beberapa macam kunci jawaban yang dapat digunakan untuk mengoreksi jawaban soal tes obyektif, yaitu : Kunci berdamping (Strip keys), Kunci sistem karbon (Carbon system keys), Kunci sistem tusukan (Pinpick system keys), Kunci berjendela (Window keys).

Adapun penjelasannya adalah sebagai berikut :

1) Kunci Berdamping (Strip keys)

Kunci jawaban ini terdiri atas jawaban-jawaban betul yang ditulis dalam satu kolom yang lurus dari atas ke bawah. Kunci jawaban jenis pertama ini

digunakan untuk memeriksa jawaban-jawaban yang ditulis pada kolom 1, yang disusun lurus dari atas ke bawah.

Adapun cara menggunakannya ialah dengan meletakkan kunci jawaban tersebut berjajar dengan lembar jawaban yang akan diperiksa. Cocokkan jawaban-jawaban yang diberikan oleh testee dengan jawaban-jawaban yang tercantum pada kunci jawaban. Jawaban yang cocok dengan kunci jawaban diisi/ ditulis dengan tanda plus (+), sedangkan jawaban yang tidak cocok dengan kunci jawaban diisi dengan tanda minus (-).

2) Kunci sistem karbon (Carbon system keys)

Kunci jawaban ini berupa lembar kunci jawaban yang telah didesain sedemikian rupa sehingga mudah digunakan sebagai untuk memeriksa atau mengoreksi.

Misalnya adalah testee diminta membubuhkan tanda silang pada abjad jawaban soal. Maka kunci jawaban diletakkan di atas lembaran jawaban yang sudah ditumpangi lembaran karbon. Pada kunci jawaban diberikan tanda (misalnya : lingkaran) pada jawaban yang betul. Jawaban yang berada diluar tanda (misalnya : lingkaran) adalah salah dan jawaban yang berada di dalam tanda (misalnya : lingkaran) adalah betul.

3) Kunci sistem tusukan (Pinpick system keys)

Pada dasarnya sama dengan sistem karbon. Perbedaannya ialah, kunci jawaban pada sistem ini adalah untuk jawaban yang betul diberikan tusukan dengan jarum atau paku, atau alat penusuk lainnya sementara lembar jawaban berada di bawahnya. Pilihan jawaban yang berlubang adalah betul dan pilihan jawaban yang tidak berlubang adalah salah.

4) Kunci berjendela (Windows keys).

Prosedur penggunaan sistem kunci berjendela (Windows keys) adalah sebagai berikut :

- a) Ambil blanko lembar jawaban yang masih kosong (belum digunakan)
- b) Pilihan jawaban yang betul diberikan lubang (bulatan) seolah-oleh seperti jendela.
- c) Lembar jawaban kita letakkan di bawah kunci berjendela.
- d) Melalui lubang-lubang (jendela-jendela) dibuat garis vertikal dengan pensil berwarna. Jika garis tersebut tepat mengenai tanda silang yang pada lembar jawaban maka jawabannya adalah betul dan pada bagian lembar jawaban yang pilihannya tidak terkena goresan berarti salah.

2. Teknik Pemeriksaan Hasil Tes Lisan

Pemeriksaan atau koreksi yang dilaksanakan dalam rangka menilai jawaban-jawaban testee pada tes hasil belajar secara lisan pada umumnya cenderung bersifat subyektif. Hal ini lebih mudah dipahami karena berhubungan langsung dengan benda hidup dan tidak berhubungan dengan lembar jawaban yang merupakan benda mati.

Dalam tes lisan, testee yang sedang di tes oleh tester kemungkinan adalah termasuk testee yang “disukai” oleh tester sehingga mendapatkan simpati dari tester, atau sebaliknya testee yang sedang dihadapi adalah termasuk testee yang “kurang disukai”, sehingga terdaat peluang bagi tester untuk bertindak tidak/kurang obyektif. Menghadapi kemungkinan-kemungkinan tersebut maka penguji harus berusaha semaksimal mungkin untuk tetap berlaku obyektif terhadap testee, dan tidak terpengaruh oleh obyek (testee) yang sedang dihadapinya.

Dalam pelaksanaan tes lisan hal-hal yang perlu diperhatikan antara lain adalah sebagai berikut :

- a. Pertahankanlah situasi evaluasi dalam pelaksanaan tes lisan.
- b. Jangan membentak testee dan mengeluarkan kata-kata yang tidak semestinya, misalnya : tolol, bodoh, dan sebagainya.
- b. Jangan berkecenderungan membantu testee yang sedang dites dengan memberikan kode-kode, isyarat atau kunci-kunci lain karena merasa kasihan, simpati dan sebagainya.

c. Siapkan rencana pertanyaan yang akan disampaikan sekaligus jawaban yang diharapkan untuk setiap pertanyaan (soal dan kunci jawaban).

d. Lakukan scoring terhadap jawaban yang diberikan oleh testee dengan teliti. Maka dari itu sebaiknya penilaian dilakukan selama tes berjalan.

Dalam hubungan ini, pemeriksaan terhadap jawaban-jawaban testee hendaknya dikendalikan oleh pedoman yang pasti, misalnya :

a. Kelengkapan jawaban yang diberikan oleh testee. Apakah jawaban testee telah memenuhi atau mencakup semua unsur yang seharusnya ada, sesuai dengan jawaban betul yang telah disusun oleh tester.

b. Kelancaran testee dalam mengemukakan jawaban-jawaban. Apakah testee cukup lancar dalam menjawab pertanyaan-pertanyaan dari tester, sehingga mencerminkan tingkat kedalaman pemahaman testee terhadap materi pertanyaan yang diajukan.

c. Kebenaran jawaban yang dikemukakan. Jawaban yang banyak dari testee belum tentu merupakan jawaban yang benar. Maka dari itu tester harus benar-benar jeli dalam menyaring jawaban dari testee, apakah jawaban tersebut mengandung kadar kebenaran atau tidak.

d. Kemampuan testee dalam mempertahankan pendapatnya. Jawaban yang disampaikan dengan ragu-ragu merupakan salah satu indikator bahwa testee kurang menguasai materi yang ditanyakan.

e. Berapa persen (%) kira-kira, pertanyaan-pertanyaan lisan yang tergolong dalam kategori sukar, sedang dan mudah dapat dijawab dengan betul oleh testee.

Penguji juga dapat menambahkan unsur-unsur lain yang di rasa perlu sebagai bahan penilaian, misalnya : penampilan, kesopanan, kerapian, kedisiplinan dan sebagainya.

3. Teknik Pemeriksaan Menilai Hasil Tes Perbuatan.

Dalam pemeriksaan tes tertulis untuk memperoleh hasilnya dilakukan dengan membaca lembar jawaban dan pada tes lisan pemeriksaan dilakukan

dengan menganalisa jawaban-jawaban lisan testee. Maka dalam tes perbuatan pemeriksaan hasilnya dilakukan dengan menggunakan observasi (pengamatan).

Sasaran yang diamati adalah tingkah laku, perbuatan, sikap dan sebagainya. Untuk dapat menilai hasil tes perbuatan diperlukan instrumen-instrumen tertentu dan setiap gejala yang muncul diberikan skor tertentu.

Teknik-Teknik Pemberian Skor

Pada hakikatnya pemberian skor (scoring) adalah proses perubahan jawaban instrumen menjadi angka-angka yang merupakan nilai kuantitatif dari suatu jawaban terhadap item dalam instrumen. Angka-angka hasil penilaian selanjutnya diproses menjadi nilai-nilai (grade).

Agar tidak terjadi kesalahan dalam memahami skor dan nilai lebih dahulu harus dipahami perbedaan antara skor dan nilai. Hal ini didasarkan dengan masih banyaknya anggapan antara skor dan nilai mengandung satu pengertian atau sama.

Skor adalah hasil pekerjaan menyekor (memberikan angka) yang diperoleh dari angka-angka dari setiap butir soal yang telah di jawab oleh testee dengan benar, dengan mempertimbangkan bobot jawaban betulnya.

Adapun yang dimaksud dengan nilai adalah angka atau huruf yang merupakan hasil ubahan dari skor yang sudah dijadikan satu dengan skor-skor lainnya, serta disesuaikan pengaturannya dengan standar tertentu. Sehingga nilai sering disebut juga dengan skor standar (Standarg Score).

Cara pemberian skor terhadap hasil tes hasil belajar pada umumnya disesuaikan dengan bentuk soal yang dikeluarkan dalam tes tersebut, tes uraian (essay) atau tes obyektif (objektive test). Penjelasannya sebagai berikut :

A. Pemberian Skor pada Tes Uraian

Pada tes uraian, pemberian skor didasarkan pada bobot (weight) yang diberikan pada setiap butir soal, didasarkan dan disesuaikan dengan tingkat

kesulitan dari soal tersebut dan atau banyak sedikitnya unsur yang terdapat dalam jawaban yang dianggap paling benar.

B. Pemberian Skor pada Tes Obyektif

Pemberian skor pada tes obyektif pada umumnya digunakan rumus correction for geussing atau di kenal dengan istilah sistem denda.

Untuk soal obyektif bentuk true-false misalnya, setiap item diberi skor maksimal 1 (satu). Apabila testee menjawab benar maka diberikan skor 1 dan apabila salah maka diberikan skor 0.

Cara menghitung skor terakhir dari seluruh item bentuk true-false, dapat digunakan dua macam rumus yaitu : Rumus yang memperhitungkan denda dan rumus yang mengabaikan atau meniadakan denda. Penggunaan rumus-rumus tersebut tergantung dari kebijakan tester.

Rumus skor akhir dengan memperhitungkan denda adalah sebagai berikut :

Keterangan :

S = Skor yang sedang di cari

R = (Right=Betul) Jumlah jawaban betul

W = (Wrong=Salah) Jumlah jawaban salah

O = Option alternative (kemungkinan jawaban), dalam tes obyektif bentuk true-false hanya ada dua kemungkinan "benar" dan "salah"

1 = Bilangan konstan (bilangan tetap)

Adapun rumus skor akhir yang tidak memperhitungkan denda adalah sebagai berikut : $S = R$

Keterangan :

S = Skor yang dicari

R = Jumlah jawaban betul

Yang perlu diperhatikan pada tes obyektif adalah karena berbentuk multiple choice maka masing-masing item soal memiliki derajat atau tingkat kesulitan masing-masing yang berbeda, jadi bobot jawaban yang benar belum tentu memiliki skor 1, melainkan bisa juga berbobot $1\frac{1}{2}$, $2\frac{1}{2}$, 5 dan sebagainya. Dalam hal ini yang dapat menentukan bobot soal adalah orang yang paling tahu dengan mengenai derajat kesulitan soal tersebut yaitu sebaiknya adalah pembuat soal itu sendiri atau tester.

Teknik Pengolahan Dan Pengkonversian Skor Menjadi Nilai

Setelah proses pemeriksaan dan pemberian skor langkah selanjutnya adalah mengolah skor tersebut menjadi nilai-nilai yang merupakan hasil akhir. Sebagaimana telah diketahui sebelumnya antara skor dan nilai adalah tidak sama.

Skor adalah hasil pekerjaan menyekor (memberikan angka) yang diperoleh dari penjumlahan angka-angka dalam setiap butir soal yang di jawab dengan benar oleh testee, dan memperhitungkan bobot jawaban, sedangkan nilai adalah angka atau huruf yang merupakan hasil konversi (rubahan) dari penjumlahan skor yang disesuaikan pengaturannya dengan standar tertentu yang pada dasarnya merupakan lambang kemampuan testee terhadap materi atau bahan yang diteskan.

Dari penjelasan diatas dapat dipahami bahwa untuk mendapatkan nilai, maka skor-skor yang telah didapat masih merupakan skor mentah dan perlu diolah dan dikonversikan sehingga skor dapat berubah menjadi nilai (menjadi skor yang sifatnya baku atau standar (Standard Score)) :

1. Pengolahan dan Pengubahan Skor Mentah Menjadi Nilai Standard (Standard Score)

Ada dua hal yang perlu diperhatikan dalam pengolahan dan pengubahan skor menjadi skor standard atau nilai yaitu :

a. Dalam pengolahan dan pengubahan skor menjadi skor standard atau nilai terdapat dua cara yang dapat ditempuh yaitu :

1) Pengolahan dan pengubahan skor mentah menjadi nilai dilakukan dengan mengacu pada kriterium (Criterion) atau sering juga disebut dengan

patokan. Cara pertama ini sering dikenal dengan istilah *criterion referenced evaluation*. Di dunia pendidikan Indonesia dikenal dengan istilah Penilaian Acuan Patokan (PAP) ada juga yang mengatakan dengan istilah Standar Mutlak.

2) Pengolahan dan pengubahan skor mentah menjadi nilai dengan mengacu pada norma atau kelompok. Cara kedua ini dikenal dengan istilah *norm referenced evaluation*. Di dalam dunia pendidikan Indonesia dikenal dengan istilah Penilaian Acuan Norma (PAN).

b. Pengolahan dan pengubahan skor mentah menjadi nilai dengan berbagai macam skala, misalnya : skala 5 (Stanfive), yaitu nilai standar berskala lima yang dikenal dengan istilah nilai huruf A, B, C, D dan F. Skala sembilan (Stanine) yaitu nilai standar berskala sembilan dimana rentang nilainya mulai dari 1 sampai dengan 9 (tidak ada nilai =0 dan >10), skala sebelas (standard eleven/ eleven points scale) rentang nilai mulai dari 0 sampai dengan 10, z score (nilai standar z), dan T score (nilai standar T).

Dalam pembahasan kali ini hanya akan dibahas mengenai pengolahan hasil belajar dengan acuan patokan dan acuan norma.

1. Penilaian Acuan Patokan (PAP)

Penilaian Acuan Patokan (*criterion referenced evaluation*) yang dikenal juga dengan standar mutlak berusaha menafsirkan hasil tes yang diperoleh siswa dengan membandingkannya dengan patokan yang telah ditetapkan. Sebelum hasil tes diperoleh atau bahkan sebelum kegiatan pengajaran dilakukan, patokan yang akan dipergunakan untuk menentukan kelulusan harus sudah ditetapkan.

Standar atau patokan tersebut memuat ketentuan-ketentuan yang dipergunakan sebagai batas-batas penentuan kelulusan testee atau batas pemberian nilai pada testee. Jika skor yang diperoleh oleh testee memenuhi batas minimal maka testee dinyatakan telah memenuhi tingkat penguasaan minimal terhadap materi yang disampaikan dan sebaliknya jika testee belum bisa memenuhi batas minimal yang ditentukan maka testee dianggap belum “lulus” atau belum

menguasai materi. Karena batasan-batasan tersebut bersifat mutlak/ pasti maka hasil yang diperoleh tidak dapat di tawar lagi.

Standar penilaian ditentukan secara mutlak, banyaknya testee yang memperoleh nilai tinggi atau jumlah kelulusan testee banyak akan mencerminkan penguasaannya terhadap materi yang disampaikan.

Pengolahan skor mentah menjadi nilai dilakukan dengan menempuh langkah-langkah sebagai berikut :

- a. Menggabungkan skor dari berbagai sumber penilaian untuk memperoleh skor akhir.
 - b. Menghitung skor minimum penguasaan tuntas dengan menerapkan prosentase Batas Minimal Penguasaan (BMP).
 - b. Menentukan tabel konversi
2. Penilaian Acuan Norma (PAN)

Penggunaan penilaian dengan norma kelompok atau norma relatif/ standar relatif ini pertama kali dikemukakan pada tahun 1908 (Cureton, 1971) dengan landasan dasar bahwa tingkat pencapaian belajar siswa akan tersebar menurut kurva normal.

Penilaian Acuan Norma (Norm Referenced Evaluation) dikenal pula dengan Standar Relatif atau Norma Kelompok. Pendekatan penilaian ini menafsirkan hasil tes yang diperoleh testee dengan membandingkan dengan hasil tes dari testee lain dalam kelompoknya. Alat pembanding tersebut yang menjadi dasar standar kelulusan dan pemberian nilai ditentukan berdasarkan skor yang diperoleh testee dalam satu kelompok. Dengan demikian, standar kelulusan baru dapat ditentukan setelah diperoleh skor dari para peserta testee.

Hal ini berarti setiap kelompok mempunyai standar masing-masing dan standar satu kelompok tidak dapat dipergunakan sebagai standar kelompok yang lain. Standar dari hasil tes sebelumnya pun tidak dapat dipergunakan sebagai standar sehingga setiap memperoleh hasil tes harus dibuat norma yang baru.

Dasar pemikiran dari penggunaan standar PAN adalah adanya asumsi bahwa di setiap populasi yang heterogen terdapat siswa dengan kelompok baik, kelompok sedang dan kelompok kurang.

Pengolahan skor dengan Penilaian Acuan Norma (PAN) mengharuskan kita menghitung dengan statistik. Perhitungan dilakukan atas skor akhir (penggabungan berbagai sumber skor), kemudian dilakukan langkah-langkah sebagai berikut :

- a. Menghitung Distribusi Angka (Da), Rata-rata Hitung (RH) dan Range of Score/ Rentang Skor (R), dengan rumus :

$$R = \text{Skor Tertinggi} - \text{Skor Terendah}$$

- b. Menentukan kelompok nilai, Kelas Interval (Ki), dan Interval Duga (RD). Dengan rumus :

$$R = \text{Rentang}$$

$$i = \text{Interval}$$

Catatan :

Jumlah kelas interval (KI) sebaiknya menggunakan antara 10 – 20.

Kelas interval (KI) menggunakan angka ganjil.

- c. Membuat Tabel Distribusi Frekuensi (TDF)
- d. Mencari rata-rata Duga (RD, Mean Duga (MD), dengan rumus :
- e. Menghitung rata-rata hitung (RH) dan Mean (M) dengan rumus :
- f. Menghitung Standar Deviasi (SD), dengan rumus :
- g. Menyusun Tabel Konversi PAN

Dalam membuat tabel harus diperhatikan adalah skala nilai yang digunakan (skala 5, skala 10, atau skala 100)

Rumus Skala 5

$$M + 1,8 SD \geq 4 \text{ atau Ahmad, hadis no.}$$

$$M + 0,6 SD = 3 \text{ atau Ibn Majah, hadis no.}$$

$$M - 0,6 SD = 2 \text{ atau C}$$

$$- 1,8 SD = 1 \text{ atau Abu Daud, hadis no.}$$

$$M - 1,8 SD \geq 0 \text{ atau E}$$

Rumus Skala 10

Skala Sigma, Skala 1-10 Nilai Kualitatif :

$M + 2,25 SD > 10$ Istimewa

$M + 1,75 SD$ 9 Baik sekali

$M + 1,25 SD$ 8 Baik

$M + 0,75 SD$ 7 Lebih dari cukup

$M + 0,25 SD$ 6 Cukup

$M - 0,25 SD$ 5 Hampir cukup

$M - 0,75 SD$ 4 Kurang

$M - 1,25 SD$ 3 Kurang sekali

$M - 1,75 SD$ 2 Buruk

$M + 2,25 SD$ 1 Buruk sekali

Kelemahan sistem PAN adalah dengan tes apapun dalam kelompok apapun dan dengan dasar prestasi yang bagaimanapun, pemberian nilai dengan sistem ini selalu dapat dilakukan. Karena itu penggunaan sistem PAN dapat dilakukan dengan baik apabila memenuhi syarat yang mendasari kurva normal, yaitu :

- a. Skor nilai terpenjar atau dapat dianggap terpenjar sesuai dengan penjaran kurva normal.
- b. Jumlah yang dinilai minimal 50 orang atau sebaiknya 100 orang ke atas.

F. Validitas dan Reliabilitas Tes

a. Validitas

1. Pengertian Validitas

Menurut Azwar (1986) Validitas berasal dari kata validity yang mempunyai arti sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya.

Suatu skala atau instrumen pengukur dapat dikatakan mempunyai validitas yang tinggi apabila instrumen tersebut menjalankan fungsi ukurnya, atau memberikan hasil ukur yang sesuai dengan maksud dilakukannya pengukuran tersebut. Sedangkan tes yang memiliki validitas rendah akan menghasilkan data yang tidak relevan dengan tujuan pengukuran.

Terkandung di sini pengertian bahwa ketepatan validitas pada suatu alat ukur tergantung pada kemampuan alat ukur tersebut mencapai tujuan pengukuran yang dikehendaki dengan tepat. Suatu tes yang dimaksudkan untuk mengukur variabel A dan kemudian memberikan hasil pengukuran mengenai variabel A, dikatakan sebagai alat ukur yang memiliki validitas tinggi.

Suatu tes yang dimaksudkan mengukur variabel A akan tetapi menghasilkan data mengenai variabel A' atau bahkan B, dikatakan sebagai alat ukur yang memiliki validitas rendah untuk mengukur variabel A dan tinggi validitasnya untuk mengukur variabel A' atau B.

Sisi lain dari pengertian validitas adalah aspek kecermatan pengukuran. Suatu alat ukur yang valid tidak hanya mampu menghasilkan data yang tepat akan tetapi juga harus memberikan gambaran yang cermat mengenai data tersebut.

Cermat berarti bahwa pengukuran itu dapat memberikan gambaran mengenai perbedaan yang sekecil-kecilnya mengenai perbedaan yang satu dengan yang lain. Sebagai contoh, dalam bidang pengukuran aspek fisik, bila kita hendak mengetahui berat sebuah cincin emas maka kita harus menggunakan alat penimbang berat emas agar hasil penimbangannya valid, yaitu tepat dan cermat. Sebuah alat penimbang badan memang mengukur berat, akan tetapi tidaklah cukup cermat guna menimbang berat cincin emas karena perbedaan berat yang sangat kecil pada berat emas itu tidak akan terlihat pada alat ukur berat badan.

Menggunakan alat ukur yang dimaksudkan untuk mengukur suatu aspek tertentu akan tetapi tidak dapat memberikan hasil ukur yang cermat dan teliti akan menimbulkan kesalahan atau eror. Alat ukur yang valid akan memiliki tingkat

kesalahan yang kecil sehingga angka yang dihasilkannya dapat dipercaya sebagai angka yang sebenarnya atau angka yang mendekati keadaan yang sebenarnya.

Validitas suatu instrumen banyak dijelaskan dalam konteks penelitian sosial yang variabelnya tidak dapat diamati secara langsung, seperti sikap, minat, persepsi, motivasi, dan lain sebagainya. Untuk mengukur variabel yang demikian sulit, untuk mengembangkan instrumen yang memiliki validitas yang tinggi karena karakteristik yang akan diukur dari variabel yang demikian tidak dapat diobservasi secara langsung, tetapi hanya melalui indikator (petunjuk tak langsung) tertentu.

Sekiranya peneliti menggunakan kuesioner di dalam pengumpulan data penelitian, maka kuesioner yang disusunnya harus mengukur apa yang ingin diukurnya. Setelah kuesioner tersebut tersusun dan teruji validitasnya, dalam praktek belum tentu data yang dikumpulkan adalah data yang valid. Banyak hal-hal lain yang akan mengurangi validitas data; misalnya apakah si pewawancara yang mengumpulkan data betul-betul mengikuti petunjuk yang telah ditetapkan dalam kuesioner.

2. Jenis-jenis Validitas

Ebel (dalam Nazirz 1988) membagi validitas menjadi :

- Concurrent Validity adalah validitas yang berkenaan dengan hubungan antara skor dengan kinerja.
- Construct Validity adalah validitas yang berkenaan dengan kualitas aspek psikologis apa yang diukur oleh suatu pengukuran serta terdapat evaluasi bahwa suatu konstruk tertentu dapat menyebabkan kinerja yang baik dalam pengukuran.
- Face Validity adalah validitas yang berhubungan apa yang nampak dalam mengukur sesuatu dan bukan terhadap apa yang seharusnya hendak diukur.

- Factorial Validity dari sebuah alat ukur adalah korelasi antara alat ukur dengan faktor-faktor yang bersamaan dalam suatu kelompok atau ukuran-ukuran perilaku lainnya, di mana validitas ini diperoleh dengan menggunakan teknik analisis faktor.

- Empirical Validity adalah validitas yang berkenaan dengan hubungan antara skor dengan suatu kriteria. Kriteria tersebut adalah ukuran yang bebas dan langsung dengan apa yang ingin diramalkan oleh pengukuran.

- Intrinsic Validity adalah validitas yang berkenaan dengan penggunaan teknik uji coba untuk memperoleh bukti kuantitatif dan objektif untuk mendukung bahwa suatu alat ukur benar-benar mengukur apa yang seharusnya diukur.

- Predictive Validity adalah validitas yang berkenaan dengan hubungan antara skor suatu alat ukur dengan kinerja seorang di masa mendatang.

- Content Validity adalah validitas yang berkenaan dengan baik buruknya sampling dari suatu populasi.

- Curricular Validity adalah validitas yang ditentukan dengan cara menilik isi dari pengukuran dan menilai seberapa jauh pengukuran tersebut merupakan alat ukur yang benar-benar mengukur aspek-aspek sesuai dengan tujuan instruksional.

Validitas isi suatu instrumen berkaitan dengan kesesuaian antara karakteristik dari variabel yang dirumuskan pada definisi konseptual dan operasionalnya. Apabila semua karakteristik variabel yang dirumuskan pada definisi konseptualnya dapat diungkap melalui butir-butir suatu instrument, maka instrument itu dinyatakan memiliki validitas isi yang baik.

Akan tetapi, hal itu mungkin tidak akan pernah tercapai karena sulitnya untuk mendefinisikan keseluruhan karakteristik itu. Selain itu, dari seluruh karakteristik yang dirumuskan pada definisi konseptual suatu variabel seringkali sulit untuk mengembangkan butir-butir yang valid untuk mengungkap atau mengukurnya.

Validitas isi dapat dianalisis dengan cara memperhatikan penampakan luar dari instrument dan dengan menganalisis kesesuaian butir-butirnya dengan karakteristik yang dirumuskan pada definisi konseptual variabel yang diukur. Validitas yang dianalisis dengan memperhatikan penampilan luar instrument itu disebut validitas tampak (face validity). Validitas tampak dievaluasi dengan membaca dan menyelidiki butir-butir instrument serta sekaligus membandingkannya dengan definisi konseptual mengenai variabel yang akan diukur.

Validitas yang dianalisis dengan memperhatikan kerepresentativan butir-butir instrument disebut validitas penyampelan (sampling validity) atau kurikulum (curriculum validity). Validitas tampak maupun penyampelan disebut juga sebagai validitas teoritis karena penganalisisannya lazim dilakukan tanpa didasarkan pada data empiris. Alat yang digunakan untuk menganalisis validitas itu adalah logika dari orang yang menganalisisnya.

Selanjutnya, validitas isi terbagi lagi menjadi dua tipe (Saifuddin Azwar), yaitu:

1. Face Validity (Validitas Muka) adalah tipe validitas yang paling rendah signifikansinya karena hanya didasarkan pada penilaian selintas mengenai isi alat ukur. Apabila isi alat ukur telah tampak sesuai dengan apa yang ingin diukur maka dapat dikatakan maka validitas muka telah terpenuhi.

2. Logical Validity (Validitas Logis) disebut juga sebagai Validitas Sampling (Sampling Validity) adalah validitas yang menunjuk pada sejauh mana isi alat ukur merupakan representasi dari aspek yang hendak diukur.

Validitas logis sangat penting peranannya dalam penyusunan prestasi dan penyusunan skala, yaitu dengan memanfaatkan blue-print atau table spesifikasi.

- Construct validity (Validitas konstruk) adalah tipe validitas yang menunjukkan sejauh mana alat ukur mengungkap suatu trait atau konstruk teoritis yang hendak diukurnya. (Allen & Yen, dalam Azwar 1986).

Pengujian validitas konstruk merupakan proses yang terus berlanjut sejalan dengan perkembangan konsep mengenai trait yang diukur.

Menurut Saifuddin Azwar, validitas konstruk adalah seberapa besar derajat tes mengukur hipotesis yang dikehendaki untuk diukur. Konstruk adalah perangai yang tidak dapat diamati, yang menjelaskan perilaku. Menguji validitas konstruk mencakup uji hipotesis yang dideduksi dari suatu teori yang mengajukan konstruk tersebut.

- Criterion-related validity (Validitas berdasar kriteria). Validitas ini menghendaki tersedianya kriteria eksternal yang dapat dijadikan dasar pengujian skor alat ukur. Suatu kriteria adalah variabel perilaku yang akan diprediksi oleh skor alat ukur.

Dilihat dari segi waktu untuk memperoleh skor kriterianya, prosedur validasi berdasar kriteria menghasilkan dua macam validitas (Saifuddin Azwar), yaitu:

1. Validitas Prediktif. Validitas Prediktif sangat penting artinya bila alat ukur dimaksudkan untuk berfungsi sebagai predictor bagi kinerja di masa yang akan datang. Contoh situasi yang menghendaki adanya prediksi kinerja ini antara lain adalah dalam bimbingan karir; seleksi mahasiswa baru, penempatan karyawan, dan sebagainya.

Menurut Saifuddin Azwar, validitas prediktif adalah seberapa besar derajat tes berhasil memprediksi kesuksesan seseorang pada situasi yang akan datang. Validitas prediktif ditentukan dengan mengungkapkan hubungan antara skor tes dengan hasil tes atau ukuran lain kesuksesan dalam satu situasi sasaran.

2. Validitas Konkuren. Apabila skor alat ukur dan skor kriterianya dapat diperoleh dalam waktu yang sama, maka korelasi antara kedua skor tersebut merupakan koefisien validitas konkuren.

b. Reliabilitas

1. Pengertian Reliabilitas

Walizer (1987) menyebutkan pengertian Reliability (Reliabilitas) adalah keajegan pengukuran.

Menurut John M. Echols dan Hasan Shadily (2003: 475) reliabilitas adalah hal yang dapat dipercaya. Popham (1995: 21) menyatakan bahwa reliabilitas adalah "...the degree of which test score are free from error measurement"

Menurut Masri Singarimbun, realibilitas adalah indeks yang menunjukkan sejauh mana suatu alat ukur dapat dipercaya atau dapat diandalkan. Bila suatu alat pengukur dipakai dua kali – untuk mengukur gejala yang sama dan hasil pengukuran yang diperoleh relative konsisten, maka alat pengukur tersebut reliable. Dengan kata lain, realibitas menunjukkan konsistensi suatu alat pengukur di dalam pengukur gejala yang sama.

Reliabilitas, atau keandalan, adalah konsistensi dari serangkaian pengukuran atau serangkaian alat ukur. Hal tersebut bisa berupa pengukuran dari alat ukur yang sama (tes dengan tes ulang) akan memberikan hasil yang sama, atau untuk pengukuran yang lebih subjektif, apakah dua orang penilai memberikan skor yang mirip (reliabilitas antar penilai). Reliabilitas tidak sama dengan validitas. Artinya pengukuran yang dapat diandalkan akan mengukur secara konsisten, tapi belum tentu mengukur apa yang seharusnya diukur.

Dalam penelitian, reliabilitas adalah sejauh mana pengukuran dari suatu tes tetap konsisten setelah dilakukan berulang-ulang terhadap subjek dan dalam kondisi yang sama. Penelitian dianggap dapat diandalkan bila memberikan hasil yang konsisten untuk pengukuran yang sama. Tidak bisa diandalkan bila pengukuran yang berulang itu memberikan hasil yang berbeda-beda.

Pengukuran reliabilitas dapat dilakukan dengan menggunakan berbagai alat statistik (Feldt & Brennan, 1989: 105)

Berdasarkan sejarah, reliabilitas sebuah instrumen dapat dihitung melalui dua cara yaitu kesalahan baku pengukuran dan koefisien reliabilitas (Feldt &

Brennan: 105). Kedua statistik di atas memiliki keterbatasannya masing-masing. Kesalahan pengukuran merupakan rangkuman inkonsistensi peserta tes dalam unit-unit skala skor sedangkan koefisien reliabilitas merupakan kuantifikasi reliabilitas dengan merangkum konsistensi (atau inkonsistensi) diantara beberapa kesalahan pengukuran.

Dalam kerangka teori tes klasik, suatu tes dapat dikatakan memiliki reliabilitas yang tinggi apabila skor tampak tes tersebut berkorelasi tinggi dengan skor murninya sendiri. Interpretasi lainnya adalah seberapa tinggi korelasi antara skor tampak pada dua tes yang paralel.

Reliabilitas alat ukur tidak dapat diketahui dengan pasti tetapi dapat diperkirakan. Dalam mengestimasi reliabilitas alat ukur, ada tiga cara yang sering digunakan yaitu (1) pendekatan tes ulang, (2) pendekatan dengan tes paralel dan (3) pendekatan satu kali pengukuran.

Pendekatan tes ulang merupakan pemberian perangkat tes yang sama terhadap sekelompok subjek sebanyak dua kali dengan selang waktu yang berbeda. Asumsinya adalah bahwa skor yang dihasilkan oleh tes yang sama akan menghasilkan skor tampak yang relatif sama.

Estimasi dengan pendekatan tes ulang akan menghasilkan koefisien stabilitas. Untuk memperoleh koefisien reliabilitas melalui pendekatan tes ulang dapat dilakukan dengan menghitung koefisien korelasi linear antara distribusi skor subyek pada pemberian tes pertama dengan skor subyek pada pemberian tes kedua. Pendekatan tes ulang sangat sesuai untuk mengukur ketrampilan terutama ketrampilan fisik.

2. Jenis-jenis Reliabilitas

Walizer (1987) menyebutkan bahwa ada dua cara umum untuk mengukur reliabilitas, yaitu:

1. Reliabilitas stabilitas. Menyangkut usaha memperoleh nilai yang sama atau serupa untuk setiap orang atau setiap unit yang diukur setiap saat anda

mengukurnya. Reliabilitas ini menyangkut penggunaan indikator yang sama, definisi operasional, dan prosedur pengumpulan data setiap saat, dan mengukurnya pada waktu yang berbeda. Untuk dapat memperoleh reliabilitas stabilitas setiap kali unit diukur skornya haruslah sama atau hampir sama.

2. Reliabilitas ekuivalen. Menyangkut usaha memperoleh nilai relatif yang sama dengan jenis ukuran yang berbeda pada waktu yang sama. Definisi konseptual yang dipakai sama tetapi dengan satu atau lebih indikator yang berbeda, batasan-batasan operasional, peralatan pengumpulan data, dan / atau pengamat-pengamat.

Menguji reliabilitas dengan menggunakan ukuran ekuivalen pada waktu yang sama bias menempuh beberapa bentuk. Bentuk yang paling umum disebut teknik belah-tengah. Cara ini seringkali dipakai dalam survai.

Apabila satu rangkaian pertanyaan yang mengukur satu variable dimasukkan dalam kuesioner, maka pertanyaan-pertanyaan tersebut dibagi dua bagian persis lewat cara tertentu. (Pengacakan atau pengubahan sering digunakan untuk teknik belah tengah ini.) Hasil masing-masing bagian pertanyaan diringkas ke dalam skor, lalu skor masing-masing bagian tersebut dibandingkan. Apabila dalam skor kemudian skor masing-masing bagian tersebut dibandingkan. Apabila kedua skor itu relatif sama, dicapailah reliabilitas belah tengah.

Reliabilitas ekuivalen dapat juga diukur dengan menggunakan teknik pengakuan yang berbeda. Kecemasan misalnya, telah diukur dengan laporan pulsa. Skor-skor relatif dari satu indikator macam ini haruslah sesuai dengan skor yang lain. Jadi bila seorang subyek nampak cemas pada "ukuran gelisah" orang tersebut haruslah menunjukkan tingkatan kecermatan relatif yang sama bila tekanan darahnya yang diukur.

3. Metode pengujian reliabilitas

Tiga tehnik pengujian realibilitas instrument antara lain :

a. Teknik Paralel (Paralel Form atau Alternate Form)

Teknik paralel disebut juga teknik "double test double trial". Sejak awal peneliti harus sudah menyusun dua perangkat instrument yang parallel (ekuivalen), yaitu dua buah instrument yang disusun berdasarkan satu buah kisi-kisi. Setiap butir soal dari instrument yang satu selalu harus dapat dicarikan pasangannya dari instrumen kedua. Kedua instrumen tersebut diujicobakan semua. Sesudah kedua uji coba terlaksana, maka hasil instrumen tersebut dihitung korelasinya dengan menggunakan rumus product moment (korelasi Pearson).

b. Teknik Ulang (Test Re-test)

Disebut juga teknik "single test double trial". Menggunakan sebuah instrument, namun dites dua kali. Hasil atau skor pertama dan kedua kemudian dikorelasikan untuk mengetahui besarnya indeks reliabilitas. Teknik perhitungan yang digunakan sama dengan yang digunakan pada teknik pertama yaitu rumus korelasi Pearson.

Menurut Saifuddin Azwar, realibilitas tes-retest adalah seberapa besar derajat skor tes konsisten dari waktu ke waktu. Realibilitas diukur dengan menentukan hubungan antara skor hasil penyajian tes yang sama kepada kelompok yang sama, pada waktu yang berbeda.

Metode pengujian reliabilitas stabilitas yang paling umum dipakai adalah metode pengujian tes-kembali (test-retest). Metode test-retest menggunakan ukuran atau "test" yang sama untuk variable tertentu pada satu saat pengukuran yang diulang lagi pada saat yang lain.

Cara lain untuk menunjukkan reliabilitas stabilitas, bila kita menggunakan survai, adalah memasukkan pertanyaan yang sama di dua bagian yang berbeda dari kuesioner atau wawancara. Misalnya the Minnesota Multiphasic Personality Inventory (MPPI) mengecek reliabilitas test-retest dalam satu kuesionernya dengan mengulang pertanyaan tertentu di bagian-bagian yang berbeda dari kuesioner yang panjang.

Kesulitan terbesar untuk menunjukkan reliabilitas stabilitas adalah membuat asumsi bahwa sifat/ variable yang akan diukur memang benar-benar bersifat stabil sepanjang waktu. Karena kemungkinan besar tidak ada ukuran yang andal dan sah yang tersedia. Satu-satunya faktor yang dapat membuat asumsi-asumsi ini adalah pengalaman, teori dan/atau putusan terbaik. Dalam setiap kejadian, asumsi ini selalu ditantang dan sulit rasanya mempertahankan asumsi tersebut atas dasar pijakan yang obyektif.

c. Teknik Belah Dua (Split Halve Method)

Disebut juga teknik “single test single trial”. Peneliti boleh hanya memiliki seperangkat instrument saja dan hanya diujicobakan satu kali, kemudian hasilnya dianalisis, yaitu dengan cara membelah seluruh instrument menjadi dua sama besar. Cara yang diambil untuk membelah soal bisa dengan membelah atas dasar nomor ganjil-genap, atas dasar nomor awal-akhir, dan dengan cara undian.

Menurut Saifuddin Azwar, realibilitas ini diukur dengan menentukan hubungan antara skor dua paruh yang ekuivalen suatu tes, yang disajikan kepada seluruh kelompok pada suatu saat. Karena reliabilitas belah dua mewakili reliabilitas hanya separuh tes yang sebenarnya, rumus Spearman-Brown dapat digunakan untuk mengoreksi koefisien yang didapat.

G. Tes Standar dan Tes Non Standar

a. Pengertian Tes

Secara harfiah, kata “tes” berasal dari bahasa Perancis kuno: *testum* dengan arti: piring untuk menyisihkan logam-logam mulia (maksudnya dengan menggunakan alat berupa piring itu akan diperoleh jenis-jenis logam mulia yang nilainya sangat tinggi), dalam bahasa Inggris ditulis dengan *test* yang dalam bahasa Indonesiannya diterjemahkan dengan makna “tes, ujian atau percobaan”.

Sedangkan Sumadi Suryabrata, mengartikan tes adalah: “pertanyaan-pertanyaan yang harus dijawab dan atau perintah-perintah yang harus dijalankan, yang mendasarkan harus bagaimana testee menjawab pertanyaan-pertanyaan atau

melakukan perintah-perintah itu, penyelidik mengambil kesimpulan dengan cara membandingkan dengan standar atau testee lainnya”.

Dari kedua pengertian diatas, diambil pengertian, tes adalah alat pengukuran berupa pertanyaan, perintah dan petunjuk yang ditujukan kepada testee unruk mendapatkan respon sesuai dengan petunjuk itu. Atas dasar respon tersebut ditentukan tinggi rendahnya skor dalam bentuk kuantitatif, selanjutnya dibandingkan dengan standar yang telah ditentukan untuk ditarik kesimpulan yang bersifat kualitatif.

b. Tes Standar

1. Pengertian Tes Standar

Pengertian tes standar secara sempit adalah tes yang disusun oleh satu tim ahli, atau disusun oleh lembaga yang khusus menyelenggarakan secara professional. Tes tersebut diketahui memenuhi syarat sebagai tes yang baik. Tes ini dapat digunakan dalam waktu yang relatif lama, dapat diterapkan pada beberapa obyek mencakup wilayah yang luas.

Istilah “standar” dalam tes dimaksudkan bahwa semua siswa menjawab pertanyaan-pertanyaan yang sama dari sejumlah besar pertanyaan dikerjakan dengan menggunakan petunjuk yang sama dan dalam batasan waktu yang sama pula. Dengan demikian maka seolah-olah ada suatu standar atau ukuran sehingga diperoleh suatu standar penampilan (performance) dan penampilan kelompok lain dapat dibandingkan dengan penampilan kelompok standar tersebut.

Istilah “standar” tidak mengandung arti bahwa tes tersebut mengukur apa yang harus dan dapat diajarkan pada suatu tingkat tertentu atau bahwa tes itu menyiapkan suatu standar prestasi dimana siswa harus dan dapat mencapai suatu tingkat tertentu. Sekali lagi, tes standar dipolakan untuk penampilan prestasi sekarang (yang ada) yang dilaksanakan secara seragam, diusahakan dalam kondisi yang seragam, baik itu diberikan kepada siswa dalam pelaksanaan perseorangan maupun siswa sebagai anggota dari suatu kelompok.

2. Kegunaan Tes Standar

Secara singkat dapat dikemukakan bahwa kegunaan tes standar adalah:

- a. Jika ingin membuat perbandingan.
- b. Jika banyak orang yang akan memasuki suatu sekolah tetapi tidak tersedia data tentang calon ini.

Secara garis besar kegunaan tes standar adalah:

- a. Membandingkan prestasi belajar dengan pembawaan individual atau kelompok.
- b. Membandingkan tingkat prestasi siswa dalam keterampilan di berbagai bidang studi untuk individual atau kelompok.
- c. Membandingkan prestasi siswa berbagai sekolah atau kelas.
- d. Mempelajari perkembangan siswa dalam suatu periode atau waktu tertentu.

3. Kelengkapan Tes Standar

Sebuah tes yang sudah distandarisasikan dan sudah dapat disebut sebagai tes standar, biasanya dilengkapi dengan sebuah manual. Manual ini memuat keterangan-keterangan atau petunjuk-petunjuk yang perlu terutama yang menjelaskan tentang pelaksanaan, mengskor, dan mengadakan interpretasi.

Secara garis besar manual tes standar ini memuat:

- a. Ciri-ciri mengenai tes.

Misalnya menyebutkan tingkat validitas. Tingkat reliabilitas dan sebagainya.

- b. Tujuan serta keuntungan-keuntungan dari tes.

Misalnya yang disebutkan untuk siapa tes tersebut diberikan dan untuk tujuan apa.

c. Proses standarisasi tes.

Misalnya mengenai hal-hal yang berhubungan dengan sampel.

-Besarnya sampel,

-Teknik sampling,

-Kelompok mana yang diambil sebagai sampel(sifat sampel).

d. Petunjuk-petunjuk tentang cara melaksanakan tes.

Misalnya: dilaksanakan dengan lisan atau tertulis, waktu yang digunakan untuk mengerjakan setiap bagian, boleh tidaknya mencoba keluar jika sudah selesai mengerjakan soal itu dan sebagainya.

e. Petunjuk-petunjuk bagaimana cara menskor.

Misalnya: untuk beberapa skor tiap-tiap soal/unit, menggunakan sistem hukuman atau tidak, bagaimana cara menghitung nilai akhir dan sebagainya.

f. Petunjuk-petunjuk untuk menginterpretasikan hasil.

4. Jenis-jenis Tes Standar

Tes standar diklasifikasikan menjadi 7 jenis utama dan beberapa contoh dari setiap jenis yang sering digunakan dalam penelitian:

1) Tes kecerdasan

Tes kecerdasan memberikan perkiraan tingkat intelektual umum dengan sampling kinerja seseorang pada berbagai tugas. Tugas-tugas ini dapat mencakup definisi kata, pemecahan masalah matematika, pengetahuan umum, dan memori jangka pendek.

Tes kecerdasan sangat dijunjung tinggi oleh para peneliti pendidikan dan personil sekolah karena keberhasilan mereka dalam memprediksi prestasi sekolah. Bahkan sering disebut tes bakat skolastik karena mayoritas dari tes tersebut mengukur spek-aspek kecerdasan yang tampaknya diperlukan untuk sukses dalam belajar di sekolah.

2) Tes bakat

Tes bakat ditujukan untuk prediksi kinerja siswa dalam perilaku spesifik. Tes yang tersedia untuk mengukur bakat bagi banyak mata pelajaran tertentu seperti bahasa asing, seni, musik, dan matematika. Sebuah kecenderungan utama dalam pengujian pendidikan adalah perkembangan dari pengujian yang mengukur berbagai macam bakat yang berkaitan dengan keberhasilan kejuruan dan skolastik.

3) Tes prestasi

Tes prestasi banyak tersedia. Beberapa tes dimaksudkan untuk mengukur pengetahuan siswa mengenai fakta yang spesifik, pemahaman siswa dan penguasaan prinsip-prinsip dasar. Meskipun tes prestasi telah dikritik atas dasar sosial, tes tersebut mungkin yang paling valid, dapat diandalkan, dan berguna untuk peneliti pendidikan.

Administrasi waktu untuk tes prestasi yang berbeda sangat bervariasi, beberapa tes mengambil sedikitnya 30 menit, sedangkan yang lain memerlukan dua hari pengujian. Tes prestasi juga berbeda dalam subjek-materi cakupan. Dalam memilih suatu tes prestasi untuk proyek penelitian, harus terlebih dahulu memutuskan bidang prestasi yang akan diukur dan kemudian mengevaluasi tes yang dimaksudkan untuk mengukur prestasi di bidang ini.

4) Tes diagnostic

Tes diagnostik adalah bentuk tes prestasi. Namun, tes prestasi biasanya menghasilkan skor tunggal menunjukkan tingkat umum prestasi siswa dalam mata pelajaran tertentu. Beberapa tes diagnostik digunakan umum adalah Stanford

Diagnostic Reading Test, Stanford Diagnostic Mathematics Test, dan Diagnostic Screening Test.

Ada beberapa keuntungan untuk administrasi tes diagnostik dalam program penelitian atau perbaikan. Pertama, siswa yang memiliki kekurangan tertentu dalam subjek dapat diidentifikasi. Kedua, penggunaan tes diagnostik sangat membantu dalam perencanaan perbaikan instruksi individual, pendekatan yang dianjurkan oleh banyak pendidik. Kerugian dari beberapa tes diagnostik adalah bahwa sub scores memiliki reliabilitas yang rendah dan sangat saling berkorelasi dengan satu sama lain.

5) Mengukur kreatifitas

Cara untuk mengidentifikasi dan melatih kreatifitas seseorang merupakan masalah utama dalam penelitian pendidikan selama dua dekade ini. Sangat mengejutkan bahwa ternyata telah banyak berkembang pula cara untuk mengukur kreatifitas pada waktu yang dua dekade ini. Alasan mendasar dalam melakukan penelitian dalam kreativitas adalah meningkatnya ketertarikan para pendidik terhadap factor-factor non-intelektual, seperti kreatifitas dan personalitas, dalam pencapaian siswa di sekolah.

Kebanyakan pengukuran tentang kreativitas ditekankan pada penilaian terhadap kemampuan belajar siswa dan sifat-sifat siswa yang dapat berkontribusi pada pencapaian kreatif siswa. Hal tersebut tidak mengukur secara langsung tingkat kreatifitas siswa.

6) Self-report dalam pengukuran kepribadian

Pengukuran ini digunakan secara berkelanjutan dalam penelitian pendidikan untuk menggambarkan karakteristik kepribadian kelompok yang sesuai dengan kebutuhan peneliti; seperti kelompok minoritas, kelompok anak-anak kurang beruntung dsb. Mereka juga menggunakan cara tersebut dalam penelitian yang dilakukannya untuk melihat keterkaitan antara karakteristik

seseorang dengan faktor-faktor lainnya, seperti tingkat kecerdasan, prestasi dalam kelompoknya atau popularitas.

Keuntungan dari tes tersebut adalah murah dan mudah untuk diatur dan dinilai. Pertanyaan yang dikemukakan pada tes tersebut sebagian besar berupa pertanyaan objektif; seperti pertanyaan ya-tidak atau pertanyaan berbentuk pilihan ganda.

Kekurangan dari tes ini adalah tes ini biasanya merupakan self-report dari individu yang dites. Seperti self-report pada umumnya, tes tersebut akan akurat apabila persepsi diri dari individu tersebut juga akurat dan juga terkait dengan kejujuran individu tersebut dalam mengisi tes tersebut. Hal tersebut merupakan masalah bagi banyak penelitian di bidang pendidikan dan psikologi.

7) Teknik proyektif

Istilah teknik-teknik proyektif ini dipopulerkan oleh Frank LK. Salah satu keuntungan yang diklaim sebagai teknik proyektif atas diri sendiri-laporan persediaan adalah bahwa mereka kurang tunduk pada berpura-pura. Teknik-teknik proyektif yang paling banyak digunakan adalah Uji Rorschach dan Test Apperception Tematik (TAT). Di Rorschach, subjek menanggapi satu set percikan tinta, sedangkan dalam TAT subjek menanggapi satu set gambar.

Terlepas dari popularitas tes proyektif kami akan merekomendasikan bahwa mereka akan digunakan dengan hati-hati dalam proyek penelitian, karena sebagai aturan, ukuran ini membutuhkan pelatihan yang ekstensif dan pengalaman untuk mengelola, skor, dan menafsirkan. Jika Anda memutuskan untuk menggunakan Rorschach atau teknik proyektif serupa dalam proyek Anda, anda harus mempekerjakan orang-orang yang memenuhi syarat untuk mengelola, skor dan menginterpretasikan hasil.

c. Tes Nonstandar

1. Pengertian Tes Nonstandar

Tes nonstandar adalah kebalikan tes standar, yaitu tes yang disusun oleh seorang pendidik yang belum memiliki keahlian professional dalam penyusunan tes, atau mereka yang memiliki keahlian tetapi tidak sempat menyusun tes secara baik, mengujicobakan, melakukan analisis sehingga validitas dan reliabilitas belum dapat dipertanggungjawabkan.

Tes buatan guru memang memiliki beberapa kekhususan, bisa jadi syarat kualitatif belum terpenuhi, tetapi ia memiliki kelebihan lebih cocok untuk mengukur hal-hal khusus yang tidak dapat distandarisasikan, seperti formatif, tes diagnostik, hasilnya lebih realistik. Sebab tes ini dirancang sesuai dengan keadaan peserta didik.

2. Kegunaan Tes NonStandar

- 1) Untuk menentukan seberapa baik siswa telah menguasai bahan pelajaran yang diberikan dalam waktu tertentu.
- 2) Untuk menentukan apakah sesuatu tujuan telah tercapai.
- 3) Untuk memperoleh suatu nilai.

Selanjutnya baik tes standard an tes buatan guru dianjurkan dipakai jika hasilnya akan digunakan untuk:

- 1) Mengadakan diagnosis terhadap ketidak mampuan siswa.
- 2) Menentukan tempat siswa dlam suatu kelas atau kelompok.
- 3) Memberikan bimbingan kepada siswa dalam pendidikan dan pemilihan jurusan.
- 4) Memilih siswa untuk program-program khusus.

3. Menilai Tes Nonstandar

Ada 4 cara untuk menilai tes yaitu :

1) Cara pertama meneliti secara jujur soal-soal yang sudah disusun, kadang-kadang dapat diperoleh jawaban tentang ketidakjelasan perintah atau bahasa, taraf kesukaran, dan lain-lain keadaan soal tersebut.

Pertanyaan-pertanyaan tersebut, antara lain:

- Ø Apakah banyaknya soal untuk tiap topik sudah seimbang?
- Ø Apakah semua soal menanyakan bahan yang telah diajarkan?
- Ø Apakah soal yang kita susun tidak merupakan pertanyaan yang membingungkan?
- Ø Apakah soal itu tidak sukar untuk dimengerti?
- Ø Apakah soal itu tidak dapat dikerjakan oleh sebagian siswa?

2) Cara kedua adalah mengadakan analisis soal (item analysis). Analisis soal adalah suatu prosedur yang sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes yang kita susun.

Faedah mengadakan analisis soal:

- Ø Membantu kita dalam mengidentifikasi butir-butir soal yang jelek.
- Ø Memperoleh informasi yang akan dapat digunakan untuk menyempurnakan soal-soal untuk kepentingan lebih lanjut.
- Ø Memperoleh gambaran secara selintas tentang keadaan yang kita susun.

3) Cara ketiga adalah mengadakan checking validitas.

Validitas yang paling penting dari tes buatan guru adalah validitas kurikuler. Untuk mengadakan checking validitas kurikuler, kita harus merumuskan tujuan setiap bagian pelajaran secara khusus dan jelas sehingga setiap soal dapat kita jodohkan dengan setiap tujuan khusus tersebut.

Dalam hal ini Terry D. Ten Brink, mengemukakan pendapatnya demikian:

Ø Untuk tes yang dirancang akan menggunakan norm-referenced tidak harus menuliskan setiap tujuan khusus, tetapi cukup dengan tujuan-tujuan yang esensial saja.

Ø Untuk tes yang dirancang akan menggunakan criterion referenced, maka setiap tujuan khusus harus dicantumkan dalam tabek spesifikasi.

Ø Cara keempat adalah dengan mengadakan checking reliabilitas.

Salah satu indikator untuk tes yang mempunyai reliabilitas yang tinggi adalah bahwa kebanyakan dari soal-soal tes itu mempunyai daya pembeda yang tinggi.

H. Analisis hasil evaluasi dan pengembangannya

1. Analisa Hasil Evaluasi Dan Pengembangannya

Pada akhir penggal waktu proses pembelajaran, antara lain akhir catur wulan, akhir semester, akhir tahun ajaran, akhir jenjang pendidikan diperlukan suatu laporan kemajuan peserta didik yang selanjutnya merupakan laporan kemajuan lembaga pendidikan.

Laporan ini akan memberikan bukti sejauhmana tujuan pendidikan yang diharapkan oleh anggota masyarakat, khususnya orang tua siswa dapat tercapai. Agar anggota masyarakat dapat menilai kemajuan sekolah secara objektif, seyogyanya setiap lembaga pendidikan membuka diri untuk memberikan informasi secara berkala. Pemberian informasi ini dapat berupa Laporan Umum dan Laporan Khusus tentang prestasi yang dapat dicapai oleh sekolah.

Analisis ialah proses untuk mengetahui informasi yang telah dikumpulkan. Analisis termasuk mengolah data yang telah dikumpulkan untuk menentukan kesimpulan yang telah didukung data tersebut, seberapa banyak ia mendukung dan seberapa banyak ia tidak mendukung. Tujuan dari analisis ialah membuat singkatan dari data dan menyimpulkan pesan-pesan yang ada di dalamnya sebagai

informasi yang dapat dipakai sebagai dasar yang tentatif untuk mengambil suatu keputusan.

2. Pengolahan Hasil Evaluasi

Langkah pertama yang perlu dilakukan dalam mengolah hasil evaluasi adalah mengadakan penyekoran terhadap jawaban siswa. Ada beberapa cara yang bisa dilakukan dalam penyekoran hasil tes, sesuai dengan bentuk-bentuk tes yang digunakan.

1. Pemberian skor untuk tes bentuk objektif

Secara sederhana, cara memberikan skor terhadap tes hasil objektif dapat dilakukan dengan memberikan skor 1 untuk jawaban benar dan memberikan skor 0 untuk jawaban salah. Total skor diperoleh dengan menjumlahkan skor yang diperoleh dari semua soal.

Beberapa jenis tes bentuk objektif, antara lain tes benar salah (true false), pilihan ganda (multiplechoice), menjodohkan (matching), melengkapi isian (completion), dan jawaban singkat (short answer). Cara pengolahan terhadap masing-masing jenis tersebut memiliki karakteristik tersendiri.

2. Pemberian skor tes bentuk esai.

Ada dua cara memberikan hasil skor terhadap hasil tes esai, yaitu cara penyekoran analitik (analytical scoring method) dan cara penyekoran holistic (holistic scoring method). Cara penyekoran analitik adalah cara penyekoran yang mengacu pada elemen-elemen jawaban ideal.

Tinggi rendahnya skor jawaban siswa, bergantung pada lengkap tidaknya elemen yang dituju. Sedangkan cara penyekoran holistic adalah cara penyekoran yang didasarkan pada keluasan respon jawaban yang diberikan. Tinggi rendahnya skor jawaban siswa bergantung pada kualitas keseluruhan jawaban siswa.

Ada dua tahap yang dilakukan dalam melakukan pengolahan hasil evaluasi, yaitu memberikan skor (scoring) dan memberikan penilaian (grading). Untuk mengolah data hasil evaluasi bisa menggunakan beberapa teknik analisis data. Interpretasi data bisa dilakukan dengan menggunakan pendekatan penilaian acuan patokan (criterion referenced interpretation) atau penilaian acuan kelompok (norm referenced interpretation).

Penilaian acuan patokan adalah penilaian yang dalam menginterpretasi hasil pengukuran secara langsung didasarkan pada standar performansi tertentu yang ditetapkan. Untuk analisis data bisa menggunakan analisis presentasi ketercapaian.

Sedangkan penilaian acuan kelompok adalah penilaian yang dalam menginterpretasi data hasil pengukuran didasarkan pada prestasi anggota kelompok lainnya. Beberapa teknik analisis yang bisa digunakan untuk mengolah data dengan pendekatan acuan kelompok adalah deviasi standar, mean, standar skor, rank, jenjang persentil dan sejenisnya.

Untuk mengolah hasil pengukuran dalam evaluasi pembelajaran, banyak teknis analisis data yang bisa digunakan. Analisis data pada hakikatnya adalah mengolah angka-angka yang diperoleh dari skor mentah menjadi suatu skor yang mudah di baca dan disimpulkan. Beberapa teknis analisis data yang banyak digunakan untuk mengolah data hasil evaluasi pembelajaran adalah sebagai berikut :

1. Tendensi sentral atau ukuran kecenderungan memusat. Ada tiga teknik utama yang digunakan untuk mengukur tendensi sentral, yakni mean, media dan mode.
2. Variabilitas adalah keanekaragaman angka-angka dalam suatu distribusi skor. Variabilitas merupakan variasi sebaran skor dari mean.
3. Skor standar
4. Skor komposit
5. Persentil dan jenjang persentil

6. Penentuan nilai akhir
7. Analisis hubungan
8. Analisis data kualitati

3. Pelaporan Data Hasil Evaluasi

Menurut Suharsimi Arikunto (Arikunto, 1999), laporan hasil evaluasi ini berupa catatan yang secara garis besarnya dibuat 2 macam, yakni:

1. Catatan Lengkap

Catatan lengkap adalah catatan tentang siswa yang berisi baik prestasi maupun aspek-aspek pribadi yang lain, misalnya: kejujuran, kebersihan, kerajinan, sikap sosial, kebiasaan bekerja, kepercayaan terhadap diri sendiri, disiplin ketelitian dan sebagainya. Tentang isi catatannya, ada yang hanya dinyatakan dengan kata singkat “Baik”, “Sedang”, “Kurang” atau dengan keterangan yang lebih terperinci.

2. Catatan tidak lengkap

Catatan tidak lengkap adalah catatan tentang siswa yang berisi gambaran tentang prestasi siswa, dan hanya sedikit saja menyinggung tentang kepribadian. Tentang catatan prestasi belajar siswa itu sendiri dapat dibedakan atas 2 cara: pertama, dengan pernyataan lulus-belum lulus, kedua, dengan nilai siswa.

Menurut Ridwan Sakni, bahwa laporan hasil evaluasi ini ada 2 (dua) bentuk laporan, yaitu:

1. Laporan Kemajuan Umum

Dikatakan laporan umum dikarenakan informasi tersebut diberikan untuk siapa saja yang berminat dengan sasaran utamanya adalah orang tua, peserta didik, dan masyarakat di sekitar sekolah. Laporan secara umum diberikan secara berkala, terutama pada akhir program sekolah, masyarakat diberi informasi tentang kegiatan yang telah dilaksanakan. Laporan kemajuan umum ini dapat berbentuk laporan fisik dan laporan melalui media.

Laporan kemajuan umum yang berbentuk fisik dapat dilaksanakan melalui berbagai kegiatan seperti pameran, dan lomba pameran yang di isi dengan:

- a. menunjukkan karya ilmiah peserta didik selama waktu tertentu, karya ilmiah ini mungkin berupa pekerjaan rumah, laporan berupa kunjungan ke tempat-tempat yang ada hubungannya dengan pendidikan, laporan pekerjaan laboratorium, laporan penemuan baru dan sebagainya.
- b. menunjukkan karya seni, baik seni lukis, seni tari, seni drama, seni karya bengkel dan lain sebagainya.
- c. mengadakan petunjuk olahraga, baik dalam bentuk pertandingan maupun dalam bentuk hiburan atau kesegaran.

Laporan kemajuan umum yang berbentuk media, selain laporan resmi kepala sekolah kepada atasannya, yang ditulis rutin, juga perlu dikembangkan laporan yang dapat dibaca masyarakat baik dalam bentuk media cetak maupun media elektronik.

2. Laporan Kemajuan Khusus

Dikatakan laporan khusus karena hanya disampaikan kepada orang tua dan peserta didik, karena laporan ini banyak menyangkut masalah pribadi yang tabu untuk diketahui oleh orang lain. Paling tidak ada dua jenis wadah yang dapat digunakan untuk menyampaikan laporan ini yaitu melalui :

- Pertemuan dengan orang tua siswa

Pertemuan dengan orang tua siswa merupakan kegiatan yang tak pernah terpisahkan dengan buku raport siswa. Dengan adanya pertemuan tatap muka ini kedua belah pihak akan membagi dan saling melengkapi informasi tentang pribadi peserta didik. Melalui pertemuan ini masalah yang dihadapi di sekolah ataupun yang terjadi di rumah akan dapat dicari jalan keluarnya demi keberhasilan siswa.

Pertemuan dengan orang tua siswa akan memberikan hasil yang bermakna, apabila direncanakan dengan baik. Melaksanakan pertemuan sejenis ini memerlukan keterampilan khusus, oleh karena itu latihan melaksanakan

pertemuan dengan orang tua siswa merupakan suatu mata tataran dalam berbagai pertemuan.

- Buku laporan kemajuan atau buku raport
Dengan membaca dan mencermati hasil yang dicapai oleh siswa melalui raport yang diterimanya, maka siswa dan orang tuanya dapat menentukan sikap dan mengambil langkah seperlunya untuk mengatasi kesulitan belajar anaknya, atau paling tidak selalu memberikan dorongan kepada anaknya untuk lebih meningkatkan lagi prestasi belajar yang telah dicapainya.

3. Pemanfaatan Data Hasil Evaluasi

Data hasil evaluasi mempunyai beberapa manfaat, baik bagi siswa ataupun guru diantaranya:

a. Manfaat bagi siswa

- Digunakan untuk mengetahui apakah siswa sudah menguasai bahan program secara menyeluruh.
- Merupakan penguatan bagi siswa.
Dengan mengetahui bahwa tes yang dikerjakan sudah menghasilkan skor yang tinggi sesuai dengan yang diharapkan, maka siswa merasa mendapat “anggukan kepala” dari guru, dan ini merupakan suatu tanda bahwa apa yang sudah dimiliki merupakan pengetahuan yang benar. Dengan demikian maka pengetahuan itu akan bertambah membekas diingatan. Disamping itu tanda keberhasilan suatu pelajaran akan memperbesar motivasi siswa untuk belajar lebih giat, agar dapat mempertahankan nilai yang sudah baik itu atau memperoleh lebih baik itu.
- Usaha perbaikan.
Dengan umpan balik (feed back) yang diperoleh setelah melakukan tes, siswa mengetahui kelemahan-kelemahannya. Bahkan dengan teliti siswa mengetahui bab atau bagian dari bahan yang mana yang belum diketahui/dikuasainya. Dengan demikian akan ada motivasi untuk meningkatkan penguasaan.
- Sebagai diagnose.

Bahan pelajaran yang sedang dipelajari oleh siswa merupakan serangkaian pengetahuan, keterampilan atau konsep. Dengan mengetahui hasil tes formatif, siswa dengan jelas dapat mengetahui bagian mana dari bahan pelajaran yang masih dirasakan sulit.

b. Manfaat bagi guru

Dengan telah mengetahui data hasil evaluasi yang diadakan, maka guru:

- Mengetahui sampai sejauh mana bahan yang diajarkan sudah dapat diterima oleh siswa. Hal ini akan menentukan pula apakah guru itu harus menggantikan cara menerangkan (strategi mengajar) atau tetap dapat menggunakan cara (strategi) yang lama.
- Mengetahui bagian-bagian mana dari bahan pelajaran yang belum dikuasai oleh siswa. Apabila bagian yang belum dikuasai kebetulan merupakan bahan prasyarat bagi bagian pelajaran yang lain, maka bagian itu harus diterangkan lagi, dan barangkali memerlukan cara atau media lain untuk memperjelas. Apabila bahan ini tidak diulangi, maka akan mengganggu kelancaran pemberian bahan pelajaran selanjutnya, dan siswa akan semakin tidak dapat menguasainya.
- Dapat meramalkan sukses dan tidaknya seluruh program yang akan diberikan.

RANGKUMAN

Kedudukan Evaluasi Pembelajaran dalam PBM/KBM Proses belajar mengajar atau kegiatan belajar mengajar merupakan sebuah sistem yg terdiri dari beberapa sub sistem/komponen/unsur dimana satu sub sistem/komponen/unsur yg satu dgn yg lain saling berhubungan dan saling mempengaruhi dan merupakan satu kesatuan yg tidak bisa dipisahkan.

Sub sistem/komponen/unsur tersebut yaitu: 1. Tujuan pembelajaran (RPP: SK, KD, indikator) 2. Sumber belajar (guru, pendidik, fasilitator (non formal), media, dll) 3. Peserta didik (penentu berlangsung tidaknya PBM/KBM) 4. Pencanaan yg merupakan bagian dari kurikulum 5. Strategi belajar mengajar (metode) 6. Media pembelajaran 7. Evaluasi Pembelajaran.

Evaluasi menuntut alat-alat ukur yang akurat dan bermakna untuk mengumpulkan informasi yang dibutuhkan guna membuat keputusan. Proses evaluasi umumnya berpusat pada siswa, ini berarti evaluasi dimaksudkan: a. Mengamati hasil belajar siswa dan berupaya menentukan bagaimana menciptakan kesempatan belajar. b. Mengamati peranan guru. c. Strategi pengajaran khusus. d. Materi kurikulum. e. Prinsip-prinsip belajar untuk diterapkan pada pengajaran.

Inti setiap evaluasi adalah pengadaan informasi bagi pihak pengelola proses belajar mengajar untuk membuat macam-macam keputusan. Evaluasi mempunyai beberapa fungsi, yaitu:

- a. Untuk mengetahui tercapai-tidaknya tujuan pengajaran secara komprehensif yang meliputi aspek pengetahuan, sikap dan tingkah laku (kognitif, afektif, psikomotorik).
- b. Sebagai umpan balik yang berguna bagi tindakan berikutnya dimana segi-segi yang dapat merugikan sebanyak mungkin dihindari.
- c. Bagi pendidik, evaluasi berguna untuk mengukur keberhasilan proses belajar mengajar, bagi peserta didik berguna untuk mengetahui bahan pelajaran yang diberikan dan dikuasainya dan bagi masyarakat untuk mengetahui berhasil-tidaknya program-program yang dilaksanakan.

DAFTAR PUSTAKA

- Arifin, Muhammad, Ilmu Pendidikan Islam, (Jakarta: Bumi Aksara, 1989)
- Arifin, Zainal, Evaluasi Pembelajaran, Bandung: Remaja Rosda Karya, 2012.
- Arikunto, Arikunto, Dasar-dasar Evaluasi Pendidikan ,Jakarta: Bumi Aksara, 2010.
- Arikunto, suharsimi. 1999. Dasar-Dasar Evaluasi Pendidikan. Jakarta: PT. Bumi Aksara.
- Djiwandono, Wuryani, Esti, Sri, Psikologi Pendidikan, (Jakarta: Grasindo, 2002)
- Farida Yusuf. 2008. Evaluasi Program dan Instrumen Evaluasi untuk program pendidikan dan penelitian. Jakarta: Rineka Cipta.
- Mudjijo. 1990. Tes Hasil Belajar. Jakarta: PT. Bumi Aksara
- Nurkencana, Wayan dan Sunartana, Evaluasi pendidikan, Surabaya: Usaha Nasional, 1986.
- Purwanto, Ngalim, Prinsip-prinsip dan teknik evaluasi pembelajaran , Bandung: Remaja Rosda Karya, 2009.
- Ramayulis, Ilmu Pendidikan Islam, (Jakarta: Kalam Mulia, 1992)
- Ramayulis, Metodologi Pendidikan Islam, (Jakarta: Kalam Mulia, 2002)
- Slameto, Drs. 1998. Evaluasi Pendidikan. Jakarta: PT. Bumi Akasara.
- Sukardi, Evalusi Pendidikan Prinsip dan Operasionalnya , Jakarta Timur: Bumi Aksara, 2008.
- Syah, Muhubbin, Psikologi Belajar, (Jakarta: PT.Rajawali Pers, 2009)
- Widyoko, Eko Putro, Evaluasi Program Pembelajaran , Yogyakarta: Pustaka Pelajar, 2010.
- Zuhairini, dkk, Metodik Khusus Pendidikan Agama, (Surabaya: Usaha Nasional, 1983)

JUDUL BUKU: Pengembangan sistem evaluasi pendidikan agama islam

Penulis ; Dr. Andi Warisno, M.MPd

PENERBIT; Lintang Rasi Aksara Books

ISBN: 978-602-7802-39-1