



Doni, Sindu, BG Phalguna, & Yogi

**BETA**

# **EVALUASI PENDIDIKAN**

**EVALUASI PENDIDIKAN**

**D'SBY**

**BETA**

**BETA**



# **EVALUASI PENDIDIKAN**

**Disusun Oleh**

**I Nyoman Doni Pramana  
Ngakan Putu Sindu Wija Putra  
Komang Wahyu Phalguna BG  
Ketut Yogi Nugraha  
(D'SBY)**



## **KATA PENGANTAR**

Puji syukur kami panjatkan ke hadapan Tuhan Yang Maha Esa, yang telah melimpahkan kekuatan lahir dan batin kepada diri kami, sehingga setelah melalui proses yang cukup panjang, pada akhirnya buku ini dapat dicetak.

Penyusunan buku ini bertujuan untuk kelengkapan tugas mata kuliah Evaluasi Pendidikan. Selain itu buku ini juga bertujuan untuk membantu memberikan pengetahuan dan keterampilan kepada calon guru maupun para guru yang sudah bertugas.

Harapan penyusun semoga buku ini dapat memberikan bimbingan sesuai dengan tugas yang diemban oleh guru, yakni mengajar, mendidik, dan di dalamnya termasuk penilaian.

Disadari bahwa buku ini jauh dari sempurna, maka dari itu kritik dan saran untuk perbaikan sangat diharapkan. Semoga buku ini memberikan manfaat bagi pembaca dalam memahami konsep evaluasi, khususnya evaluasi dalam bidang pendidikan.

Penyusun

## DAFTAR ISI

<b>Kata Pengantar .....</b>	<b>i</b>
<b>Daftar Isi .....</b>	<b>ii</b>

<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
A. Pengertian Evaluasi dan Evaluasi Pendidikan .....	1
B. Fungsi Evaluasi Pendidikan .....	3
C. Tujuan Evaluasi Pendidikan.....	14
D. Obyek dan Subyek Evaluasi Pendidikan .....	16
E. Ruang Lingkup Evaluasi Pendidikan di Sekolah ..	19

<b>BAB II TEKNIK EVALUASI HASIL BELAJAR.....</b>	<b>22</b>
A. Prinsip-prinsip Dasar Evaluasi Hasil Belajar .....	22
B. Ciri-ciri Evaluasi Hasil Belajar .....	25
C. Ranah Kognitif, Afektif dan Psikomotorik sebagai Obyek Evaluasi Hasil Belajar .....	27
D. Langkah-langkah Pokok dalam Evaluasi Hasil Belajar.....	37

<b>BAB III TEKNIK PENYUSUNAN DAN PELAKSANAAN TES HASIL BELAJAR... 39</b>	
A. Teknik Tes .....	39
B. Teknik Nontes .....	44
C. Ciri-ciri Tes yang Baik .....	50
D. Prinsip-prinsip Dasar dalam Penyusunan Tes Hasil Belajar.....	54

E. Bentuk-bentuk Tes Hasil Belajar dan Teknik Penyusunannya.....	57
--	----

<b>BAB IV TEKNIK PENGUJIAN VALIDITAS TES DAN VALIDITAS ITEM TES HASIL BELAJAR .....</b>	<b>64</b>
A. Teknik Pengujian Validitas Tes Hasil Belajar .....	66
B. Cara Mengetahui Validitas Alat Ukur .....	72
C. Teknik Pengujian Validitas Item Tes Hasil Belajar.....	75

<b>BAB V TEKNIK PENGUJIAN RELIABILITAS TES HASIL BELAJAR .....</b>	<b>82</b>
A. Teknik Pengujian Reliabilitas Tes Hasil Belajar Bentuk Uraian .....	84
B. Teknik Pengujian Reliabilitas Tes Hasil Belajar Bentuk.....	89

<b>BAB VI TEKNIK PEMERIKSAAN, PEMBERIAN SKOR, DAN PENGOLAHAN HASIL TES HASIL BELAJAR.....</b>	<b>109</b>
A. Teknik Pemeriksaan Hasil Tes Hasil Belajar .....	109
B. Teknik Pemberian Skor Hasil Tes Hasil Belajar .....	118
C. Teknik Pengolahan dan Pengubahan Skor Tes Hasil Belajar Menjadi Nilai.....	122

<b>BAB VII ANALISIS ITEM TES HASIL BELAJAR .....</b>	<b>133</b>
--	------------

A. Analisis Derajat Kesukaran Item .....	134
B. Analisis Daya Pembeda Item .....	143
C. Analisis Fungsi Distraktor .....	146

## **BAB VIII TEKNIK PENENTUAN NILAI AKHIR**

### **DAN PENYUSUNAN RANKING .....**

A. Fungsi Nilai Akhir .....	148
B. Faktor-faktor yang Turut Diperhitungkan dalam Penilaian .....	151
C. Teknik Penentuan Nilai Akhir .....	152
D. Teknik Penyusunan Ranking .....	157

## **Daftar Pustaka**

## **BAB I**

### **PENDAHULUAN**

#### **A. Pengertian Evaluasi dan Evaluasi Pendidikan**

Evaluasi berasal dari bahasa Inggris *evaluation* yang berarti *penilaian*. Evaluasi diartikan sebagai suatu proses penilaian untuk mengambil keputusan yang menggunakan seperangkat hasil pengukuran dan berpatokan kepada tujuan yang telah dirumuskan.

Ada beberapa definisi evaluasi menurut para ahli, diantaranya,

1. *Blom et. al* (1971) (dalam Daryanto, 1999:1)

*"Evaluation, as we see it, is the systematic collection of evidence to determine whether in fact certain changes are taking place in the learners as well as to determine the amount or degree of change in individual students."*

Artinya: Evaluasi, sebagaimana kita lihat, adalah pengumpulan kenyataan secara sistematis untuk menetapkan apakah dalam kenyataannya terjadi perubahan dalam diri siswa dan menetapkan mana tingkat perubahan dalam pribadi siswa.

2. *Stufflebeam et. al* (1971) (dalam Daryanto, 1999:1)

*"Evaluation is the process of delineating, obtaining, and providing useful information for judging decision alternatives."*

Artinya: Evaluasi merupakan proses menggambarkan, memperoleh dan menyajikan informasi yang berguna untuk menilai alternatif keputusan.

3. Robert L. Thorndike dan Elizabeth Hagen (1961) (dalam Iskandar, 2011), menjelaskan evaluasi tersebut dengan mengatakan bahwa evaluasi itu berhubungan dengan pengukuran. Dalam beberapa hal evaluasi lebih luas, karena dalam evaluasi juga termasuk penilaian formal dan penilaian intuitif mengenai kemajuan peserta didik. Evaluasi juga mencakup penilaian tentang apa yang baik dan apa yang diharapkan. Dengan demikian hasil pengukuran yang benar merupakan dasar yang kokoh untuk melakukan evaluasi.

Dari beberapa definisi ahli diatas dapat disimpulkan pengertian evaluasi sebagai kegiatan identifikasi untuk melihat apakah suatu program yang telah direncanakan telah tercapai atau belum, berharga atau tidak berharga, dan dapat pula untuk melihat tingkat efisiensi pelaksanaannya.

Sedangkan Evaluasi Pendidikan merupakan proses untuk menentukan tujuan pendidikan dibandingkan tujuan yang telah ditentukan (Sudijono, 2009), atau secara umum dapat diartikan bahwa evaluasi pendidikan sebagai suatu kegiatan penilaian yang dilakukan didalam dunia pendidikan.



## **B. Fungsi Evaluasi Pendidikan**

Menurut H. Daryanto dalam bukunya yang berjudul *Evaluasi Pendidikan*, ada beberapa fungsi evaluasi diantaranya:

### **1. Evaluasi berfungsi selektif**

Dengan cara mengadakan evaluasi guru mempunyai cara untuk mengadakan seleksi terhadap siswanya. Seleksi itu sendiri mempunyai berbagai tujuan, antara lain:

- a) Untuk memilih siswa yang dapat diterima di sekolah tertentu.
- b) Untuk memilih siswa yang dapat naik ke kelas atau tingkat berikutnya.
- c) Untuk memilih siswa yang seharusnya mendapat beasiswa.
- d) Untuk memilih siswa yang sudah berhak meninggalkan sekolah dan sebagainya.

### **2. Evaluasi berfungsi diagnostik**

Apabila alat yang digunakan dalam evaluasi cukup memenuhi syarat, maka dengan melihat hasilnya, guru akan mengetahui kelemahan siswa. Di samping itu diketahui pula sebab-sebab kelemahan itu. Jadi dengan mengadakan evaluasi, sebenarnya guru mengadakan diagnosis kepada siswa tentang kebaikan dan kelemahannya. Dengan diketahuinya sebab-sebab kelemahan ini, akan lebih mudah dicari cara untuk mengatasi.

### 3. Evaluasi berfungsi sebagai penempatan

Sistem baru yang kini banyak dipopulerkan di negara Barat, adalah sistem belajar sendiri. Belajar sendiri dapat dilakukan dengan cara mempelajari sebuah paket belajar, dilakukan dengan cara mempelajari sebuah paket belajar, baik itu berbentuk modul maupun paket belajar yang lain. Sebagai alasan dari timbulnya sistem ini adalah adanya pengakuan yang besar terhadap kemampuan individual. Setiap siswa sejak lahirnya telah membawa bakat sendiri-sendiri sehingga pelajaran akan lebih efektif apabila disesuaikan dengan pembawaan yang ada. Pendekatan yang lebih bersifat melayani perbedaan kemampuan adalah pengajaran secara kelompok. Untuk dapat menentukan dengan pasti di kelompok mana seorang siswa harus ditempatkan, digunakan suatu evaluasi. Sekelompok siswa mempunyai hasil evaluasi yang sama, akan berada dalam kelompok yang sama dalam belajar.

### 4. Evaluasi berfungsi sebagai pengukuran keberhasilan

Fungsi keempat dari evaluasi ini dimaksudkan untuk mengetahui sejauh mana suatu program berhasil diterapkan. Keberhasilan program ditentukan oleh beberapa faktor yaitu faktor guru, metode mengajar, kurikulum, sarana dan sistem kurikulum.

Kemudian jika dalam proses pengembangan sistem pendidikan, evaluasi berfungsi untuk:

a) Perbaikan sistem

Disini peran evaluasi lebih bersifat konstruktif, karena informasi hasil penilaian dijadikan input bagi perbaikan-perbaikan yang diperlukan di dalam sistem pendidikan yang sedang dikembangkan. Di sini evaluasi lebih merupakan kebutuhan yang datang dari dalam sistem itu sendiri karena evaluasi itu dipandang sebagai faktor yang memungkinkan dicapainya hasil pengembangan yang optimal dari sistem yang bersangkutan.

b) Pertanggung jawaban kepada pemerintah dan masyarakat

Selama dan terutama pada akhir fase pengembangan sistem pendidikan, perlu adanya semacam pertanggung jawaban dari pihak pengembangan kepada berbagai pihak yang berkepentingan. Pihak-pihak yang dimaksudkan mencakup baik pihak yang mensponsori kegiatan pengembangan sistem tersebut maupun pihak yang akan menjadi konsumen dari sistem yang telah dikembangkan. Dengan kata lain, pihak-pihak tersebut mencakup pemerintah, masyarakat, orang tua, petugas-petugas pendidikan dan pihak-pihak lainnya yang ikut mensponsori kegiatan pengembangan sistem yang bersangkutan.

Bagi pihak pengembang, tujuan yang kedua ini tidak dipandang sebagai kebutuhan dari dalam melainkan lebih merupakan suatu "keharusan" dari luar. Sekalipun demikian hal ini tidak bisa kita hindarkan karena

persoalan ini mencakup pertanggungjawaban sosial, ekonomi dan moral, yang sudah merupakan suatu konsekuensi logis dalam kegiatan pembaruan pendidikan.

Dalam pertanggung jawaban hasil yang telah dicapainya, pihak pengembang perlu mengemukakan kekuatan dan kelemahan dari sistem yang sedang dikembangkannya serta usaha lebih lanjut yang diperlukan untuk mengatasi kelemahan-kelemahan tersebut. Untuk menghasilkan informasi mengenai kekuatan dan kelemahan tersebut di atas itulah diperlukan kegiatan evaluasi.

c) Penentuan tindak lanjut hasil pengembangan.

Tindak lanjut hasil pengembangan sistem pendidikan dapat berbentuk jawaban atas dua kemungkinan pertanyaan: Pertama, apakah sistem baru tersebut akan atau tidak akan disebarluaskan? Kedua, dalam kondisi yang bagaimana dan dengan cara yang bagaimana pula sistem baru tersebut akan disebarluaskan?

Ditinjau dari proses pengembangan sistem yang sudah berlangsung, pertanyaan pertama dipandang tidak tepat untuk diajukan pada akhir fase pengembangan. Pertanyaan tersebut hanya mempunyai dua kemungkinan jawaban, ya atau tidak. Secara teoretik dapat saja terjadi bahwa jawaban yang diberikan itu adalah tidak, dan apabila hal ini terjadi, kita akan dihadapkan pada situasi yang tidak menguntungkan; biaya, tenaga dan waktu yang telah dikerahkan selama ini ternyata terbuang

dengan percuma; murid-murid yang telah menggunakan cara baru tersebut selama fase pengembangan telah terlanjur dirugikan; sekolah-sekolah tempat proses pengembangan itu berlangsung harus kembali menyesuaikan diri lagi kepada cara yang lama; dan lambat laun akan timbul sikap skeptik di kalangan orang tua dan masyarakat terhadap pembaruan pendidikan dalam bentuk apa pun.

Pertanyaan yang kedua dipandang lebih tepat untuk diajukan pada akhir fase pengembangan. Pertanyaan tersebut mengimpilkasikan sekurang-kurangnya tiga anak pertanyaan; aspek-aspek mana dari sistem tersebut yang masih perlu diperbaiki ataupun disesuaikan, strategi penyebaran yang bagaimana yang sebaiknya ditempuh, dan persyaratan-persyaratan apa yang perlu dipersiapkan terlebih dahulu di lapangan. Pertanyaan-pertanyaan dirasakan lebih bersifat konstruktif dan lebih dapat diterima, ditinjau dari segi sosial, ekonomi, moral maupun teknis.

Untuk menghasilkan informasi yang diperlukan dan menjawab pertanyaan yang kedua itulah diperlukan kegiatan evaluasi.

Sedangkan menurut Anas Sudijono dalam bukunya yang berjudul *Pengantar Evaluasi Pendidikan*, fungsi evaluasi pendidikan dibedakan secara umum dan khusus. Secara umum, evaluasi sebagai suatu tindakan atau proses setidaknya-tidaknya memiliki tiga macam fungsi pokok yaitu:

(1) mengukur kemajuan, (2) menunjang penyusunan rencana, dan (3) memperbaiki atau melakukan penyempurnaan kembali.

Seperti telah dikemukakan dalam pembicaraan terdahulu, evaluasi merupakan kegiatan atau proses untuk mengukur dan selanjutnya menilai, sampai sejauh mana tujuan yang telah dirumuskan sudah dapat dilaksanakan. Apabila tujuan yang telah dirumuskan itu direncanakan untuk dicapai secara bertahap, maka dengan evaluasi yang berkesinambungan akan dapat dipantau, tahapan manakah yang sudah dapat diselesaikan, tahapan manakah yang berjalan dengan mulus, dan mana pula tahapan yang mengalami kendala dalam pelaksanaannya. Sehingga dengan evaluasi terbuka kemungkinan bagi evaluator untuk mengukur seberapa jauh atau seberapa besar kemajuan atau perkembangan program yang dilaksanakan dalam rangka pencapaian tujuan yang telah dirumuskan.

Setidaknya ada dua kemungkinan hasil yang diperoleh dari kegiatan evaluasi, yaitu:

- 1) Hasil evaluasi itu ternyata menggembirakan, sehingga dapat memberikan rasa lega bagi evaluator, sebab tujuan yang telah ditentukan dapat dicapai sesuai dengan yang direncanakan.
- 2) Hasil evaluasi itu ternyata tidak menggembirakan atau bahkan mengkhawatirkan, dengan alasan bahwa berdasar hasil evaluasi ternyata dijumpai adanya penyimpangan-penyimpangan, hambatan

atau kendala, sehingga mengharuskan evaluator untuk bersikap waspada.

Evaluasi yang dilaksanakan secara berkesinambungan, akan membuka peluang bagi evaluator untuk membuat perkiraan, apakah tujuan yang telah dirumuskan akan dapat dicapai pada waktu yang telah ditentukan, atau tidak. Apabila berdasar data hasil evaluasi itu dipekirakan bahwa tujuan tidak akan dapat dicapai sesuai dengan rencana, maka evaluator akan berusaha untuk mencari dan menemukan faktor-faktor penyebabnya, serta mencari dan menemukan jalan keluar atau cara-cara pemecahannya. Bukan tidak mungkin bahwa atas dasar data hasil evaluasi itu evaluator perlu mengadakan perubahan-perubahan, penyempurnaan-penyempurnaan atau perbaikan-perbaikan, baik perbaikan yang menyangkut organisasi, tata kerja, dan bahkan mungkin juga perbaikan terhadap tujuan organisasi itu sendiri. Jadi kegiatan evaluasi pada dasarnya juga dimaksudkan untuk melakukan perbaikan atau penyempurnaan usaha. Perbaikan usaha tanpa didahului oleh kegiatan evaluasi adalah tidak mungkin, sebab untuk mengadakan perbaikan terlebih dahulu harus diketahui apa yang harus diperbaiki, dan mengapa hal itu perlu diperbaiki. Kegiatan evaluasi yang tidak menghasilkan titik tolak untuk perbaikan adalah hampa dan tidak ada artinya sama sekali.

Adapun secara khusus, fungsi evaluasi dalam dunia pendidikan dapat ditilik dari tiga segi, yaitu:

- 1) Segi psikologis.
- 2) Segi didaktik.
- 3) Segi administratif.

Secara psikologis, kegiatan evaluasi dalam bidang pendidikan di sekolah dapat disoroti dari dua sisi, yaitu dari sisi peserta didik dan dari sisi pendidik.

Bagi peserta didik, evaluasi pendidikan secara psikologis akan memberikan pedoman atau pegangan batin kepada mereka untuk mengenal kapasitas dan status dirinya masing-masing di tengah-tengah kelompok atau kelasnya. Dengan dilakukannya evaluasi terhadap hasil belajar siswa, maka para siswa yang berkemampuan tinggi, berkemampuan rata-rata, ataukah berkemampuan rendah. Demikian pula dengan dilakukannya evaluasi hasil belajar tersebut maka para siswa yang bersangkutan akan menjadi tahu atau mengerti: dimanakah posisi (letak) dirinya di tengah teman-temannya. Apakah dia termasuk siswa kelompok atas (pandai), tengah (sedang), ataukah termasuk dalam kelompok bawah (bodoh).

Bagi pendidik, evaluasi pendidikan akan memberikan kepastian atau ketetapan hati kepada diri pendidik tersebut, sudah sejauh manakah kiranya usaha yang telah dilakukannya selama ini telah membawa hasil, sehingga ia secara psikologis memiliki pedoman atau pegangan batin yang pasti guna menentukan langkah-langkah apa saja yang dipandang perlu dilakukan selanjutnya. Misalnya, dengan menggunakan metode-



metode mengajar tertentu, hasil-hasil belajar siswa telah menunjukkan adanya peningkatan daya serap terhadap materi yang telah diberikan kepada para siswa tersebut, karena itu penggunaan metode-metode mengajar tadi akan terus dipertahankan. Sebaliknya, apabila hasil-hasil belajar siswa ternyata tidak menggembirakan, maka pendidik akan berusaha melakukan perbaikan-perbaikan dan penyempurnaan sehingga hasil belajar siswa menjadi lebih baik.

Bagi peserta didik, secara didaktik evaluasi pendidikan akan dapat memberikan dorongan kepada mereka untuk dapat memperbaiki, meningkatkan dan mempertahankan prestasinya. Evaluasi hasil belajar itu misalnya, akan menghasilkan untuk masing-masing siswa. Ada siswa yang nilainya jelek, karena itu siswa tersebut terdorong untuk memperbaikinya, agar untuk waktu-waktu yang akan datang nilai hasil belajarnya tidak sejelek sekarang. Ada siswa yang nilainya tidak jelek, tetapi belum dapat dikatakan baik atau memuaskan, karena itu siswa tersebut akan memperoleh dorongan untuk meningkatkan prestasi belajarnya pada masa-masa yang akan datang. Adapula siswa yang nilainya baik, dengan nilai yang sudah baik itu, siswa yang bersangkutan akan termotivasi untuk mempertahankan prestasi yang tinggi itu, agar tidak mengalami penurunan pada masa-masa yang akan datang.

Bagi pendidik, secara didaktik evaluasi pendidikan itu setidaknya-tidaknya memiliki lima macam fungsi. yaitu:

- 1) Memberikan landasan untuk menilai hasil usaha (prestasi) yang telah dicapai oleh peserta didiknya.

Disini evaluasi diaktakan berfungsi memeriksa pada bagian-bagian manakah para peserta didik mengalami kesulitan dalam mengikuti pembelajaran, untuk selanjutnya dapat dicari jalan keluar untuk menagtasinya. Jadi disini evaluasi mempunyai sifat diagnostik.

- 2) Memberikan informasi yang sangat berguna, untuk mengetahui posisi masing-masing peserta didik di tengah-tengah kelompoknya.

Dalam hubungan ini, evaluasi sangat diperlukan untuk dapat menentukan secara pasti, pada kelompok manakah kiranya seorang peserta didik seharusnya ditempatkan. Jadi disini evaluasi memiliki fungsi *placement*.

- 3) Memberikan bahan yang penting untuk memilih dan kemudian menetapkan status peserta didik.

Dalam hubungan ini, evaluasi pendidikan dilakukan untuk menetapkan, apakah seorang peserta didik dapat dinyatakan lulus atau tidak lulus, naik kelas atau tidak. Sehingga evaluasi memiliki fungsi selektif.

- 4) Memberikan pedoman untuk mencari dan menemukan jalan keluar bagi peserta didik yang memang memerlukannya.

Berlandaskan pada hasil evaluasi, pendidik dimungkinkan untuk dapat memberikan petunjuk dan bimbingan kepada para peserta didik. Dalam keadaan ini evaluasi memiliki fungsi bimbingan.

- 5) Memberikan petunjuk tentang sudah sejauh manakah program pengajaran yang telah ditentukan telah dapat dicapai.

Disini evaluasi dikatakan memiliki fungsi instruksional, yaitu melakukan perbandingan antara tujuan instruksional khusus (TIK) yang telah ditentukan untuk masing-masing mata pelajaran dengan hasil-hasil belajar yang telah dicapai oleh peserta didik bagi masing-masing mata pelajaran tersebut, dalam jangka waktu yang telah ditentukan

Adapun secara administratif, evaluasi pendidikan setidaknya memiliki tiga macam fungsi, yaitu:

- 1) Memberikan Laporan

Dengan melakukan evaluasi, akan dapat disusun dan disajikan laporan mengenai kemajuan dan perkembangan peserta didik setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu. Laporan mengenai perkembangan dan kemajuan belajar peserta didik biasanya tertuang dalam bentuk raport (untuk siswa), atau KHS (untuk mahasiswa), yang selanjutnya disampaikan ke pada orang tua peserta didik tersebut pada setiap akhir semester.

2) Memberikan Bahan-bahan Keterangan (Data)

Dalam hubungan ini, nilai-nilai hasil belajar peserta didik yang diperoleh dari kegiatan evaluasi, adalah merupakan data yang sangat penting untuk keperluan pengambilan keputusan pendidikan dan lembaga pendidikan, apakah seseorang peserta didik dapat dinyatakan tamat belajar, dapat dinyatakan naik kelas, tinggal kelas, lulus atau tidak lulus.

3) Memberikan Gambaran

Dari kegiatan evaluasi hasil belajar yang telah dilakukan untuk berbagai jenis mata pelajaran misalnya, akan dapat tergambar bahwa dalam mata pelajaran tertentu (misalnya: Matematika dan IPA) pada umumnya kemampuan peserta didik masih sangat memprihatinkan. Sebaliknya, untuk mata pelajaran pendidikan pancasila, hasil belajar siswa pada umumnya sangat menggembirakan. Gambaran tersebut dapat diketahui melalui data yang berupa IPK, Nilai Ebtanas Murni atau lainnya.

### **C. Tujuan Evaluasi Pendidikan**

Tujuan evaluasi pendidikan terbagi atas tujuan umum dan khusus, dimana dijelaskan sebagai berikut.

1. Tujuan Umum

Secara umum, tujuan evaluasi dalam bidang pendidikan ada dua, yaitu:

- a) Untuk menghimpun bahan-bahan keterangan yang akan dijadikan sebagai bukti mengenai taraf perkembangan atau taraf kemajuan yang dialami oleh para peserta didik, setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu. Dengan kata lain tujuan umum dari evaluasi dalam pendidikan adalah untuk memperoleh data pembuktian, yang akan menjadi petunjuk sampai dimana tingkat kemampuan dan tingkat keberhasilan peserta didik dalam pencapaian tujuan-tujuan kurikuler, setelah mereka menempuh proses pembelajaran dalam jangka waktu yang telah ditentukan.
- b) Untuk mengukur dan menilai sejauh mana efektivitas mengajar dan metode-metode mengajar yang telah diterapkan atau dilaksanakan oleh pendidik, serta kegiatan belajar yang dilaksanakan oleh peserta didik.

## 2. Tujuan Khusus

Adapun tujuan khusus dari kegiatan evaluasi dalam bidang pendidikan adalah sebagai berikut:

- a) Untuk merangsang kegiatan peserta didik dalam menempuh program pendidikan. Tanpa adanya evaluasi maka tidak mungkin timbul kegairahan atau rangsangan pada diri peserta didik untuk memperbaiki dan meningkatkan prestasinya masing-masing.

- b) Untuk mencari dan menemukan faktor-faktor penyebab keberhasilan dan ketidakberhasilan peserta didik dalam mengikuti program pendidikan, sehingga dapat dicari dan ditemukan jalan keluar atau cara-cara perbaikannya.

## **D. Obyek dan Subyek Evaluasi Pendidikan**

### **1. Obyek Evaluasi Pendidikan**

Obyek atau sasaran evaluasi pendidikan ialah segala sesuatu yang bertalian dengan kegiatan atau proses pendidikan, yang dijadikan titik pusat perhatian atau pengamatan, karena pihak penilai (evaluator) ingin memperoleh informasi tentang kegiatan atau proses pendidikan tersebut. Salah satu cara untuk mengenal atau mengetahui obyek evaluasi pendidikan adalah dengan jalan menyurutinya dari tiga segi, yaitu dari segi input, transformasi dan output, dimana input kita anggap sebagai "bahan mentah yang akan diolah", transformasi kita anggap sebagai "dapur tempat mengolah bahan mentah", dan output kita anggap sebagai "hasil pengolahan yang dilakukan di dapur dan siap untuk dipakai".

Dalam dunia pendidikan, khususnya dalam proses pembelajaran di sekolah, input adalah para calon peserta didik, transformasi adalah sekolah tempat kita mendidik calon peserta didik, dan output adalah peserta didik yang telah berhasil menimba ilmu di sekolah tersebut. Dari segi

input, maka obyek evaluasi pendidikan meliputi tiga aspek, yaitu:

a) Aspek Kemampuan

Untuk dapat diterima sebagai calon peserta didik dalam rangka mengikuti program pendidikan tertentu, maka para calon peserta didik itu harus memiliki kemampuan yang sesuai atau memadai, sehingga dalam mengikuti proses pembelajaran pada program pendidikan tertentu itu nantinya, peserta didik tidak akan mengalami banyak hambatan atau kesulitan.

b) Aspek Kepribadian

Sebelum mengikuti program pendidikan tertentu, para calon peserta didik perlu terlebih dahulu dievaluasi kepribadiannya masing-masing, sebab baik buruknya kepribadian mereka secara psikologis akan dapat memperngaruhi keberhasilan mereka dalam mengikuti program pendidikan tertentu. Evaluasi yang dapat dilakukan untuk mengetahui kepribadian seseorang adalah dengan jalan menggunakan tes kepribadian (*personality test*).

c) Aspek Sikap

Sikap, pada dasarnya adalah merupakan bagian dari tingkah laku manusia. Aspek sikap perlu dinilai atau dievaluasi terlebih dahulu bagi para calon peserta didik sebelum mengikuti program pendidikan tertentu. Untuk menilai sikap tersebut digunakan alat berupa tes sikap.

Selanjutnya, apabila disoroti dari segi transformasi, maka obyek dari evaluasi pendidikan itu meliputi:

- 1) Kurikulum atau materi pelajaran.
- 2) Metode mengajar dan teknik penilaian.
- 3) Sarana atau media pendidikan.
- 4) Sistem administrasi.
- 5) Guru dan unsur-unsur personil lainnya yang terlibat dalam proses pendidikan.

Transformasi yang dapat diibaratkan sebagai “mesin pengolahan yang bertugas mengubah bahan mentah menjadi bahan jadi”, akan memegang peranan yang sangat penting. Ia dapat menjadi faktor penentu yang dapat menyebabkan keberhasilan atau kegagalan dalam upaya pencapaian tujuan pendidikan yang telah ditentukan, karena itu obyek-obyek evaluasi yang termasuk dalam transformasi itu perlu dinilai atau dievaluasi secara berkesinambungan.

Adapun dari segi output, yang menjadi sasaran evaluasi pendidikan adalah tingkat pencapaian atau prestasi belajar yang berhasil diraih oleh masing-masing peserta didik, setelah mereka terlibat dalam proses pendidikan selama jangka waktu yang telah ditentukan. Untuk mengetahui seberapa jauh tingkat pencapaian atau prestasi belajar yang diraih oleh para peserta didik itu, dipergunakan alat berupa *Tes Belajar* atau *Tes Hasil Belajar* yang dikenal dengan tes pencapaian (*achievement test*).



## 2. Subyek Evaluasi Pendidikan

Subyek atau pelaku evaluasi pendidikan disini ialah orang yang melakukan pekerjaan evaluasi dalam bidang pendidikan.

Dalam kegiatan evaluasi pendidikan di mana sasaran evaluasinya adalah prestasi belajar, maka subyek evaluasinya adalah guru atau dosen yang mengasuh mata pelajaran tertentu. Jika evaluasi yang dilakukan itu sasarannya adalah sikap peserta didik, maka subyek evaluasinya adalah guru atau petugas yang sebelum melaksanakan evaluasi tentang sikap itu, terlebih dahulu telah memperoleh pendidikan atau latihan mengenai cara-cara menilai sikap seseorang. Jika sasaran yang dievaluasi kepribadian peserta didik, di mana pengukuran tentang kepribadian itu dilakukan dengan menggunakan instrument berupa tes yang sifatnya baku, maka subyek evaluasinya adalah seorang psikolog, karena psikolog merupakan seseorang yang memang telah dididik untuk menjadi tenaga ahli yang professional di bidang psikologi.

## **E. Ruang Lingkup Evaluasi Pendidikan di Sekolah**

Secara umum, ruang lingkup dari evaluasi dalam bidang pendidikan di sekolah mencakup tiga komponen utama, yaitu:

- 1) evaluasi mengenai program pengajaran,

- 2) evaluasi mengenai proses pelaksanaan pengajaran,
- 3) evaluasi mengenai hasil belajar (hasil pengajaran).

*1. Evaluasi Program Pengajaran*

Evaluasi atau penilaian terhadap program pengajaran akan mencakup tiga hal, yaitu:

- a) Evaluasi terhadap tujuan pengajaran.
- b) Evaluasi terhadap isi program pengajaran.
- c) Evaluasi terhadap strategi belajar mengajar.

*2. Evaluasi Proses Pelaksanaan Pengajaran*

Evaluasi mengenai proses pelaksanaan pengajaran akan mencakup:

- a) Kesesuaian antara proses belajar mengajar yang berlangsung, dengan garis-garis besar program pengajaran yang telah ditentukan.
- b) Kesiapan guru dalam melaksanakan program pengajaran.
- c) Kesiapan siswa dalam mengikuti proses pembelajaran.
- d) Minat atau perhatian siswa di dalam mengikuti pelajaran.
- e) Keaktifan atau partisipasi siswa selama proses pembelajaran berlangsung.
- f) Peranan bimbingan dan penyuluhan terhadap siswa yang memerlukannya.

- g) Komunikasi dua arah antara guru dan murid selama proses pembelajaran berlangsung.
- h) Pemberian tugas-tugas kepada siswa dalam rangka penerapan teori-teori yang diperoleh di dalam kelas.
- i) Upaya menghilangkan dampak negatif yang timbul sebagai akibat dari kegiatan-kegiatan yang dilakukan di sekolah.

### *3. Evaluasi Hasil Belajar*

Evaluasi terhadap hasil belajar peserta didik ini mencakup:

- a) Evaluasi mengenai tingkat penguasaan peserta didik terhadap tujuan-tujuan khusus yang ingin dicapai dalam unit-unit program pengajaran yang bersifat terbatas.
- b) Evaluasi mengenai tingkat pencapaian peserta didik terhadap tujuan-tujuan umum pengajaran.

## **BAB II**

### **TEKNIK EVALUASI HASIL BELAJAR**

Pada bab ini membahas mengenai prinsip-prinsip dasar evaluasi belajar, ciri-ciri evaluasi belajar, ranah kognitif, ranah afektif dan ranah psikomotorik sebagai obyek evaluasi hasil belajar, langkah-langkah pokok dalam evaluasi hasil belajar dan teknik-teknik evaluasi hasil belajar di sekolah. Deskripsi lebih lanjut mengenai bab ini sebagai berikut.

#### **A. Prinsip-prinsip Dasar Evaluasi Hasil Belajar**

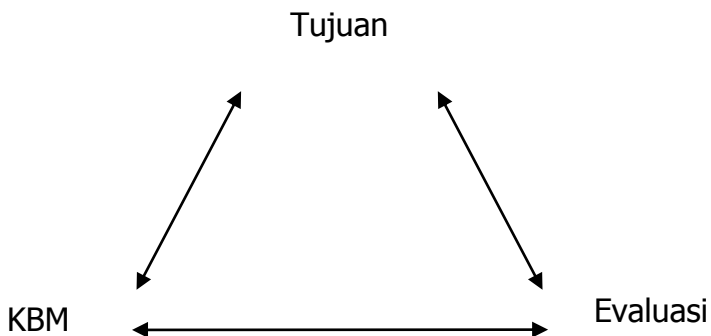
Sub-bab pertama pada BAB II ini menjelaskan bahwa evaluasi hasil belajar dapat dikatakan sudah baik pelaksanaannya jika sudah terdapat tiga prinsip dasar yaitu, prinsip Keseluruhan (*comprehensive*), dimana evaluasi hasil belajar harus dilakukan secara utuh dan menyeluruh, tidak boleh dilakukan secara terpisah dan setengah-setengah. Dengan demikian maka akan diperoleh suatu informasi mengenai perkembangan subyek didik yang sedang dinilai. Kemudian yang kedua ada prinsip Kesiambungan (*continuity*), disini dijelaskan bahwa evaluasi hasil belajar yang baik merupakan evaluasi yang dilakukan secara teratur atau berkesinambungan dari waktu ke waktu dimana evaluator akan bisa memperoleh informasi mengenai

perkembangan peserta didik dari awal hingga akhir. Sehingga nantinya evaluator akan dapat menentukan langkah-langkah selanjutnya yang harus diambil agar Tujuan Instruksional khusus dapat tercapai. Lalu yang ketiga ada prinsip Obyektivitas (*obyectivity*), dimana evaluasi hasil belajar dikatakan baik jika sudah terlepas dari faktor subyektif, karena faktor tersebut dapat menodai pekerjaan evaluasi atau penilaian tersebut.

Menurut Arikunto (2012: 38-40) menyebutkan bahwa prinsip umum dan penting dalam kegiatan evaluasi, yaitu adanya triagulasi atau hubungan erat tiga komponen, yaitu:

- a. Tujuan pembelajaran
- b. Kegiatan pembelajaran atau KBM
- c. Evaluasi

Triagulasi tersebut dapat digambarkan dalam gambar sebagai berikut.



Triagulasi ini memiliki hubungan terkait. Penjelasan mengenai hubungan tersebut sebagai berikut.

**a. Hubungan antara Tujuan dengan KBM**

Kegiatan belajar mengajar yang direncanakan dalam bentuk rencana mengajar disusun oleh guru dengan mengacu pada tujuan yang hendak dicapai. Dengan demikian, anak panah yang menunjukkan antara keduanya mengarah pada tujuan dengan makna bahwa KBM mengacu pada tujuan, tetapi juga mengarah dari tujuan ke KBM, menunjukkan langkah dari tujuan dilanjutkan pemikiran ke KBM.

**b. Hubungan antara Tujuan dengan Evaluasi**

Evaluasi adalah kegiatan pengumpulan data untuk mengukur sejauh mana tujuan sudah dicapai. Dengan makna demikian maka anak panah berasal dari evaluasi menuju ke tujuan. Di sisi lain, jika dilihat dari langkah, dalam menyusun alat evaluasi ia mengacu pada tujuan yang sudah dirumuskan.

**c. Hubungan antara KBM dengan Evaluasi**

Seperti yang sudah disebutkan dalam nomor (1), KBM dirancang dan disusun dengan mengacu pada tujuan yang telah dirumuskan. Telah disebutkan pula dalam nomor (2) bahwa alat evaluasi juga disusun dengan mengacu pada tujuan. Selanjutnya mengacu pada tujuan, evaluasi juga harus mengacu atau disesuaikan dengan KBM yang dilaksanakan. Sebagai

hasil, jika kegiatan belajar mengajar dilakukan oleh guru dengan menitikberatkan pada keterampilan, evaluasinya juga harus mengukur tingkat keterampilan siswa. Bukannya aspek pengetahuan.

Kecenderungan yang terdapat dalam praktek sekarang ini adalah bahwa evaluasi hasil belajar hanya dilakukan dengan tes tulis, menentukan aspek pengetahuan saja. Hal-hal yang berkaitan dengan aspek-aspek lain, kurang mendapat perhatian dalam evaluasi saat ini.

Selanjutnya pada bab berikut akan dijelaskan beberapa alat evaluasi berupa tes dan bukan tes (non tes).

## **B. Ciri-ciri Evaluasi Hasil Belajar**

Pada buku ini dijelaskan terdapat lima ciri evaluasi hasil belajar, yang pertama dimana evaluasi yang dilakukan adalah secara tidak langsung. Ketika pendidik ingin mengetahui orang itu pandai atau tidak, maka yang dinilai adalah berdasarkan indikatornya. Disini juga dijelaskan bahwa *Carl Witherington* menjelaskan ada 6 indikator yang bisa dijadikan tolak ukur orang tersebut pandai atau tidak, (1) kemampuan bekerja dengan angka atau bilangan, (2) penggunaan bahasa dengan baik dan benar, (3) kemampuan menangkap sesuatu yang baru, ketika cepat mengikuti pembicaraan seseorang, (4) kemampuan mengingat sesuatu, (5) kemampuan memahami hubungan antar gejala, (6) kemampuan

berpikir abstrak. Kemudian untuk ciri yang kedua, menilai keberhasilan peserta didik biasanya menggunakan penilaian kuantitatif atau simbol-simbol angka, lalu angka tersebut dianalisis dengan metode statistik dan diberikan hasil secara kualitatif. Kemudian untuk ciri ketiga kegiatan evaluasi menggunakan satuan-satuan tetap berdasarkan teori setiap populasi peserta didik bersifat heterogen. Untuk ciri yang keempat dijelaskan bahwa hasil-hasil evaluasi pada peserta didik umumnya tidak selalu terdapat kesamaan karena peserta didik merupakan makhluk hidup yang sewaktu-waktu dapat berubah tergantung keadaan disekitarnya. Kemudian untuk ciri kelima dijelaskan bahwa dalam evaluasi hasil belajar tidak menutup kemungkinan terjadinya kesalahan pendidik dalam memberikan penilaian dan pengukuran. *J.P. Guilford* sendiri menuturkan beberapa sumber kekeliruan pengukuran tersebut dilihat dari *sampling* (kekeliruan tester menentukan item soal dari sekian banyak materi ajar), *scoring* (kekeliruan tester dalam menentukan nilai/score), *ranking* (kekeliruan tester dalam pemberian peringkat pada peserta didik), dan *guessing* (kekeliruan yang terjadi akibat dari tes yang biasanya hanya tes obyektif/tebak terka).

Dijelaskan pula bahwa kekeliruan tersebut dapat terjadi karena empat faktor yaitu, faktor alat pengukur yang tidak tepat, faktor evaluator yang dikarenakan suasana batinnya, sifat evaluator yang pemurah atau pelit dalam pemberian skor, mudahnya evaluator



terpengaruh berita mengenai peserta didiknya yang akan dinilai dan kesan yang dialami evaluator dengan peserta didiknya pada masa lalu, kemudian faktor kekeliruan dalam diri peserta didik dapat berupa faktor *Psikis* (kejiwaan), *Fisik* (jasmani), *Nasib*, dan faktor yang terakhir adalah faktor situasi di saat terjadinya evaluasi hasil belajar tersebut, dimana lingkungan testee dapat mempengaruhi nilai dari testee tersebut.

### **C. Ranah Kognitif, Afektif dan Psikomotorik sebagai Obyek Evaluasi Hasil Belajar**

Sub-bab selanjutnya mencakup materi ranah *kognitif*, *afektif* dan *psikomotor* sebagai obyek evaluasi belajar, dimana ketiga ranah tersebut termuat dalam taksonomi tujuan pendidikan dari *Benjamin S.Bloom*. Ranah kognitif adalah ranah yang mencakup kegiatan mental (otak). Menurut Bloom, segala upaya yang menyangkut aktivitas otak adalah termasuk dalam ranah kognitif. Ranah kognitif berhubungan dengan kemampuan berfikir, termasuk didalamnya kemampuan menghafal, memahami, mengaplikasi, menganalisis, mensintesis, dan kemampuan mengevaluasi. Dalam ranah kognitif itu terdapat enam aspek atau jenjang proses berfikir, mulai dari jenjang terendah sampai dengan jenjang yang paling tinggi. Keenam jenjang atau aspek yang dimaksud adalah:

## 1. Pengetahuan (knowledge)

Pengetahuan (knowledge) adalah kemampuan seseorang untuk mengingat-ingat kembali (*recall*) atau mengenali kembali tentang nama, istilah, ide, rumus-rumus, dan sebagainya, tanpa mengharapkan kemampuan untuk menggungkannya. Pengetahuan atau ingatan adalah merupakan proses berfikir yang paling rendah. Contoh:

- Mengemukakan arti
- Menentukan lokasi
- Mendriskripsikan sesuatu
- Menceritakan apa yang terjadi
- Menguraikan apa yang terjadi

## 2. Pemahaman

Pemahaman adalah kemampuan seseorang untuk mengerti atau memahami sesuatu setelah sesuatu itu diketahui dan diingat. Dengan kata lain, memahami adalah mengetahui tentang sesuatu dan dapat melihatnya dari berbagai segi. Seseorang peserta didik dikatakan memahami sesuatu apabila ia dapat memberikan penjelasan atau memberi uraian yang lebih rinci tentang hal itu dengan menggunakan kata-katanya sendiri. Pemahaman merupakan jenjang kemampuan berfikir yang setingkat lebih tinggi dari ingatan atau hafalan. Contoh:

- Mengungkapkan gagasan dan pendapat dengan kata-kata sendiri
- Membedakan atau membandingkan
- Mengintepretasi data
- Mendriskripsikan dengan kata-kata sendiri
- Menjelaskan gagasan pokok
- Menceritakan kembali dengan kata-kata sendiri

### 3. Aplikasi

Aplikasi adalah kesanggupan seseorang untuk menerapkan atau menggunakan ide-ide umum, tata cara ataupun metode-metode, prinsip-prinsip, rumus-rumus, teori-teori dan sebagainya, dalam situasi yang baru dan kongkret. Penerapan ini adalah merupakan proses berfikir setingkat lebih tinggi ketimbang pemahaman. Peserta didik harus mampu memecahkan masalah atau menerapkan dalam kehidupan sehari-hari. Contoh:

- Menghitung kebutuhan
- Melakukan percobaan
- Membuat peta
- Membuat model
- Merancang strategi

#### 4. Analisis

Analisis adalah kemampuan seseorang untuk merinci atau menguraikan suatu bahan atau keadaan menurut bagian-bagian yang lebih kecil dan mampu memahami hubungan di antara bagian-bagian atau faktor-faktor yang satu dengan faktor-faktor lainnya. Jenjang analisis adalah setingkat lebih tinggi ketimbang jenjang aplikasi. Peserta didik dapat merenung dan memikirkan dengan baik tentang wujud nyata dari kedisiplinan seorang siswa di rumah, disekolah, dan dalam kehidupan sehari-hari di tengah-tengah masyarakat. Contoh:

- Mengidentifikasi faktor penyebab
- Merumuskan masalah
- Mengajukan pertanyaan untuk mencari informasi
- Membuat grafik
- Mengkaji ulang

#### 5. Sintesis

Sintesis adalah kemampuan berfikir yang merupakan kebalikan dari proses berfikir analisis. Sintesis merupakan suatu proses yang memadukan bagian-bagian atau unsur-unsur secara logis, sehingga menjelma menjadi suatu pola yang berstruktur atau bebrbentuk pola baru. Jenjang sintesis kedudukannya setingkat lebih tinggi daripada

jenjang analisis. Analisis juga merupakan menggabungkan berbagai informasi menjadi satu kesimpulan/konsep atau meramu/merangkai berbagai gagasan menjadi suatu hal yang baru

Contoh:

- Membuat desain
- Menemukan solusi masalah
- Menciptakan produksi baru, dst.

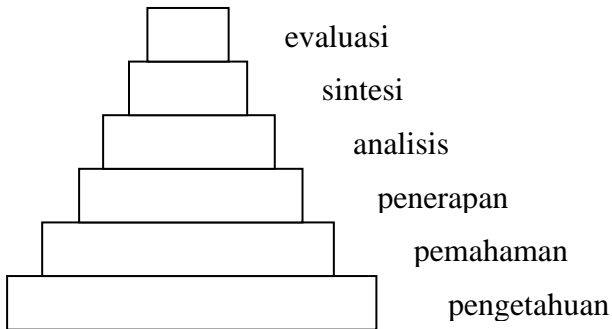
## 6. Evaluasi

Evaluasi adalah merupakan jenjang berpikir paling tinggi dalam ranah kognitif dalam taksonomi Bloom. Penilaian/evaluasi disini merupakan kemampuan seseorang untuk membuat pertimbangan terhadap suatu kondisi, nilai atau ide, misalkan jika seseorang dihadapkan pada beberapa pilihan maka ia akan mampu memilih satu pilihan yang terbaik sesuai dengan patokan-patokan atau kriteria yang ada. Salah satu contoh hasil belajar kognitif jenjang evaluasi adalah: peserta didik mampu menimbang-nimbang tentang manfaat yang dapat dipetik oleh seseorang yang berlaku disiplin dan dapat menunjukkan mudharat atau akibat-akibat negatif yang akan menimpa seseorang yang bersifat malas atau tidak disiplin.

Contoh:

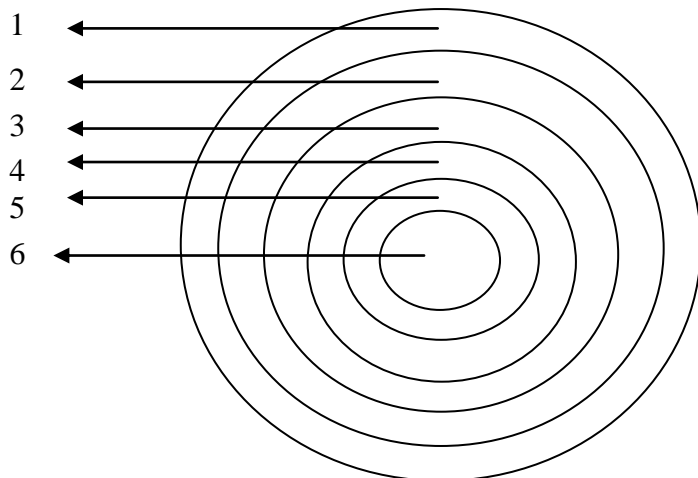
- Mempertahankan pendapat
- Membahas suatu kasus
- Memilih solusi yang lebih baik
- Menulis laporan,dst.

Keenam jenjang berpikir terdapat pada ranah kognitif me pyramidal nurut taksonomi Bloom, jika diurutkan secara hierarki.seperti gambar dibawah ini:



Keenam jenjang berpikir

pada ranah kognitif ini juga bersifat kontinum dan overlap ( tumpang tindih ), dimana ranah yang lebih tinggi meliputi semua ranah yang ada dibawahnya. Seperti gambar dibawah ini:



Keterangan:

- A. Pengetahuan
- B. Pemahaman
- C. Penerapan
- D. Analisis
- E. Sintesis
- F. Evaluasi

Selanjutnya ranah afektif dijelaskan sebagai ranah yang berkaitan dengan sikap dan nilai. Taksonomi untuk daerah afektif mula-mula dikemukakan oleh David R. Krathwohl dkk (1974) dalam buku *Taxonomy of Educational Objectives : Affective Domain*. Ranah afektif adalah ranah yang berkaitan dengan sikap dan nilai. Ciri-

ciri belajar afektif akan tampak pada peserta didik dalam berbagai tingkah laku, seperti : perhatiannya terhadap mata pelajaran pendidikan agama Islam, kedisiplinannya dalam mengikuti pelajaran agama di sekolah, motivasinya yang tinggi untuk tahu lebih banyak mengenai pelajaran agama Islam yang diterimanya, penghargaan atau rasa hormatnya terhadap guru pendidikan agama, dsb.

Ranah afektif oleh Krathwohl (1974), dkk ditaksonomikan ke dalam lima jenjang yaitu.

### *1. Receiving*

Receiving atau attending adalah kepekaan seseorang dalam menerima rangsangan dari luar yang dating kepada dirinya dalam bentuk masalah, situasi, gejala dll.

Contohnya adalah peserta didik menyadari bahwa disiplin wajib ditegakkan sifat malas dan tidak berdisiplin harus disingkirkan jauh-jauh.

### *2. Responding*

Responding mengandung arti adanya partisipasi aktif, yaitu kemampuan yang dimiliki seseorang untuk mengikutsertakan dirinya secara aktif dalam fenomena tertentu dan membuat reaksi terhadapnya dengan salah satu cara.



Contohnya adalah peserta didik tumbuh hasratnya untuk mempelajari lebih jauh atau menggali lebih dalam lagi, ajaran-ajaran Islam tentang disiplin.

### *3. Valuing*

Valuing atau menilai artinya memberikan nilai atau memberikan penghargaan terhadap suatu kegiatan atau obyek. Peserta didik tidak hanya mau menerima nilai yang diajarkan tetapi mereka telah berkemampuan untuk menilai konsep, yaitu baik atau buruk.

Contohnya adalah tumbuhnya kemauan yang kuat pada diri peserta didik untuk berlaku disiplin, baik di sekolah di rumah maupun di tengah-tengah masyarakat.

### *4. Organization*

Organization artinya mempertemukan perbedaan nilai sehingga terbentuk nilai baru yang lebih universal, yang membawa kepada perbaikan umum.

Contohnya adalah peserta didik mendukung penegakan disiplin nasional.

### *5. Characterization by a value or value complex*

Characterization by a value or value complex atau karakterisasi dengan suatu nilai atau kompleks nilai, yaitu keterpaduan semua sistem nilai yang telah dimiliki seseorang, yang mempengaruhi pola kepribadian dan tingkah lakunya. Jadi pada jenjang ini peserta didik telah memiliki sistem nilai yang mengontrol tingkah lakunya

untuk suatu waktu yang cukup lama, sehingga membentuk karakteristik “pola hidup” tingkah lakunya menetap, konsisten dan dapat diramalkan.

Contohnya adalah siswa telah memiliki kebulatan sikap wujudnya peserta didik menjadikan perintah Allah sebagai pegangan hidupnya.

Ranah terakhir yang dibahas dalam sub-bab ini adalah ranah psikomotor yang merupakan ranah yang berkaitan dengan keterampilan (*skill*) atau kemampuan bertindak setelah seseorang tersebut menerima pengalaman belajar, jadi bisa dikatakan bahwa hasil belajar psikomotor merupakan kelanjutan dari hasil belajar kognitif (memahami sesuatu) dan hasil belajar afektif (kecenderungan dalam berperilaku).

Ranah Psikomotor adalah ranah yang berkaitan dengan keterampilan seseorang menerima pengalaman belajar tertentu. Menurut Simpson (1956) hasil belajar psikomotor ini tampak dalam bentuk keterampilan (*skill*) dan kemampuan bertindak individu. Wujud nyata dari hasil belajar psikomotor yang merupakan kelanjutan dari hasil belajar afektif dan kognitif adalah.

1. Peserta didik bertanya kepada guru tentang contoh kedisiplinan.
2. Peserta didik mencari dan membaca buku tentang kedisiplinan.

3. Peserta didik menjelaskan kepada temannya tentang pentingnya kedisiplinan.
4. Peserta didik menganjurkan temannya agar berlaku disiplin.
5. Peserta didik dapat memberikan contoh kedisiplinan di sekolah.
6. Peserta didik dapat memberikan contoh kedisiplinan di rumah.
7. Peserta didik dapat memberikan contoh kedisiplinan di masyarakat

#### **D. Langkah-langkah Pokok dalam Evaluasi Hasil Belajar**

Sub-bab keempat membahas mengenai langkah-langkah pokok dalam evaluasi hasil belajar, dimana dijelaskan enam langkah pokok, yaitu:

- a. Menyusun rencana hasil belajar, dimana didalamnya dimuat 6 jenis kegiatan, yaitu, merumuskan tujuan dilaksanakannya evaluasi, menetapkan aspek yang dievaluasi, memilih teknik yang dipergunakan, menyusun alat pengukuran hasil belajar, menentukan tolak ukur, dan menentukan frekuensi dari kegiatan evaluasi tersebut.

- b. Menghimpun data dimana dilakukan dengan pengukuran baik menggunakan tes ataupun instrument lainnya.
- c. Melakukan verifikasi data, proses penyaringan data yang layak digunakan sebagai penilaian dan tidak layak.
- d. Mengolah dan menganalisis data yang berguna untuk memberi makna untuk data yang telah berhasil dikumpulkan dalam kegiatan evaluasi dan data tersebut disusun sedemikian rupa agar bisa dipahami.
- e. Memberikan interpretasi dan menarik kesimpulan data yang telah diolah sebelumnya, dan kesimpulan harus sesuai dengan tujuan evaluasi tersebut.
- f. Tindak lanjut evaluasi yang dilakukan setelah sebelumnya telah dilakukan langkah-langkah sebelumnya, kemudian disini harus diketahui makna yang terkandung di dalamnya tersebut sehingga nantinya evaluator dapat memberi penilaian dan mengambil keputusan dan langkah apa yang dianggap perlu untuk kegiatan evaluasi tersebut.

### **BAB III**

## **TEKNIK PENYUSUNAN DAN PELAKSANAAN TES HASIL BELAJAR**

Secara kodrati semua individu terlahir secara unik. Unik berarti individu terlahir dengan kondisi yang berbeda-beda satu sama lainnya. Akibat adanya perbedaan yang mendasar antar manusia satu dengan yang lainnya, maka secara otomatis tugas seorang guru sebagai subyek evaluasi memerlukan bentuk alat ukur untuk mendiagnosis atau mengukur keadaan individu yang dimaksud (siswa). Alat ukur yang dimaksud adalah penggunaan tes. Dengan alat ukur berupa tes, maka orang akan berhasil memenuhi adanya perbedaan antar individu. Karena adanya aspek psikis yang berbeda-beda antar satu individu dengan individu yang lainnya, maka timbul pula bermacam-macam tes yang beragam pula. Berbagai tes yang dimaksud akan dijelaskan sebagai berikut.

### **A. Teknik Tes**

#### *a. Pengertian Tes*

Istilah tes diambil dari kata *testum* (Prancis) yang diartikan sebagai piring untuk menyisihkan logam-logam mulia. Ada pula yang menyebutkan sebagai sebuah piring yang terbuat dari tanah. Sementara itu istilah tes

pertama kali diperkenalkan oleh seorang ahli bernama James Ms. Cattell pada tahun 1890 kepada khalayak umum melalui bukunya yang berjudul "*Mental Test and Measurement*". Kemudian berkembang di Amerika yang selanjutnya secara berkesinambungan berkembang dengan tempo yang pesat sampai saat ini.

Pada buku karya Anas Sudijono, secara garis besar Anas menyebutkan bahwa tes didefinisikan sebagai alat atau prosedur yang dipergunakan dalam rangka pengukuran dan penilaian. Beberapa istilah lain mengenai tes seperti *testing*, *tester*, *testees* dan sebagainya memiliki definisi sendiri yang berbeda dengan konsep tes itu sendiri. Kalau dikaitkan dengan evaluasi pendidikan, tes adalah cara (yang dapat dipergunakan) atau prosedur (yang perlu ditempuh) dalam rangka pengukuran dan penilaian di bidang pendidikan, yang berbentuk pemberian tugas atau serangkaian tugas (baik berupa pertanyaan-pertanyaan (yang bisa di jawab), atau perintah-perintah (yang harus dikerjakan) sehingga dapat dihasilkan nilai yang melambangkan tingkah laku.

Arikunto menyebutkan sebelum sampai kepada uraian yang lebih jauh, dijelaskan terlebih dahulu dijabarkan definisi dari beberapa istilah terkait dengan tes, yaitu:

a. Tes

Tes merupakan alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu

dalam suasana, dengan cara dan aturan-aturan yang sudah ditentukan. Untuk mengerjakan tes tergantung dari petunjuk yang memberikan misalnya: melingkari salah satu huruf di depan pilihan jawaban, menerangkan, mencorat jawaban yang salah, melakukan tugas atau suruhan, menjawab secara lisan, dan sebagainya.

**b. Testing**

Testing merupakan saat pada waktu tes itu disampaikan atau dilaksanakan. Atau dapat disederhanakan dengan maksud bahwa testing adalah saat pengambilan tes.

**c. Testee**

Testee adalah responden yang sedang mengerjakan tes. Orang-orang inilah yang akan dinilai atau diukur, baik mengenai kemampuan, bakat, pencapaian, dan sebagainya.

**d. Tester**

Tester merupakan orang yang disertai untuk melaksanakan pengambilan tes terhadap para responden. Dengan kata lain tester adalah subjek evaluasi (tetapi adakalanya any orang yang ditunjuk oleh subjek evaluasi untuk melaksanakan tugasnya). Tugas tester antara lain:

1. Mempersiapkan ruangan dan perlengkapan yang diperlukan.
2. Membagikan lembaran tes dan alat-alat lain untuk mengerjakan tes.
3. Menerangkan cara mengerjakan tes.
4. Mengawasi responden mengerjakan tes.
5. Memberikan tanda-tanda waktu.
6. Mengumpulkan pekerjaan responden.
7. Mengisi berita acara atau laporan yang dilakukan (jika ada).

Dari beberapa istilah yang diatas, diharapkan akan mempermudah pemahaman dalam pelaksanaan bentuk evaluasi dalam kegiatan pembelajaran. Hal ini juga dapat mempermudah pemahaman pada bab berikutnya yang akan lebih diperkaya dengan istilah-istilah tersebut. Untuk itu, penting dipahaminya mengenai beberapa istilah yang dimaksudkan.

#### *a. Fungsi Tes*

Pada Bab ini, dibahas mengenai fungsi tes yang secara umu terdiri atas 2 (dua) macam, yaitu:

- A. Sebagai alat pengukur terhadap peserta didik. Hal yang diukur dalam hal ini ini berupa tingkat perkembangan atau kemajuan yang telah dicapai oleh peserta didik setelah menempuh proses belajar.
- B. Sebagai alat pengukur keberhasilan program pengajaran karena dapat diketahui sejauh mana



program pengajaran telah dicapai oleh peserta didik.

*b. Pengolahan Tes*

Sebagai sebuah alat ukur, tes digolongkan kedalam beberapa golongan besar yang dapat dijabarkan sebagai berikut.

1. Berdasarkan fungsinya sebagai alat pengukur perkembangan/kemajuan belajar peserta didik, tes dijabarkan menjadi 6 (enam) jenis yaitu tes seleksi (dilaksanakan dalam rangka penerimaan calon siswa baru); tes awal (mengetahui pengetahuan materi awal sebelum materi disampaikan); tes akhir (untuk mengetahui ketercapaian materi yang sudah diajarkan); tes diagnostik (untuk mengetahui kesukaran suatu materi pelajaran oleh peserta didik); tes formatif (mengetahui sejauh mana peserta didik sudah terbentuk dari materi yang diajarkan) dan tes sumatif (dilakukan setelah semua materi ajar selesai dilaksanakan).
2. Berdasarkan aspek psikologi diungkap digolongkan atas tes intelegensi (mengetahui tingkat kecerdasan siswa); tes kemampuan (mengungkap kemampuan dasar atau bakat khusus siswa); tes sikap (mengetahui kecenderungan respons terhadap lingkungan sekitar); tes kepribadian (mengungkap ciri khas yang bersifat lahiriah) dan tes hasil belajar (untuk mengungkap pencapaian atau prestasi belajar).

## **B. Teknik NonTes**

Pada bab terdahulu telah dikemukakan bahwa kegiatan “mengukur” atau “melakukan pengukuran” adalah merupakan kegiatan yang paling umum dilakukan dan merupakan tindakan yang mengawali kegiatan evaluasi dalam penilaian hasil belajar. Dengan teknik non tes dilakukan dengan tanpa “menguji” peserta didik, melainkan dilakukan dengan melakukan pengamatan secara sistematis (observation), melakukan wawancara (interview), menyebarkan angket (questionnaire, dan memeriksa atau meneliti dokumen-dokumen (documentary analysis).

### *a. Pengamatan (Observation)*

Observasi adalah metode pengumpulan data melalui pengamatan langsung atau peninjauan secara cermat dan langsung di lapangan atau lokasi penelitian. Dalam hal ini, peneliti dengan berpedoman kepada desain penelitiannya perlu mengunjungi lokasi penelitian untuk mengamati langsung berbagai hal atau kondisi yang ada di lapangan. Penemuan ilmu pengetahuan selalu dimulai dengan observasi dan kembali kepada observasi untuk membuktikan kebenaran ilmu pengetahuan tersebut.

Pengamatan didefinisikan sebagai suatu cara menghimpun bahan-bahan keterangan yang dilakukan dengan mengadakan pengamatan dan pencatatan secara sistematis terhadap fenomena-fenomena yang sedang dijadikan sasaran pengamatan. Diantara segi

kebaikan yang dimiliki oleh observasi itu misalnya data yang diperoleh secara langsung di lapangan, serta datanya mencakup berbagai aspek kepribadian masing-masing individu peserta didik. Adapun dari segi kelemahannya seperti observasi sebagai salah satu alat evaluasi hasil belajar tidak selalu dapat dilakukan dengan baik dan benar oleh para pengajar, kepribadian (personality) dari observer atau evaluator juga acapkali mewarnai atau menyelip masuk ke dalam penilaian yang dilakukan dengan cara observasi dan data yang diperoleh dari kegiatan observasi umumnya baru dapat mengungkap “kulit luar” nya saja.

Melalui observasi kita dapat memperoleh gambaran tentang kehidupan sosial yang sukar untuk diketahui dengan metode lainnya. Observasi dilakukan untuk menjajaki sehingga berfungsi eksploitasi. Dari hasil observasi kita akan memperoleh gambaran yang jelas tentang masalahnya dan mungkin petunjuk-petunjuk tentang cara pemecahannya. Jadi, jelas bahwa tujuan observasi adalah untuk memperoleh berbagai data konkret secara langsung di lapangan atau tempat penelitian.

Berdasarkan pelaksanaan, observasi dapat dibagi dalam dua jenis, yaitu observasi partisipasi dan observasi non partisipasi.

- a) Observasi partisipasi adalah observasi yang melibatkan peneliti atau observer secara langsung dalam kegiatan pengamatan di lapangan. Jadi, peneliti bertindak sebagai observer, artinya peneliti merupakan bagian dari kelompok yang ditelitinya. Keuntungan cara ini adalah peneliti merupakan bagian yang integral dari situasi yang dipelajarinya sehingga kehadirannya tidak memengaruhi situasi penelitian. Kelemahannya, yaitu ada kecenderungan peneliti terlampau terlibat dalam situasi itu sehingga prosedur yang berikutnya tidak mudah dicek kebenarannya oleh peneliti lain.
- b) Observasi non partisipasi adalah observasi yang dalam pelaksanaannya tidak melibatkan peneliti sebagai partisipasi atau kelompok yang diteliti. Cara ini banyak dilakukan pada saat ini. Kelemahan cara ini antara lain kehadiran pengamat dapat memengaruhi sikap dan perilaku orang yang diamatinya.

Layaknya jenis pengukuran yang lainnya, observasi memiliki langkah-langkah dalam pelaksanaannya, yaitu:

- a) Harus diketahui di mana observasi itu dapat dilakukan.
- b) Harus ditentukan dengan pasti siapa saja yang akan diobservasi.

- c) Harus diketahui dengan jelas data-data apa saja yang diperlukan.
- d) Harus diketahui bagaimana cara mengumpulkan data agar berjalan mudah dan lancar.
- e) Harus diketahui tentang cara mencatat hasil observasi, seperti telah menyediakan buku catatan, kamera, tape recorder, dan alat-alat tulis lainnya.

*b. Wawancara (interview)*

Secara umum yang dimaksud dengan wawancara adalah cara menghimpun bahan-bahan keterangan yang dilaksanakan dengan melakukan tanya jawab lisan secara sepihak, berhadapan muka, dan dengan arah serta tujuan yang telah ditentukan. Ada dua jenis wawancara yang dapat dipergunakan sebagai alat evaluasi yaitu wawancara terpimpin dan wawancara tidak terpimpin. Adapun kelebihan yang dimiliki oleh wawancara adalah, bahwa dengan melakukan wawancara, pewawancara sebagai evaluator (dalam hal ini guru, dosen dan lain-lain) dapat melakukan kontak langsung dengan peserta didik yang akan dinilai, sehingga dapat diperoleh hasil penilaian yang lebih lengkap dan mendalam.

*c. Angket (Questionnaire)*

Angket (questionnaire) juga dapat digunakan sebagai alat bantu dalam rangka penilaian hasil belajar. Berbeda dengan wawancara dimana penilai

(evaluator) berhadapan secara langsung (face of face) dengan peserta didik atau dengan pihak lainnya, maka dengan menggunakan angket, pengumpulan data sebagai bahan penilaian hasil belajar jauh lebih praktis, menghemat waktu dan tenaga.

*d. Pemeriksaan Dokumen (documentary Analysis)*

Evaluasi mengenai kemajuan, perkembangan atau keberhasilan belajar peserta didik tanpa menguji (teknik non tes) juga dapat dilengkapi atau diperkaya dengan cara melakukan pemeriksaan terhadap dokumen-dokumen, misalnya dokumen yang memuat informasi mengenai riwayat hidup (auto biografi). Dari uraian tersebut di atas dapatlah dipahami, bahwa dalam rangka evaluasi itu tidak harus semata-mata dilakukan dengan menggunakan alat berupa tes-tes hasil belajar. Teknik non tes juga menempati kedudukan yang penting dalam rangka evaluasi hasil belajar, lebih-lebih evaluasi yang berhubungan dengan kondisi kejiwaan peserta didik, seperti persepsinya terhadap mata pelajaran tertentu, persepsinya terhadap guru, minatnya, bakatnya, tingkah laku atau sikapnya dan sebagainya, yang kesemuanya itu tidak mungkin dievaluasi dengan menggunakan tes sebagai alat pengukurnya.

Kuesioner (questionnaire) juga sering dikenal sebagai angket (daftar pertanyaan). Pada dasarnya, kuesioner adalah sebuah daftar pertanyaan yang harus

diisi oleh responden (objek ukur). Ditinjau dari siapa yang menjawab, ada kuesioner langsung dan tidak langsung.

Kuesioner adalah suatu teknik pengumpulan informasi yang memungkinkan analis mempelajari sikap-sikap, keyakinan, perilaku, dan karakteristik beberapa orang utama di dalam organisasi yang bisa terpengaruh oleh sistem yang diajukan atau oleh sistem yang sudah ada.

Dengan menggunakan kuesioner, analis berupaya mengukur apa yang ditemukan dalam wawancara, selain itu juga untuk menentukan seberapa luas atau terbatasnya sentimen yang diekspresikan dalam suatu wawancara. Penggunaan kuesioner tepat bila:

1. Responden (orang yang merenpons atau menjawab pertanyaan) saling berjauhan.
2. Melibatkan sejumlah orang di dalam proyek sistem, dan berguna bila mengetahui berapa proporsi suatu kelompok tertentu yang menyetujui atau tidak menyetujui suatu fitur khusus dari sistem yang diajukan.
3. Melakukan studi untuk mengetahui sesuatu dan ingin mencari seluruh pendapat sebelum proyek sistem diberi petunjuk-petunjuk tertentu.
4. Ingin yakin bahwa masalah-masalah dalam sistem yang ada bisa diidentifikasi dan dibicarakan dalam wawancara tindak lanjut.

### **C. Ciri- ciri Tes Belajar yang Baik**

Ada 4 ciri-ciri test hasil belajar yang baik menurut Anas (2007) yaitu: valid, reliabel, obyektif, dan praktis. Valid dapat diartikan sebagai ketepatan, kebenaran, keshahihan. Reliabel yaitu keajegan, kemantapan. Obyektif maksudnya jika suatu tes itu menurut apa adanya dari sumber. Praktis artinya suatu tes tersebut harus mudah dilaksanakan karena test itu harus bersifat sederhana dan lengkap. Senada dengan itu Arikunto juga menunjukkan setidaknya ada 5 ciri-ciri tes hasil belajar yang baik, yaitu memiliki:

1. Validitas
2. Reliabilitas
3. Objektifitas
4. Praktikabilitas
5. Ekonomis

Kemudian selanjutnya akan dibahas secara lebih terperinci mengenai ciri-ciri yang dimaksud sebagai berikut.

#### **a. Validitas**

Pembicaraan mengenai evaluasi pada umumnya sering dikenal istilah "valid" untuk alat evaluasi atau instrumen evaluasi. Sebuah data atau informasi dapat dikatakan valid apabila sesuai dengan keadaan nyatanya. Sebagai contoh misalnya informasi tentang



seseorang bernama A menyebutkan bahwa si A pendek karena tingginya tidak lebih dari 140 sentimeter. Data tentang A ini dikatakan valid apabila memang sesuai dengan kenyataan, yakni bahwa tinggi si A kurang dari 140 sentimeter.

Jika data yang dihasilkan dari sebuah instrumen valid, maka dapat dikatakan bahwa instrumen tersebut valid, karena dapat memberikan gambaran tentang data secara benar sesuai dengan kenyataan atau keadaan sesungguhnya. Berkaitan dengan tes, sebuah tes dikatakan valid apabila tes itu dapat mengukur apa yang hendak diukur.

#### b. **Reliabilitas**

Kata reliabilitas dalam bahasa Indonesia diambil dari kata *reliability* dalam bahasa Inggris, berasal dari kata *reliable* yang artinya dapat dipercaya. Seperti halnya istilah validitas dan valid, kekacauan penggunaan "reliabilitas" sering muncul dan dikacaukan dengan istilah "reliabel". "Reliabilitas" merupakan kata benda, sedangkan "reliabel" merupakan kata sifat atau keadaan.

Seseorang dapat dikatakan dapat dipercaya jika orang tersebut selalu berbicara ajeg, tidak berubah-ubah pembicaraannya dari waktu ke waktu. Sama pula dengan sebuah tes dapat dikatakan dipercaya jika memberikan hasil yang tetap apabila diteskan berkali-kali. Sebuah tes dikatakan reliabel apabila

hasil-hasil tes tersebut menunjukkan ketepatan. Dengan kata lain, jika kepada siswa yang diberikan tes yang sama pada waktu yang berlainan, maka setiap siswa akan tetap berada dalam urutan (ranking) yang sama dalam kelompoknya.

Walaupun tampaknya hasil tes pada pengtesan kedua lebih baik, akan tetapi karena kenaikan dialami oleh siswa, maka tes yang digunakan dapat dikatakan memiliki reliabilitas yang tinggi. Kenaikan tes kedua barangkali disebabkan oleh adanya "pengalaman" yang diperoleh pada waktu mengerjakan tes pertama. Dalam keadaan seperti ini dikatakan bahwa ada *carry-over* atau *practice-effect*, yaitu adanya akibat yang dibawa karena siswa telah mengalami suatu kegiatan.

### **c. Objektifitas**

Pengertian sehari-hari sering ditemukan bahwa objektif berarti tidak ada unsur pribadi yang mempengaruhi. Sebuah tes dikatakan objektif apabila dalam melaksanakan tes itu tidak ada faktor subjektif yang mempengaruhi. Hal ini terutama terjadi pada sistem skoring yang akan dijelaskan pada bab berikutnya.

Apabila dikaitkan dengan reliabilitas maka objektifitas menekankan ketetapan (*consistency*) pada sistem skoring, sedangkan reliabilitas menekankan pada ketetapan dalam hasil tes.

#### **d. Praktikabilitas**

Sebuah tes dikatakan memiliki praktikabilitas yang tinggi apabila tes tersebut bersifat praktis, mudah pengadministrasiannya. Tes dikatakan praktis apabila tes tersebut:

1. Mudah dilaksanakan, misalnya tidak menuntut peralatan yang banyak dan memberikan kebebasan kepada siswa untuk mengerjakan terlebih dahulu bagian yang dianggap mudah oleh siswa.
2. Mudah pemeriksaannya, artinya bahwa tes itu dilengkapi dengan kunci jawaban; jawaban maupun pedoman skoringnya. Untuk soal bentuk objektif, pemeriksa akan lebih mudah dilakukan jika dikerjakan oleh siswa dalam lembar jawaban.
3. Dilengkapi dengan petunjuk-petunjuk yang jelas sehingga dapat diberikan/diawali oleh orang lain.

#### **e. Ekonomis**

Yang dimaksud ekonomis di sini ialah bahwa pelaksanaan tes tersebut tidak dibutuhkan ongkos/biaya yang mahal, tenaga yang banyak dan waktu yang lama.

## **D. Prinsip- prinsip Dasar dalam Penyusunan Tes Hasil Belajar**

Ada beberapa prinsip dasar yang perlu dicermati di dalam menyusun tes hasil belajar agar tes tersebut dapat mengukur tujuan instruksional khusus untuk mata pelajaran yang telah diajarkan, atau mengukur kemampuan dan keterampilan peserta didik yang diharapkan, setelah mereka menyelesaikan suatu unit pelajaran tertentu.

Pertama, tes hasil belajar harus dapat mengukur secara jelas hasil belajar (learning outcomes) yang telah ditetapkan sesuai dengan tujuan instruksional. Kejelasan mengenai pengukuran hasil belajar yang dikehendaki akan memudahkan bagi guru dalam menyusun butir-butir soal tes hasil belajar.

Kedua, butir-butir soal tes hasil belajar harus merupakan sampel yang representatif dari populasi bahan pelajaran yang telah diajarkan, sehingga dapat dianggap mewakili seluruh performance yang telah diperoleh selama peserta didik mengikuti suatu unit pelajaran.

Ketiga, bentuk soal yang dikeluarkan dalam tes hasil belajar harus dibuat bervariasi, sehingga betul-betul cocok untuk mengukur hasil belajar yang diinginkan sesuai dengan tujuan tes itu sendiri. Untuk mengukur hasil belajar yang berupa keterampilan misalnya, tidak tepat kalau hanya menggunakan soal-soal yang

berbentuk essay test yang jawabannya hanya menguraikan dan bukan melakukan atau mempraktekan sesuatu. Demikian pula untuk mengukur kemampuan menganalisis suatu prinsip, tidak cocok jika digunakan butir-butir soal yang berbentuk objective test yang pada dasarnya hanya mengungkap daya ingat peserta didik.

Keempat, tes hasil belajar harus didesain sesuai dengan kegunaannya untuk memperoleh hasil yang diinginkan. Pernyataan tersebut mengandung makna, bahwa desain tes hasil belajar harus disusun relevan dengan kegunaan yang dimiliki oleh masing-masing jenis tes. Desain dari placement test (tes yang digunakan untuk untuk penentuan penempatan siswa dalam suatu jenjang atau jenis program pendidikan tertentu) sudah barang tentu akan berbeda dengan desain dari formative test (tes yang digunakan untuk mencari umpan balik guna memperbaiki proses pembelajaran, baik bagi guru maupun bagi siswa) dan summatif test (tes yang digunakan untuk mengukur atau menilai sampai dimana pencapaian siswa terhadap bahan pelajaran yang telah diajarkan dan selanjutnya untuk menentukan kenaikan tingkat atau kelulusan siswa yang bersangkutan). Demikian pula desain dari diagnostic test (tes yang digunakan dengan tujuan untuk mencari sebab-sebab kesulitan belajar siswa, seperti latar belakang psikologis, fisik, dan lingkungan sosial ekonomi siswa) tentu akan berbeda pula dengan tiga jenis tes yang telah disebutkan sebelumnya.

Kelima, tes hasil belajar harus memiliki reliabilitas yang dapat diandalkan. Artinya setelah tes hasil belajar itu dilaksanakan berkali-kali terhadap subyek yang sama, hasilnya selalu sama atau relatif sama. Dengan demikian tes hasil belajar itu hendaknya memiliki keajegan hasil pengukuran yang tidak diragukan lagi.

Keenam, tes hasil belajar disamping harus dapat dijadikan alat untuk mencari informasi yang berguna untuk memperbaiki cara belajar siswa dan cara mengajar guru itu sendiri.

Ada beberapa prinsip dasar dalam penyusunan test hasil belajar.

1. Tes hasil belajar harus dapat mengukur secara jelas hasil belajar (learning outcomes).
2. Butir-butir soal tes hasil belajar harus merupakan sampel yang representatif dari populasi pelajaran yang telah disampaikan.
3. Bentuk soal yang dibuat dalam test hasil belajar harus didesain secara bervariasi sehingga betul-betul cocok untuk mengukur hasil belajar yang diinginkan sesuai dengan tujuan test itu sendiri.
4. Test hasil belajar harus didesain sesuai dengan kegunaannya untuk mendapatkan hasil yang diinginkan.
5. Tes hasil belajar harus memiliki reliabilitas (keajegan) yang andal. Artinya bahwa setelah tes diberikan berkali-kali terhadap subjek yang sama, hasilnya selalu sama atau relatif sama.

6. Tes hasil belajar disamping harus dapat dijadikan alat ukur keberhasilan belajar peserta didik, juga harus dapat dijadikan alat untuk mencari informasi yang berguna memperbaiki cara belajar dan cara mengajar guru.

## **E. Bentuk-bentuk Tes Hasil Belajar dan Teknik Penyusunannya**

### **1. Tes hasil belajar bentuk uraian**

Tes uraian memiliki karakteristik: 1, berbentuk pertanyaan atau perintah yang menghendaki jawaban berupa uraian/paparan kalimat panjang. 2, bentuk pertanyaan menuntut testee untuk memberikan penjelasan/komentar. 3, jumlah soalnya terbatas. 4, umumnya di awal dengan kata jelaskan, mengapa, bagaimana, uraikan.

#### **a. Penggolongan tes uraian**

- 1) Tes uraian bentuk bebas/terbuka, yaitu tes yang menghendaki jawaban dari testee sepenuhnya.
- 2) Tes uraian bentuk terbatas yaitu, tes yang menghendaki jawaban yang sudah terarah.

Tes uraian dipergunakan apabila pembuat soal menghendaki seberapa jauh pemahaman testee pada materi tertentu.

b. Kelebihan dan kekurangan

Kelebihan: 1) Tes yang dapat dibuat dengan cepat dan mudah, 2) dapat di cegah spekulasi pada testee, 3) dapat mengetahui pemahaman testee, 4) testee akan terdorong untuk berani mengemukakan pendapat

Kekurangan: 1) kurang bias mencakup isi materi keseluruhan, 2) cara mengoreksi jawaban sulit, 3) tester cenderung subyektif, 4) pekerjaan koreksi sulit diserahkan kepada orang lain, 5) daya mengukur dan keajegan cenderung rendah.

c. Petunjuk operasional penyusunan tes uraian

- 1) Diusahakan agar butir-butir soal tes uraian dapat mencakup materi yang telah di ajarkan
- 2) Untuk menghindari kecurangan, susunan alimatsoal dibuat berlainan dengan kalimat di buku.
- 3) Setelah membuat tes, hendaknya dirumuskan dengan tegas
- 4) Jangan membuat dengan perintah seragam
- 5) Kalimat soal hendaknya disingkat secara ringkas



- 6) Hendaknya di kemukakan pedoman dalam menjawab tes
2. Tes hasil belajar bentuk objektif (objective test)

Tes objektif merupakan dengan jawaban pendek yang terdiri dari butir-butir soal yang dapat dijawab oleh testee dengan jalan memilih salah satu atau lebih diantara beberapa kemungkinan jawaban pada masing-masing item, atau menuliskan jawaban berupa kata maupun symbol pada tempat yang telah disediakan

  - a. Penggolongan tes objektif
    - 1) Tes Obyektif bentuk benar salah
    - 2) Tes obyektif bentuk menjodohkan
    - 3) Tes obyektif bentuk melengkapi
    - 4) Tes obyektif bentuk isian
    - 5) Tes obyektif bentuk pilihan ganda

#### F. Teknik Pelaksanaan Tes Hasil Belajar

##### 1. Teknik pelaksanaan tes tertulis

Dalam mengerjakan tes tertulis ada beberapa hal yang perlu diperhatikan:

- a. Agar dalam mengerjakan soal tes para peserta mendapat ketenangan

- b. Ruangan tes cukup longgar
  - c. Ruangan tes sebaiknya memiliki sistem pencahayaan dan pertukaran udara yang baik
  - d. Jika tidak tersedia meja tulis atau kursi, hendaknya sudah disiapkan alas tulis pengganti
  - e. Testee mengerjakan soal secara bersamaan
  - f. Dalam mengawasi testee, hendaknya jangan berperilaku wajar
  - g. Sebelum tes, hendaknya ditentukan sanksi bagi testee yang curang
  - h. Sebagai bukti mengikuti tes, disiapkan daftar hadir
  - i. Jika waktu habis, testee dipersilakan mengumpulkan jawaban dan meninggalkan ruangan
  - j. Untuk mencegah terjadinya kesulitan dikemudian hari, hendaknya ditulis keadaan testee di berita acara.
2. Teknik pelaksanaan tes lisan
- a. Sebelum tes lisan dilaksanakan, seyogyanya tester sudah melakukan inventarisasi berbagai jenis soal yang akan diajukan kepada testee dalam tes lisan tersebut

- b. Setiap butir soal yang telah ditetapkan untuk diajukan dalam tes lisan itu, juga harus disiapkan sekaligus pedoman atau ancar-ancar jawaban betulnya.
- c. Jangan sekali-kali menentukan skor atau nilai hasil tes lisan setelah seluruh testee menjalani tes lisan.
- d. Tes hasil belajar yang dilaksanakan secara lisan hendaknya jangan sampai menyimpang atau berubah arah dari evaluasi menjadi diskusi.
- e. Dalam rangka menegakkan prinsip obyektivitas dan prinsip keadilan, dalam tes yang dilaksanakan secara lisan itu, tester hendaknya jangan sekali-kali "memberikan angin segar" atau "memancing-mancing" dengan kata-kata, kalimat-kalimat atau kode-kode tertentu yang sifatnya menolong testee
- f. Tes lisan harus berlangsung secara wajar. Pernyataan tersebut mengandung makna bahwa tes lisan itu jangan sampai menimbulkan rasa takut, gugup atau panik di kalangan testee.
- g. Sekalipun acapkali sulit untuk dapat diwujudkan, namun sebaiknya tester mempunyai pedoman atau ancar-ancar yang

pasti, berapa lama atau berapa waktu yang disediakan bagi tiap peserta tes dalam menjawab soal-soal atau pertanyaan-pertanyaan pada tes lisan tersebut.

- h. Pertanyaan-pertanyaan yang diajukan dalam tes lisan hendaknya dibuat bervariasi, dalam arti bahwa sekalipun inti persoalan yang ditanyakan itu sama, namun cara pengajuan pertanyaannya dibuat berlainan atau beragam.
- i. Sejauh mungkin dapat diusahakan agar tes lisan itu berlangsung secara individual (satu demi satu).

### 3. Teknik Pelaksanaan Tes Perbuatan

Tes perbuatan pada umumnya digunakan untuk mengukur taraf kompetensi yang bersifat keterampilan (psikomotorik), dimana penilaiannya dilakukan terhadap proses penyelesaian tugas dan hasil akhir yang dicapai oleh testee setelah melaksanakan tugas tersebut. Tes ini hendaknya dilakukan secara individual.

Dalam melaksanakan tes perbuatan itu, ada beberapa hal yang perlu diperhatikan oleh tester, yaitu:

- a. Tester harus mengamati dengan secara teliti, cara yang ditempuh oleh testee dalam menyelesaikan tugas yang telah ditentukan
- b. Agar dapat dicapai kadar obyektivitas setinggi mungkin, hendaknya tester jangan berbicara atau berbuat sesuatu yang dapat mempengaruhi testee yang sedang mengerjakan tugas tersebut.

Dalam mengamati testee yang sedang melaksanakan tugas itu, hendaknya tester telah menyiapkan instrumen berupa lembar penilaian yang di dalamnya telah ditentukan hal-hal apa sajakah yang harus diamati dan diberikan penilaian.

## **BAB IV**

### **TEKNIK PENGUJIAN VALIDITAS TES DAN VALIDITAS ITEM TES HASIL BELAJAR**

Setelah mempelajari teknik penyusunan dan pengolahan evaluasi hasil belajar siswa, perlu juga dipelajari tentang pengujian tes apakah valid atau tidak. Bab 5 ini menjelaskan tentang teknik pengajuan validitas tes dan validitas item tes hasil belajar.

Sebelum mulai dengan penjelasan perlu kiranya dipahami terlebih dahulu perbedaan arti istilah "validitas" dengan "valid". Validitas merupakan sebuah kata benda, sedangkan valid merupakan kata sifat. Dari pengalaman sehari-hari tidak sedikit siswa atau guru mengatakan: "Tes ini baik karena sudah validitas", jelas kalimat tersebut tidak tepat. Yang benar adalah: "Tes ini sudah baik karena sudah valid" atau "Tes ini baik karena memiliki validitas yang tinggi".

Dalam pembicaraan evaluasi pada umumnya orang banyak mengenal istilah "valid" untuk alat evaluasi atau instrument evaluasi. Hingga saat ini belum banyak buku yang menerapkan istilah "valid" untuk data. Dalam buku ini dicoba menjelaskan asal pengertian valid untuk instrument dimulai dari pengertian "Valid" untuk data.

Sebuah data atau informasi dapat dikatakan valid apabila sesuai dengan data yang sebenarnya. Sebagai contoh, informasi tentang seorang bernama A menyebutkan bahwa si A pendek karena tingginya tidak lebih dari 140 sentimeter. Data tentang A ini dikatakan valid apabila memang sesuai dengan kenyataan, yakni bahwa tinggi A kurang dari 140 sentimeter. Contoh lain data B yang diperoleh dari cerita orang lain menunjukkan bahwa ia pembohong. Bukti bahwa si B pembohong diperoleh dari kenyataan bahwa si B sering berbicara tidak benar, tidak sesuai dengan kenyataan. Dengan demikian maka data tentang B tersebut valid dan cerita orang tersebut benar.

Jika data yang dihasilkan dari sebuah instrument valid, maka dapat dikatakan bahwa instrument tersebut valid, karena dapat memberikan gambaran tentang data secara benar sesuai dengan kenyataan atau keadaan sesungguhnya. Dari sedikit uraian dan contoh tersebut dapat disimpulkan bahwa:

Jika data yang dihasilkan oleh instrument benar dan valid, sesuai kenyataan, maka instrument yang digunakan tersebut juga valid.

Sebuah tes disebut valid apabila tes itu dapat tepat mengukur apa yang hendak diukur. Istilah "valid", sangat sukar dicari gantinya. Ada istilah baru yang mulai diperkenalkan, yaitu sah sehingga validitas diganti menjadi kesahihan. Walaupun istilah "tepat" belum dapat

mencangkup semua arti yang tersirat dalam kata “valid”, dan kata “tepat” kadang-kadang digunakan dalam konteks yang lain, akan tetapi tambahan kata “tepat” dalam menerapkan kata “valid” dapat memperjelas apa yang dimaksud.

## **A. Teknik Pengujian Validitas Tes Hasil Belajar**

Teknik pengujian validitas tes hasil belajar terdiri atas dua, yaitu pengujian secara rasional dan pengujian secara empirik.

### **1. Pengujian Validitas Tes Secara Rasional**

Pengujian validitas tes hasil belajar secara rasional merupakan sebuah tes dimana ketika dilakukan penganalisan secara rasional ternyata memiliki daya ketepatan dalam pengukuran. Validitas rasional diperoleh atas dasar pemikiran yang kritis dan berpikir secara logis. Validitas rasional adalah validitas yang diperoleh atas dasar hasil pemikiran, validitas yang diperoleh dengan berfikir secara logis. Dengan demikian maka suatu tes hasil belajar dapat dikatakan telah memiliki validitas rasional, apabila setelah dilakukan penganalisan secara rasional ternyata bahwa tes hasil belajar memang (secara rasional) dengan tepat telah dapat mengukur apa yang seharusnya diukur. Validitas rasional dapat



ditelusuri dari dua segi yaitu dari segi isinya (*content*) dan susunannya (*construct*).

#### a. Validitas isi

Validitas isi merupakan validitas yang diperoleh setelah penganalisisan terhadap isi yang terkandung dalam tes hasil belajar. Validitas isi sering disebut validitas kurikuler karena materi yang diajarkan pada umumnya tertuang pada GBPP. Validitas isi dapat diketahui dengan membandingkan isi yang terkandung dalam tes hasil belajar dengan tujuan instruksional khusus yang telah ditentukan. Selain membandingkan, validitas isi juga dapat diketahui dengan cara melaksanakan diskusi panel. Para pakar diminta untuk memberikan pendapat dan rekomendasinya terhadap isi tes hasil belajar. Penganalisisan validitas isi dapat dilakukan sebelum maupun sesudah tes dilaksanakan.

#### b. Validitas Konstruksi

Sebuah tes dikatakan memiliki validitas konstruksi apabila butir-butir soal yang membangun tes tersebut mengukur setiap aspek berfikir seperti yang disebutkan oleh Tujuan Instruksional Khusus. Dengan kata lain jika butir-butir soal mengukur aspek berfikir tersebut sudah

sesuai dengan aspek berfikir yang menjadi tujuan intruksional.

Sebagai contoh jika rumusan Tujuan Intruksional Khusus (TIK) "siswa dapat membandingkan antara efek biologis dan efek psikologis", maka butir soal pada tes merupakan perintah agar siswa membedakan antara dua efek tersebut. Sekarang TIK dikenal dengan Indikator.

"Konstruksi" dalam pengertian ini bukanlah "susunan" seperti yang sering dijumpai dalam teknik, tetapi merupakan rekaan yang dibuat oleh para ahli Ilmu Jiwa yang dengan suatu cara tertentu "memperinci" isi jiwa atas beberapa aspek seperti: Ingatan (pengetahuan), pemahaman, aplikasi, dan seterusnya. Dalam hal ini, mereka menganggap seolah-olah jiwa dapat dibagi-bagi, tetapi sebenarnya tidak demikian. Pembagian ini hanya merupakan tindakan sementara untuk mempermudah mempelajari.

Seperti halnya validitas isi, validitas konstruksi dapat diketahui dengan cara memperinci dan memasang setiap butir soal dengan setiap aspek dalam indikator. Pengerjaannya dilakukannberdasarkan logika, bukan pengalaman. Dalam pembicaraan mengenai penyusunan tes hal ini akan disinggng lagi

Validitas konstruksi merupakan validitas yang diperoleh dengan cara melihat tes hasil belajar dari segi susunan atau kerangkanya. Tes dikatakan valid apabila tes tersebut sudah mencerminkan suatu konstruksi dalam

teori psikologis. Konstruksi dalam teori psikologis (penjelasan dari beberapa ahli psikologis) meliputi tiga ranah yaitu kognitif, afektif, dan psikomotor. Tes akan dianggap valid apabila butir-butir soal yang membangun tes tersebut telah minimal dapat mengukur ketiga ranah tersebut. Validitas konstruksi juga dapat diketahui dengan menyelenggarakan diskusi panel. Validitas isi dapat dilakukan sebelum atau sesudah pelaksanaan tes.

## 2. Pengujian Validitas Tes Secara Empiris

Istilah “validitas empiris” memuat kata “empiris” yang artinya “pengalaman”. Sebuah instrument dapat dikatakan memiliki validitas empiris apabila sudah diuji dari pengalaman. Sebagai contoh sehari-hari seseorang dapat diakui jujur oleh masyarakat apabila dalam pengalaman dibuktikan bahwa orang tersebut memang jujur. Pengujian validitas tes hasil belajar secara empiris merupakan sebuah tes dimana validitas tes didapatkan atau bersumber dari pengamatan langsung di lapangan. Tes hasil belajar dikatakan memiliki validitas empiris apabila analisis terhadap data yang didapatkan dengan pengamatan di lapangan dapat diukur tanpa adanya pelaksanaan tes kembali. Validitas tes secara empiris dapat ditelusuri dari dua segi, yaitu dari segi daya ketepatan ramalan dan daya ketepatan bandingan.

#### a. Validitas Ramalan

Validitas ramalan dari sebuah tes merupakan suatu kondisi yang menunjukkan sejauh mana sebuah tes dapat menunjukkan kemampuannya untuk memberikan ramalan terhadap kondisi yang akan terjadi pada masa yang akan datang. Misalnya, tes masuk Perguruan Tinggi adalah sebuah tes yang diperkirakan mampu meramalkan keberhasilan peserta tes dalam mengikuti kuliah dimasa yang akan datang. Calon yang tersaring berdasarkan hasil tes diharapkan mencerminkan tinggi rendahnya kemampuan mengikuti kuliah. Jika nilai tesnya tinggi, tentu menjamin keberhasilannya kelak. Sebaliknya, seorang calon dikatakan tidak lulus tes karena memiliki nilai tes yang rendah jadi diperkirakan akan tidak mampu mengikuti perkuliahan yang akan datang.

Sebuah tes dikatakan memiliki validitas ramalan apabila terdapat korelasi yang positif dan signifikan antara tes hasil belajar yang sedang diuji dengan kriterium yang ada. Korelasi tersebut dapat diketahui dengan menggunakan teknik analisis korelasi *product moment* dari Karl Pearson.

#### b. Validitas Bandingan

Validitas bandingan merupakan suatu cara yang dilakukan untuk mendapatkan validitas tes hasil belajar

dengan cara membandingkan. Sebuah tes dapat dikatakan memiliki validitas bandingan apabila tes tersebut dalam kurun waktu yang sama dapat menunjukkan hubungan searah antara tes pertama dengan tes berikutnya. Validitas bandingan ditinjau dari pengukur lain yang dijadikan kriterium. Untuk mengetahui ada tidaknya hubungan yang searah antara tes pertama dengan tes berikutnya dapat digunakan teknik analisis korelasi *product moment* dari Karl Pearson. Yang dimaksud searah adalah jika tes tersebut bernilai positif dan signifikan.

Validitas hasil belajar tidak terlepas dari berbagai pengaruh. Salah satu yang berpengaruh dalam kevaliditasan dari tes hasil belajar adalah dari item tes hasil belajar itu sendiri. Validitas item adalah ketepatan mengukur yang dimiliki oleh sebutir item dalam sebuah tes hasil belajar. Setiam butir item merupakan bagian yang tak terpisahkan dari tes hasil belajar (suatu totalitas). Semakin banyak item yang dapat dijawab, maka skor tes hasil belajar semakin tinggi, begitu pula sebaliknya. Validitas tes akan sangat dipengaruhi oleh validitas item. Validitas dari masing-masing butir item yang membangun sebuah tes, akan dapat diketahui besar kecilnya dukungan yang diberikan oleh masing-masing butir item terhadap tes hasil belajar.

## B. Cara Mengetahui Validitas Alat Ukur

Tes dikatakan memiliki validitas jika hasilnya sesuai dengan kriteria, dalam artian memiliki kesejajaran antara hasil tes tersebut dengan kriteria. Teknik yang digunakan untuk mengetahui kesejajaran adalah teknik korelasi *product moment* yang dikemukakan oleh Pearson.

Rumus korelasi *product moment* ada 2 (dua) macam, yaitu:

- a. Korelasi *Product moment* dengan simpangan, dan
- b. Korelasi *Product moment* dengan angka kasar.

### Rumus korelasi *product moment* dengan simpangan:

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

di mana:

$r_{xy}$  = koefisien korelasi antara variabel  $X$  dan variabel  $Y$ , dua variabel yang dikorelasikan ( $x = X - \bar{X}$  dan  $y = Y - \bar{Y}$ ).

$\sum xy$  = jumlah perkalian  $x$  dengan  $y$

$\sum x^2$  = kuadrat dari  $x$

$\sum y^2$  = kuadrat dari  $y$

**Rumus korelasi *product moment* dengan angka kasar:**

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

$r_{xy}$  = koefisien korelasi antara variabel X dan Variabel Y, dua variabel yang dikorelasikan

Table 01. Skor Skala Sikap terhadap Mata Kuliah Statistik

Testee	Nomor Butir Soal										Total Skor
	1	2	3	4	5	6	7	8	9	10	
A	2	5	4	3	4	5	4	4	5	3	39
B	4	5	5	3	4	4	3	5	5	4	42
C	2	3	4	2	3	3	4	3	4	4	32
D	2	3	4	3	3	3	3	3	3	3	30
E	3	4	5	3	4	4	3	4	3	4	37
F	1	2	3	2	1	2	1	2	3	2	19
G	1	3	4	2	1	3	2	2	3	2	23

Tabel 02. Cara Menghitung Skor Butir 6 dengan Skor Total

Responden	X	Y	XY	X <sup>2</sup>	Y <sup>2</sup>
A	5	39	195	25	1521
B	4	42	168	16	1764
C	3	32	96	9	1024
D	3	30	90	9	900
E	4	37	148	16	1369
F	2	19	38	4	361
G	3	23	69	9	529
TOTAL	24	222	804	88	7468

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

$$r_{xy} = \frac{7.804 - (24)(222)}{\sqrt{[7.88 - (24)^2][7.7468 - (222)^2]}}$$

$$r_{xy} = \frac{5628 - 5328}{\sqrt{[616 - (576)][52276 - (49284)]}}$$

$$r_{xy} = \frac{300}{\sqrt{[40][2992]}}$$

$$r_{xy} = \frac{300}{\sqrt{[119680]}}$$

$$r_{xy} = \frac{300}{345,948}$$

$$r_{xy} = 0,867$$



Jadi korelasi skor butir 6 dengan skor total adalah  
0,867

### **C. Teknik Pengujian Validitas Item Tes Hasil Belajar**

Apa yang sudah dibicarakan di atas adalah validitas soal secara keseluruhan tes. Disamping mencari validitas soal perlu juga dicari validitas item. Jika seorang peneliti atau seorang guru mengetahui bahwa validitas soal tes misalnya terlalu rendah dan rendah saja, maka selanjutnya ingin mengetahui butir-butir tes manakah yang menyebabkan soal secara keseluruhan tersebut jelek karena memiliki validitas rendah. Untuk kepentingan inilah dicari validitas butir soal.

Pengertian umum untuk validitas item adalah demikian sebuah item dikatakan valid apabila mempunyai dukungan yang besar terhadap skor total. Skor pada item menyebabkan skor total menjadi tinggi atau rendah. Dengan kata lain dapat dikemukakan di sini bahwa item memiliki validitas yang tinggi jika skor pada item mempunyai kesejajaran dengan skor total. Kesejajaran ini dapat diartikan dengan kolerasi sehingga untuk mengetahui validitas item digunakan rumus kiolerasi seperti sudah diterangkan di atas.

Sebutir item dapat dikatakan valid apabila skor-skor pada butir item yang bersangkutan memiliki kesesuaian dengan skor totalnya. Skor total merupakan variabel terikat dan skor item merupakan variabel bebas. Teknik analisis yang digunakan untuk menguji sebuah item valid atau tidak yaitu teknik korelasi. Sebuah item dapat dinyatakan valid apabila memiliki korelasi positif dan signifikan dengan skor totalnya.

### 1. Pengertian validitas item

Validitas item dari sebuah tes adalah ketepatan mengukur yang dimiliki oleh sebuah item (yang merupakan bagian yang tak terpisahkan dari tes bagian suatu totalitas), dalam mengukur apa yang seharusnya diukur lewat butir item tersebut.

Apabila kita mau memperhatikan secara cermat, maka tes-tes hasil belajar yang dibuat akan disusun oleh para pengajar, baik guru, dosen staf pengajar lainnya, sebenarnya adalah merupakan kumpulan dari sekian banyak butir-butir item; dengan item mana para penyusun tes ingin mengukur atau mengungkap hasil belajar yang telah dicapai oleh masing-masing individu peserta didik, setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu. Pernyataan itu mengandung makna, bahwa sebenarnya setiap butir item yang ada dalam tes hasil belajar itu adalah

merupakan bagian tak terpisahkan dari tes hasil belajar tersebut sebagai totalitas.

Erat hubungannya antara butir item dengan tes hasil belajar sebagai suatu totalitas itu kiranya dapat dipahami dari kenyataan, bahwa semakin banyak butir-butir item yang dapat dijawab dengan betul oleh testee, maka skor-skor total hasil tes tersebut akan semakin tinggi. Sebaliknya, semakin sedikit butir-butir item yang dapat dijawab dengan betul oleh testee, maka skor-skor total hasil tes itu akan semakin rendah atau semakin menurun.

## 2. Teknik Pengujian Validitas Item Tes Hasil Belajar

Berdasarkan uraian diatas maka cukup jelas bahwa sebutir item dapat dikatakan telah memiliki validitas tinggi atau dapat dinyatakan valid, jika skor-skor pada butir item yang bersangkutan memiliki kesesuaian atau kesejajaran arah dengan skor totalnya, skor total disini berkedudukan sebagai variable terikat sedangkan variable item merupakan variable bebasnya. Kalau demikian, maka untuk sampai pada kesimpulan bahwa item-item yang ingin diketahui validitasnya, yaitu valid ataukah tidak, kita dapat menggunakan teknik kolerasi sebagai teknik analisisnya. Sebutir item dapat dinyatakan valid apabila skor item yang bersangkutan terbukti

mempunyai kolerasi positif yang signifikan dengan skor totalnya.

Permasalahannya adalah bagaimana Memilih dan menentukan jenis tehnik dalam rangka menguji validitas item itu. Seperti yang diketahui pada tes objektif maka hanya ada dua kemungkinan yaitu betul atau salah.

Setiap butir soal yang dapat dijawab dengan benar diberikan skor 1 ( satu ) sedangkan untuk setiap jawaban yang salah diberikan skor 0 ( nol ) jenis data seperti ini biasanya merupakan tes benar – salah, ya – tidak dan sejenisnya dalam ilmu statistic dikenal dengan diskret murni atau data dikotomik. Sedangkan, skor total yang dimiliki oleh masing-masing testee adalah merupakan penjumlahan dari setiap skor itu merupakan data kontinyu.

Berdasarkan teori yang ada apabila variable I berupa data dikotomik sedangkan variable II data kontinyu maka, teknik korelasi yang tepat untuk digunakan dalam mencari korelasi dua variable adalah teknik korelasi point biserial, diman angka indeks korelasi diberi lambang  $r_{pbi}$  dapat diperoleh dengan rumus :

$$R_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

Dimana :

- $R_{pbi}$  = koefisien korelasi point berserial yang melambangkan kekuatan korelasi antara variable I dan II yang dalam hal ini sebagai koefisien validitas item.
- $M_p$  = skor rata-rata hitung yang dimiliki oleh testee.
- $M_t$  = skor rata-rata dari skor total.
- $SD_t$  = standar deviasi dari skor total.
- $p$  = proporsi testee yang menjawab denagn benar terhadap butir item yang diuji validitasnya..
- $q$  = proporsi testee yang menjawab salah terhadap butir item yang diuji validitasnya

Table 03. Cara menghitung Validitas Butir Instrumen dengan Korelasi Point Biserial

Testee	Nomor Butir Soal										Skor Total
	1	2	3	4	5	6	7	8	9	10	
A	1	0	1	0	1	1	1	1	1	1	8
B	0	0	1	0	1	0	0	1	1	1	5
C	0	1	0	0	0	1	0	1	0	1	4

D	1	1	0	0	1	1	0	0	1	0	5
F	1	1	1	1	1	1	0	0	0	0	6
F	1	0	1	0	1	0	1	0	0	0	4
G	1	1	1	1	1	1	1	0	0	0	7
H	0	1	0	1	1	1	1	1	1	1	8
I	0	0	0	1	1	1	0	1	1	1	6
J	1	1	0	1	1	1	0	1	1	0	7

Misalnya diuji Validitas Butir Soal no 6, maka perhitungannya sebagai berikut

$$1) M_p = (8+4+5+6+7+8+6+7) : 8 = 51 : 8 = 6,375$$

$$2) M_t = (8 + 5 + 4 + 5 + 6 + 4 + 7 + 8 + 6 + 7) : 10 = 60 : 10 = 6$$

$$3) S_i = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} = \sqrt{\frac{(10 \times 380) - (60)^2}{10(10-1)}} = \sqrt{\frac{(3800) - (3600)}{10(9)}} = \sqrt{\frac{200}{90}} = \sqrt{2,222} = 1,491$$

$$4) p = 8 : 10 = 0,8$$

$$5) q = 2 : 10 = 0,2 \text{ atau } (1 - 0,8) = 0,2$$

6) Masukkan ke dalam rumus:

$$\begin{aligned} r_{\text{pbi}} &= \frac{M_P - M_t}{S_t} \sqrt{\frac{p}{q}} \\ &= \frac{6,375 - 6}{1,491} \sqrt{\frac{0,8}{0,2}} \\ &= \frac{0,375}{1,491} \sqrt{4} \\ &= 0,252 \times 2 \\ &= 0,504 \end{aligned}$$

Jadi validitas butir soal no 6 dengan rumus Korelasi Point Biserial adalah 0, 504

## **BAB V**

### **TEKNIK PENGUJIAN RELIABILITAS TES HASIL BELAJAR**

Kata reliabilitas dalam bahasa Indonesia diambil dari kata *reliability* dalam bahasa Inggris yang berasal dari kata asal *reliable* yang memiliki arti dapat dipercaya. Seperti halnya istilah validitas dan valid, kekacauan dalam penggunaan istilah "reliabilitas" sering dikacaukan dengan istilah "reliable". "Reliabilitas" merupakan kata benda. Sedangkan "reliable" merupakan kata sifat atau kata keadaan.

Sudah diterangkan dalam persyaratan tes, bahwa tes yang baik adalah tes yang memiliki taraf reliabilitas tinggi. Suatu tes dapat dikatakan memiliki taraf reliabilitas yang tinggi jika tes tersebut diujikan dalam kurun waktu yang berbeda dapat memberikan hasil yang tetap. Maka pengertian reliabilitas tes berhubungan dengan masalah ketetapan hasil tes.

konsep reliabilitas ini tidak akan sulit dimengerti apabila telah memahami konsep validitas. Tuntutan bahwa instrumen evaluasi harus valid menyangkut harapan diperolehnya data yang valid, sesuai dengan kenyataan. Sedangkan reliabilitasuntutannya tidak jauh



berbeda. Jika validitas terkait dengan ketepatan objek yang tidak lain adalah tidak menyimpangnya data dari kenyataan, artinya bahwa data tersebut benar. Maka dari itu konsep reliabilitas terkait dengan pemotretan berkali-kali. Instrumen yang baik adalah instrumen yang dapat dengan ajeg memberikan data yang sesuai dengan kenyataan.

Lebih lanjut yang dimaksud reliabilitas adalah sebagai berikut.

- Se jauh mana hasil pengukuran dapat dipercaya
- Se jauh mana hasil pengukuran dapat dipercaya bila dilakukan pengukuran pada waktu yg berbeda pada kelompok subjek yg sama diperoleh hasil yg relatif sama asalkan aspek yg diukur dalam diri subjek memang belum berubah
- Tinggi/rendahnya reliabilitas secara empirik ditunjukkan oleh suatu angka yang disebut nilai koefisien reliabilitas
- Reliabilitas yg tinggi ditunjukkan dengan nilai 1.00, reliabilitas yg dianggap sudah cukup memuaskan atau tinggi adalah  $\geq 0.70$

Sehubungan dengan reliabilitas ini, Scarvia B. Anderson menyatakan bahwa persyaratan bagi tes, yaitu validitas dan reliabilitas ini penting. Dalam hal ini, validitas lebih penting, dan reliabilitas ini perlu karena menyokong terbentuknya validitas. Sebuah tes mungkin reliable tetapi tidak valid. Sebaliknya, sebuah tes yang valid biasanya reliable.

Untuk dapat memperoleh gambaran yang ajeg memang sulit karena unsur kejiwaan manusia itu sendiri tidak ajeg. Misalnya, kemampuan, kecakapan, sikap, dan sebagainya berubah-ubah dari waktu ke waktu.

Dalam bab ini dibahas tentang teknik pengujian reliabilitas dari sebuah tes hasil belajar. Dalam bab ini disajikan 2 teknik pengujian reliabilitas, yaitu teknik pengujian reliabilitas tes hasil belajar bentuk uraian dan bentuk obyektif.

### **A. Teknik Pengujian Reliabilitas Tes Hasil Belajar Bentuk Uraian**

Dalam menentukan apakah tes belajar bentuk uraian yang disusun sudah memiliki daya keajegan atau reliabilitas yang tinggi ataukah belum, digunakan sebuah rumus yang dikenal dengan *Rumus Alpha*.

$$r_{11} = \left( \frac{n}{n-1} \right) - \left( 1 - \frac{\sum s_i^2}{s_t^2} \right)$$

di mana:

$r_{11}$  = Koefisien reliabilitas tes.

$n$  = Banyaknya butir item yang dikeluarkan dalam tes.

1 = Bilangan konstan.

$\sum S_i^2$  = Jumlah varian skor dari tiap-tiap butir item.

$S_t^2$  = Varian total.

$\sum S_i^2$  diperoleh dengan menggunakan rumus di bawah.

Misalkan tes uraian yang akan ditentukan reliabilitasnya terdiri atas 5 item.

$$\text{Rumus } \sum S_i^2 = S_{i_1}^2 + S_{i_2}^2 + S_{i_3}^2 + S_{i_4}^2 + S_{i_5}^2$$

$$\text{Rumus } S_{i_1}^2 = \frac{\sum X_{i_1}^2 - \frac{(\sum X_{i_1})^2}{N}}{N}$$

$$S_{i_2}^2 = \frac{\sum X_{i_2}^2 - \frac{(\sum X_{i_2})^2}{N}}{N}, \text{ dst...}$$

Contoh

*Langkah pertama* :Menjumlahkan skor masing-masing testee  $\sum X_{i_1}$ ,  $\sum X_{i_2}$ ,  $\sum X_{i_3}$ ,  $\sum X_{i_4}$ , dan  $\sum X_{i_5}$ , dan mencari skor total yang dicapai ( $X_t$ ) serta mencari kuadrat dari skor total ( $X_t^2$ )

**Tabel 5.1 Tabel analisis dalam rangka mencari skor total untuk masing-masing butir item, skor total dan kuadrat skor total dari masing-masing testee**

Testee	Skor untuk butir item nomor :					$X_t$	$X_t^2$
	1	2	3	4	5		
A	8	6	7	7	6	34	1156
B	7	6	6	5	6	30	900
C	4	4	3	5	4	20	400
D	6	5	5	5	6	27	729
E	5	5	4	5	4	23	529
5 = N	30 $\Sigma X_{i_1}$	26 $\Sigma X_{i_2}$	25 $\Sigma X_{i_3}$	27 $\Sigma X_{i_4}$	26 $\Sigma X_{i_5}$	134= $\Sigma X_t$	3714= $\Sigma X_t^2$

*Langkah kedua* : Mencari jumlah kuadrat item

$$JK_{\text{item1}} = 8^2 + 7^2 + 4^2 + 6^5 + 5^2 = 190$$

$$JK_{\text{item2}} = 6^2 + 6^2 + 4^2 + 5^2 + 5^2 = 138$$

$$JK_{\text{item3}} = 7^2 + 6^2 + 3^2 + 5^2 + 4^2 = 135$$

$$JK_{\text{item4}} = 7^2 + 5^2 + 5^2 + 5^2 + 5^2 = 149$$

$$JK_{\text{item5}} = 6^2 + 6^2 + 4^2 + 6^2 + 4^2 = 140$$

*Langkah ketiga* : Mencari varian skor

$$S_{i_1}^2 = \frac{190 - \frac{90^2}{5}}{5} = \frac{190 - 180}{5} = 2,00$$

$$S_{i_2}^2 = \frac{138 - \frac{26^2}{5}}{5} = \frac{138 - 135,2}{5} = 0,56$$

$$S_{i_4}^2 = \frac{149 - \frac{27^2}{5}}{5} = \frac{149 - 145,8}{5} = 0,64$$

$$S_{i_5}^2 = \frac{140 - \frac{26^2}{5}}{5} = \frac{140 - 135,2}{5} = 0,96$$

$$S_{i_3}^2 = \frac{135 - \frac{25^2}{5}}{5} = \frac{138 - 135,2}{5} = 2,00$$

*Langkah keempat*: Mencari jumlah varian skor item secara keseluruhan:

$$\Sigma S_i^2 = S_{i_1}^2 + S_{i_2}^2 + S_{i_3}^2 + S_{i_4}^2 + S_{i_5}^2$$

$$\Sigma S_i^2 = 2,00 + 0,56 + 2,00 + 0,64 + 0,96$$

$$= 6,16$$

*Langkah kelima:* Mencari varian total ( $S_t^2$ )

$$S_t^2 = \frac{\sum X_t^2 - \frac{(\sum X_t)^2}{N}}{N}$$
$$= \frac{3714 - \frac{134^2}{5}}{5} = \frac{3714 - 3591,2}{5} = \frac{122,8}{5} = 24,6$$

*Langkah keenam:* Mencari koefisien reliabilitas tes

$$r_{11} = \left( \frac{n}{n-1} \right) \left( 1 - \frac{\sum S_i^2}{S_t^2} \right)$$
$$r_{11} = \left( \frac{5}{5-1} \right) \left( 1 - \frac{6,16}{24,56} \right) = (1,25) (1 - 0,251) = 1,25 \times$$
$$0,749 = 0,93625 = 0,94$$

Selanjutnya dalam pemberian interpretasi terhadap koefisien reabilitas tes ( $r_{11}$ ) pada umumnya digunakan patokan sebagai berikut:

1. Apabila  $r_{11} \geq 0,70$  maka tes hasil belajar memiliki reabilitas tinggi (reliable)
2. Apabila  $r_{11} \leq 0,70$  maka tes hasil belajar belum memiliki reabilitas tinggi (un-reliable)

Jadi  $r_{11} \geq 0,70$  maka tes hasil belajarn pada contoh di atas memiliki reliabilitas tinggi (reliabel)

## **B. Teknik Pengujian Reliabilitas Tes Hasil Belajar Bentuk Obyektif**

### **1. Pengujian realibilitas tes hasil belajar bentuk obyektif dengan menggunakan pendekatan single test – single trial**

Single test – single trial merupakan pendekatan “serba single” atau “serba satu”, yaitu satu kelompok subyek, satu jenis alat pengukur atau satu kelompok testee, dan satu kali testing.

Ada 5 jenis formula untuk menentukan reliabilitas dengan menggunakan single test- single trial.

*Pertama*, formula Brown, Flanagan, dan Rulon yang dilakukan dengan “membelah 2” tes (*split-half technique*). Namun Formula Kuder-Richardson dan C. Hoyt tidak menggunakan teknik belah dua. Formula Spearman Brown didasarkan pada korelasi, yaitu korelasi antara separoh belahan pertama dengan separoh belahan kedua dari tes. Sedangkan Formula Flanagan

didasarkan pada deviasi (selisih skor dengan mean totalnya, yaitu  $x = X - Mx$ )

*Kedua*, formula Rullon didasarkan pada selisih skor yang dimiliki oleh belahan pertama dengan belahan kedua dari tes yang dilambangkan:  $d = (X - Y)$ , dimana  $d$  = difference

*Ketiga*, formula Kuder-Richardson dilakukan dengan menganalisis skor-skor teitem tes hasil belajar. Sedangkan formula C. Hyot, sasarannya adalah interaksi antar testee dengan item tes hasil belajar. Teknik analisis tersebut adalah Teknik Analisis Varian (ANOVA).

a. Pendekatan Single Test-Single Trial dengan Formula Spearman Brown

Penentuan reliabilitas tes hasil belajar bentuk obyektif dengan menggunakan formula spearman-brown dikenal dengan istilah teknik belah dua (*split half technique*). Hal itu dikarenakan dalam penentuan reliabilitas tes, penganalisisannya dilakukan dengan jalan membelah dua butir-butir soal tes menjadi dua bagian yang sama, sehingga masing-masing testee memiliki dua macam skor. Untuk mengetahui reliabilitas tes secara



keseluruhan spearman-brown menciptakan formula sebagai berikut.

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

Keterangan :

$r_{tt}$  = Koefisien reliabilitas tes secara total (tt = total tes).

$r_{hh}$  = Koefisien kolerasi product moment antara separuh tes pertama dan separuh tes kedua (hh = half-half).

1 & 2 = Bilangan konstan.

1) Pendekatan Single Test-Single Trial Dengan Menggunakan Formula Spearman-Brown Model Gasal Genap

Langkah-langkah yang perlu ditempuh dalam penentuan reliabilitas tes dengan pendekatan single test-single trial dengan menggunakan formula spearman-brown model gasal genap adalah sebagai berikut.

- Menjumlahkan skor-skor dari butir-butir item yang bernomor ganjil yang dimiliki oleh masing-masing individu testee.
- Menjumlahkan skor-skor dari butir-butir item yang bernomor genap yang dimiliki oleh masing-masing individu testee.
- Mencari (menghitung) koefisien korelasi "r" product moment ( $r_{xy} = r_{hh} = r_{\frac{11}{12}}$ ). Dalam hal ini jumlah skor-skor dari butir-butir item yang bernomor ganjil kita anggap sebagai variabel X, sedangkan jumlah skor-skor dari butir-butir item yang bernomor genap kita anggap variabel Y, dengan menggunakan rumus:

$$r_{tt} = r_{hh} = r_{\frac{11}{12}} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

- Mencari (menghitung) koefisien reliabilitas tes ( $r_{11} = r_{tt}$ ) dengan menggunakan rumus:

$$r_{11} = r_{tt} = \frac{2 r_{\frac{11}{22}}}{1 + r_{\frac{11}{22}}}$$

- Memberikan interpretasi terhadap  $r_{11}$ .

## 2) Pendekatan Single Test-Single Trial Dengan Menggunakan Formula Spearman-Brown Model Belahan Kiri dan Kanan

Langkah-langkah yang perlu ditempuh dalam penentuan reliabilitas tes dengan pendekatan single test-single trial dengan menggunakan formula spearman-brown model belahan kiri dan kanan adalah sebagai berikut.

- Menjumlahkan skor-skor dari butir-butir item yang terletak diseparuh bagian kiri yang dimiliki oleh masing-masing individu testee.
- Menjumlahkan skor-skor dari butir-butir item yang terletak diseparuh bagian kanan yang dimiliki oleh masing-masing individu testee.
- Mencari (menghitung) koefisien kolerasi "r" product mement ( $r_{xy} = r_{hh} = \frac{7 \cdot 11}{12}$ ). Dalam hal ini jumlah skor-skor dari butir-butir item yang terletak diseparuh bagian kiri kita anggap sebagai variabel X, sedangkan jumlah skor-skor dari butir-butir item yang terletak diseparuh bagian kanan kita anggap variabel Y, dengan menggunakan rumus:

$$r_{tt} = r_{hh} = \frac{r_{11}}{12}$$
$$= \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

- Mencari (menghitung) koefisien reliabilitas tes ( $r_{11} = r_{tt}$ ) dengan menggunakan rumus:

$$r_{11} = r_{tt} = \frac{2 r_{11}}{1 + r_{11}}$$

- Memberikan interpretasi terhadap  $r_{11}$ .
- 3) Kelemahan-Kelemahan Formula Spearman-Brown
- Formula ini menghendaki agar belahan yang dicari kolerasinya, yaitu belahan gasal-genap dan belahan kiri-kanan haruslah sebanding.
  - Formula ini juga menuntut agar jumlah butir-butir item yang akan diuji reliabilitasnya haruslah genap.
  - Dengan dua buah model perhitungan tersebut, dapat terjadi bahwa koefisien reliabilitas menunjukkan bilangan yang tidak sama.

4) Ketepatan Penggunaan formula Spearman-Brown

- Tes yang diajukan cukup banyak

- Butir-butir item termasuk kategori item yang sukar
  - Materi tes bersifat menyeluruh, agar penyusun dapat membuat dua item untuk satu permasalahan yang sama.
- b. Pendekatan single test-single trial dengan menggunakan formula Flanagan

Formula:  $r_{11} = 2\left(1 - \frac{S_1^2 + S_2^2}{S_t^2}\right)$

Dimana:

$r_{11}$  = Koefisien reliabilitas tes secara total

2 & 1 = Bilangan konstan

$S_1^2$  = Jumlah varian dari skor-skor hasil tes yang termasuk belahan I

$S_2^2$  = Jumlah varian dari skor-skor hasil tes yang termasuk belahan II

$S_t^2$  = Jumlah varian total dari skor hasil tes belahan I dan II

Rumus  $S_1^2 = \frac{\sum x^2}{N}$ , deviasi  $x = X - M_x$

$$S_2^2 = \frac{\sum y^2}{N}, \text{ deviasi } y = Y - M_y$$

$$S_t^2 = \frac{\sum (x+y)^2}{N}$$

- 1) Pendekatan dengan menggunakan Formula Flanagan dengan menerapkan Model gasal-genap

*Langkah 1.* Menghitung kuadrat dari deviasi X ( $\sum x^2$ ), Y ( $\sum y^2$ ), dan jumlah kuadrat dari deviasi total X dan Y ( $\sum (x+y)^2$ )

Langkah-langkah perhitungan dalam tabel:

- Menjumlahkan skor variable X dan Y
- Mencari mean X dan Y:  $M_x = \frac{\sum x}{N}$ ,  $M_y = \frac{\sum y}{N}$
- Mencari deviasi skor X dan Y:  $x = X - M_x$ ,  $y = Y - M_y$ .  
 $\sum x$  maupun  $\sum y = 0$
- Mengkuadratkan deviasi x ( $x^2$ ) dan y ( $y^2$ )
- Mencari jumlah (x+y) dan diukudratkan  $(x+y)^2$

*Langkah 2.* Menghitung varian skor item gasal dan genap.

Rumus:  $S_1^2 = \frac{\sum x^2}{N}$  dan  $S_2^2 = \frac{\sum y^2}{N}$

Langkah 3. Mencari varian total :  $S_t^2 = \frac{\sum (x+y)^2}{N}$

Langkah 4. Mencari koefisien reliabilitas ( $r_{11}$ ) lalu memberikan interpretasi.

$$r_{11} = 2\left(1 - \frac{S_1^2 + S_2^2}{S_t^2}\right)$$

Terdapat cara yang lebih efisien menurut Flanagan tanpa menghitung mean dan deviasi ;

Pertama. Mencari  $\sum x^2$  dan  $\sum y^2$ :  $\sum x^2 = \sum Y^2 - \frac{(\sum X)^2}{N}$  dan

$$\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{N}$$

Kedua. Mencari  $\sum (x + y)^2 = \sum x_t^2 - \frac{(\sum X_t)^2}{N}$  lalu mencari  $S_1^2$ ,

$S_2^2$ , dan  $r_{11}$

2) Pendekatan dengan menggunakan Formula Flanagan dengan menerapkan Model kiri-kanan.

Langkah 1. Menghitung kuadrat dari deviasi X ( $\sum x^2$ ), Y ( $\sum y^2$ ), dan jumlah kuadrat dari deviasi total X dan Y ( $\sum (x + y)^2$ )

Langkah-langkah perhitungan dalam tabel:

- a) Menjumlahkan skor variable X dan Y
- b) Mencari mean X dan Y:  $M_x = \frac{\sum x}{N}$ ,  $M_y = \frac{\sum y}{N}$
- c) Mencari deviasi skor X dan Y:  $x = X - M_x$ ,  $y = Y - M_y$ .  
 $\sum x$  maupun  $\sum y = 0$
- d) Mengkuadratkan deviasi x ( $x^2$ ) dan y ( $y^2$ )
- e) Mencari jumlah ( $x+y$ ) dan diukudratkan  $(x+y)^2$

*Langkah 2.* Menghitung varian skor item kiri dan kanan.

Rumus:  $S_1^2 = \frac{\sum x^2}{N}$  dan  $S_2^2 = \frac{\sum y^2}{N}$

*Langkah 3.* Mencari varian total :  $S_t^2 = \frac{\sum (x+y)^2}{N}$

*Langkah 4.* Mencari koefisien reliabilitas ( $r_{11}$ ) lalu memberikan interpretasi.

$$r_{11} = 2\left(1 - \frac{S_1^2 + S_2^2}{S_t^2}\right)$$

Terdapat cara yang lebih efisien menurut Flanagan tanpa menghitung mean dan deviasi ;

Pertama. Mencari  $\sum x^2$  dan  $\sum y^2$ :  $\sum x^2 = \sum Y^2 - \frac{(\sum X)^2}{N}$  dan

$$\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{N}$$

Kedua. Mencari  $\sum (x + y)^2 = \sum x_t^2 - \frac{(\sum X_t)^2}{N}$  lalu mencari  $S_1^2$ ,

$S_2^2$ , dan  $r_{11}$



c. Pendekatan single tes-single trial dengan formula Rulon

Rumus:  $r_{11} = 1 - \frac{S_d^2}{S_t^2}$ , dimana  $S_d^2$  = Varian perbedaan testee belahan I dan II

Langkah-langkah:

➤ Menghitung  $d = (X-Y)$ , menjumlahkan  $d = \sum d$  lalu mengkuadratkan  $\sum d^2$

➤ Menghitung jumlah kuadrat perbedaan ( $\sum x_d^2$ ) dan mencari varian perbedaan ( $S_d^2$ ) skor belahan I dan II

$$\sum x_d^2 = \sum d^2 - \frac{(\sum d)^2}{N} \text{ lalu } S_d^2 = \frac{\sum x_d^2}{N}$$

➤ Mencari skor total  $X_t = (X+Y)$  lalu dijumlahkan  $X_t = \sum X_t$

➤ Mengkuadratkan  $X_t^2$  dan menjumlahkan menjadi  $\sum X_t^2$

➤ Mencari  $x_t^2 = \sum X_t^2 - \frac{((\sum X_t)^2)}{N}$  lalu mencari

$$S_t^2 = \frac{\sum X_t^2}{N}$$

➤ Lalu mencari koefisien reliabilitas:

$$r_{11} = 1 - \frac{S_d^2}{S_t^2}$$

Pendekatan formula rulon dengan menggunakan model gasal-genap dan belahan kiri-kanan menggunakan formula yang sama hanya membedakan skor item yang dipilih.

d. Pendekatan single tes-single trial dengan formula Kuder-Richardsen

Formula Kuder-Richardsen yaitu  $KR_{20}$  dan  $KR_{21}$

$$KR_{20} = \left( \frac{n}{(n-1)} \right) \left( \frac{S_t^2 - \sum p_i q_i}{S_t^2} \right)$$

Dimana:

N = banyaknya butir item

$S_t^2$  = varian total

$p_i$  = Proporsi testee yang menjawab dengan betul  
butir item tersebut

$q_i$  = Proporsi testee yang menjawab salah,  $q_i = 1 - p_i$

$\sum p_i q_i$  = Jumlah hasil perkalian  $p_i$  dan  $q_i$

$$KR_{21} = \left( \frac{n}{(n-1)} \right) \left( 1 - \frac{M_t (n-M_t)}{(n)(S_t^2)} \right)$$

Dimana:

$M_t$  = Mean total (rata-rata hitung dari skor total)

KR<sub>20</sub> dalam perhitungannya lebih teliti namun perhitungannya lebih rumit. Sedangkan KR<sub>21</sub> perhitungannya lebih sederhana namun kurang teliti.

1) Pendekatan formula Kuder-Richardson dengan Rumus KR<sub>20</sub>

Langkah-langkah:

- Mencari  $s_t^2 = \frac{\sum X_t^2}{N}$  namun sebelumnya mencari  $\sum x_t^2$   
 $= \sum X_t^2 - \left( \frac{\sum X_t}{N} \right)^2$
- Telah mengetahui  $n, s_t^2, \sum p_i q_i$ .

Akhirnya mencari  $r_{11} = \left( \frac{n}{(n-1)} \right) \left( \frac{S_t^2 - \sum p_i q_i}{S_t^2} \right)$

2) Pendekatan formula Kuder-Richardson dengan Rumus KR<sub>21</sub>

Langkah-langkah:

Pertama mencari  $M_t = \frac{\sum X_t}{N}$  setelah diketahui  $\sum X_t$  dan  $N$

Kedua substitusikan ke dalam rumus  $KR_{21}$  :

$$r_{11} = \left( \frac{n}{(n-1)} \right) \left( 1 - \frac{M_t (n - M_t)}{(n)(S_t^2)} \right)$$

e. Pendekatan single tes-single trial dengan formula C. Hyot

Keistimewaan dari teknik ANAVA:

- 1) Bisa digunakan selain single test-single trial. Misalny test retest atau alternate form
- 2) Bisa digunakan bila skor-skor hasil tesnya dikotomik tidak hanya dikotomi (betul skor 1 sedangkan salah skor 0)
- 3) Rumus:  $r_{11} = 1 - \frac{MK_g}{MK_s}$

Dimana:  $MK_g$  = Mean kuadrat interaksi antara testee dan item

$MK_s$  = Mean kuadrat antar subjek

Langkah-langkah:

- Menjumlahkan skor-skor jawaban yang betul lalu menjumlahkan  $X_t$  sehingga memperoleh  $\Sigma X$
- Menghitung jumlah kuadrat totat ( $JK_{tot}$ )

- Menghitung  $\sum X_t^2$ . Caranya mengkuadratkan semua skor hasil tes dari kiri kekanan tabel sampai habis. Misalkan  $1^2+1^2+0^2+1^2+1^2+1^2+0^2...$ dst. Sehingga mendapatkan hasil  $\sum X_t$
- Jumlah subjeknya adalah banyaknya siswa x banyak item.
- Mencari  $JK_{tot} = \sum X_t^2 - \frac{(\sum X_t)^2}{N}$  dengan diketahui  $\sum X_t^2$ ,  $\sum X_t$ , dan N
- Mengkuadratkan jawaban yang betul misalkan item/soal nomor 1 yang dijawab betul adalah oleh 15 testee. Setelah itu membagi masing-masing dengan jumlah testee. Misalkan jumlah testee adalah 20 orang. Setelah dibagi, jumlahkan lalu kurangi dengan  $\frac{(\sum X_t)^2}{N}$

Hasilnya merupakan  $JK_{ant.item}$

- Mengkuadratkan skor yang dicapai masing-masing siswa lalu dibagi dengan jumlah soal. Misalkan skor yang dicapai adalah 16 dan jumlah soal adalah 20. Jumlahkan semuanya lalu dikurangi dengan  $\frac{(\sum X_t)^2}{N}$

Hasilnya merupakan  $JK_{ant.subjek}$

- Mencari jumlah kuadrat interaksi antar item dengan subyek

$$\text{Rumus} = JK_{\text{ant.i-s}} \text{ atau } JK_e = JK_{\text{tot}} - JK_{\text{ant.it.}} - JK_{\text{ant.s.}}$$

- Mencari mean kuadrat antar item dan antar subyek

$$MK_{\text{ant.item}} = \frac{JK_{\text{ant.item}}}{db_{\text{ant.item}}}$$

dimana  $db_{\text{ant.item}} = N_1(\text{jumlah soal/item}) - 1$

$$MK_{\text{ant.subyek}} \text{ atau } MK_s = \frac{JK_{\text{ant.subyek}}}{db_{\text{ant.subyek}}} \text{ dimana } db_{\text{ant.subyek}} =$$

$N_2(\text{jumlah siswa/testee}) - 1$

- Mencari mean kuadrat interaksi antar item dengan subyek

$$MK_{\text{ant.i-s}} \text{ atau } MK_e = \frac{JK_{\text{ant.i-s}}}{db_{\text{ant.i-s}}}$$

dimana  $db_{\text{ant.i-s}} = db_{\text{ant.it}} \times db_{\text{ant.s}}$

- Mencari koefisien reliabilitas tes:  $r_{11} = 1 - \frac{MK_g}{MK_s}$  lalu memberikan interpretasi

## **2. Pengujian realibilitas tes hasil belajar bentuk obyktif dengan menggunakan pendekatan test-reetest (single test – double trial)**

Dikenal dengan pendekatan bentuk ulangan dimana tester hanya menggunakan satu seri tes tetapi percobaannya dilakukan dua kali. Misalkan seorang guru memberikan satu seri tes dalam dua kesempatan. Setelah tes selesai skor hasil tes pertama dikorelasikan dengan skor hasil tes kedua. Jika terdapat korelasi positif maka tes hasil belajar dapat dinyatakan reliabel karena skor hasil tes belajar memperlihatkan kestabilan atau keajegan.

Untuk mencari korelasi tersebut terdapat teknik korelasi rank-order (tata jenjang) dari Spearman:

$$\rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$
 dimana  $\rho$  = Koefisien korelasi antara variabel I (skor) dan II

D = Difference.  $D = R_I - R_{II}$ , N = Banyaknya subyek

Langkah-langkah:

- Merumuskan hipotesis nihil: "Tidak terdapat korelasi positif yang signifikan antara hasil tes I daengan hasil tes II
- Mencari koefisien korelasi rho:  $\rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$

- Memberikan interpretasi terhadap  $p$

Melihat nilai  $\rho$  dengan menggunakan derajat kebebasan (db) Misalkan  $N = 20$  diperoleh  $\rho_{\text{tabel}}$  pada taraf signifikansi 5% sebesar 0,450. Jika  $\rho_{\text{observasi}} \geq \rho_{\text{tabel}}$  baik pada taraf 1% maupun 5% hipotesis nihil ditolak. Berarti tes pertama dan kedua terdapat korelasi positif yang signifikan.

- Menarik kesimpulan.

Bahwa tes hasil belajar sudah memiliki reliabilitas yang tinggi (reliabel). Pengujian ulang tidak hanya sebatas satu kali saja, boleh lebih sampai penyusun tes sudah yakin dengan keajegan hasil pengukurannya.

### **3. Pengujian realibilitas tes hasil belajar bentuk obyektif dengan menggunakan pendekatan alternate form (double test – double trial)**

Mempergunakan dua tes yang diberikan tanpa adanya tenggang waktu dengan ketentuan tes harus sejenis, mengukur hal yang sama maupun tingkat kesukaran yang sama.



Alasan bahwa pendekatan alternate form (bentuk paralel) lebih baik daripada pendekatan yang sebelumnya:

- a. Terhindar dari kemungkinan testee latihan atau menghafal karena butir item yang diberikan tidak sama
- b. Terhindar akan timbulnya perbedaan situasi dan kondisi baik yang bersifat sosial maupun alami.

Apabila terdapat korelasi positif signifikan maka dapat dikatakan tes tersebut reliabel. Teknik korelasi yang digunakan antara teknik product moment dari Pearson atau teknik rank order dari Spearman(khusus untuk  $N \leq 30$ )

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

Langkah-langkah:

- Merumuskan hipotesis nihil
- Menghitung  $r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}}$
- Memberikan interpretasi terhadap  $r_{xy}$  atau  $r_{\text{observasi}}$ .  $Db = N - nr$ . Misalkan data 30 berarti  $db = 30 - 2 = 28$ , mencari  $r$  tabel. Jika  $r_{\text{tabel}}$  pada taraf 5% lebih besar

$r_{\text{observasi}}$  maka hipotesis nihil ditolak. Namun jika  $r_{\text{tabel}}$  pada taraf 1% lebih kecil daripada  $r_{\text{observasi}}$  maka hipotesis nihil disetujui.

Jika salah satu hipotesis nihil disetujui berarti hasil tes belajar memiliki reliabilitas sedang.

## **BAB VI**

### **TEKNIK PEMERIKSAAN, PEMBERIAN SKOR, DAN PENGOLAHAN HASIL TES HASIL BELAJAR**

#### **A. Teknik Pemeriksaan Hasil Tes Hasil Belajar**

Setelah dalam bab VI dibahas tentang teknik pengujian reliabilitas dari sebuah tes hasil belajar. Dalam bab selanjutnya akan dibahas mengenai teknik pemeriksaan, pemberian skor, dan pengolahan tes hasil belajar siswa. Teknik pemeriksaan tes hasil belajar terdiri atas tiga jenis teknik. Adapun tiga jenis teknik melakukan pemeriksaan hasil tes hasil belajar yaitu teknik pemeriksaan hasil tes tertulis, tes lisan, dan tes perbuatan.

##### **1. Teknik Pemeriksaan Hasil Tes Tertulis**

Pada pembicaraan terdahulu telah dikemukakan bahwa tes hasil belajar yang diselenggarakan secara tertulis dapat dibedakan menjadi dua golongan, yaitu: tes hasil belajar (tertulis) bentuk uraian (*subjective test = essay test*) dan tes hasil belajar (tertulis) bentuk objektif (*objective test*). Tes hasil belajar yang diselenggarakan secara tertulis dapat dibedakan menjadi dua golongan, yaitu tes hasil belajar bentuk uraian dan tes hasil belajar dalam bentuk objektif. Kedua tes

tersebut memiliki karakteristik yang berbeda, sudah barang tentu teknik pemeriksaan hasilnya pun berbeda.

a. Teknik Pemeriksaan Hasil Tes Hasil Belajar Bentuk Uraian

Dalam melaksanakan teknik pemeriksaan hasil tes hasil belajar bentuk uraian, sebelum melakukan pemeriksaan, sebaiknya tester segera membuat pedoman jawaban atas butir-butir soal yang telah disusun sebagai pegangan dalam pemeriksaan. pada saat pemeriksaan, tester membandingkan antara jawaban yang diberikan dengan pedoman yang sebelumnya disusun. Pedoman jawaban betul atas butir-butir soal yang telah disusun itulah yang selanjutnya akan digunakan sebagai pegangan atau patokan dalam pemeriksaan atau pengkoreksian terhadap hasil-hasil tes uraian. Sudah barang tentu, pemeriksaannya adalah dengan jalan membandingkan antara jawaban yang diberikan oleh testee dengan pedoman jawaban betul yang sebelumnya telah disusun oleh tester. Dalam pelaksanaan pemeriksaan hasil-hasil tes uraian ini ada dua hal yang perlu dipertimbangkan, yaitu: (1) apakah nantinya pengolahan dan penentuan nilai hasil tes uraian itu akan didasarkan pada standar mutlak, atau (2) apakah nantinya pengolahan dan penentuan nilai hasil tes subyektif itu akan didasarkan pada standar relatif.

Apabila nantinya pengolahan dan penentuan nilai hasil tes uraian itu akan didasarkan pada standar mutlak (dimana penentuan nilai secara mutlak akan didasarkan pada prestasi individual), maka prosedur pemeriksaannya adalah sebagai berikut:

- 1) Membaca setiap jawaban yang diberikan oleh testee untuk setiap butir soal tes uraian dan membandingkannya dengan pedoman jawaban betul yang sudah disiapkan.
- 2) Atas dasar hasil perbandingan antara jawaban testee dengan pedoman jawaban betul yang telah disiapkan itu, testee 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 (yang nantinya akan dijadikan bahan dalam pengolahan dan penentuan nilai lebih lanjut).

Adapun apabila nantinya pengolahan dan penentuan nilai akan didasarkan pada standar relatif (di mana penentuan nilai akan didasarkan pada prestasi kelompok), maka prosedur pemeriksaaannya adalah sebagai berikut:

- 1) Memeriksa jawaban atas butir-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 jawabannya termasuk lengkap, kurang lengkap, menyimpang, dan tidak memberikan jawaban sama sekali.

- 2) Memberikan skor terhadap jawaban soal nomor 1 untuk seluruh teste, 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 dari seluruh testee dapat diselesaikan, lalu dilanjutkan dengan pemeriksaan terhadap jawaban butir butir soal nomor 2, dengan cara yang sama.
- 4) Memberikan skor terhadap jawaban butir soal nomor 2 dari seluruh testee, dengan cara yang sama.
- 5) Setelah jawaban atas seluruh butir soal yang diberikan oleh seluruh testee dapat diselesaikan, akhirnya di lakukanlah penjumlahan skor (yang nantinya akan dijadikan bahan dalam pengolahan dan penentuan nilai).

#### b. Teknik Pemeriksaan Hasil Tes Hasil Belajar Bentuk Objektif

Dalam melakukan teknik pemeriksaan hasil tes hasil belajar bentuk objektif, adapun beberapa macam kunci jawaban yang dapat dipergunakan untuk mengkoreksi jawaban tes objektif yaitu kunci berdamping (*strip keys*),

adapun cara menggunakannya adalah dengan meletakkan kunci jawaban tersebut berjajar dengan lembar jawaban yang diperiksa; kunci sistem karbon (*carbon system keys*), kunci jawaban ini diletakkan di atas lembaran karbon. Pada kunci jawaban telah dibubuhi tanda berupa lingkaran-lingkaran untuk setiap jawaban yang betul. Jawaban yang berada di dalam lingkaran adalah betul, dan jawaban yang berada di luar lingkaran adalah salah; kunci sistem tusukan (*pinprick system keys*), pada dasarnya kunci sistem tusukan sama dengan kunci sistem karbon. Letak perbedaannya ialah, kunci jawaban sistem tusukan ini, untuk jawaban yang benar diberi tusukan dengan jarum atau paku; kunci berjendela (*window keys*), kunci jawaban yang kosong dan telah dilubangi dengan pilihan yang betul kemudian kita letakkan di atas kunci jawaban. Melalui lubang-lubang tadi kita buat garis-garis vertikal. Jika garis mengenai tanda silang yang dibuat testee pada lembar jawaban, maka jawaban testee dianggap betul.

Memeriksa atau mengoreksi jawaban atas soal tes objektif pada umumnya dilakukan dengan jalan menggunakan kunci jawaban, ada beberapa macam kunci jawaban yang dapat dipergunakan untuk mengoreksi jawaban soal tes objektif, yaitu sebagai berikut :

### 1) Kunci berdampingan (*strip keys*)

Kunci jawaban berdampingan ini terdiri dari jawaban – jawaban yang benar yang ditulis dalam satu kolom yang lurus dari atas kebawah, adapun cara menggunakannya adalah dengan meletakan kunci jawaban tersebut berjajar dengan lembar jawaban yang akan diperiksa, lalu cocokkan, apabila jawaban yang diberikan oleh teste benar maka diberi tanda (+) dan apabila salah diberi tanda (-).

### 2) Kunci system karbon (*carbon system key*)

Pada kunci jawaban system ini teste diminta membubuhkan tanda silang (X) pada salah satu jawaban yang mereka anggap benar kemudian kunci jawaban yang telah dibuat oleh teste tersebut diletakan diatas lembar jawaban teste yang sudah ditumpangi karbon kemudian tester memberikan lingkaran pada setiap jawaban yang benar sehingga ketika diangkat maka, dapat diketahui apabila jawaban teste yang berada diluar lingkaran berarti salah sedangkan yang berada didalam adalah benar.

### 3) Kunci system tusukan (*panprick system key*)

Pada dasarnya kunci system tusukan adalah sama dengan kunci system karbon. Letak perbedaannya ialah



pada kunci sistem ini, untuk jawaban yang benar diberi tusukan dengan paku atau alat penusuk lainnya sementara lembar jawaban testee berada dibawahnya, sehingga tusukan tadi menembus lembar jawaban yang ada dibawahnya. Jawaban yang benar akan terkena tusukan dsedangkan yang salah tidak.

#### 4) Kunci berjendela (*window key*)

Prosedur kunci berjendela ini adalah sebagai berikut :

- a) Ambil blanko lembar jawaban yang masih kosong
- b) Pilihan jawaban yang benar dilubangi sehingga seolah – olah menyerupai jendela.
- c) Lembar jawaban teste diletakan dibawah kunci berjendela.
- d) Melalui lubang tersebut kita dapat membuat garis vertical dengan pencil warna sehingga jawaban yang terkena pencil warna tersebut berarti benar dan sebaliknya.

## 2. Teknik Pemeriksaan dalam Rangka Menilai Hasil Tes Lisan

Pemeriksaan atau koreksi yang dilaksanakan dalam rangka menilai jawaban-jawaban testee pada tes hasil belajar secara lisan, pada umumnya cenderung bersifat subjektif. Hal ini kiranya mudah dipahami, sebab dalam tes lisan itu tester tidak berhadapan dengan lembar-lembar jawaban soal yang wujudnya adalah benda mati, melainkan berhadapan dengan individu-individu atau makhluk hidup yang masing-masing mempunyai ciri atau karakteristik berbeda-beda, sehingga peluang bagi tester untuk bertindak kurang atau bahkan tidak objektif.

Dalam melakukan teknik pemeriksaan dalam rangka menilai hasil tes lisan, pemeriksaan terhadap jawaban-jawaban testee hendaknya dikendalikan oleh pedoman yang pasti, misalnya:

- a. Kelengkapan jawaban yang diberikan oleh testee.

Pernyataan tersebut mengandung makna: apakah jawaban-jawaban yang diberikan oleh testee sudah memenuhi atau mencakup semua unsur yang seharusnya ada, sesuai dengan pedoman jawaban betul yang telah disusun oleh tester.

- b. Kelancaran testee dalam memberi jawaban-jawaban.

Maksudnya: apakah dalam memberikan jawaban-jawaban lisan atau soal-soal (pertanyaan-pertanyaan) yang diajukan kepada teste itu cukup lancar sehingga mencerminkan tingkat kedalaman atau tingkat pemahaman testee terhadap materi pertanyaan yang diajukan kepadanya.

c. Kebenaran jawaban yang diberikan.

Jawaban panjang lebar dan diemukakan secara lancar di hadapan tester, belum tentu merupakan jawaban yang benar. Karena itu tester harus benar-benar memperhatikan jawaban-jawaban testee tersebut, apakah jawaban testee itu mengandung kadar kebenaran yang tinggi atau sebaliknya.

d. Kemampuan testee dalam mempertahankan pendapatnya.

Maksudnya, apakah jawaban yang diberikan testee itu diberikan dengan penuh keyakinan akan kebenarannya ataukah tidak. Jawaban lisan yang disampaikan dengan nada ragu-ragu merupakan salah satu indikator bahwa testee kurang menguasai materi yang ditanyakan dalam tes lisan tersebut.

e. berapa pesen kira-kira, pertanyaan lisan yang termasuk kategori sukar, sedang, dan mudah dapat dijawab dengan benar oleh testee.

### 3. Teknik Pemeriksaan dalam Rangka Menilai Hasil Tes Perbuatan

Jika pada tes tertulis pemeriksaan hasilnya dilakukan dengan membaca lembar-lembar jawaban testee, dan pada tes lisan pemeriksaan itu dilakukan lewat jawaban-jawaban lisan yang diberikan oleh testee terhadap butir-butir soal atau pertanyaan yang diajukan secara lisan kepada mereka, maka tes perbuatan, "pemeriksaan" hasil-hasilnya dilakukan dengan menggunakan observasi (pengamatan).

Dalam teknik pemeriksaan dalam rangka menilai hasil tes perbuatan, pemeriksaan hasil-hasilnya dilakukan dengan menggunakan pengamatan. Sasaran yang diamati adalah: tingkah laku, perbuatan, sikap dan sebagainya. Untuk dapat menilai hasil tes perbuatan itu diperlukan adanya instrument tertentu dan setiap gejala yang muncul diberi skor-skor tertentu pula.

## **B. Teknik Pemberian Skor Hasil Tes Hasil Belajar**

Setelah mengetahui teknik pemeriksaan tes hasil belajar siswa, maka perlu dipelajari tentang teknik pemberian skor kepada sebuah tes hasil belajar. Cara pemberian skor pada umumnya disesuaikan dengan bentuk soal-soal yang dikeluarkan dalam tes tersebut.

### **1. Pemberian Skor pada Tes Uraian**

Pada tes uraian, pemberian skor umumnya mendasarkan diri pada bobot yang diberikan untuk setiap

butir soal, atas dasar tingkat kesukaran, atau atas dasar banyak sedikitnya unsur yang harus terdapat jawaban yang dianggap paling baik. Sebagai contoh dapat dikemukakan di sini misalnya tes subjektif mengidangkan lima butir soal. Pembuat soal telah menetapkan bahwa kelima butir soal itu mempunyai derajat kesukaran yang sama dan unsur-unsur yang terdapat pada setiap butir soal telah dibuat sama banyaknya. Atas dasar itu maka tester menetapkan bahwa testee yang dapat menjawab dengan jawaban paling betul (paling sempurna) diberikan skor 10. Jika hanya betul separuh diberikan skor 5, hamper seluruhnyabetul diberikan skor 9, dan seterusnya.

Dalam keadaan dimana butir-butir soal yang diajukan dalam bentuk tes uraian itu untuk tiap soal tidak memiliki derajat kesukaran yang sama, atau jumlah unsur yang terdapat pada setiap butir soal adalah tidak sama, maka pemberian skornya juga harus berpegang kepada derajat kesukaran dan jumlah unsur yang terdapat pada masing-masing butir soal tersebut.

Sebagai contoh, misalkan dari lima butir soal tes uraian, butir soal nomor 1 diberi skor maksimum 8, butir soal nomor 2 diberikan skor maksimum 10, butir soal nomor 3 diberi skor maksimum 6, butir soal nomor 4 diberi skor maksimum 10, dan butir soal nomor 5 diberi skor maksimum 8, maka seorang testee yang untuk butir soal nomor 1 jawabannya hanya betul separuh,

diberikanskor 4 (yaitu  $8 : 2 = 4$ ); untuk butir soal nomor 2 dari 10 unsur jawaban yang ada hanya dijawab betul sebanyak 6 unsur saja, maka kepada testee tersebut diberikan skor 6. Demikian seterusnya.

## 2. Pemberian Skor pada Tes Objektif

Pada tes objektif, setiap item diberi skor maksimum 1, jika dijawab salah maka skornya 0. Adapun cara menghitung skor terakhir dari seluruh item bentuk true false, dapat dilakukan menggunakan dua macam rumus, yaitu: rumus yang memperhitungkan denda, dan rumus yang mengabaikan denda. Penggunaan rumus-rumus itu sepenuhnya diserahkan kepada kebijaksanaan testee, apakah dalam tes hasil belajar tersebut kepada testee akan dikenai denda (bagi jawaban yang salah), ataukah tidak. Rumus skor akhir dengan memperhitungkan denda adalah sebagai berikut:

$$S = \frac{R - W}{0 - 1}$$

Adapun rumus skor akhir yang tidak memperhitungkan denda adalah sebagai berikut:

$$S = R$$

dimana:

S = skor yang sedang dicari

R = jumlah jawaban betul

W	=	jumlah jawaban yang salah
O	=	kemungkinan jawaban
1	=	bilangan konstan.

Contoh:

Dalam tes hasil belajar bidang studi Matematika yang diikuti oleh 40 orang siswa diajukan 80 butir item tes objektif; 20 butir diantaranya adalah tes objektif bentuk true-false, dengan ketentuan bahwa untuk setiap butir item yang dijawab betul diberikan bobot 1 dan untuk setiap butir item yang dijawab salah diberikan bobot 0.

dalam tes tersebut seorang siswa bernama Budi dapat menjawab dengan betul sebanyak 15 butir item ( $R = 15$ ); berarti jawaban yang salah  $= 20 - 15 = 5$  ( $W = 5$ ); sedangkan optionnya  $= 2$  ( $O = 2$ ).

Apabila terhadap jawaban salah satu itu dikenal sanksi berupa denda, maka skor akhir yang diberikan kepada Budi adalah:

$$S = \frac{R-W}{O-1} = \frac{15-5}{2-1} = 10$$

Sedangkan apabila terhadap jawaban salah itu tidak dikenakan sanksi berupa denda, maka skor yang diberikan kepada Budi itu adalah:

$$S = R = 15$$

Untuk tes objektif bentuk *matching*, *fill in* dan *completion*, perhitungan skor akhir pada umumnya tidak memperhitungkan sanksi berupa denda, sehingga rumus yang digunakan adalah:

$$S = R$$

Dengan kata lain, skor yang diberikan kepada testee adalah sama dengan jumlah jawaban betulnya.

### **C. Teknik Pengolahan dan Pengubahan Skor Tes Hasil Belajar Menjadi Nilai**

Selanjutnya setelah dapat melakukan pemberian skor terhadap tes hasil belajar siswa, perlu juga diperhatikan teknik pengolahan dan pengubahan skor tes hasil belajar menjadi nilai.

#### **1. Perbedaan Skor dan Nilai**

Sebelum sampai pada pembicaraan tentang teknik pengolahan dan pengubahan (konvensi) skor mentah hasil tes hasil belajar menjadi nilai standar, perlu dijelaskan terlebih dahulu tentang perbedaan antara skor dan nilai. Hal ini didasarkan pada pertimbangan bahwa kadang-kadang orang menganggap bahwa skor itu mempunyai pengertian yang sama dengan nilai; padahal pengertian seperti itu belum tentu benar. Skor adalah



hasil pekerjaan menyekor yang diperoleh dengan jalan menjumlahkan angka-angka bagi setiap butir item yang oleh testee telah dijawab dengan betul, dengan memperhatikan bobot jawaban betulnya. Adapun yang dimaksud dengan nilai adalah angka (bias juga huruf), yang merupakan hasil ubahan dari skor yang sudah dijadikan satu dengan skor-skor lainnya, serta disesuaikan pengaturannya dengan standar tertentu. Itulah mengapa nilai sering disebut skor standar.

Misalkan tes hasil belajar dalam bidang studi bahasa Indonesia menyajikan lima butir soal tes uraian di mana untuk setiap butir soal yang dijawab dengan betul diberikan bobot 10. Siswa bernama Budi, untuk kelima butir soal tes uraian tersebut memberikan jawaban sebagai berikut:

- Untuk butir soal nomor 1 dapat dijawab dengan sempurna, sehingga kepadanya diberikan skor 10.
- Untuk butir soal nomor 2 hanya dijawab betul separuhnya, sehingga skor yang diberikan kepada siswa tersebut adalah 5.
- Untuk butir soal nomor 3, hanya sekitar seperempat bagian saja yang dapat dijawab dengan betul, sehingga diberikan skor 2,5.
- Untuk butir soal nomor 4, dijawab betul sekitar separuhnya, sehingga diberikan skor 5.

- Untuk butir soal nomor 5, dijawab betul sekitar tiga perempatnya, sehingga diberikan skor 7,5.

Dengan demikian untuk kelima butir soal tes uraian tersebut siswa bernama Budi tersebut mendapatkan skor sebesar  $= 10 + 5 + 2,5 + 5 + 7,5 = 30$ . Angka 30 disini belum dapat disebut nilai, sebab angka 30 itu masih merupakan skor mentah (*raw score*), yang untuk dapat disebut nilai masih memerlukan pengolahan atau pengubahan (= konversi).

## 2. Pengolahan dan Pengubahan Skor Mentah Hasil Tes Hasil Belajar Menjadi Nilai Standar

Ada dua hal penting yang perlu dipahami terlebih dahulu dalam pengolahan dan pengubahan skor mentah menjadi skor standar atau nilai. Penilaian nilai dengan menggunakan standar relatif ini sangat cocok untuk diterapkan pada tes-tes sumatif (ulangan umum, ujian akhir semester, atau yang setara itu), sebab dipandang lebih adil, wajar, dan bersifat manusiawi

- a. Bahwa pengolahan dan pengubahan skor mentah menjadi nilai itu dilakukan dengan mengacu atau mendasarkan diri pada kriterium atau patokan dan norma atau kelompok.

1) Pengolahan Dan Pengubahan Skor Mentah Hasil Tes Hasil Belajar Menjadi Nilai Standar Dengan Mendasarkan Diri Atau Mengacu Pada Kriterium.

Pertama-tama harus dipahami bahwa penilaian beracuan kriterium ini berdasarkan pada asumsi, bahwa:

- a) Hal yang harus dipelajari oleh testee adalah mempunyai struktur hierakis tertentu, dan bahwa masing-masing taraf tertentu harus dikuasai secara baik sebelum testee tadi maju sampai taraf selanjutnya.
- b) Evaluator atau taster dapat mengidentifikasi masing-masing taraf itu sampai tuntas, atau setidak-tidaknya mendekati tuntas, sehingga dapat disusun alat pengukurnya.

Apabila dalam penentuan nilai tes hasil belajar itu digunakan acuan kriterium (menggunakan PAP), maka hal ini mengandung arti bahwa nilai yang akan diberikan kepada testee itu harus didasarkan pada standar mutlak, artinya, pemberian nilai kepada testee itu dilaksanakan

dengan jalan membandingkan antara skor mentah hasil tes yang dimiliki oleh masing-masing individu testee, dengan skor maksimum ideal (SMI) yang mungkin dapat dicapai oleh testee kalau saja seluruh soal tes dapat dijawab dengan betul.

karena itu maka pada penentuan nilai yang mengacu kepada kriterium atau patokan ini, tinggi rendahnya atau besar kecilnya nilai yang diberikan kepada masing-masing individu testee, mutlak ditentukan oleh besar kecil atau tinggi rendahnya skor yang dapat dicapai oleh masing-masing testee yang bersangkutan. Itulah sebabnya mengapa penentuan nilai dengan mengacu pada kriterium sering disebut sebagai penentuan nilai secara individual.

Disamping itu, karena penentuan nilai seorang testee dilakukan dengan jalan membandingkan skor mentah hasil belajar dengan skor maksimum idealnya, maka penentuan nilai yang beracuan pada kriterium ini juga sering dikenal dengan istilah penentuan nilai secara teoritik, atau penentuan nilai secara *das sollen*. Dengan istilah teoritik dimaksudkan disini adalah bahwa secara teoritik seorang siswa berhasil mendapatkan nilai 100 misalnya apabila keseluruhan butir soal tes dapat dijawab dengan betul oleh siswa tersebut.

Penilaian beracuan patokan (PAP) ini sangat baik diterapkan pada tes-tes formatif, dimana tester (guru, dosen, dan lain-lain) ingin mengetahui sudah sampai sejauh manakah peserta didiknya sudah terbentuk setelah mereka mengikuti program pengajaran dalam jangka waktu tertentu, dengan menggunakan *criterion referenced evaluation* dimana guru atau dosen dapat mengetahui berapa orang siswa atau mahasiswa yang tingkat penguasaannya tinggi, cukup, rendah, maka guru atau dosen tersebut dapat melakukan upaya-upaya yang dipandang perlu agar tujuan pengajaran dapat tercapai dengan optimal.

## 2) Pengolahan Dan Pengubahan Skor Mentah Hasil Tes Hasil Belajar Menjadi Nilai Standar Dengan Mendasarkan Diri Atau Mengacu Pada Norma Tau Kelompok.

Penilaian beracuan norma atau penilaian beracuan kelompok ini sering dikenal dengan istilah penentuan nilai secara relative, atau penilaian dengan mendasarkan diri pada standar relative. Dikatakan demikian, sebab dalam penentuan nilai hasil tes, skor mentah hasil tes yang dicapai oleh seorang peserta tes diperbandingkan dengan skor mentah hasil tes yang dicapai oleh peserta oleh peserta tes yang lain, sehingga kualitas yang

dimiliki oleh seorang peserta tes akan sangat tergantung kepada atau sangat ditentukan oleh kualitas kelompoknya.

Penilaian acuan norma diistilahkan dengan penentuan nilai secara empiris, penentuan nilai secara aktual atau penentuan nilai secara bas sein.

Dikatakan penentuan nilai secara aktual sebab didasarkan pada distribusi skor yang secara aktual yang dicapai oleh testee dalam suatu test belajar. Dikatakan penentuan nilai secara empiris karena penentuan nilai dilakukan dengan memperhatikan hasil-hasil test secara empiris yaitu skor hasil test berdasarkan pengamatan praktek dilapangan. Penentuan nilai berdasarkan standar relatif sangat cocok diterapkan pada tes-tes somatif sebab dipandang lebih adil, wajar dan manusiawi. Penilaian beracuan pada kelompok ini mendasarkan diri pada asumsi sebagai berikut:

- a) Pada setiap poulasi peserta didik yang sifatnya heterogen, akan selalu didapati kelompok baik/tinggi, kelompok sedang/tengah, dan kelomok kurang/rendah.
- b) Bahwa tujuan evaluasi hasil belajar adalah untuk menentukan posisi relatif dari para peserta tes

dalam hal yang sedang dievaluasi itu, yaitu apakah peserta tes posisi relatifnya berada di atas, di tengah, ataukah dibawah.

Dalam pembicaraan yang berhubungan dengan nilai standar kiranya perlu diketahui bahwa dalam dunia evaluasi pendidikan, khususnya evaluasi hasil belajar dikenal berbagai jenis nilai standar, seperti:

a) Pengubahan Skor Mentah Hasil Tes Menjadi Nilai Standar Berskala Lima (Stanfive)

Nilai standar berskala lima ,yang sering dikenal dengan istilah nilai huruf , yaitu nilai, A ,B ,C ,D , dan E. Pengubahan skor mentah hasil tes menjadi nilai standar berskala lima atau nilai huruf ,menggunakan patokan sebagai berikut :

$$\text{Mean} + 1,5 \text{ SD} = A$$

$$\text{Mean} + 0,5\text{SD} = B$$

$$\text{Mean} - 0.5 \text{ SD} = C$$

$$\text{Mean} - 1,5 \text{ SD} = E$$

b) Pengubahan Skor Mentah Hasil Tes Menjadi Nilai Standar Berskala Sembilan ( Stannine)

Nilai standar berskala sembilan ,yaitu rentangan atau skala nilai yang bergerak mulai dari 1 sampai 9.

Nilai standar berskala sembilan adalah nilai standar yang meniadakan nilai 0 dan 10 . nilai standar tersebut tidak bisa di pakai di gunakan di Indonesia

c) Pengubahan Skor Mentah Hasil Tes Menjadi Nilai Standar Yang Berskala Sebelas ( Standar Eleven)

Ada lagi standar nilai lain, yaitu yang selanjutnya dikembangkan yang disesuaikan dengan system penilaian di Indonesia. Dengan Standar eleven ini, system penilaian membagi skala menjadi 11 golongan, yaitu angka-angka 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, yang satu sama lain berjarak sama. Tiap-tiap angka menempati interval sebesar 0,55 SD, bertitik tolak dari mean = 5 yang menempati jarak - 0,275 SD sampai +0,275 SD. Seluruh jarak yang digunakan adalah dari -3,025 Sd sampai +3,025 SD.

Bilangan-bilangan persentil untuk menentukan titik dalam standar eleven ini adalah:  $P_1$ ,  $P_3$ ,  $P_8$ ,  $P_{21}$ ,  $P_{39}$ ,  $P_{61}$ ,  $P_{79}$ ,  $P_{92}$ ,  $P_{97}$ , dan  $P_{99}$ . Bagaimana menentukan  $P_1$ ,  $P_3$ ,  $P_8$ , dan seterusnya silakan membuka buku statistik.

Dasar pemikiran untuk standar eleven ini adalah bahwa jarak praktis dalam kurva normal adalah 6 SD yang terbagi atas 11 skala.

$$\begin{aligned} 11 \text{ Skala} &= 6 \text{ SD} \\ 1 \text{ Skala} &= \frac{6}{11} \text{ SD} \end{aligned}$$



$$= 0,55 \text{ SD}$$

d) Pengubah Skor Mentah Hasil Tes Menjadi Nilai Standar Z ( Z Score )

Nilai standar z atau z score umumnya dipergunakan untuk mengubah skor-skor mentah yang diperoleh dari berbagai jenis pengukuran yang berbeda-beda. Untuk pengubahan score mentah jadi score z, diperlukan adanya nilai bersifat baku ( standar ),di mana dengan nilai standar kita dapat mengetahui kedudukan relatif ( standar position ) untuk peserta testee, dengan cara menggunakan rumus berikut :

$$z = \frac{x}{SDx}$$

dimana:

z = z score

x = deviasi skor x, yaitu selisih antara skor x dengan  $M_x$ .

$SD_x$  = Deviasi standar dari skor-skor x

Dengan menggunakan nilai standar z ini maka testee yang dipandang memiliki kemampuan lebih tinggi adalah testee yang skornya bertanda positif (+). Adapun testee yang skornya bertanda negatif

(-) dipandang sebagai testee yang kemampuannya lebih lemah jika dibandingkan dengan testee lainnya.

e) Pengubahan Skor Mentah Hasil Tes Menjadi Nilai Standar T(T Score )

Dimaksud dengan T score adalah angka skala yang menggunakan mean sebesar 50 ( $M=50$ ) dan standar deviasi sebesar 10 ( $SD=10$ ). T skor dapat di peroleh dengan jalan memperkalikan z score dengan angka 10 kemudian ditambah dengan 50. T score dicari dengan maksud untuk meniadakan tanda minus yang terdapat di depan nilai standar z.

$$T \text{ score} = 10z + 50 \text{ atau}$$

$$T \text{ score} = 50 + 10z$$

T score dicari atau dihitung dengan maksud untuk meniadakan tanda minus yang terdapat di depan nilai standar z, sehingga lebih mudah dipahami oleh mereka yang masih asing atau awam terhadap ukuran-ukuran statistik.

## **BAB VII**

### **ANALISIS ITEM TES HASIL BELAJAR**

Salah satu tugas penting yang seringkali dilupakan oleh staf pengajar (guru, dosen, dan lain-lain) adalah tugas melakukan evaluasi terhadap alat ukur yang telah digunakan untuk mengukur keberhasilan belajar dari peserta didiknya. Alat ukur yang dimaksud adalah tes hasil belajar, yang sebagaimana telah kita ketahui, batang tubuhnya terdiri dari kumpulan butir-butir soal (item). Hendaknya staf pengajar perlu melakukan penganalisisan terhadap tes hasil belajar yang telah dijadikan alat ukur dalam rangka mengukur keberhasilan belajar dari para peserta didik tersebut. Di sini staf pengajar selaku tester perlu melakukan penelusuran dan pelacakan secara cermat, terhadap butir-butir soal atau item yang merupakan bagian yang tidak terpisahkan dari tes hasil belajar sebagai suatu totalitas.

Penelusuran atau pelacakan yang dimaksudkan di atas bertujuan untuk mengetahui, apakah butir-butir item yang membangun tes hasil belajar itu sudah dapat menjalankan fungsinya sebagai alat pengukur hasil belajar yang memadai ataukah belum. Identifikasi terhadap setiap butir item tes hasil belajar itu dilakukan dengan harapan akan menghasilkan berbagai informasi berharga, yang pada dasarnya akan dijadikan sebagai

umpan balik (*feed back*) guna melakukan perbaikan, pembenahan, dan penyempurnaan kembali terhadap butir-butir item yang telah dikeluarkan dalam tes hasil belajar, sehingga pada masa-masa yang akan datang tes hasil belajar yang telah disusun atau dirancang oleh staf pengajar (guru, dosen, dan lain-lain) selaku tester itu betul-betul dapat menjalankan fungsinya sebagai alat pengukur hasil belajar yang memiliki kualitas yang tinggi. Rangkaian kegiatan seperti inilah yang dikenal dengan istilah analisis item (*item analysis*).

Penganalisisan terhadap butir-butir item tes hasil belajar dapat dilakukan dari tiga segi, yakni derajat kesukaran item, daya pembeda item, dan fungsi distratornya. Untuk lebih jelasnya akan diuraikan pada pembahasan berikut.

### **A. Analisis Derajat Kesukaran Item**

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

Bertitik tolak dari pernyataan tersebut, maka butir-butir item tes hasil belajar dimana seluruh testee tidak dapat menjawab dengan betul (karena terlalu sukar)

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

Cara apa yang digunakan untuk mengetahui butir-butir item tes hasil belajar tertentu yang dapat dikatakan sudah memiliki derajat kesukaran yang memadai? Witherington dalam bukunya yang berjudul *Psychological Education*, mengatakan bahwa, sudah atau belum memadainya derajat kesukaran item tes hasil belajar dapat diketahui dari besar kecilnya angka yang melambangkan tingkat kesulitan dari item tersebut. Angka yang dapat memberikan petunjuk mengenai tingkat kesulitan item tersebut dikenal dengan istilah *difficulty index* (angka indeks kesukaran item), yang dalam dunia evaluasi hasil belajar umumnya dilambangkan dengan huruf P yang merupakan singkatan dari proporsi.

Menurut Witherington, angka indeks kesukaran item tersebut besarnya berkisar antara 0,00 sampai dengan 0,01. Artinya angka indeks kesukaran paling rendah adalah 0,00 dan paling tinggi adalah 1,00. Angka indeks 0,00 merupakan petunjuk bagi tester bahwa item butir-butir tes hasil belajar tersebut terlalu sukar. Sebaliknya, jika angka indeks 0,01 merupakan bahwa butir item yang bersangkutan terlalu mudah.

Untuk memperoleh angka indeks tersebut menggunakan

$$\text{rumus : } P = \frac{N_p}{N}$$

Dimana :

P = proporsi atau angka indeks kesukaran (P kotor)

Np = jumlah testee yang menjawab benar

N = jumlah testee

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

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

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

Kedua, untuk butir-butir item yang termasuk dalam kategori terlalu sukar, ada tiga kemungkinan tidak lanjut, yaitu : (1) Butir item tersebut dibuang atau didrop dan tidak akan dikeluarkan lagi dalam tes-tes hasil belajar yang akan datang. (2) diteliti ulang, dilacak dan ditelusuri sehingga dapat diketahui factor yang menyebabkan butir

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

Ketiga, untuk butir-butir item yang termasuk dalam kategori terlalu mudah, juga ada tiga kemungkinan tindak lanjut, yaitu: (1) butir item tersebut dibuang atau didrop dan tidak akan dikeluarkan lagi dalam tes-tes hasil belajar yang akan datang. (2) diteliti ulang, dilacak dan ditelusuri secara cermat guna mengetahui factor yang menyebabkan butir item tersebut dapat dijawab betul oleh hamper seluruh testee; ada kemungkinan option atau alternative yang dipasangkan pada butir-butir item

yang bersangkutan terlalu mudah diketahui oleh testee, mana item yang merupakan kunci jawaban item dan mana option yang berfungsi sebagai pengecoh atau distractor. Disini tester harus berusaha memperbaiki atau menggantinya dengan option yang lain sedemikian rupa sehingga antara kunci jawaban dengan pengecoh sulit dibedakan oleh testee. Setelah dilakukan perbaikan, item yang bersangkutan dicoba untuk dikeluarkan lagi pada tes hasil belajar selanjutnya. Tujuannya adalah untuk mengetahui apakah derajat kesukaran item itu menjadi lebih baik atau tidak daripada yang sebelumnya. (3) seperti halnya butir-butir item yang terlalu sukar, butir-butir item yang terlalu mudah juga masih mengandung manfaat, yaitu bahwa butir-butir item yang termasuk dalam kategori ini dapat dimanfaatkan pada tes-tes yang sifatnya longgar, dalam arti bahwa sebagian besar dari testee akan dinyatakan lulus dalam tes seleksi tersebut. Dalam kondisi seperti ini sangat bijaksana apabila butir-butir item yang dikeluarkan dalam tes seleksi tersebut adalah butir-butir item yang termasuk dalam kategori terlalu mudah, sehingga tes seleksi itu boleh dikatakan hanya sebagai formalitas saja.

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



item dengan derajat kesukaran dan derajat kemudahan yang tinggi.

Kembali pada pembicaraan tentang rumus untuk mencari atau menghitung P yang telah dikemukakan sebelumnya, maka proporsia itu sebenarnya adalah merupakan rata-rata hitung yang berhasil diraih oleh seluruh testee untuk butir item yang bersangkutan, dimana rumus itu  $Np$  tidak lain adalah  $\sum X$  yaitu jumlah skor-skor jawaban benar dari tiap-tiap individu testee. Itulah sebabnya mengapa indeks kesukaran item P sering dikenal dengan istilah angka indeks kesukaran rata-rata inilah yang paling sering digunakan sebagai dasar untuk menentukan derajat kesukaran item.

Namun demikian, angka indeks ini bukanlah merupakan angka indeks yang tanpa cacat. Kelemahan utama yang terdapat pada angka indeks kesukaran rata-rata P ialah, adanya hubungan yang terbalik antara derajat kesukaran item dengan angka indeks itu sendiri. Seperti yang sudah disampaikan sebelumnya, makin rendah angka indeks kesukaran item yang dimiliki oleh sebutir item akan semakin tinggi derajat kesukaran item tersebut. Sebaliknya semakin tinggi angka indeks kesukaran yang dimiliki oleh sebutir item, maka derajat kesukaran item tersebut semakin rendah. Jadi hubungan diantara keduanya merupakan hubungan yang berlawanan arah. Itulah sebabnya mengapa orang cenderung untuk mengatakan bahwa istilah angka indeks kesukaran item itu lebih tepat diganti dengan istilah

angka indeks kemudahan item atau angka indeks fasilitas item, dimana dengan angka indeks kemudahan item itu terdapat kesejajaran arah antara angka indeks kemudahan item dengan derajat kemudahannya, yaitu semakin tinggi angka indeks kemudahan item maka derajat kemudahan item juga semakin rendah, sehingga lebih rendah adanya hubungan langsung yang sifatnya searah antara tinggi rendahnya angka indeks kemudahan item dengan derajat kemudahan item yang bersangkutan.

Cara kedua untuk menghitung indeks adalah dengan menggunakan skala kesukaran linier. Skala kesukaran linier ini disusun dengan cara mentransformasikan nilai P menjadi nilai z, dimana perubahan dari P ke z itu dilakukan dengan berkonsultasi pada tabel nilai z yang pada umumnya dilampirkan pada buku-buku statistik.

Dengan menggunakan cara yang kedua, maka memperhatikan langkah-langkah sebagai berikut.

Langkah pertama dengan menggunakan rumus :

$$P_b = \frac{aP_k - 1}{a - 1}$$

Dimana :

P<sub>b</sub> = P bersih

P<sub>k</sub> = P kotor

A = option yang digunakan

1 = bilangan konstan

Sebelum dilakukan perhitungan untuk memperoleh P bersih terlebih dahulu perlu dikemukakan bahwa angka indeks kesukaran item (P) seperti yang sudah diuraikan sebelumnya masih merupakan P yang bersifat kotor, karena P diperoleh tanpa mempertimbangkan option atau alternative yang dipasang pada setiap butir item yang sedang dianalisis derajat kesukarannya. Dengan P bersih maka derajat kesukaran item itu telah dikoreksi dengan jalan memperhatikan option yang dipasang pada setiap butir item yang bersangkutan.

Langkah kedua mentransformasikan nilai p bersih menjadi nilai z, dengan berkonultasi pada tabel kurva normal. Kita ambil sebagai contoh P bersih yang dimiliki oleh butir item nomor. Butir item nomor 9 memiliki P bersih sebesar 0,75. Untuk mentransformasikan P bersih sebesar 0,75 itu menjadi nilai z, kita cari angka sebesar 0,75 itu di dalam tabel kurva normal terlampir.

Berdasarkan hasil konsultasi pada tabel kurva normal, maka dengan P bersih sebesar 0,75 diperoleh harga z sebesar 0,6745. Dengan berpegang pada Patoka yang diberikan oleh Robert L. Thorndike dan Elizabeth Hagen dalam bukunya yang berjudul *Measurement and Evaluation in Psychology and Education*, seperti yang telah diungkapkan sebelumnya, maka dengan angka indeks kesukaran item sebesar 0,6745 kita dapat menyatakan bahwa butir item nomor 9 itu dinyatakan

dalam kategori item yang telah memiliki derajat kesukaran yang cukup.

Cara lain selain berkonsultasi pada tabel kurva normal dengan menggunakan rumus :

$$P_b = \frac{B - \frac{S}{a-1}}{B + S}$$

Dimana:

$P_b$  = angka indeks kesukaran item setelah dikoreksi (P bersih)

$B$  = jumlah testee yang menjawab benar

$S$  = jumlah testee yang menjawab salah

$A$  = opsi yang diberikan pada setiap item

Langkah ketiga mencari atau menghitung angka indeks kesukaran item ialah dengan menggunakan angka indeks davis. Angka indeks Davis dapat diperoleh dengan rumus :

$$D = 21,063 z + 50$$

Keuntungan dari penggunaan indeks Davis ini adalah bahwa kita akan dapat terhindar dari tanda negative atau tanda minus, seperti yang dimungkinkan terjadi apabila kita menggunakan skala kesukaran linier.

Cara menggunakan indeks Davis sebagai petunjuk tentang derajat kesukaran item adalah seperti dikemukakan pada contoh berikut ini. Misalkan sebuah item memiliki P kotor sebesar 0,265. Dengan berkonsultasi pada tabel kurva normal diperoleh z sebesar 0,6280. Dengan menggunakan rumus angka indeks Davis, maka :

$$\begin{aligned} D &= 21,063 z + 50 \\ &= (21,063) (0,6280) + 50 \\ &= 13,227564 + 50 \\ &= 63,2274564 \\ &= 63,23 \end{aligned}$$

Karena rentangan angka indeks Davis adalah antara 0 sampai 100, maka dengan D sebesar 63,23 kita dapat menyatakan bahwa butir item yang bersangkutan memiliki derajat kesukaran yang cukup/sedang (D sebesar 63,23 itu berada antara 30-70).

Jika kita ingin memperoleh angka indeks Davis secara cepat, maka kita dapat menggunakan sebuah tabel yang disebut tabel untuk mengestimasi nilai D (indeks Davis).

## **B. Analisis Daya Pembeda Item**

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

kemampuan rendah tersebut sebagian besar tidak dapat menjawab item dengan betul.

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

Sejalan dengan pernyataan di atas maka kegiatan analisis terhadap daya pembeda item itu ditunjukkan untuk menjawab pertanyaan "apakah testee yang kita anggap pandai jawabannya pada umumnya betul, dan apakah testee yang kita anggap bodoh itu pada umumnya jawabannya salah?". Jika jawaban atas pertanyaan itu adalah "ya", maka butir item yang bersangkutan dapat kita anggap sebagai butir item yang baik, dalam arti bahwa butir item tersebut telah menunjukkan kemampuannya didalam membedakan antara testee yang termasuk dalam kategori pandai dengan testee yang termasuk dalam kategori bodoh.

Sebaliknya, jika jawaban atas pertanyaan tersebut adalah "tidak", maka butir item yang bersangkutan dapat dinyatakan sebagai butir item jelek, sebab hasil yang dicapai dalam tes itu justru bertentangan atau berlawanan dengan tujuan tes itu sendiri.

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

Untuk mengetahui besar kecilnya angka indeks deskriminasi dapat dilakukan dengan dua macam rumus antara lain:

Rumus pertama

$$D = PA - PB \text{ atau } D = PH - PL$$

Rumus kedua

$$\theta = \frac{P_H - P_L}{2\sqrt{(p)(q)}}$$

Untuk mengetahui angka indeks deskriminasi item D, ditempuh langkah-langkah sebagai berikut: (1) mengelompokkan testee menjadi beberapa kelompok yaitu kelompok atas (pandai) dan kelompok bawah (bodoh), (2) memberi kode-kode terhadap hasil pengelompokkan testee, (3) menghitung BA, BB, PA, PB, (4) memberikan penafsiran mengenai kualitas daya pembeda item yang dimiliki oleh satu butir item tes hasil belajar.

### **C. Analisis Fungsi Distraktor**

Pada saat membicarakan tentang tes obyektif bentuk multiple choice item telah dikemukakan bahwa pada tes bentuk multiple choice item tersebut untuk setiap butir item yang dikeluarkan dalam tes hasil belajar sudah dilengkapi dengan beberapa kemungkinan jawaban, atau yang sering dikenal dengan istilah option.

Option atau alternatif ini jumlahnya berkisar antara tiga sampai dengan lima buah, dan dari kemungkinan-kemungkinan jawaban yang dipasang pada setiap butir item itu salah satu merupakan jawaban yang betul, dan sisanya itu merupakan jawaban yang salah. Jawaban-jawaban yang salah itulah yang bisa dikenal dengan istilah distractor.

Tujuan utama dari pemasangan distractor pada setiap butir item yaitu agar dari sekian banyak testee yang mengikuti tes hasil belajar ada yang tertarik untuk memilihnya, sebab mereka akan mengira bahwa jawaban yang merepilih adalah jawaban betul. Jadi mereka akan terkecoh, menganggap bahwa distractor yang terpasang pada item itu sebagai kunci jawaban, padahal bukan. Dengan kata lain distractor baru dapat menjalankan fungsinya dengan baik jika distraktor telah memiliki daya rangsang yang membuat testee terkecoh. Menganalisis fungsi distractor sering dikenal dengan istilah menganalisis pola penyebaran jawaban item.

Menganalisis fungsi distractor sering dikenal dengan istilah lain, yaitu: menganalisis pola penyebaran jawaban



item. Adapaun yang dimaksud dengan pola penyebaran jawaban item adalah suatu pola yang dapat menggambarkan bagaimana testee menentukan pilihan jawabannya terhadap kemungkinan-kemungkinan jawaban yang telah dipasangkan pada setiap butir item.

Dalam hubungannya dengan pebicaraan tentang pengecoh atau distractor, cara yang dapat digunakan untuk menentukan apakah suatu distractor tersebut sudah dapat menjalankan fungsinya dengan baik atau tidak, ada beberapa hal yang perlu ditekankan. Kelaziman yang berlaku dalam dunia evaluasi hasil belajar ialah, bahwa distractor dinyatakan telah dapat menjalankan fungsinya dengan baik apabila distractor tersebut sekurang-kurangnya sudah dipilih oleh 5% dari seluruh peserta. Misalnya tes hasil belajar diikuti oleh 100 testee. Distractor yang dipasang pada item tersebut dapat dinyatakan berfungsi apabila minimal 5 orang dari 100 testee sudah terkecoh untuk memilih distractor tersebut.

## **BAB VIII**

### **TEKNIK PENENTUAN NILAI AKHIR DAN PENYUSUNAN RANKING**

#### **A. Fungsi Nilai Akhir**

Bagi seorang siswa, nilai merupakan sesuatu yang sangat penting karena nilai merupakan cermin dari keberhasilan belajar. Namun, bukan hanya siswa sendiri saja yang memerlukan cermin keberhasilan belajar ini; guru dan orang lainpun memerlukannya.

Secara garis besar, nilai mempunyai 4 (empat) fungsi sebagai berikut:

1. Fungsi Instruksional,
2. Fungsi Informatif,
3. Fungsi bimbingan, dan
4. Fungsi Administratif.

#### **1. Fungsi Instruksional**

Tidak ada tujuan yang lebih penting dalam proses belajar mengajar kecuali mengusahakan agar perkembangan dan belajar siswa mencapai tingkat optimal. Pemberian nilai merupakan salah satu cara dalam usaha ke arah tujuan itu, asal dilakukan dengan hati-hati dan bijaksana.

Pemberian nilai merupakan suatu pekerjaan yang bertujuan untuk memberikan suatu balikan (umpan balik)

yang mencerminkan seberapa jauh seorang siswa telah mencapai tujuan yang ditetapkan dalam pengajaran atau sistem instruksional.

Apabila pemberian nilai dapat dilakukan dengan cermat dan terperinci, maka akan lebih mudah diketahui pula keberhasilan dan kegagalan siswa disetiap bagian tujuan. Oleh karenanya, penggabungan nilai dari berbagai nilai sehingga menjadi nilai kahir, kadang-kadang dapat menghilangkan arti dari petunjuk yang semula telah disajikan secara teliti.

Nilai rendah yang diperoleh oleh seorang atau beberapa siswa, jika disajikan dalam keadaan yang terperinci akan dapat membantu siswa dalam usaha memperbaiki dan memberi motivasi peningkatan prestasi berikutnya. Bagi pengelola pengajaran, sajian terperinci nilai siswa dapat berfungsi menunjukkan bagian-bagian proses pengajaran mana yang diperbaiki.

## 2. Fungsi Informatif

Meberikan nilai siswa kepada orang tuanya mempunyai arti bahwa orang tua siswa tersebut menjadi tahu akan kemajuan dan prestasi putranya di sekolah. Catatan ini akan sangat berguna, terutama bagi orang tua yang ikut serta menyadari tujuan sekolah dan perkembangan putranya. Dengan catatan nilai untuk orang tua, maka:

- a) Orang tua menjadi sadar akan keadaan putranya untuk kemudahan lebih baik memberikan berupa perhatian, dorongan, atau bimbingan, dan
- b) Hubungan antara orang tua dengan sekolah menjadi baik.

### 3. Fungsi Bimbingan

Pemberian nilai kepada siswa akan mempunyai arti besar bagi pekerjaan bimbingan. Dengan perincian gambaran nilai siswa, petugas bimbingan akan segera tahu bagian-bagian mana dari usaha siswa di sekolah yang masih memerlukan bantuan. Catatan lengkap yang juga mencakup tingkat (rating) dalam kepribadian siswa serta sifat-sifat yang berhubungan dengan rasa sosial akan sangat membantu siswa dalam pengarahannya sebagai pribadi seutuhnya.

### 4. Fungsi Administratif

Yang dimaksud dengan fungsi administrasi dalam penilaian antara lain mencakup:

- a) Menentukan kenaikan dan kelulusan siswa.
- b) Memindahkan atau menempatkan siswa.
- c) Memberikan beasiswa.
- d) Memberikan rekomendasi untuk melanjutkan belajar, dan
- e) Memberi gambaran tentang prestasi tentang prestasi siswa/lulusan kepada para calon pemakai tenaga.

## **B. Faktor-Faktor yang Turut Diperhitungkan dalam Penilaian**

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

1. Prestasi/pencapaian (*achievement*).
2. Usaha (*effort*).
3. Aspek pribadi dan sosial (*personal and social characteristics*), dan
4. Kebiasaan bekerja (*working habits*).

### **1. Prestasi/Pencapaian**

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

### **2. Usaha**

Terpisah dan nilai prestasi, guru dapat menyampaikan laporannya kepada orang tua siswa. Laporan atau nilai tidak boleh dicampuri dengan nilai prestasi sama sekali. Yang sering terjadi adalah

kecenderungan dari guru untuk menilai unsur usaha ini lebih rendah bagi anak yang restasinya rendah dan sebaliknya.

### 3. Aspek Pribadi dan Sosial

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

### 4. Kebiasaan Bekerja

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

## **C. Teknik Penentuan Nilai Akhir**

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

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

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

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

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

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

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

di mana:

$N_A$  = Nilai Akhir

$F_1$  = Nilai hasil tes formatif ke-1

$F_2$  = Nilai hasil tes formatif ke-2

$F_3$  = Nilai hasil tes formatif ke-3

$F_4$  = Nilai hasil tes formatif ke-n

$n$  = Banyaknya kali tes formatif dilaksanakan

2 & 3 = Bilangan konstan (2=bobot tes formatif,

3= bobot tes secara keseluruhan)

Contoh:

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

- Nilai tes formatif I = 8
- Nilai tes formatif II = 7,5



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

Dengan demikian nilai akhir yang dapat diberikan kepada A :

$$N_A = \frac{\frac{(8 + 7,5 + 6,5 + 7)}{4} + (2 \times 8)}{3}$$
$$= \frac{7,25 + 16}{3}$$

= 7,75 = 8 (dibulatkan ke atas).

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

$$N_A = \frac{\frac{(\sum H)}{N} + 2 E}{3}$$

Contoh:

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

$$\begin{aligned} N_A &= \frac{\frac{(7 + 8 + 9)}{3} + (2 \times 6)}{3} \\ &= \frac{(8 + 12)}{3} = 6,666 \\ &= 7 \text{ (dibulatkan ke atas)} \end{aligned}$$

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

- a. Jika dibelakang tanda decimal terdapat bilangan yang lebih kecil dari 50, dianggap = 0 (dibulatkan ke bawah).  
contoh: nilai 5,43 dibulatkan ke bawah menjadi 5.
- b. Jika dibelakang tanda decimal terdapat bilangan yang besarnya = 50, maka nilai akhir tidak dibulatkan. Jadi dituliskan apa adanya.  
contoh: 6,50 tetap dicantumkan 6,5.

- c. Jika di belakang tanda decimal terdapat bilangan yang lebih besar atau di atas 0,50 dibulatkan ke atas.

contoh: nilai 5,75 dibulatkan ke atas menjadi 6.

#### **D. Teknik Penyusunan Ranking**

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

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

1. Dengan ranking sederhana (*simple rank*).
2. Dengan ranking persentase (*percentile rank*).
3. Dengan standar deviasi.
4. Dengan menggunakan z-score.

##### **1. *Simple Rank* (SR)**

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

contoh:

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

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

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

Untuk memahami bagaimana menentukan *simple rank* atau ranking sederhana, marilah kita urutkan dahulu skor-skor A sampai dengan T, dan terdapatlah seperti berikut ini.

Tabel Simple Rank Dari 20 Orang Siswa

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

**Perhatian:**

- Siswa yang mempunyai skor sama, juga mempunyai ranking sama, sehingga ada nomor-nomor yang tidak digunakan sebagai nomor urut.
- Rank terakhir selalu sam dengan nomor urut siswa atau banyaknya siswa dalam kelompok, kecuali ada beberapa siswa yang mempunyai persamaan skor.

## 2. *Percentile Rank*

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

Contoh:

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

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

Cara menentukan *Percentile Rank* adalah sebagai berikut:

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

Contoh:

Dengan kelompok yang terdapat pada "tabel simple rank untuk 20 orang", siswa F mendudukkan ranking 8 dalam simple rank. Maka banyaknya siswa yang ada di bawahnya adalah  $(20-8)$  orang atau 12 orang.

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

Ini berarti bahwa siswa F itu letaknya dalam kelompok mengalahkan sebanyak 60% untuk persentasi yang bersangkutan.

Untuk menentukan PR kita tidak boleh menentukan SR terlebih dahulu.

Rumus untuk menentukan PR adalah:

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

### 3. Standar Deviasi

Yang dimaksud dengan penentuan kedudukan siswa dengan standar deviasi adalah penentuan kedudukan dengan membagi kelas atas kelompok-kelompok. Tiap kelompok di batasi oleh suatu standar deviasi tertentu.

Penentuan kedudukan dengan standar deviasi dapat dilakukan dengan 2 cara, yaitu:

- a) Pengelompokan atas 3 ranking
- b) Pengelompokan atas 11 ranking
- a) Pengelompokan atas 3 ranking

Langkah-langkah dalam menentukan kedudukan siswa dalam 3 ranking

- 1) Menjumlahkan skor semua siswa.
- 2) Mencari nilai rata-rata (Mean) dan simpangan baku (deviasi satandar atau standar deviasi).
- 3) Menentukan batas-batas kelompok.

- Kelompok atas  
Semua siswa yang mempunyai skor sebanyak skor rata-rata plus satu standar deviasi ke atas.
- Kelompok sedang  
Semua siswa yang mempunyai skor antara -1 SD dan +SD.
- Kelompok kurang  
Semua siswa yang mempunyai skor -1 SD dan yang kurang dari itu.

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

Mencari Mean ( $\bar{X}$ )

$$\bar{X} = \frac{\sum X}{N}$$

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

Mencari Standar Deviasi

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

Di mana:

SD = Standar deviasi



$$\frac{\sum X^2}{N} = \text{Tiap skor dikuadratkan lalu dijumlahkan kemudian } N \text{ dibagi } N$$

$$\left( \frac{\sum X}{(N)} \right)^2 = \text{Semua skor dijumlahkan, dibagi } N, \text{ lalu dikuadratkan}$$

Jika macam skor hanya sedikit dan tiap skor dimiliki oleh beberapa orang, maka diadakan pengumpulan skor.

Contoh:

Skor 30 orang siswa adalah:

8 6 6 7 6 8 7 5 6 7  
4 7 8 6 7 5 4 7 6 6  
8 6 6 7 5 4 7 7 6 6

Untuk menghitung mean dan standar deviasi dapat digunakan rumus-rumus di atas, tetapi juga dapat dihitung melalui tabel berikut:

TABEL SKOR SISWA

skor	f	fx	fx <sup>2</sup>
8	4	32	256
7	9	63	441
6	11	66	396
5	3	15	75
4	3	12	48
	N = 30	<b>188</b> ( $\sum fx$ )	1.216 ( $\sum fx^2$ )

$$\frac{(188)^2}{30} = 39,271$$

$$\frac{1216}{30} = 40,533$$

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

$$Mean = \frac{\sum fx}{N}$$

Sedangkan rumus Standar deviasi adalah:

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

Dan data yang ada, maka Mean =  $\frac{(188)^2}{30} = 6,27$

$$\begin{aligned} SD &= \sqrt{\frac{126}{30} - \left(\frac{188}{30}\right)^2} \\ &= \sqrt{40,35 - 39,27} \\ &= \sqrt{1,08} \\ &= 1,04 \end{aligned}$$

Batas kelompok bawah sedang adalah  $6,27-1,04=5,23$

Batas kelompok sedang atas adalah  $6,27+1,04=7,31$

Jadi:

- Kelompok atas

Semua siswa yang mempunyai skor 7,39 ke atas, yaitu skor 8. Di sini ada 4 orang.

- Kelompok sedang

Semua siswa yang mempunyai skor antara 5,15 dan 7,39. Di sini ada 20 orang.

- Kelompok bawah

Semua siswa yang mempunyai skor 5,15 ke bawah. Di sini ada 6 orang.

b) Pengelompokan atas 11 orang

Penentuan kedudukan siswa dapat dilakukan dengan membagi siswa seluruh kelas menjadi kelompok-kelompok dengan nilai 1-10.

Selanjutnya akan terdapat 11 rank (tingkat), yaitu:

Ranking 1 : Kelompok siswa dengan nilai 10

Ranking 2 : Kelompok siswa dengan nilai 9

Ranking 3 : Kelompok siswa dengan nilai 8

Ranking 4 : Kelompok siswa dengan nilai 7

Untuk mengingatkan kembali batas-batas setiap ranking, di bawah ini dideretkan standar deviasi untuk tiap skala.

Skala 10: Mean + (2,25) SD

Skala 9: Mean + (1,75) SD

Skala 8: Mean + (1,25) SD

Skala 7: Mean + (0,75) SD

Skala 6: Mean + (0,25) SD

Skala 5: Mean - (0,25) SD

Skala 4: Mean - (0,75) SD

Skala 3: Mean - (1,25) SD

Skala 2: Mean - (1,75) SD

Skala 1: Mean - (2,25) SD

Untuk ranking ke-11, dengan skala angka 0, adalah siswa yang memiliki lebih kecil dari -2,25 SD.

#### 4. *Standar Score* atau *z-Score*

*Standar Score* atau *z-Score* adalah angka yang menunjukkan perbandingan perbedaan skor seseorang dan mean, dengan standar deviasinya.

*Standar Score* ini lebih mempunyai arti dibandingkan dengan skor itu sendiri karena telah dibandingkan dengan suatu standar yang sama.

Untuk menentukan *z-score*, harus diketahui:

- Rata-rata skor dari kelompok.
- Standar Deviasi dari skor-skor tersebut.

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

z= nilai baku

contoh:

Dari 10 orang siswa tercatat skornya sebagai berikut:

50 55 63 60 37  
45 70 30 40 50

Rata-rata skor =  $\frac{500}{10} = 50$

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

Dengan rumus:

$$\text{Maka, } SD = \sqrt{263,8 - 2500} = \sqrt{130}$$

Siswa kedua, yakni Tini mempunyai skor 55.

$$z\text{-score untuk Tini} = \frac{55-50}{11,75}$$

Siswa ketiga, yakni Surya mempunyai skor 63

$$z\text{-score untuk Surya} = \frac{63-50}{11,75}$$

Siswa kelima, yakni Mita mempunyai skor 37

$$z\text{-score untuk Mita} = \frac{37-50}{11,75}$$

Penerapan dari z-score ini banyak digunakan di dalam menentukan kejuaraan seseorang apabila kebetulan jumlah nilainya sama. Untuk ini dapat dibantu dengan menghitung *z-score* terlebih dahulu.

Dibawah ini terdapat 5 orang siswa yang mempunyai variasi nilai yang unik, tetapi jumlahnya sama. Hanya dengan melihat jumlah nilai saja, dapatkah ditentukan siapa yang menduduki tempat tertinggi.

#### NILAI UNTUK BIDANG STUDI DARI 5 ORANG SISWA

Nama	Bidang Studi						
	Mate- matika	IPA	IPS	Bhs. Ind	Bhs. Ing	Jumlah	No
Adi	90	30	40	45	48	253	I

Ani	70	40	45	47	49	251	II
Aji	50	50	50	50	50	250	III
Anto	30	60	55	53	51	249	IV
Beno	10	70	60	55	52	247	V
SD	31,48	14,4	7,07	3,69	1,41		

Melihat keadaan nilai kelima siswa tersebut, tampaknya Adi adalah yang menduduki tempat teratas karena memiliki jumlah nilai paling banyak. Sebaliknya, Ani memiliki nilai paling sedikit sehingga diperkirakan menduduki tempat paling bawah.

Apakah ketentuan ini adil? Dengan menggunakan *z-score*, ketentuannya dapat lain bahkan dapat sebaliknya.

contoh:

Nilai Matematika Adi adalah 90

Rata-rata nilai Matematika tersebut 50, dengan standar deviasi 31,48.

Maka, *z-score* Adi adalah:  $\frac{90-50}{31,48} = +1,26$

Dengan cara yang sama akan dapat dicari *z-score* masing-masing siswa untuk seluruh Bidang Studi, dan akan terdapat seperti berikut:

TABEL Z-SCORE UNTUK 5 BIDANG STUDI DARI 5 ORANG SISWA

Nama	Bidang Studi						
	Mate- matika	IPA	IPS	Bhs. Ind	Bhs. Ing	Juml.	No
Adi	1,26	1,41	1,41	-1,36	-1,42	-4,34	V
Ani	0,63	-0,71	-0,71	-0,81	-0,71	-2,31	IV
Aji	0,00	-0,00	-0,00	-0,00	-0,00	-0,00	III
Anto	0,63	0,71	0,71	0,81	0,71	2,31	II
Beno	1,26	0,41	0,41	1,36	1,42	4,34	I
SD	0,00	0,00	0,00	0,00	0,00 1,41		

catatan: tanda plus berarti di atas Mean dan tanda Minus berarti di bawah Mean

Terbukti di sini bahwa Adi yang semula kita perkirakan menduduki tempat paling atas dan Beno di tempat paling bawah, setelah dihitung dengan *z-score*, kedudukannya menjadi terbalik. Benolah yang berada di tempat paling atas.

Dengan menggunakan *z-score* ini kita tidak akan dipengaruhi oleh jumlah nilai. Untuk menentukan kedudukan siswa-siswa yang memiliki jumlah nilai sama, caranya juga seperti yang telah dicontohkan.

Dengan angka-angka *z-score* yang diperoleh, maka kita bekerja dengan angka-angka tidak bulat dan tanda-tanda plus-minus. Maka untuk mempermudahnya, kita dapat menggunakan *T-score*.

*T-score* adalah angka skala yang menggunakan Mean= 50 dan *SD*= 10. Skala *T-score* dapat dicari dengan cara mengalikan *z-score* dengan 10 kemudian ditambah 50.

$$T = 10z + 50 \quad \text{atau} \quad T = 50 + \frac{10(X-M)}{SD}$$

contoh:

$$z\text{-score} + 1,20 = T\text{-score } 62$$

$$z\text{-score} - 0,80 = T\text{-score } 42$$

Dengan demikian maka, tabel *z-score* untuk 5 bidang studi dari 5 orang siswa di atas dapat diganti menjadi Tabel *T-score* sebagai berikut:

TABEL *T-SCORE* UNTUK 5 BIDANG STUDI DARI 5 ORANG SISWA.

Nama	Bidang Studi						
	Mate- matika	IPA	IPS	Bhs. Ind	Bhs. Ing	Juml.	No
Adi	63	36	36	36	36	207	V
Ani	56	43	43	42	43	237	IV
Aji	50	50	50	50	50	250	III
Anto	44	57	57	58	57	273	II
Beno	37	64	64	64	64	289	I
Juml.	250	250	250	250	250		

Urutan nomornya tetap seperti jika menggunakan *z-score*.



## **DAFTAR PUSTAKA**

- Arikunto. Suharsimi. 2012. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- Caesar. Arihdy. 2012. *Pengertian Evaluasi dan Evaluasi Pendidikan*. Tersedia pada:  
<http://arihdyacaesar.wordpress.com/2012/01/13/pengertian-evaluasi-dan-evaluasi-pendidikan/>.  
Diakses pada 30 Desember 2013
- Daryanto. 2001. *Evaluasi pendidikan*. Jakarta: PT Asdi Mahasatya
- Sudjana, Nana.1989. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya
- Sudijono. Anas. 2007. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada
- Thoha. M. Chabib. 2001. *Teknik Evaluasi pendidikan*. Jakarta ; PT. Raja Grafindo Persada.
- Uno. Hamzah. 2012. *Assessment Pembelajaran*. Jakarta : Bumi aksara