

PENGEMBANGAN TEST PRESTASI



NINI SRIWAHYUNI, S.PSI

**FAKULTAS PSIKOLOGI
UNIVERSITAS MEDAN AREA
M E D A N
2 0 0 3**

UNIVERSITAS MEDAN AREA

PENYUSUNAN INSTRUMEN PENILAIAN HASIL BELAJAR

PENDAHULUAN

Untuk mengetahui seberapa jauh tingkat penguasaan siswa terhadap tujuan materi pelajaran yang diajarkan (TIK), diperlukan instrumen/alat ukur yang memadai. Instrumen yang dipergunakan untuk mengukur TIK dapat berupa tes dan nontes. Tentang jenis-jenis kedua instrumen tersebut telah dibicarakan pada bagian modul sebelumnya.

Pada modul ini akan diuraikan mengenai "Bagaimana menyusun instrumen untuk mengukur hasil belajar siswa". Uraian materi ini mencakup: 1) Persyaratan instrumen hasil belajar; 2) Cara penyusunan instrumen penilaian hasil belajar; dan 3) Menilai instrumen tes hasil belajar.

Dengan mengkaji ketiga pokok materi tersebut di atas secara umum Anda diharapkan dapat menyusun dan menilai instrumen tes hasil belajar dengan memperhatikan persyaratannya. Secara khusus diharapkan Anda dapat:

- a. menyebutkan persyaratan instrumen tes hasil belajar yang baik,
- b. menjelaskan pengertian tes hasil belajar yang reliabel,
- c. menjelaskan pengertian tes hasil belajar yang valid,
- d. menjelaskan pengertian tingkat kesukaran suatu tes,
- e. menjelaskan pengertian kemampuan diskriminasi suatu tes,
- f. menjelaskan bahwa instrumen penilaian hasil belajar harus praktis,
- g. menjelaskan bahwa suatu tes yang reliabel belum tentu valid,
- h. membuat daftar klasifikasi tujuan-tujuan instruksional ranah kognitif untuk keperluan suatu unit pengajaran pelajaran tertentu,
- i. menyusun suatu tabel spesifikasi tes sumatif dari suatu pelajaran tertentu,
- j. menyebutkan beberapa saran yang harus diperhatikan dalam penyiapan suatu tes hasil belajar,
- k. menjelaskan bahwa tes esei memiliki validitas dan reliabilitas yang rendah,
- l. membuat beberapa pertanyaan tes esei yang mengungkap beberapa aspek kognitif,
- m. menyebutkan beberapa bentuk tes yang termasuk dalam tes esei,
- n. menyebutkan beberapa bentuk tes yang termasuk dalam tes objektif,
- o. membuat beberapa contoh bentuk tes objektif yang baik,
- p. menjelaskan kelemahan dan kelebihan tes objektif secara umum,
- q. menyusun suatu alat non-tes untuk mengukur hasil belajar siswa,
- r. menjelaskan persyaratan-persyaratan dalam uji coba tes,
- s. menghitung besarnya skor tes objektif dengan formula pembentukan (*correction for chance*), untuk:
 1. tes objektif dengan option jawaban,
 2. tes objektif dengan 4 option jawaban, dan
 3. tes objektif dengan 2 option jawaban.
- t. mencari validitas suatu tes/item tes hasil belajar,
- u. mencari reliabilitas tes hasil belajar,
- v. mencari tingkat kesukaran suatu butir tes,
- w. mencari kemampuan diskriminasi suatu butir tes,
- x. menafsirkan apakah suatu butir tes itu dapat digunakan atau tidak sebagai alat ukur.

1. PERSYARATAN INSTRUMEN PENILAIAN HASIL BELAJAR

Pada bagian modul sebelumnya telah dijelaskan mengenai instrumen penilaian hasil belajar jenis tes dan nontes secara panjang lebar. Pada bagian modul ini akan dibicarakan tentang persyaratan yang harus dimiliki oleh instrumen tersebut. Ada beberapa persyaratan yang dituntut pada perangkat instrumen hasil belajar ini, yaitu suatu instrumen penilaian hasil belajar (tes) harus (1) valid, (2) reliabel, (3) memiliki tingkat kesukaraan yang memadai dan memiliki kemampuan diskriminasi yang baik, dan (4) bersifat praktis untuk diterapkan (*practicable*).

INSTRUMEN TES PENILAIAN HASIL BELAJAR HARUS VALID

Instrumen yang hendak digunakan untuk mengukur keberhasilan belajar siswa harus valid. Jika instrumen penilaian digunakan untuk mengukur tingkat pencapaian suatu tujuan instruksional, maka instrumen harus dirancang untuk mengukur hal itu. Umpama saja seorang guru hendak mengukur tujuan instruksional dengan mengacu pada rumusan perilaku yang dikembangkan oleh Bloom, apakah itu jenis perilaku aspek kognitif, afektif ataupun psikomotor, maka instrumen yang dibuat harus sesuai dengan tingkat dan jenis aspek perilaku tersebut. Sebagai contoh, jika tujuan instruksional yang hendak diukur adalah aspek kognitif jenis aplikasi maka instrumen penilaiannya harus dapat meneliti apakah siswa telah menunjukkan prestasi dalam kemampuan ini.

Untuk mengukur aspek kognitif jenis aplikasi di atas dapat dibuatkan pertanyaan tes uraian seperti "Hitunglah validitas internal instrumen tes bidang studi IPS, berdasarkan data hasil uji-cobanya di bawah ini?" Dalam hal ini siswa akan menerapkan (mengaplikasikan) pengetahuan dan pemahamannya terhadap konsep validitas internal, serta menerapkan rumus statistik untuk mengolah data, sekaligus membuat kesimpulan atas hasil dari perhitungannya. Jika pertanyaan yang disusun bukan untuk mengungkap kemampuan aplikasi siswa, maka tes tersebut tidak valid, karena siswa tidak memberikan macam kemampuan yang dimaksud.

Suatu tes sumatif yang direncanakan untuk mengukur beberapa jenis aspek kognitif dan isi tujuan instruksional, maka tes harus berupa pertanyaan-pertanyaan atau tugas-tugas yang menuntut prestasi dengan menampakkan aspek perilaku dan isi yang memadai. Dengan kata lain instrumen penilaian harus dapat mengungkap kemampuan siswa baik berupa aspek perilaku maupun materi ajar. Sebagai contoh adalah pertanyaan "*Jelaskan* bahwa salah satu syarat tes hasil belajar harus valid!" Pertanyaan dengan kata kerja "*jelaskan*" adalah kata yang operasional yang menuntut siswa untuk menampakkan kemampuan memberikan informasi verbal/tulis tentang aspek kognitif "memahami". Pemahaman siswa menyangkut materi tertentu yaitu dalam mata kuliah Penilaian Hasil Belajar atau dalam Metodologi Penelitian, berarti sesuai dengan aspek isi tujuan instruksional dalam mata kuliah tersebut. Tetapi pertanyaan tersebut tidak akan sesuai untuk meneliti jenis perilaku atau jenis prestasi aspek kognitif yang lainnya, seperti "analisis". Karena untuk meneliti kemampuan analisis perlu instrumen butir tes yang lain. Misalnya untuk mengungkap aspek analisis ini digunakan tes esei, maka bentuk pertanyaannya dapat seperti ini "Bandingkanlah efektivitas belajar sendiri dengan belajar kelompok!". Dalam hal ini siswa akan mengidentifikasi unsur-unsur, yang ada pada cara belajar sendiri dan belajar kelompok, mencari kelemahan dan kelebihan masing-masing baru dapat membandingkannya.

Dengan demikian dapat dikatakan bahwa suatu instrumen penilaian hasil belajar disebut valid, jika instrumen tersebut benar-benar dapat mengukur apa yang seharusnya diukur. Ada beberapa jenis validitas antara lain validitas tampak (*face validity*) yang berarti bahwa suatu instrumen yang disusun/dibuat tampak dari luar sudah valid; validitas logika (*logical validity*), yaitu instrumen yang dibuat diperkirakan akan mengukur apa yang hendak diukur. Kedua jenis validitas tersebut sebenarnya sama, hanya yang disebut pertama kurang enak didengarkan. Jenis validitas yang lain, yaitu validitas isi yang juga dapat disebut dengan validitas kurikulum, sebab instrumen yang disusun sebagai alat evaluasi didasarkan pada materi atau isi

kurikulum. Selanjutnya adalah validitas empiris (*empirical validity*) yaitu alat/instrumen yang disusun setelah diuji-cobakan hasilnya dibandingkan dengan hasil pengukuran dengan instrumen lain yang telah valid. Dekat dengan jenis validitas empiris ini adalah validitas statistik. Disebut dengan validitas statistik (*statistical validity*) karena dalam mencari seberapa tingkat validitas suatu instrumen, digunakan analisis statistik. Jenis validitas statistik ini akan dijelaskan pada bagian berikutnya. Terakhir adalah validitas konsep/konstruksi (*construct validity*), yaitu instrumen yang dibuat diasumsikan telah dapat mengukur konstruk psiko-sosial tertentu. Misalkan konstruk psikologis yang hendak diukur adalah "Kedisiplinan siswa", maka instrumen yang dibuat benar-benar harus mengukur prestasi disiplin siswa. Misalkan faktor-faktor disiplin tersebut adalah disiplin dalam belajar, mengerjakan tugas sekolah, bekerja di rumah, melakukan kegiatan dimasyarakat dan sebagainya. Faktor-faktor disiplin tersebut harus dapat diungkap dengan instrumen penilaian disiplin. Bila faktor-faktor tersebut telah dirumuskan dalam bentuk pertanyaan-pertanyaan maka instrumen tersebut telah dianggap memiliki validitas konstruk.

INSTRUMEN PENILAIAN HASIL BELAJAR HARUS RELIABEL

Reliabilitas suatu alat/instrumen penilaian hasil belajar adalah untuk menghasilkan gambaran tentang tingkat prestasi belajar yang benar-benar dapat dipercaya (terandalkan). Jika alat penilaian hasil belajar berbentuk tes objektif yang jumlah itemnya cukup banyak, maka hasil tes pertama kali yang diperoleh siswa, harus hampir sama dengan hasil tes di waktu mendatang dari alat tes yang sama. Maksudnya jika suatu perangkat tes objektif dikerjakan oleh sekelompok siswa kemudian hasil (skor-skor) siswa tersebut diranking dari yang paling rendah ke yang paling tinggi selanjutnya dibandingkan dengan skor-skor siswa yang sama dari hasil mengerjakan tes yang sama tetapi pada waktu yang berbeda, maka hasil perbandingan tersebut harus relatif sama. Selain menggunakan cara membandingkan ranking skor tes pertama dengan tes kedua (cara yang paling sederhana) untuk mencari atau mengetahui reliabilitas tes, ada cara lain yaitu dengan analisis statistik. Cara menghitung koefisien reliabilitas ini, yaitu dengan mengkorelasikan skor tes pertama tersebut dengan skor tes kedua dengan rumus korelasi product moment. Hasil dari perhitungan tersebut disebut dengan *koefisien stabilitas* atau test-retest reliability. Koefisien reliabilitas tersebut biasa dipenuhi oleh tes-tes terstandar seperti tes bakat, inteligensi atau tes kepribadian, namun juga tidak salah jika dimiliki oleh tes hasil belajar.

Jenis koefisien reliabilitas tes yang lain adalah *koefisien ekuivalen* dan *konsistensi internal*. Suatu alat tes memiliki koefisien reliabilitas ekuivalen, jika ada dua jenis perangkat tes yang paralel memiliki korelasi yang tinggi. Jika tes tersebut adalah tes hasil belajar maka keparalelan yang dimaksud adalah dari aspek *isi (materi)*, *tingkat kesukaran*, dan *tingkat atau jenis aspek kognisi yang diukur*. Hal ini biasa dilakukan karena sangat luasnya materi, sehingga tidak mungkin keseluruhan materi tersebut diukur dengan suatu tes yang terdiri atas ratusan butir/item soal dalam suatu waktu. Maka dengan membuat tes yang paralel, salah satunya dapat digunakan atau digunakan satu untuk kelas A sedangkan yang lain untuk kelas B.

Koefisien reliabilitas terakhir yaitu konsistensi internal. Suatu alat tes memiliki reliabilitas jenis konsistensi internal karena angka koefisien korelasinya (r) diperoleh dari korelasi satu perangkat tes yang *dibelah dua*. Misalkan suatu perangkat tes dibelah menjadi dua, yaitu kelompok tes yang bernomor genap dan kelompok lain adalah butir tes yang bernomor ganjil. Karena koefisien korelasi dari perhitungan ini baru diperoleh koefisien reliabilitas separoh tes, maka perlu dikoreksi dengan rumus lain untuk mendapatkan koefisien korelasi yang utuh.

Penghitungan tentang koefisien reliabilitas baik untuk koefisien stabilitas, ekuivalen maupun konsistensi internal akan dijelaskan pada bagian selanjutnya.

INSTRUMEN PENILAIAN HASIL BELAJAR HARUS MEMILIKI TINGKAT KESUKARAN TERTENTU DAN KEMAMPUAN DISKRIMINASI YANG BAIK

Suatu persyaratan tes yang lain adalah jika tes tersebut memiliki tingkat kesukaran tertentu dan memiliki kemampuan diskriminasi. Untuk mengetahui tingkat kesukaran suatu tes maupun kemampuan diskriminasi-

sinya, dilakukan dengan analisis soal-soal atau item-item soal. Hasil dari analisis keduanya dapat dibuat kesimpulan "apakah suatu item telah memenuhi fungsinya atau belum. Jika suatu item soal belum bisa memenuhi fungsinya dengan baik perlu direvisi, bahkan bila perlu tidak lagi digunakan. Sebenarnya analisis terhadap item-item tes hasil belajar ini merupakan suatu keharusan, namun di lapangan, guru yang melakukannya dapat dikatakan seribu-satu. Dengan kata lain soal-soal tes yang digunakan guru di sekolah baik atau buruk, memenuhi syarat atau tidak, tetap dianggap baik dan memenuhi syarat. Akibatnya anak yang selalu dikambinghitamkan, artinya jika anak mendapat skor jelek dalam suatu tes dia dianggap kurang mampu (goblok = bodoh) meskipun yang salah kadang tesnya yang tidak baik.

Untuk pekerjaan analisis item memang tidak ringan bagi guru, tetapi jika kita ingat bahwa konsekuensi dari penggunaan berhubungan dengan nasib siswa, seharusnya tetap dilakukan. Dengan kata lain kita memiliki tanggung jawab moral untuk melakukan analisis item terhadap soal-soal yang kita gunakan. Bahkan untuk soal tes terstandar, analisis item harus dan wajib dilakukan, sebab tes jenis ini akan digunakan secara umum dan oleh orang banyak. Tes terstandar seperti tes inteligensi, bakat maupun kepribadian ini, dapat diibaratkan sebagai alat ukur seperti *timbangan* atau *meteran* yang dipasarkan dan digunakan oleh masyarakat, sehingga pembuat alat tersebut harus ikut bertanggung jawab terhadap alat ukur tersebut akan kesahannya.

Pengertian *tingkat kesukaran* suatu soal tes adalah berkaitan dengan jumlah siswa yang dapat mengerjakannya dengan benar, yang sebetulnya juga dapat disebut *tingkat kemudahannya*, sehingga dalam Bahasa Inggris diistilahkan dengan *level of difficulty* atau *degree of succes*. Suatu tes dapat dikatakan baik jika tes tersebut tidak terlalu sukar dan tidak terlalu mudah. Tes yang terlalu sukar tidak dapat mengungkap apa yang telah diketahui siswa, demikian juga tes yang terlalu mudah tidak mampu mengungkap apa yang belum diketahui siswa. Dengan kata lain tes yang terlalu sukar atau terlalu mudah tidak dapat melakukan fungsinya, sehingga guru tidak dapat membuat kesimpulan apa-apa dari hasil tes tersebut.

Persyaratan lain bagi suatu tes, yaitu apabila suatu tes digunakan mampu membedakan siswa yang pandai dengan siswa yang tidak pandai, inilah yang disebut dengan *kemampuan diskriminasi*. Untuk mencari besarnya persentase tingkat kesukaran tes maupun kemampuan diskriminasi akan mudah dilakukan terhadap tes-tes bentuk objektif. Analisis terhadap tingkat kesukaran dan diskriminasi ini akan dibicarakan pada bagian selanjutnya.

INSTRUMEN PENILAIAN HASIL BELAJAR HARUS BERSIFAT PRAKTIS

Suatu instrumen penilaian hasil belajar khususnya tes harus cukup praktis untuk digunakan. Meskipun demikian kepraktisan ini tidak harus mengalahkan persyaratan validitas dan reliabilitas suatu instrumen. Kepraktisan suatu alat atau instrumen ini dikaitkan dengan kemudahan penggunaannya termasuk juga dalam proses penskoran atau penilaiannya, serta pada penggunaan hasilnya, yaitu apakah memadai atau tidak sebagai alat ukur.

Berdasarkan berbagai persyaratan tersebut di atas dapat dikatakan, bahwa:

- 1) Suatu alat/instrumen penilaian hasil belajar yang digunakan harus sesuai dengan aspek prestasi yang ingin dihasilkan. Suatu alat yang sesuai untuk mengukur aspek kognitif, tidak mesti cocok untuk mengukur aspek afektif atau psikomotor, begitu pula sebaliknya.
- 2) Instrumen atau alat penilaian hasil belajar harus dapat mengungkap tujuan-tujuan instruksional secara representatif. Bila suatu tujuan instruksional yang ingin dicapai begitu banyak, maka soal-soal atau tugas-tugas yang terdapat pada soal tes harus merupakan pertanyaan-pertanyaan yang terpilih, sehingga representatif bagi tujuan-tujuan instruksional yang ada. Pertanyaan-pertanyaan yang representatif ini disebut dengan istilah *sampling adequacy*.

3) Setiap tugas atau pertanyaan pada instrumen penilaian hasil belajar harus dibuat sedemikian rupa sehingga menghasilkan prestasi yang harus ada pada diri siswa. Prestasi tersebut dapat dinampakkan oleh siswa sebagai kemampuan internal yang telah dimilikinya. Dalam hal ini satu pertanyaan atau tugas dalam suatu tes hasil belajar harus mempergunakan kata-kata kerja yang operasional,

- 4) Setiap tugas atau pertanyaan instrumen penilaian hasil belajar yang telah dikerjakan siswa, harus dapat dikoreksi dengan pedoman tertentu. Andai kata ada satu pertanyaan tes uraian yang dikerjakan oleh seorang siswa kemudian dikoreksi oleh dua orang korektor, maka koreksi tersebut harus menghasilkan skor yang sama. Hal seperti ini tidak akan bisa terjadi bila pedoman untuk mengoreksi butir pertanyaan tersebut tidak dipersiapkan terlebih dahulu.

Selanjutnya untuk memantapkan pemahaman Anda terhadap uraian materi ini, cobalah kerjakan latihan-latihan berikut ini:

- 1) a. Di antara beberapa jenis validitas, yang mana yang harus dimiliki oleh instrumen tes hasil belajar? Jelaskan!
b. Bagaimana agar tes tersebut memiliki validitas yang dimaksud?
- 2) Jelaskan bahwa suatu alat tes yang reliabel belum pasti valid?
- 3) Apa artinya jika item tes memiliki tingkat kesukaran 95%?
- 4) Salah satu persyaratan instrumen penilaian hasil belajar adalah harus cukup praktis. Jelaskan apa maksudnya!
- 5) Kadang-kadang guru ingin mengukur hasil belajar seperti "keaktifan dalam praktek di laboratorium"; maka instrumen jenis apa yang cocok digunakan serta harus memiliki validitas apa instrumen tersebut?

Bila Anda telah mengerjakan latihan-latihan di atas, cobalah cocokkan dengan petunjuk-petunjuk berikut ini:

- 1) Suatu alat ukur dituntut harus valid, artinya harus mengukur aspek yang hendak diukur. Tes hasil belajar adalah alat ukur untuk mengukur hasil belajar siswa terhadap *materi* atau *bahan* yang telah dia pelajari. Maka tes hasil belajar harus dapat mengungkap *isi materi* yang biasanya telah dirumuskan dalam tujuan instruksional. Apabila pertanyaan atau tugas yang harus dijawab atau dikerjakan oleh siswa telah secara representatif mengukur isi materi atau tujuan instruksional yang dirumuskan berarti instrumen tersebut telah valid, dan validitas semacam ini dinamakan dengan validitas isi atau validitas kurikulum (*content or curriculum validity*). Dengan kata lain instrumen tes hasil belajar harus memiliki validitas isi.
- 2) Untuk menjawab latihan no. 2 ini, coba Anda pahami ilustrasi berikut ini:
"Pak Ali (Guru kelas III-SMA) menyusun instrumen tes hasil belajar Matematika, yang materinya diambil dari GBPP bidang pelajaran Matematika kelas III SMA. Setelah tes tersebut digunakan untuk siswa-siswa kelas III SMA yang dia ajar hasilnya adalah 95% siswa dapat mengerjakan dengan benar. Karena tes yang dia susun dianggap terlalu mudah kemudian dia cobakan lagi pada kelas III SMA yang lain hasilnya juga sama, yaitu sebagian besar siswa dapat mengerjakan benar di atas 95%. Hal ini dia cobakan lagi pada kelas lain yang paralel, hasilnya tidak jauh berbeda. Akhirnya dia mencoba mempelajari soal-soal yang telah dibuatnya tersebut. Ternyata soal-soal yang ada pada tes tersebut hanya berupa pertanyaan-pertanyaan tentang pokok bahasan tertentu saja pada GBPP Matematika tersebut. Dengan kata lain pertanyaan-pertanyaan tes tersebut kurang representatif dalam mengukur bidang pelajaran Matematika yang telah diajarkan". Selanjutnya, berdasarkan ilustrasi di atas apa yang Anda dapat simpulkan?
- 3) Suatu item tes memiliki tingkat kesukaran 95%, berarti 95% siswa yang dites atau yang mengerjakan item tes tersebut dapat menjawabnya dengan benar. Item tes yang demikian tidak dapat membedakan mana kelompok siswa yang pandai dan yang tidak pandai. Dengan kata lain tes yang demikian tersebut tidak dapat berfungsi dengan baik. Terutama bila tes tersebut untuk tujuan seleksi, tetapi tidak begitu berarti untuk tujuan diagnosis. Tes yang terlalu mudah ataupun yang terlalu sukar mempunyai

kelemahan yang sama, yaitu tidak dapat membedakan mana siswa yang pandai dan tidak pandai. Tes dianggap baik bila memiliki tingkat kesukaran mendekati 50%.

- 4) Instrumen penilaian hasil belajar harus cukup praktis digunakan (*practicable*), artinya mudah digunakan dan dapat memperoleh hasil yang diharapkan. Biasanya alat penilaian hasil belajar yang bersifat *fracticable* ini dimiliki oleh tes-tes bentuk objektif. Tes-tes bentuk objektif ini mudah dalam penskorannya, dalam proses pemberian nilai serta dapat mengukur materi/tujuan instruksional secara lebih representatif. Hal ini berbeda dengan jenis tes uraian. Tes uraian hanya mengukur sedikit materi atau tujuan intruksional yang dirumuskan. Kemungkinan besar terjadi ketidakkonsistenan dalam memberi skor antara siswa satu dengan lainnya.
- 5) "Keaktifan dalam bekerja di laboratorium" adalah suatu konstruk psikologis. Gejala-gejala konstruk psikologis tersebut dapat dikenali melalui kegiatan-kegiatan atau perilaku motorik, seperti: mulai bekerja tepat pada waktunya; tidak banyak istirahat selama jam praktek belum selesai; membantu teman-teman lain yang memerlukannya; menunjukkan hasil kerja yang lebih baik daripada teman-temannya dan sebagainya.
Oleh sebab itu untuk mengukur aspek tersebut akan cocok bila digunakan alat atau instrumen nontes yaitu dengan "form observasi dengan skala-rating (*rating scale*)".
Karena yang diukur merupakan konstruk psikologis maka instrumennya harus memiliki validitas konstruk. Cara menyusun validitas konstruk tersebut yaitu dengan menjabarkannya ke dalam indikator-indikator yang selanjutnya diuraikan menjadi item-item pertanyaan atau pernyataan. Untuk lebih jelasnya Anda pelajari tentang cara menyusun alat nontes pada bagian berikut dari modul ini.

RANGKUMAN

- Menyusun instrumen penilaian yang baik merupakan suatu keharusan bagi setiap guru di sekolah.
- Instrumen penilaian hasil belajar dikatakan baik, bila instrumen tersebut valid, reliabel, memiliki tingkat kesukaran antara 10 - 90%, serta praktis untuk digunakan.
- Untuk instrumen penilaian hasil belajar, validitas yang harus dimiliki adalah validitas isi atau kurikulum sebab instrumen yang disusun harus dapat mengungkap prestasi siswa akan penguasaan materi atau isi kurikulum.

TES FORMATIF 1

Petunjuk : Silanglah salah satu *option* A, B, C, atau D yang Anda anggap paling benar pada setiap item soal berikut ini.

- 1) Suatu instrumen tes dibuat untuk tujuan meramalkan keberhasilan atau kegagalan seseorang pada studi lanjut, maka dipersyaratkan tes tersebut memiliki validitas.
A. kurikulum
B. empiris
C. konstruk
D. logik.
- 2) Tes-tes jenis inteligensi atau bakat khusus harus memiliki validitas

UNIVERSITAS MEDAN AREA

- A. prediktif
 - B. isi
 - C. konstruk
 - D. logik.
- 3) Berikut ini ciri item tes yang reliabel tetapi tidak valid, *kecuali* item tes tersebut
- A. memiliki tingkat kesukaran kurang dari 10%
 - B. memiliki tingkat kesukaran lebih dari 90%
 - C. disusun berdasarkan salah satu tujuan instruksional yang dirumuskan
 - D. disusun dengan mengacu pada materi untuk siswa kelas yang lebih tinggi.
- 4) Seorang guru ingin mengungkap kemampuan *kognitif pengetahuan* siswanya, maka pertanyaan yang valid adalah
- A. Hitunglah tingkat kesukaran item soal no 5, dengan mendasarkan pada data hasil uji coba berikut ini!
 - B. Bagaimanakah hubungan antara pengukuran dengan penilaian hasil belajar?
 - C. Kemukakan definisi pengukuran menurut Stake!
 - D. Uraikan unsur-unsur penting yang terdapat pada definisi pengukuran yang dikemukakan oleh Stake!
- 5) Apabila seorang guru ingin mengungkap kemampuan kognitif analisis siswa maka bentuk pertanyaan berikut inilah yang valid
- A. Hitunglah tingkat kesukaran item soal no. 5, dengan mendasarkan pada data hasil uji coba berikut ini!
 - B. Bagaimana hubungan antara pengukuran dengan penilaian hasil belajar?
 - C. Kemukakan definisi pengukuran menurut Stake!
 - D. Uraikan unsur-unsur penting yang terdapat pada definisi pengukuran menurut Stake!
- 6) Reliabilitas tes diperoleh dengan cara mengkorelasikan 2 tes yang paralel, maka koefisien yang diperoleh adalah
- A. stabilitas
 - B. ekuivalen
 - C. konsistensi
 - D. reliabilitas isi.
- 7) Suatu alat tes akan memiliki koefisien korelasi konsistensi internal jika kita mengkorelasikan
- A. suatu perangkat tes dengan tes lain yang telah terstandar
 - B. suatu perangkat tes dengan tes lain yang dianggap sama
 - C. antara-item tes
 - D. hasil test-retest.
- 8) Suatu item tes dapat dikerjakan dengan benar oleh 75% siswa kelompok bawah dan 20% siswa kelompok atas; artinya item tes tersebut
- A. memiliki tingkat kesukaran yang rendah
 - B. tidak valid
 - C. memiliki tingkat kesukaran yang tinggi
 - D. tidak memiliki kemampuan diskriminasi.
- 9) Suatu item tes dapat dikerjakan dengan benar oleh 75% siswa kelompok atas dan 20% untuk siswa kelompok bawah; artinya item tes tersebut
- A. memiliki kemampuan diskriminasi yang baik
 - B. tidak memiliki kemampuan diskriminasi

- C. memiliki tingkat kesukaran antara 20 - 70%
- D. memiliki reliabilitas yang tinggi.

- 10) Tes untuk mengukur materi pelajaran yang sangat luas tidak harus berisikan item-item soal yang banyak tetapi dapat dilakukan dengan cara menyusun item-item tes :
- A. berupa pertanyaan-pertanyaan yang mengukur pokok-pokok materi saja
 - B. harus memiliki tingkat kesukaran yang tinggi
 - C. yang paralel
 - D. harus mengungkap masing-masing bahasan secara representatif.

Bila Anda telah selesai mengerjakan tes formatif tersebut selanjutnya cocokkanlah dengan kunci jawaban yang tersedia di bagian akhir modul ini. Kemudian gunakanlah rumus berikut ini, untuk mengetahui tingkat penguasaan Anda terhadap materi ini.

Rumus:

$$\text{Tingkat penguasaan} = \frac{\text{jumlah jawaban benar}}{10} \times 100\%$$

Makna klasifikasi tingkat penguasaan materi

90%	-	100%	=	baik sekali
80%	-	89%	=	baik
79%	-	79%	=	sedang/cukup
<	-	70%	=	kurang.

Jika tingkat penguasaan Anda adalah 80% ke atas, berarti Anda dapat meneruskan mempelajari materi modul berikutnya. Bagus! Tetapi jika tingkat penguasaan Anda terhadap materi ini kurang dari 80%, Anda harus mengulang lagi materi ini, terutama pada bagian yang belum dikuasai.

2. CARA MENYUSUN INSTRUMEN PENILAIAN HASIL BELAJAR

Pada bagian yang lalu telah dibicarakan mengenai persyaratan-persyaratan suatu instrumen penilaian hasil belajar khususnya tes secara umum. Pada bagian ini akan dilanjutkan dengan membahas bagaimana cara menyusun instrumen penilaian hasil belajar khususnya instrumen tes, dan sedikit tentang instrumen nontes. Pertimbangan ini dibuat karena alat/instrumen tes lebih sering dan banyak digunakan sebagai teknik untuk mengukur hasil belajar siswa.

Terdapat empat langkah penting penyusunan instrumen tes hasil belajar, yaitu (1) merencanakan tes; (2) menyiapkan tes. (3) mencobakan tes; dan (4) menilai tes. Kedua langkah yang disebut pertama dan kedua dibicarakan pada bagian ini, selanjutnya diikuti dengan cara membuat tes hasil belajar. Tentang langkah ketiga dan keempat akan dijelaskan pada bagian/pokok bahasan selanjutnya.

MERENCANAKAN TES HASIL BELAJAR

UNIVERSITAS MEDAN AREA

Jika suatu instrumen tes diharapkan dapat berfungsi sesuai dengan tujuannya, haruslah tes tersebut direncanakan dengan seksama.

Hal-hal yang harus diperhatikan dalam merencanakan suatu perangkat tes hasil belajar, adalah: *Jelas apa yang hendak diukur ; untuk apa skor-skor hasil pengukuran tersebut ; dan dalam kondisi yang bagaimana* suatu tes diberikan. Apa yang hendak diukur berkaitan dengan tujuan yang hendak dicapai melalui penilaian, sedangkan untuk apa skor-skor tersebut berkaitan dengan keputusan yang hendak dibuat dari kegiatan penilaian, dan yang terakhir berkaitan dengan penyelenggaraannya.

Untuk dapat memenuhi hal-hal di atas ada ketentuan yang harus dipenuhi oleh tes hasil belajar. *Pertama*, tes harus dapat mengukur semua aspek hasil belajar yang penting. *Kedua*, tes harus mengukur lebih banyak aspek-aspek yang ditekankan dalam suatu pelajaran. *Ketiga*, sifat tes harus sesuai dengan tujuannya. *Terakhir*, sifat tes harus mencerminkan kondisi-kondisi di mana tes diberikan.

TES HARUS MENGUKUR SEMUA ASPEK HASIL BELAJAR

Seorang guru harus tahu persis mengenai tujuan lembaga pendidikan di mana dia bekerja, terutama tujuan-tujuan dari setiap bidang mata pelajaran. Kemajuan-kemajuan atau perubahan-perubahan tingkah laku siswa apa sajakah yang diharapkan dicapai oleh suatu mata pelajaran. Hal semacam ini seharusnya tidak saja ada dalam kesadaran guru melainkan juga harus terumuskan secara teliti. Perubahan-perubahan tingkah laku tersebut dapat menyangkut aspek-aspek pengetahuan, keterampilan, sikap dan kebiasaan-kebiasaan siswa. Mengenai aspek-aspek ini Anda telah mempelajarinya pada modul 2 mata kuliah ini, yaitu tentang tujuan-tujuan instruksional yang dikemukakan oleh Bloom dan kawan-kawan. Selanjutnya agar tujuan-tujuan itu dapat diukur keberhasilannya setelah proses belajar mengajar dengan instrumen tes, hendaknya rumusan tujuan instruksionalnya operasional, konkret, realistik, dan tidak muluk-muluk.

Bagaimana merencanakan suatu tes untuk sebuah unit atau pokok-pokok bahasan yang telah diajarkan? Berikut ini saran-saran atau anjuran yang dapat dilakukan oleh para guru.

Sebuah tes yang meliputi suatu unit atau pokok-pokok bahasan tertentu hendaknya disiapkan oleh dua orang guru atau lebih. Langkah-langkah yang harus dilakukan sebagai berikut:

- a) Dua orang guru atau lebih bekerja secara sendiri-sendiri mengklasifikasi tujuan ke dalam tiga ranah/ aspek menurut Bloom (kognitif, afektif, dan psikomotor). Masih bekerja sendiri-sendiri dia mengklasifikasi tujuan-tujuan kognitif ke dalam skema yang formatnya dapat seperti berikut ini.

Format : Klasifikasi tujuan-tujuan ranah kognitif mata pelajaran "x"

Klasifikasi Taksonomi Bloom	Bobot Persentase Tujuan	Tujuan Instruksional	Jumlah item tes objektif	Persentase item
Kognitif				
C ₁	...	1, 2, 3 ... n
C ₂	...	1, 2, 3 ... n
C ₃	...	1, 2, 3 ... n
↓	...	↓
C ₄	...	1, 2, 3 ... n
...
...
...
	100%		Σ ⁿ	100%

Bila dijumpai tujuan-tujuan yang bersifat umum dan masih bersifat kompleks hendaknya lebih dispesifikasikan. Tujuan-tujuan yang masih ganda hendaknya dapat ditulis kembali menjadi 2-3 tujuan. Selanjutnya para guru dapat saling membandingkan atau mencocokkan tujuan-tujuan yang telah dia klasifikasi. Untuk berikutnya diadakan kesepakatan tentang klasifikasi tujuan tersebut. Kegiatan semacam ini akan cocok dilakukan oleh guru-guru kelas paralel atau guru-guru yang mengajar bidang studi yang sama.

- b) Setelah tujuan diklasifikasikan, guru selanjutnya mempertimbangkan alokasi waktu untuk mencapai tiap-tiap tujuan. Waktu keseluruhan untuk satu unit pelajaran termasuk kegiatan belajar kelompok di kelas atau mengerjakan pekerjaan rumah dinyatakan dengan 100 persen. Dengan demikian setiap tujuan mendapatkan persentase yang ekuivalen dengan kepentingan atau penekannya. Mempertimbangkan alokasi waktu untuk suatu tujuan ini, siswa dapat diminta pendapatnya.
- c) Guru-guru bekerja ke arah pencapaian tujuan, dengan cara-cara atau prosedur yang dianggap paling cocok untuk menghasilkan perubahan tingkah laku siswa (kognitif, efektif atau psikomotor). Langkah ini sangat penting dalam proses belajar mengajar.
- d) Langkah yang terakhir adalah guru harus memikirkan prosedur-prosedur penilaian yang sesuai untuk mengukur keberhasilan pencapaian tujuan yang ditetapkan. Idealnya dengan prosedur-prosedur ini diharapkan tidak tercurah pada ranah kognitif saja, akan tetapi juga ranah-ranah yang lain. Untuk kejelasannya sementara dibicarakan dahulu mengenai tujuan ranah kognitif.

Soal-soal dalam tes untuk mengungkap tingkat pencapaiannya dalam ranah kognitif, harus berupa pertanyaan atau tugas yang sesuai dengan tingkat tujuan yang telah dirumuskan sebelumnya.

Sebaiknya dalam menyusun tes ini dikerjakan oleh dua orang guru yang bekerja secara terpisah. Hal ini memberi kesempatan pada guru lain untuk mengoreksi apakah tujuan yang telah ditetapkan sudah terukur oleh pertanyaan-pertanyaan dalam tes yang dibuat. Manfaat dari cara ini (*Cross Classification*) adalah :

- (a) untuk dapat saling memeriksa kembali sesuai tidaknya soal dalam mengukur tujuan;
- (b) untuk memastikan bahwa antara tujuan dengan soal ada pada tingkat kognitif yang sama dan
- (c) untuk mendapatkan beberapa soal tes untuk suatu tujuan tertentu.

Berdasarkan uraian di atas beberapa kesimpulan tentang langkah-langkah untuk penulisan tes-tes prestasi belajar dapat disebutkan di sini:

- 1) Susunlah tujuan-tujuan pengajaran dalam suatu daftar.
- 2) Klasifikasikanlah tujuan-tujuan tersebut sesuai dengan bidang dan tingkatannya dalam suatu form.
- 3) Tentukanlah bobot relatif setiap tujuan tersebut .
- 4) Tentukanlah pengalaman-pengalaman belajar yang dapat membawa ke arah tercapainya tujuan tersebut
- 5) Pilihlah prosedur-prosedur penilaian yang sesuai dengan bidang dan tingkatan tujuan yang hendak dicapai.

TES HARUS MENGUKUR LEBIH BANYAK ASPEK-ASPEK YANG DITEKANKAN DALAM SUATU PELAJARAN

Agar diperoleh keseimbangan antara jumlah soal dan tujuan yang hendak dicapai perlu dibuat *job analysis* atau *tabel of specifications* yang akan menjadi pedoman dalam menulis soal-soal tes.

Dalam pembuatan tabel spesifikasi hendaknya diperhatikan tujuan-tujuan yang ditekankan untuk dicapai. Misalnya dalam mata kuliah *Penilaian hasil belajar* pada pokok bahasan tertentu menekankan pada menghafal fakta dan memahami konsep-konsep. Maka pada tabel spesifikasi pada pokok bahasan tersebut harus dituliskan beberapa item tes yang mengukur tingkat kognitif tersebut. Jenis soal, bobot skor/item atau

UNIVERSITAS MEDAN AREA

Mengenai tipe soal (esai, pilihan, mencari pasangan, melakukan sesuatu dan sebagainya) biasanya tergantung pada tujuan yang hendak dicapai. Tabel spesifikasi juga harus menggambarkan perbandingan

Aspek yang ditekankan (kognitif, afektif, psikomotor). Mengenai form tabel spesifikasi ini dapat dicontohkan sebagai berikut:

Tabel Spesifikasi : Penilaian Hasil Belajar
Pokok Bahasan : Penyusunan Instrumen Penilaian Hasil Belajar

MATERI	ASPEK YANG DIUKUR	JENIS TES DAN JUMLAH BUTIR SOAL			JUMLAH TOTAL
		MEMASANGKAN	PILIHAN GANDA	ESEI	
I. Persyaratan Inst. Penilaian Hasil Belajar	Kognitif (C)				
	C ₁	2	3	-	5
	C ₂	1	3	-	4
	C ₃	-	1	1	2
	C ₄	-	2	-	2
	C ₅	-	-	-	-
II. Cara menyusun Instrumen Penilaian Hasil Belajar	C ₁	2	-	-	2
	C ₂	2	3	-	5
	C ₃	-	4	-	4
	C ₄	-	2	-	2
	C ₅	-	-	-	-
	C ₆	-	-	2	2
III. Menilai alat tes Penilaian Hasil Belajar	C ₁	2	-	-	2
	C ₂	1	3	-	4
	C ₃	-	1	-	1
	C ₄	-	-	-	-
	C ₅	-	2	-	2
	C ₆	-	1	2	4
Jumlah Soal		10	25	5	40
Bobot skor/item → jumlah		1 → 10	1 → 25	5 → 25	60 = MES*)
Alokasi waktu/item → jmlah		1,5' → 15'	1,5' → 37,5'	5' → 100'	152,5'

*) MES = Maximum, Expected Scores (skor maksimum yang diharapkan)

Bila kita perhatikan tabel spesifikasi di atas terlihat kerangka tes yang hendak disusun yaitu: pokok-pokok materi, aspek kognitif yang hendak diukur, bentuk dan jumlah soal, alokasi waktu yang dijabarkan untuk mengerjakan tes, juga skor maksimum yang ditargetkan.

SIFAT TES HARUS MENCERMINKAN TUJUANNYA

Jika tujuan tes untuk promosi atau kenaikan kelas, maka tes tersebut harus dapat menyuruh siswa untuk menunjukkan prestasi dalam tes tersebut. Berbeda, jika tes itu bertujuan untuk diagnosis, maka tes tersebut harus dapat menunjukkan kelemahan-kelemahan tertentu dari setiap prestasi siswa.

Tes diagnosis meliputi bidang yang terbatas akan tetapi tujuan yang hendak dicapai lebih mendetail. Tingkat diskriminasi atau tingkat kesukaran soal tidak begitu menjadi pertimbangan. Tes untuk promosi atau

seleksi biasanya menekankan pada usaha mengungkap ketercapaian tujuan yang dipentingkan, sehingga tes tersebut secara detail mengukur setiap tujuan. Tingkat kesukaran dan kemampuan diskriminasi sangat dipentingkan di dalam menyusun tes untuk tujuan seleksi ini. Meskipun demikian sampling tujuan yang diukur juga harus representatif.

SIFAT TES HARUS MENCERMINKAN KONDISI-KONDISI DI MANA TES TERSEBUT DIBERIKAN

Dalam merencanakan sebuah perangkat tes tidak hanya memperhatikan usia atau pengalaman belajar siswa, akan tetapi juga faktor-faktor seperti waktu, tempat penyelenggaraan tes juga perbanyak dan harga tes perlu dipertimbangkan.

MENYIAPKAN TES

Langkah kedua adalah mempersiapkan tes. Saran-saran dari hasil pengalaman di bawah ini sangatlah membantu dalam mempersiapkan tes.

1) *Mulailah dengan segera membuat konsep pendahuluan menyusun tes*

Membuat daftar pertanyaan-pertanyaan yang dapat diajukan dalam tes selama pengajaran akan/sedang berlangsung. Hal ini dapat membantu pelaksanaan kegiatan belajar-mengajar. Dengan daftar pertanyaan tersebut guru akan selalu ingat apa yang harus dipahami oleh siswa.

2) *Suatu tes dapat terdiri dari beberapa tipe soal*

Suatu tes yang terdiri dari beberapa tipe soal tidak akan membosankan. Hal ini terutama untuk tes yang terdiri dari banyak item soal. Keragaman tipe soal sering juga sebagai tuntutan dari tujuan yang hendak diukur maupun bentuk materi.

3) *Sebagian besar item soal dalam sebuah tes hendaknya mempunyai tingkat kesukaran kira-kira 50 persen*

Suatu item tes yang memiliki tingkat kesukaran 50 persen berarti hanya separo dari kelompok yang dites dapat menjawab dengan benar. Jangan digunakan soal yang hanya dapat dikerjakan benar oleh siswa yang dites kurang dari 10 persen atau lebih dari 90 persen. Tingkat kesukaran ini dapat diketahui setelah tes tersebut diujicobakan.

Bila masing-masing item soal tes telah diketahui tingkat kesukarannya, kemudian susunlah dari yang paling mudah ke yang paling sukar.

Meskipun tingkat kesukaran soal tes perlu diperhatikan namun untuk hal-hal tertentu perlu juga sedikit dikesampingkan. Bila suatu tes untuk mengukur *kecepatan* bukan untuk jumlah *pengetahuan* seperti *kecepatan menghitung*, maka soal dibuat tidak sukar. Juga tes-tes untuk mengukur *tingkat penguasaan bahan* dan tes-tes untuk diagnosis isinya harus ditentukan oleh pentingnya bahan bukan oleh tingkat kesukarannya.

4) *Hendaknya konsep-konsep soal dibuat lebih banyak dari yang dibutuhkan dalam bentuk akhir*

Dengan tersedianya item-item soal yang banyak kita dapat membuang item-item soal yang tidak baik. Disarankan item soal dibuat lebih banyak 25 - 50 persen dari yang dibutuhkan.

5) *Setelah soal tes diberlakukan beberapa waktu perlu diperiksa kembali*

Dengan pemeriksaan kembali terhadap item soal-soal tes, kemungkinan dijumpai soal tes tidak sesuai lagi. Sebaiknya juga, disarankan soal-soal tes yang telah jadi dapat dibaca oleh guru lain dalam bidang studi yang sama. Dengan cara ini kemungkinan dijumpai soal-soal tes yang sukar dimengerti atau timbul interpretasi yang berbeda antara guru satu dengan yang lainnya.

- 6) *Soal-soal tes harus disusun sebaik mungkin, sehingga isi dan bukan bentuk pertanyaan yang menentukan jawabannya.*

Perlu dihindari kalimat-kalimat yang merupakan petunjuk ke arah jawaban yang tepat. Untuk soal tes bentuk benar-salah kata-kata khusus seperti "selalu", "tidak pernah", "seluruhnya" dan yang sejenisnya mengarah pada jawaban yang *salah*.

Perkataan seperti "mungkin", "kadang-kadang", "biasanya", pada umumnya menunjukkan pada jawaban yang benar. Hindarkan pula penggunaan kata-kata yang persis sama dengan kata-kata pada materi/bahan pelajaran. Sebab bagi siswa yang baik ingatannya, meskipun tidak tahu akan isi bahan pelajaran, dengan mengingat kata-katanya saja mereka dapat menjawabnya dengan benar.

Juga perlu dihindari penggunaan kata secara berulang, yakni digunakan pada kalimat stem juga dipergunakan dalam salah satu atau beberapa option jawabannya.

- 7) *Penunjuk-petunjuk yang diberikan hendaknya sejelas, selengkap, dan sesingkat mungkin*

Kepada siswa harus diberitabukan bagaimana dan di mana dia harus mengerjakan soal tes yang diberikan. Pemberian contoh cara mengerjakan soal tes juga dipentingkan dalam hal ini.

- 8) *Jika mungkin suatu soal tes objektif yang berbentuk benar-salah dihindari.*

Lebih baik suatu soal tes sediakan 4 alternatif jawaban. Hal ini akan mengurangi kemungkinan memberi jawaban menerka tetapi benar.

Kedelapan hal di atas merupakan saran-saran yang dapat diikuti dalam mempersiapkan tes hasil belajar.

MEMBUAT TES ESEI

Pada modul sebelumnya telah dibicarakan mengenai beberapa jenis alat tes hasil belajar termasuk juga mengenai tes esei/uraian.

Sebelum dibicarakan mengenai cara membuat tes esei ini, terlebih dahulu dikemukakan sedikit mengenai kekurangan dan kelebihan tes esei.

KEKURANGAN-KEKURANGAN TES ESEI

- a) *Validitas rendah*

Rendahnya validitas tes esei ini terutama pada validitas isinya. *Giles Ruch*, seorang ahli dalam pengukuran, mengemukakan bahwa tes esei hanya mengukur kurang dari separo pengetahuan yang sebenarnya dimiliki oleh siswa. Lagi pula untuk mengerjakan tes esei memakan waktu yang cukup lama dibandingkan dengan tes objektif. Selain itu bagi korektor menghadapi banyak faktor yang tidak relevan seperti penggunaan kata-kata/kalimat, tulisan dan sebagainya, dalam hal ini tidak ada rumus atau patokan untuk mengoreksinya.

Tes esei hanya menekankan pada bagaimana cara menyatakan apa yang diketahui siswa, bukan tentang apa saja yang diketahuinya. Hal-hal itulah yang membuat tes esei bervaliditas rendah.

b) *Reliabilitas rendah*

Karena untuk menjawab tes esei siswa harus mengorganisasi jawabannya dalam bentuk ceritera maka membutuhkan waktu yang lama. Dengan tes esei siswa hanya dapat menjawab pertanyaan-pertanyaan relatif sedikit dibandingkan bila digunakan tes objektif. Sedikitnya jumlah pertanyaan pada tes esei ini menyebabkan reliabilitasnya rendah. Hal yang lebih penting lagi yang dapat menjadikan tes esei reliabilitasnya rendah adalah dalam proses pemberian skor. Subjektivitas dalam proses pemberian skor ini sangat mungkin terjadi. Suatu jawaban seorang siswa terhadap pertanyaan tes esei akan mungkin terjadi mendapat skor yang berbeda jika:

- dikoreksi oleh orang/guru yang berbeda;
- dikoreksi seorang guru tapi berbeda waktunya;
- kondisi mental guru saat mengoreksi jawaban siswa terhadap tes esei juga berpengaruh terhadap hasil koreksinya.

Dalam sebuah penelitian yang dilakukan di Universitas West Virginia ternyata dalam memberikan skor/ nilai siswa dari hasil mengerjakan tes esainya, dapat terjadi hal-hal berikut:

1. Skor/nilai siswa bergantung pada skor/nilai siswa lain yang pekerjaannya telah diperiksa sebelumnya

Suatu pekerjaan siswa sebenarnya hanya mendapat skor 60, tetapi mendapat 70, gara-gara dikoreksi setelah pekerjaan siswa yang jelek sekali. Dapat pula skor siswa tersebut menjadi 50, gara-gara dikoreksi setelah pekerjaan siswa lain yang sangat baik.

2. Dalam penelitian ini disimpulkan pula bahwa keberhasilan atau kegagalan seseorang dalam mengikuti tes esei 40% bukan disebabkan oleh tahu atau tidaknya seseorang terhadap bahan, tetapi disebabkan oleh *siapa* yang mengoreksinya, selain itu 10% yang lain disebabkan oleh *kapan* tes itu diperiksa.

Ternyata kebanyakan penelitian mengenai reliabilitas tes-tes esei adalah mengenai kesepakatan antara mereka dalam mengoreksinya, *bukan* mengenai reliabilitas tes esei tersebut. Hasil dari penelitian tentang kesepakatan dalam mengoreksi antara dua orang guru dari data hasil pengukuran dengan alat tes esei yang sama menunjukkan koefisien korelasi yang tinggi. Akan tetapi penelitian yang berkaitan dengan reliabilitas tes itu tetap tidak mendapatkan hasil yang memuaskan. Rendahnya reliabilitas itu disebabkan keterbatasan jumlah/sedikitnya soal-soal yang dipergunakan dalam tes esei.

c) *Kesulitan dalam penggunaan*

Hasil pengukuran dengan tes esei biasanya tidak sebanding dengan jerih payah dan waktu yang digunakan oleh guru dan siswa. Oleh karena itu penggunaan tes esei harus betul-betul dipertimbangkan. Maka dapat dikatakan, penggunaan tes esei bisa dibenarkan jika benar-benar mendapatkan hasil yang selaras dengan waktu dan tenaga yang dikeluarkan.

KELEBIHAN-KELEBIHAN TES ESEI

Sebenarnya kelebihan tes esei hampir tidak ada jika dibandingkan dengan tes objektif. Bila dikatakan tes esei punya kelebihan yaitu penggunaan waktu yang sedikit dalam membuat soal itupun semu sebab dia juga akan memerlukan banyak waktu yang dipergunakan dalam mengoreksi hasilnya. Bahkan dalam penulisan soal tes esei juga memerlukan waktu yang cukup lama bila kita mempertimbangkan tujuannya/kemanfaatannya.

Meskipun demikian ada kelebihan dari tes esei jika benar-benar disusun dengan seksama yaitu: ia dapat memiliki *kelebihan* dalam mengukur tujuan tertentu yang kadang sulit dilakukan dengan tes objektif. Tes esei lebih mampu mengukur proses mental yang tinggi dibandingkan dengan tes objektif. Bila tujuan penelitian untuk mendapatkan prestasi siswa mengenai kemampuannya dalam mengorganisasi bahan dan jawaban yang

sistematis maka tes esei yang cocok digunakan. Akan tetapi jika yang diukur adalah tentang ingatan siswa tentang fakta, prinsip atau hal-hal tertentu, tes objektif akan tepat digunakan.

BEBERAPA CARA MENYUSUN SOAL TES ESEI

Beberapa jenis pertanyaan yang dapat digunakan dalam tes esei antara lain:

- a) *Mengingat suatu fakta* misalnya tentang apa, siapa, kapan, di mana, yang mana.
- b) *Mengingat secara selektif*, misalnya tentang ;
 - Sebutkan persyaratan utama dari sebuah tes!
- c) *Mengingat dengan menilai*, misalnya,
 - Sebutkan 2 orang tokoh yang sangat berperan dalam mendalangi G 30 S PKI!
- d) *Membandingkan dua hal*, misalnya:
 - Bandingkan penggunaan tes esei dengan tes objektif sebagai dua alat ukur hasil belajar!
- e) *Mengemukakan sebab akibat*, misalnya:
 - Hal-hal apakah yang menyebabkan tes objektif lebih banyak digunakan dalam tes-tes di sekolah?
- f) *Menjelaskan arti suatu kata/kalimat*, misalnya:
 - Apa arti "validitas prediktif" dari suatu tes?
- g) *Membuat rangkuman dari sebuah artikel atau tulisan*, misalnya:
 - Buatlah rangkuman mengenai kelemahan dan kelebihan tes esei! (tidak lebih dari 1 halaman (300 kata).

Perlu dijelaskan di sini, bila seorang siswa menuliskan rangkumannya terlalu detail dan tidak memperhatikan hal-hal yang penting berarti nilainya kurang.
- h) *Mengadakan analisis*, misalnya:
 - Mengapa banyak SMA dalam program penjurusannya menggunakan jasa tes psikologi?
- i) *Menyatakan adanya hubungan*, misalnya:
 - Mengapa alat tes esei cenderung dikatakan sebagai alat yang mengukur kemampuan berbahasa?
- j) *Memberikan contoh-contoh*, misalnya:
 - Buatlah suatu contoh mencari validitas tes, dengan mendasarkan pada data hasil tes berikut ini!
- k) *Mengadakan klasifikasi*, misalnya:
 - Kesalahan apa yang terdapat pada pernyataan "....."?
- l) *Menerapkan prinsip-prinsip atau aturan-aturan*, misalnya:
 - Apa akibatnya jika suatu tes hasil belajar disusun tidak memperhatikan validitasnya?
- m) *Memberikan pembahasan atau uraian*:
 - Bahaslah mengenai pandangan Siniski Suzuki tentang cara mengembangkan bakat anak!
- n) *Menuliskan tujuan dari sebuah artikel yang ditinjau oleh penulisnya*, misalnya:
 - Mengapa dalam modul ini tidak dibahas mengenai alat non-tes secara mendetail?

- o) *Membuat bagan*, misalnya:
 - Buatlah bagan dari langkah-langkah pokok membuat tes hasil belajar!
- p) *Mengadakan reorganisasi fakta-fakta*, (artinya: memberikan pertanyaan-pertanyaan yang bersifat melihat kembali untuk melatih kemampuan menyusun), misalnya:
 - Sebutkan 5 macam saran-saran praktis yang dikemukakan dalam buku ini, yang dapat Anda terapkan dalam tugas Anda sendiri!
- q) *Mengemukakan cara baru atau saran-saran tentang suatu prosedur*, misalnya:
 - Bagaimana menurut Anda tentang langkah-langkah menyusun tes buatan guru?
- r) *Menarik kesimpulan*, misalnya:
 - Kesimpulan apa yang bisa Anda buat berdasarkan hasil perhitungan analisis item berikut ini?

Berdasarkan uraian tersebut di atas beberapa kesimpulan mengenai tes esei dapat dikatakan sebagai berikut:

Banyak guru mengira bahwa pembuat tes esei lebih mudah dibandingkan dengan tes objektif. Sebenarnya menyusun tes esei yang berkualitas lebih sulit dibandingkan dengan cara membuat tes objektif yang berkualitas.

Dalam membuat tes esei yang baik harus diadakan pemikiran yang mendalam, terutama jika tes esei ditujukan untuk mengukur suatu proses mental yang pantas diukur, tidak sekedar pengetahuan mengenai fakta atau hafalan saja.

MEMBUAT TES OBJEKTIF

Pada modul sebelumnya juga telah dibicarakan mengenai tes objektif ada beberapa jenis tes objektif namun mereka dapat dibedakan menjadi tes jawaban singkat (*Completion item*) dan memilih alternatif jawaban yang tersedia. Termasuk dalam tes jenis memilih alternatif jawaban yang tersedia yaitu: tes benar - salah (*true - false*); tes menjodohkan (*matching*); dan pilihan ganda (*multiple choice*). Kedua jenis tes tersebut akan dicontohkan sebagai berikut:

TES JAWABAN SINGKAT

a) *Variasi pertanyaan langsung*

Pada tes ini testee diminta langsung untuk menjawab pertanyaan secara singkat. Perhatikan contoh berikut ini.

- Bilamanakah tes-tes inteligensi pertama kali diperkenalkan?
- Hasil korelasi *test-retest* akan diperoleh koefisien korelasi reliabilitas jenis apa?
- Instrumen tes hasil belajar yang disusun berdasarkan isi materi pelajaran akan memiliki jenis validitas apa?

b) *Variasi penyempurnaan kalimat*

Pada jenis ini testee diminta untuk melengkapi atau menyelesaikan kalimat yang belum selesai. Perlu diperhatikan dalam pembuatan tes ini yaitu:

- (1) Kalimat yang dipergunakan jangan persis sama dengan yang ada pada bahan pelajaran.
- (2) Kata-kata yang dihilangkan jangan terlalu banyak, sehingga kalimat menjadi tidak berarti, contoh:

"..... diperoleh dari antara tes dengan"

- (3) Kata-kata yang dihilangkan hendaknya yang berada di belakang kalimat dan jangan di tengah atau di depan kalimat. Contoh "Proses pemberian harga atau angka dengan kriteria tertentu disebut". Bandingkan dengan "..... pemberian angka dengan kriteria tertentu disebut dengan pengukuran".

Untuk tes variasi penyempurnaan kalimat di atas dapat dicontohkan antara lain:

- item tes yang dikerjakan benar oleh banyak testee, berarti memiliki tingkat kesukaran
- tes yang dapat membedakan antara testee kelompok pandai dan yang tidak pandai berarti tes tersebut memiliki

c) *Variasi berbentuk asosiasi*

Pada tes jenis ini testee dituntut mengadakan asosiasi. Dapat dicontohkan untuk jenis tes ini yaitu: Tulislah jenis validitas apa yang dituntut dari tes berikut ini:

- Tes Intelligensi :
- Tes Bakat :
- Tes Hasil Belajar :
- Tes Sikap Mandiri :

TES DENGAN ALTERNATIF JAWABAN YANG DISEDIAKAN

TES BENAR - SALAH (TRUE-FALSE)

Terdapat beberapa variasi dalam tes benar-salah yaitu:

- (1) Tes benar-salah berbentuk pernyataan yang harus dinyatakan benar atau salah. Perhatikan contoh berikut ini:

"Tulislah huruf "B" di dalam tanda (. .) bila pernyataan di belakangnya Anda anggap benar dan huruf "S" bila pernyataan tersebut Anda anggap salah"

- (. . .) 1. Nilai hasil belajar merupakan hasil dari proses pengukuran.
(. . .) 2. Suatu alat tes hasil belajar harus memiliki validitas isi.

- (2) Tes benar-salah yang menuntut alasan. Variasi jenis tes ini selain seperti di atas juga menuntut testee untuk memberikan alasan mengapa *benar* atau mengapa *salah*.

"Tulislah huruf "B" jika pernyataan di belakangnya Anda anggap benar dan huruf "S" jika pernyataan di belakangnya Anda anggap salah. Selanjutnya berilah alasan mengapa Anda menjawab benar atau salah tersebut.

- (. . .) 1. Nilai hasil belajar adalah hasil dari proses pengukuran, . . .
(. . .) 2. Suatu alat penilaian hasil belajar harus memiliki validitas isi, . . .

- (3) Tes benar salah berganda. Pada variasi jenis tes ini suatu induk persoalan menghasilkan anak-anak persoalan. Contoh:

"Suatu tes hasil belajar yang terstandar"

- (a) dapat digunakan untuk meramalkan keberhasilan belajar siswa di kemudian hari (. . .)
(b) dapat untuk mengetahui inteligensi siswa (. . .)
(c) untuk mengetahui tingkat penguasaan siswa terhadap materi yang telah diajarkan (. . .)

Untuk tes benar-salah perlu diperhatikan hal-hal berikut:

- (1) Hindarkan penggunaan kata-kata, seperti: biasanya, kadang-kadang, semua, mungkin dan sebagainya. Sebab kata-kata tersebut dapat mengkaburkan kebenaran atau kesalahan.

UNIVERSITAS MEDIAN AREA berisikan satu pokok persoalan, sehingga dapat dikatakan hal itu 100% benar dan 100% salah.

- 5) Urutan jawaban hendaknya jangan disusun dengan suatu pola tertentu, misalkan:
- (a) BS BS BS BS BS BS dan seterusnya
 - (b) BB SS BB SS BB SS dan seterusnya
 - (c) SS SS BB BB SS SS dan seterusnya

Atau pola-pola lain yang menyebabkan testee dapat menerka dengan benar jawabannya, hanya dengan memperhatikan susunan pola jawaban tersebut.

VARIASI MENJODOHKAN ATAU MEMASANGKAN (MATCHING)

Variasi tes menjodohkan ini menuntut testee memasangkan kata-kata dalam suatu kelompok/kolom dengan kata-kata pada kolom lain. Perintah cara mengerjakan tes ini dapat dibuat demikian "Pasangkanlah kata-kata yang terdapat pada kolom I dengan kata-kata pada kolom II".

- | I | II |
|---------------------------|-------------------------|
| 1. Pengukuran | a. kemampuan membedakan |
| 2. Penilaian | b. pemberian nilai |
| 3. koefisien reliabilitas | c. pemberian skor |
| 4. reliabilitas | d. konsistensi |
| 5. diskriminasi | e. keterandalan |
| | f. prestasi belajar. |

atau dapat disusun sebagai berikut:

- | I | II |
|------------------------|---------------------|
| 1. Tes esai | a. validitas rendah |
| 2. Tes objektif | b. validitas tinggi |
| 3. Tes inteligensi | |
| 4. Tes bakat skolastik | |

- Sehubungan dengan tes memasangkan kata ini disarankan beberapa hal antara lain:
- (1) isi masing-masing kolom harus berupa kata-kata yang homogen, misalnya konsep-konsep; arti-arti konsep, nama-nama orang/kata dan sebagainya;
 - (2) dasar pemasangan harus jelas, seperti contoh yang pertama adalah tentang konsep dipasangkan dengan arti atau makna dari konsep tersebut.

VARIASI ALTERNATIF PILIHAN (MULTIPLE CHOICE)

Soal tes tipe pilihan ini menghadapkan kepada testee pada sejumlah alternatif jawaban. Alternatif jawaban yang umum disediakan berkisar 3 - 5 alternatif untuk masing-masing soal.

Agar lebih jelas pemahaman kita terhadap istilah-istilah yang biasa digunakan dalam tes pilihan ini dapat dijelaskan terlebih dahulu di sini.

Stem adalah bagian pokok soal yang merumuskan isi soal. *Stem* ini berbentuk pertanyaan, perintah, maupun kalimat tidak sempurna. Alternatif jawaban yang menyertai *stem* disebut dengan *options* atau bila diterjemahkan secara tegas pilihan-pilihan. Alternatif atau *options* yang benar disebut dengan kunci jawaban atau *key*, sedangkan *options* yang bukan kunci jawaban disebut distraktor atau pengganggu-pengganggu.

*Manakah di antara pengertian pengukuran berikut ini yang tepat? (Stem)

- A. proses pemberian angka atau harga terhadap gejala atau barang dengan patokan ukuran tertentu,
—————→ Key/kunci.
- B. proses memberikan nilai terhadap pekerjaan siswa
C. proses memberikan kualitas prestasi seseorang akan suatu pekerjaannya
D. proses membandingkan gejala atau barang.
- } distraktor

Ada beberapa variasi dalam tes pilihan ini, dan masing-masing dapat dicontohkan seperti berikut ini.

(1) Variasi menemukan satu jawaban yang benar

Pada jenis tes variasi ini alternatif kunci jawaban merupakan satu-satunya jawaban yang benar, baik berdasarkan logika, perhitungan matematika, secara historis maupun fakta.

Contoh:

- Presiden Filipina sekarang adalah:
A. Ferdinan Marcos
B. Enrile
C. Corason Aquino
D. Romulo
E. Quirino.

(2) Variasi memilih jawaban yang terbaik

Pada jenis tes ini testee diminta untuk memilih secara selektif alternatif jawaban yang benar. Sebab alternatif-alternatif yang disediakan merupakan jawaban-jawaban yang benar, tetapi mempunyai tingkatan dari *paling benar* ke *agak benar*, dari *paling lengkap* ke *agak lengkap* dan sebagainya.

Contoh:

- Penilaian hasil belajar bertujuan untuk maksud
A. seleksi siswa
B. promosi siswa
C. diagnosa kesulitan siswa
D. penempatan siswa
E. membuat keputusan tentang siswa.

(3) Variasi memilih lebih dari satu jawaban yang benar

Pada jenis tes ini testee dapat menentukan pilihan jawaban yang benar lebih dari satu untuk setiap soal.

Contoh:

- Ketidak fungsinya tes buatan guru dikarenakan oleh:
A. kemampuan guru dalam menyusun tes sangat terbatas
B. validitas tes yang rendah
C. tingkat kesukaran tes relatif tinggi
D. reliabilitas tes yang rendah
E. tidak praktis digunakan.

(4) Variasi dengan soal kalimat negatif

Pada jenis tes ini testee diharapkan dapat memilih option jawaban yang tidak termasuk dalam rumpun dasar pemikiran.

Contoh:
UNIVERSITAS MEDAN AREA

Manakah di antara pernyataan di bawah ini bukan termasuk tujuan penilaian hasil belajar:

- A. membuat diagnosa kesulitan belajar siswa

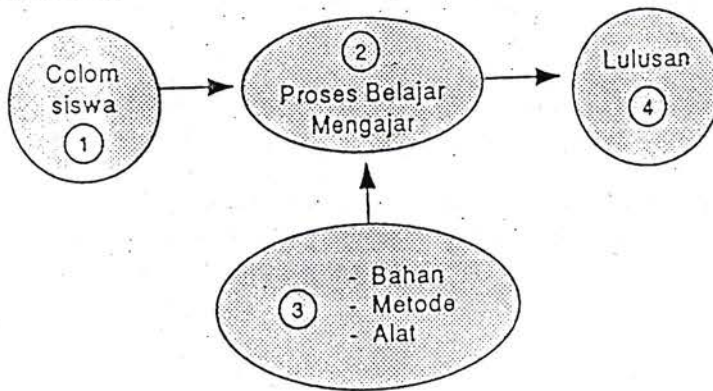
- B. mengetahui kelemahan dan kelebihan jalannya kegiatan belajar mengajar
- C. mengetahui kemampuan umum siswa
- D. menempatkan siswa pada jenis program tertentu
- E. menyeleksi/penerimaan siswa baru.

5) Variasi menggunakan gambar, bagan, peta dan lain-lain

Pada tes jenis ini persoalan tidak dirumuskan dalam suatu kalimat, melainkan dapat berupa gambar, grafik, skema atau peta. Testee harus memahami gambar, peta atau skema yang ada, sebelum dia menjawabnya.

Contoh:

- Berikut ini disajikan suatu skema sistem pendidikan di suatu sekolah.



Berdasarkan skema tersebut nomor berapakah yang termasuk/disebut dengan instrumental input?

- A. 1
- B. 2
- C. 3
- D. 4

Contoh-contoh tes pilihan di atas hanyalah merupakan jenis pokok saja. Sebenarnya mengenai tes pilihan ini fleksibel artinya bergantung pada kreativitas pembuatnya. Meskipun demikian ada beberapa petunjuk yang dapat dipakai sebagai pedoman membuat tes objektif khususnya tes pilihan.

- (1) *Stem harus jelas dan spesifik.* Dengan membaca stem dalam item tes, testee harus sudah memperoleh gambaran kemana arah persoalannya, sehingga dia tahu, bagaimana dia diharapkan untuk menjawab. Testee jangan dihadapkan pada persoalan untuk menafsirkan stem yang tidak jelas arah persoalannya. Dia tidak perlu memikirkan lagi, apa yang dimaksudkan oleh setiap stem yang ada.
- (2) *Kunci jawaban harus tidak dapat diperdebatkan lagi.* Dengan kata lain, apa yang dianggap oleh penyusun soal tentang kunci jawaban harus dapat disepakati oleh para ahli atau guru yang lain. Jangan sampai membuat suatu option jawaban yang semuanya dapat dibenarkan sebagai kunci jawabannya.
- (3) *Alternatif-alternatif jawaban atau options yang dicantumkan sebagai distraktor/pengganggu tidak boleh terlalu jelas kelihatannya "salahnya" atau "kejanggalannya".* Hal ini bila terjadi akan mengurangi fungsi dari tes pilihan itu sendiri, sebab dapat mengurangi options yang disediakan 5 dapat menjadi 4, 3 atau 2 saja.
- (4) *Jangan sampai terjadi dalam tes pilihan ini, testee merasa mendapat arahan untuk menjawab dengan benar, yang dikarenakan ada kata-kata yang sama antara yang terdapat pada stem dengan yang ada pada options jawaban.*

- (5) *Hindari penggunaan kata-kata yang sama dalam suatu tes dengan yang ada pada bahan pelajaran.*
- (6) *Hindari options jawaban seperti misalnya: A dan B benar; A, B, C benar; bukan semua itu; salah semua dan sebagainya.*
Gunakanlah *options* jawaban antara 4 - 5 options tetapi tidak seperti *options* tersebut di atas. Terlalu sedikit jumlah *options* menjadikan soal tes mudah dan lebih dari 5 *options* akan membuat *option* yang keenam dan seterusnya terlalu diada-adakan, sehingga tidak punya daya distraktor.
- (7) *Setiap soal harus berdiri sendiri*, artinya bahwa jawaban terhadap soal yang satu tidak boleh membantu dalam menjawab soal yang lain.
- (8) *Hindari soal-soal yang menuntut terlalu banyak detail*, lebih-lebih jika detail-detail tersebut merupakan fakta-fakta yang tidak penting maknanya apalagi fakta-fakta tersebut menuntut hafalan yang bersifat mekanis.
- (9) *Options yang digurukan sebagai alternatif pilihan jawaban hendaknya dirumuskan secara sederhana (to the point)*. Caranya dengan menghindari susunan kata yang berbelit-belit.

Berdasarkan hasil penelitian untuk tes variasi pilihan (*multiple choice*) ini dapat disarankan 12 model pertanyaan. Meskipun demikian ke 12 pertanyaan tentang hal-hal di bawah ini hanya merupakan petunjuk umum, bukan barang yang mati/tidak bisa ditawar-tawar.

- (1) *Definisi:*
- Pertanyaan mana yang mengandung pengertian
 - Kesimpulan apa yang dapat ditarik dari pernyataan "....."
- (2) *Tujuan:*
- Tujuan apa yang dapat dicapai dengan
 - Mengapa hal berikut ini dilakukan
 - Apakah maksud dari
- (3) *Efek/akibat:*
- Apakah akibat dari
 - Jika hal berikut ini dilakukan apa yang terjadi?
- (4) *Sebab:*
- Apakah penyebab dari timbulnya
 - Manakah syarat-syarat di bawah ini yang paling benar?
- (5) *Asosiasi:*
- Apa yang akan terjadi bila hal ini dilakukan?
 - Mana yang tepat bila hal ini terjadi?
- (6) *Melihat suatu kesalahan:*
- Hal berikut ini mana yang merupakan kesalahan?
 - Mana yang tidak tepat dari langkah-langkah (prosedur tertentu) berikut ini?
- (7) *Mengidentifikasi kesalahan:*
- Kesalahan jenis apakah ini?
 - Apakah nama kesalahan ini?
 - Persyaratan mana yang telah dilanggar dari

- (8) *Evaluasi:*
- Pendekatan mana yang terbaik untuk melaksanakan (berhubungan dengan tujuan)?
 - Mengapa *hal ini* dapat diterima oleh sebagian besar para guru?
- (9) *Perbedaan*
- Berikut ini merupakan perbedaan antara dengan
 - Perbedaan pokok apa yang ada antara dengan
- (10) *Persamaan:*
- Apakah kesamaan yang penting antara
 - Kesamaan apa yang ada antara
- (11) *Susunan:*
- Susunan manakah yang tepat (untuk tujuan tertentu)?
 - Manakah susunan berikut ini yang berada pada urutan pertama?
- (12) *Prinsip umum:*
- Data tes berikut ini mana yang tidak valid?
 - Soal berikut ini mana yang harus diganti?

KELEMAHAN DAN KELEBIHAN TES OBJEKTIF

Secara umum tes objektif lebih baik dibandingkan dengan tes esei. Tes objektif lebih valid dan lebih reliabel dibandingkan dengan tes esei. Hal ini bisa dipahami, karena suatu tes hasil belajar harus memiliki validitas isi/kurikulum. Hal ini bisa dipenuhi bila pertanyaan-pertanyaan dalam tes tersebut mengukur materi yang telah diajarkan atau mengukur TIK yang telah dirumuskan. Untuk ini diperlukan cukup banyak pertanyaan sesuai dengan materi yang diajarkan. Dengan tes bentuk objektif sebagian besar atau secara representatif TIK yang hendak dicapai dapat diukur dalam waktu yang relatif singkat. Berbeda dengan tes esei yang dalam pengerjaannya membutuhkan waktu yang relatif lama untuk setiap soalnya. Oleh sebab itu tes esei jika dibandingkan dengan tes objektif dalam segi efektivitasnya sebagai alat ukur, akan kalah, sebab dalam waktu yang sama tes objektif dapat secara representatif mengukur TIK sedangkan tes esei kurang/tidak representatif. Hal ini juga dari segi pemberian skor, tes objektif lebih mudah dan lebih objektif dibandingkan dengan tes esei.

Meskipun demikian bukan berarti tes objektif bebas dari kelemahan-kelemahan sebagai alat ukur. Berikut ini secara khusus dikemukakan kelemahan dan kelebihan tes-tes bentuk objektif.

Kelemahan

- Kemungkinan menerka kecil sekali sebab testee dihadapkan banyak sekali pilihan yang dipikirkan sendiri.
- Baik untuk soal-soal matematika, IPA yang berhubungan dengan hitung menghitung.
- Baik pula untuk soal-soal yang berbentuk peta, tabel, yang dalam hal ini testee tinggal menuliskan jawabannya pada tempat yang tersedia.

Kelebihan

I Tes Jawaban singkat 1. Variasi pertanyaan langsung

- Cenderung hanya mengukur pengetahuan faktual, karena hanya dengan memberikan jawaban pendek. Akibatnya merupakan pengetahuan yang terpotong-potong.
- Mengalami kesulitan dalam memberikan skor terhadap jawaban testee, misalnya testee menuliskan jawabannya salah ejaannya, apakah jawaban tersebut dibenarkan atau disalahkan. Hal semacam ini belum disepakati adanya

2. Variasi penyempurnaan kalimat

- Pembuatan soal mudah.
- Dapat diterapkan untuk berbagai mata pelajaran.
- Dapat mengukur pengetahuan yang sesungguhnya.
- Sedikit kesulitan dalam pemberian skor, terutama jika soal berbentuk kalimat-kalimat panjang yang dihilangkan beberapa kata-kata nya.

3. Variasi berbentuk asosiasi

- Kelebihannya hampir sama dengan variasi kedua di atas.
- Hanya mengukur pengetahuan/asosiasi testee.
- Cukup sulit dalam pembuatan soalnya.

II. Tes dengan alternatif jawaban yang disediakan

1. Tes benar-salah

- Pembuatan soal mudah.
- Dapat digunakan untuk berbagai mata pelajaran.
- Kemungkinan mengetes sample pengetahuan lebih representatif, karena tes ini biasanya terdiri dari banyak item.
- Testee cenderung menerka jawaban yang sebenarnya dia tidak tahu.
- Adanya pengaruh sugesti dari pernyataan-pertanyaan, sehingga dapat mengarahkan testee untuk menjawab benar/salah.
- Tidak dapat digunakan untuk mendiagnosis kemampuan siswa.
- Cenderung untuk mengukur aspek ingatan saja.

2. Variasi menjodohkan

- Dapat digunakan untuk berbagai mata pelajaran.
- Baik untuk mengukur aspek pengetahuan dan pengertian testee.
- Mudah dalam pemberian skor.
- Dapat membuat jenuh testee karena dia harus membaca berulang-ulang dari kolom I ke kolom yang II.
- Sulit mencari kata-kata yang serumpun terutama yang berkaitan dengan konsep.
- Cenderung mengukur aspek pengetahuan dan pemahaman saja.

3. Variasi pilihan (*multiple choice*)

- Lebih baik dari pada tes-tes objektif yang lain.
- Mudah dalam pemberian nilai/skor.
- Lebih representatif untuk mengukur sample pengetahuan yang diajarkan.
- Memiliki validitas yang baik.
- Dapat untuk mengukur berbagai aspek dan tingkatan kognisi.
- Perlu persiapan yang cukup lama dalam pembuatan soal tes.
- Ada kemungkinan siswa/testee menjawab soal yang tidak diketahui dengan hanya menerka.

Kita telah membicarakan mengenai berbagai hal tentang instrumen penilaian hasil belajar dalam bentuk tes. Untuk melengkapi instrumen penilaian hasil belajar tersebut perlu dijelaskan juga sedikit tentang instrumen yang berbentuk non-tes. Tentang instrumen nontes ini hanya akan diberikan 1 contoh mengenai cara penyusunannya/pembuatannya, sedangkan jenis-jenis instrumen non tes telah dibicarakan pada bagian

MEMBUAT INSTRUMEN NONTES HASIL BELAJAR

Instrumen nontes hasil belajar biasa digunakan untuk mengukur aspek nonkognitif. Instrumen nontes ini dalam bidang pendidikan sering digunakan untuk mengukur konsep psikologis seperti *keaktifan, kerajinan, keterampilan, kedisiplinan, ketaatan* dan yang lainnya.

Cara untuk menyusun alat nontes untuk konsep-konsep psikologis tersebut yaitu dengan merinci konsep yang hendak diukur ke dalam indikator-indikatornya. Selanjutnya, atas dasar indikator tersebut item-item pertanyaan atau pernyataan dibuat.

Berikut ini dicontohkan mengenai pengukuran "kemampuan kerja di laboratorium".

Daftar Cek "Kemampuan Kerja Siswa di Laboratorium"

Nama siswa yang dinilai : -----

Nama penilai : -----

Tanggal : -----

No.	Kemampuan Kerja	Skala penilaian				
		Tidak pernah	Jarang	Kadang-kadang	Sering	Selalu
0	1	2	3	4	5	6
1.	Menguasai masalah dalam tugas.					
2.	Mampu memecahkan sendiri masalah-masalah dalam tugas.					
3.	Kritis dalam menerima tugas					
4.	Bekerja dengan teliti.					
5.	Bekerja dengan rapi, bersih.					
6.	Bekerja cepat.					
7.	Terampil menggunakan alat.					
8.	Terampil memelihara alat.					
9.	Mampu bekerja sesuai dengan rencana.					

Catatan : Berilah tanda cek pada kolom yang disediakan.

Untuk mengetahui prestasi kemampuan kerja siswa dalam bekerja di laboratorium, dilakukan pembotan skor pada setiap skala penilaian di atas.

Misalkan, skala penilaian:

- tidak pernah = 0
- jarang = 1
- kadang-kadang = 2
- sering = 3
- selalu = 4

Dengan menjumlah skor-skor pada setiap siswa akan diperoleh skor total siswa mengenai kemampuan

UNIVERSITAS MEDAN AREA

kerja. Atas dasar skor ini selanjutnya dapat dianalisis berapa nilai setiap siswa tersebut.

Akhir dari uraian mengenai materi merencanakan tes; menyiapkan tes; cara membuat tes esei dan objektif; serta cara membuat instrumen nontes, Anda harus mengerjakan latihan berikut ini. Latihan ini merupakan usaha cara mendalami materi-materi di atas.

- 1) Buatlah daftar tujuan yang pokok-pokok dari materi yang telah Anda pelajari ini (cara menyusun instrumen penilaian hasil belajar), minimal 15 rumusan tujuan!
- 2) Berdasarkan soal latihan 1 tersebut selanjutnya buatlah tabel spesifikasinya untuk pedoman penyusunan soal tes!
- 3) Buatlah alat penilaian nontes tentang "Kemandirian siswa dalam tugas sekolah"!
- 4) Buatlah contoh soal tes tipe:
 - a) menjodohkan/memasangkan;
 - b) melengkapi;
 - c) benar-salah;
 - d) alternatif pilihan (*multiple choice*);
 - e) esei.Masing-masing satu soal.

Untuk dapat menjawab latihan soal-soal di atas ikuti petunjuk berikut ini.

Latihan 1:

Anda telah mengetahui bahwa materi pokok bahasan ini terdiri dari sub-sub pokok bahasan yaitu: (1) Merencanakan tes, (2) Menyiapkan tes, (3) Cara membuat tes esei dan tes objektif, (4) Cara membuat alat nontes hasil belajar. Agar rumusan tujuan yang Anda tulis dapat mewakili keseluruhan materi maka Anda harus merumuskan tujuan-tujuan dari subpokok bahasan tersebut. Hal ini sangat membantu dalam menyusun alat tes yang valid. Ingat bahwa tujuan yang Anda rumuskan harus spesifik dan operasional. Tidak boleh terjadi pada setiap butir rumusan tujuan yang Anda buat berisi tujuan ganda. Sekarang coba lakukan hal ini untuk menjawab latihan 1.

Latihan 2:

Cara membuat tabel spesifikasi tes tidak dicontohkan pada uraian modul ini. Meskipun demikian apa yang telah dicontohkan tersebut hanya merupakan satu model tabel spesifikasi. Anda dapat membuat model tabel spesifikasi yang lain, namun beberapa persyaratan harus Anda penuhi.

Persyaratan yang dimaksud antara lain untuk: materi yang hendak diukur; ranah dan tingkatannya yang hendak diukur; jumlah item soal untuk setiap pokok materi; bentuk/jenis tes yang dipergunakan dan sebagainya.

Sekarang berdasarkan beberapa persyaratan tersebut dan mengacu pada apa yang telah dicontohkan, Anda dapat mencoba menjabarkan tujuan-tujuan yang telah Anda buat pada latihan 1 ke dalam tabel spesifikasi tesnya.

Latihan 3:

Pada latihan ke-3 ini Anda harus menjabarkan konsep-konsep "kemandirian dalam tugas sekolah" ke dalam indikator-indikatornya. Ada dua konsep psikologis dalam hal ini; Pertama adalah kemandirian dan kedua tugas-tugas sekolah. Tentang konsep kemandirian Anda dapat mencari literatur atau definisinya atau juga dapat mendiskusikan apa itu "kemandirian" dan bagaimana "indikatornya". Bila konsep yang pertama sudah terjawab lanjutkan mencari/mendiskusikan tentang indikator-indikator tugas sekolah. Misalkan Anda membagi tugas-tugas sekolah menjadi tugas akademis dan non akademis. Tugas akademis termasuk di dalamnya mengerjakan PR, mencatat keterangan guru, diskusi kelompok, tugas terstruktur dan sebagainya. Sedangkan termasuk tugas non akademis misalnya: pramuka, upacara bendera, kebersihan kelas dan sebagainya. Di bagian ini Anda sudah dapat Anda kemukakan, coba buatlah tabel spesifikasinya. Anda dapat menggunakan contoh tabel spesifikasi berikut ini:

Tugas-tugas Sekolah Kemandirian	Akademis			Non Akademis		
	Indikator I	Indikator II	Indikator III	Indikator IV	Indikator V	Indikator VI
Indikator A	1, 2, 3	10, 11	17, 18	22, 23	28, 29	35, 36
Indikator B	4, 5, 6	12, 13, 14	19	24, 25	30, 31, 32	37, 38, 39
Indikator C	7, 8, 9	15, 16	20, 21	26, 27	32, 33, 34	39, 40

Keterangan:

- Indikator A, B dan C adalah indikator konsep kemandirian.
- Indikator I, II, III adalah indikator tugas akademis
- Indikator IV, V dan VI adalah indikator tugas non akademis.
- Butir angka 1 sampai dengan 40 adalah nomor butir pernyataan atau pertanyaan untuk mengukur konsep "kemandirian dalam tugas-tugas sekolah".

Contoh: butir no 1 yaitu pertanyaan atau pernyataan tentang indikator 4 dalam kaitannya dengan tugas akademis indikator I.

Selanjutnya berdasarkan keterangan atau petunjuk di atas silakan Anda coba menyusun daftar pertanyaan atau pernyataan tentang "Kemandirian dalam tugas-tugas sekolah" untuk mengukur siswa. Anda dapat membuatnya dalam bentuk "angket" atau "pedoman observasi" seperti dicontohkan pada bagian uraian modul ini.

Latihan 4:

Mengenai petunjuk latihan no 4 saya kira Anda telah tahu maksudnya sebab masing-masing jenis tipe soal tersebut tabel dijelaskan di muka dan pula telah ada beberapa contohnya. Hanya saja dalam latihan 4 ini Anda diminta membuat contoh-contoh soal yang materinya adalah materi yang ada pada modul ini, sehingga mempunyai manfaat ganda. Manfaat yang pertama Anda bisa menyusun soal dengan berbagai jenis/ tipe soal. Kedua, Anda akan lebih memahami materi modul ini, lewat membuat pertanyaan-pertanyaan/soal-soalnya.

RANGKUMAN

Dalam menyusun instrumen penilaian hasil belajar khususnya yang berupa tes harus memperhatikan hal-hal: (1) aspek-aspek apa yang hendak diukur; (2) apa tujuan diadakan pengukuran terhadap aspek-aspek tersebut; (3) kapan, di mana dan bagaimana tes itu diselenggarakan. Untuk dapat mencapai maksud tersebut maka perlu adanya perencanaan dan persiapan yang matang dalam penyusunan tes tersebut.

Instrumen penilaian hasil belajar untuk mengukur prestasi hasil belajar siswa maka alat yang digunakan harus memiliki validitas materi/isi yang baik. Agar alat tes memiliki validitas isi ini, maka sebelum perangkat tes ditulis perlu terlebih dahulu dilakukan "Job Analysis" dengan membuat tabel spesifikasi.

Selain validitas isi dipersyaratkan bagi tes hasil belajar yang berarti bahwa tes itu harus dapat mengukur materi yang telah diajarkan juga masalah aspek kognitif dan tingkatannya perlu diperhatikan dalam penyusunan tes hasil belajar. Erat kaitannya dengan aspek kognitif berikut tingkatannya ini adalah tipe/bentuk tes yang dipergunakan. Suatu bentuk tes tertentu cocok untuk mengukur aspek kognitif jenis tertentu pula. Meskipun dalam hal ini tidak dapat secara tegas dibuat

gratis pemisahannya. Oleh karena itu guru perlu memiliki pengetahuan dan keterampilan dalam membuat tes dengan berbagai tipe atau bentuknya. Di samping itu dia juga harus tahu kelebihan dan kekurangan masing-masing tipe tes.

Selanjutnya untuk melengkapi alat/instrumen penilaian hasil belajar dalam bentuk tes, perlu juga alat nontes dipergunakan. Alat nontes dapat digunakan untuk mengukur aspek-aspek perkembangan siswa di luar aspek kognitif. Alat nontes ini dipersyaratkan untuk memiliki validitas konstruk jika akan digunakan.

TES FORMATIF 2

Pemunjuk: Silangkan salah satu alternatif jawaban yang Anda anggap benar pada setiap soal di bawah ini

- 1) Manfaat tabel spesifikasi suatu tes adalah sebagai berikut, *kecuali*:
 - A. untuk mendapatkan gambaran tentang validitas tes
 - B. untuk memperoleh gambaran bentuk-bentuk tes yang akan digunakan
 - C. untuk memperoleh informasi tentang lamanya pelaksanaan tes
 - D. untuk memperoleh gambaran tentang reliabilitas tes.
- 2) Manakah di antara pernyataan berikut ini yang *bukan* merupakan ciri tes esei?
 - A. validitas rendah
 - B. reliabilitas rendah
 - C. dapat mengukur proses mental yang kompleks
 - D. mempunyai tingkat kesukaran yang tinggi.
- 3) Manakah di antara jenis-jenis tes berikut ini yang memiliki validitas yang lebih baik?
 - A. tes esei
 - B. tes memasangkan (*matching*)
 - C. tes benar - salah (*true-false*)
 - D. tes alternatif pilihan (*multiple choice*)
- 4) Manakah di antara pertanyaan berikut ini yang mengukur aspek kognitif *analisis*?
 - A. Kemukakan dengan bahasamu sendiri mengenai definisi penialain!
 - B. Jelaskan unsur-unsur penting yang terdapat pada pengertian pengukuran berikut ini!
 - C. Buatlah karangan yang di dalamnya membuat unsur-unsur pendahuluan, pembahasan dan kesimpulan!
 - D. Coba koreksilah karangan/artikel di bawah ini!
- 5) Manakah di antara pernyataan-pernyataan tes *benar-salah* berikut ini yang perlu diganti?
 - A. Suatu tes yang reliabel pasti valid (. . .)
 - B. Suatu tes alternatif pilihan (*multiple choice*) tepat digunakan untuk tes hasil belajar dan tidak tepat untuk potensi akademis (. . .)
 - C. Tes yang baik bila dapat dikerjakan benar oleh kurang lebih 50% orang yang dites (. . .)
 - D. Tes esei lebih valid daripada tes objektif (. . .)
- 6) Berikut ini adalah persyaratan tes untuk diagnosis, *kecuali*:
 - A. tingkat kesukaran soal tidak dipentingkan
 - B. harus dapat mengukur pokok-pokok materi yang penting
 - C. jumlah soal relatif banyak
 - D. harus dapat mengukur sebagian besar materi.

- Untuk mengukur "Tingkat kedisiplinan belajar" dapat digunakan alat berikut ini kecuali:
- pedoman observasi
 - daftar cek
 - tes atau tugas
 - pedoman interview.
- Validitas alat ukur yang dipersyaratkan untuk mengukur "Tingkat kedisiplinan siswa dalam kegiatan di sekolah" adalah:
- validitas isi
 - validitas prediktif
 - validitas konstruk
 - validitas logik.
- Untuk mengurangi terjadinya usaha menerka jawaban pada tes objektif perlu dilakukan hal-hal berikut, kecuali:
- memperingatkan agar siswa tidak menerka jawaban bila memang tidak tahu
 - perlu dijelaskan tentang cara pemberian skor untuk menetralsir mereka yang menerka jawaban
 - tidak menggunakan kata-kata yang sama pada stem dan alternatif jawabannya
 - menyuruh siswa kerja dengan teliti.
- 10) Pertanyaan-pertanyaan manakah di bawah ini yang mengukur aspek kognitif pengetahuan ?
- Di manakah tes inteligensi pertama kali diperkenalkan?
 - Validitas apa yang harus dimiliki oleh tes hasil belajar?
 - Hitunglah berapa koefisien reliabilitas tes ini berdasarkan data hasil uji cobanya!
 - Mana yang lebih dipentingkan, reliabilitas tes yang tinggi atau tingkat kesukaran yang berkisar 50%?

Bila tes formatif telah Anda kerjakan selanjutnya koreksilah dengan kunci jawaban yang telah disediakan pada bagian akhir modul ini. Untuk menghitung berapa persen tingkat penguasaan Anda tentang materi ini, gunakan rumus berikut:

Rumus:

$$\text{Tingkat Penguasaan} = \frac{\text{Jumlah Jawaban Benar}}{10} \times 100\%$$

Arti klasifikasi persentase:

90%	-	100%	=	baik sekli
80%	-	89%	=	baik
70%	-	79%	=	cukup
<		70%	=	kurang

Bila tingkat penguasaan Anda di atas 80% berarti Anda dapat mempelajari bagian modul selanjutnya. Tetapi jika tingkat penguasaan Anda berada di bawah 80%, maka Anda harus mempelajari lagi materi ini, khususnya yang belum Anda kuasai.

3. MENILAI INSTRUMEN TES HASIL BELAJAR

Pada pembahasan mengenai "Cara Menyusun Instrumen Penilaian Hasil Belajar" dikemukakan bahwa materi tentang "Mencoba" dan "Menilai Tes" hendak dibicarakan pada bagian modul selanjutnya. Maka pada modul inilah kedua hal tersebut hendak dijelaskan.

Perlu kita ingat kembali bahwa mencobakan dan menilai tes hasil belajar merupakan langkah ketiga dan keempat dari langkah-langkah cara menyusun instrumen penilaian hasil belajar.

Untuk materi tentang *mencobakan instrumen tes hasil belajar* secara berurutan hendak dijelaskan secara garis besar tentang: (1) kondisi pelaksanaan tes; (2) waktu untuk mengerjakan tes; dan (3) prosedur pemberian skor dan peraturan-peraturannya. Dalam materi tentang menilai tes hasil belajar, hendak dijelaskan mengenai: (1) tingkat kesukaran tes dan cara penghitungannya; (2) kemampuan diskriminasi tes dan cara penghitungannya; serta (3) validitas dan reliabilitas tes dan cara penghitungannya.

MENCOBAKAN TES HASIL BELAJAR

Setelah seperangkat tes dipersiapkan sesuai dengan rencana, langkah selanjutnya adalah mengujicobakan tes tersebut. Prinsip-prinsip yang harus diperhatikan dalam uji-coba tes adalah sebagai berikut.

1) *Kondisi pelaksanaan uji coba tes*

Kondisi pelaksanaan uji coba tes adalah hal yang sangat penting untuk menjadi perhatian. Sebab jawaban-jawaban yang diberikan siswa tidak saja bergantung pada keadaan tes itu sendiri, melainkan juga dapat dipengaruhi oleh keadaan sekitarnya. Hendaknya dalam pelaksanaan tes dipilih tempat yang tidak asing bagi siswa yang dites (misal: ruang kelas siswa sendiri). Untuk menghindarkan siswa "nyontek" atau berbuat curang lainnya, maka harus diusahakan adanya pengawasan yang ketat. Mengatur tempat duduk siswa saat tes dengan memberikan selang satu kursi/bangku kosong di antara mereka juga dianjurkan.

2) *Waktu penyelenggaraan tes*

Waktu yang disediakan untuk mengerjakan tes hendaknya cukup lama, sebab waktu yang cukup lama ini memberikan kesempatan pada siswa untuk mencoba menjawab semua pertanyaan dalam tes. Hal ini erat kaitannya dengan usaha untuk mengetahui tingkat kesukaran soal tes berikut kemampuan diskriminasi masing-masing butir tes. Meskipun demikian dalam pelaksanaan tes yang sebenarnya lama waktu mengerjakan tes banyak ditentukan oleh tujuan diselenggarakannya tes tersebut. Untuk tes yang bertujuan mengukur kecepatan berhitung, waktu mengerjakannya harus tidak begitu lama, sehingga siswa yang paling cepat berhitungpun tidak akan dapat selesai mengerjakan tes tersebut. Sebaiknya tes yang digunakan untuk tujuan mendiagnosis kesulitan belajar perlu waktu yang lebih lama dibandingkan dengan untuk tujuan di atas, maupun untuk tujuan tes-tes prestasi belajar yang lain. Juga perlu diperhatikan mengenai lama waktu mengerjakan tes ini yaitu untuk tes-tes yang hanya mengungkap fakta-fakta waktunya relatif lebih singkat dibandingkan dengan tes-tes yang mengungkap proses mental yang lebih tinggi.

Perlu pula dicatat, apakah setelah waktu uji coba tes selesai masih banyak dijumpai siswa yang belum selesai mengerjakan. Beberapa saran dalam menerapkan lama waktu mengerjakan tes adalah sekurang-kurangnya 90% siswa yang dites mempunyai waktu mengerjakan keseluruhan tes. Dengan kata lain mereka dapat mencoba semua soal tes yang mereka dapat kerjakan. Soal-soal yang hanya mengungkap "fakta-fakta pendek" untuk 2 - 3 butir tes cukup diberi waktu 1 menit, sedangkan soal tes yang mengungkapkan aspek kognitif di atas aspek pengetahuan dapat diberi waktu 1 menit untuk 1 soal (bentuk tes objektif). Untuk tes-tes uraian yang memerlukan jawaban panjang dapat diberi waktu 3 - 5 menit untuk setiap soal.

3) Proses pemberian skor dan peraturan-peraturannya

Mengenai ketentuan dalam pemberian skor pada setiap jenis tes disarankan memberi bobot yang sama untuk setiap jawaban benar pada soal tes yang mengungkapkan aspek kognitif sejenis. Misalnya, untuk soal berbentuk alternatif jawaban pilihan mengungkapkan aspek kognitif ingatan diberi bobot 1 (satu) untuk setiap jawaban yang benar. Dalam praktek di lapangan pembobotan dalam pemberian skor seperti di atas jarang sekali dilakukan oleh para guru.

Meskipun demikian ada jalan tengah yang dapat dilakukan oleh para guru dalam hal memberikan bobot dalam proses *scoring*. Misalnya untuk 1 soal tes benar-salah yang dikerjakan benar diberi bobot satu karena cenderung hanya mengungkapkan aspek kognitif yang rendah. Tetapi untuk jenis tes alternatif jawaban pilihan -5, untuk 1 soal jawaban yang dikerjakan benar dapat diberi bobot skor 1, 5, dan untuk tes esei dapat diberi bobot skor 3 - 5 untuk setiap soal yang dikerjakan benar.

Hal semacam di atas rasional, sebab untuk setiap jenis/bentuk tes memiliki kecenderungan untuk mengungkapkan aspek kognitif tingkat tertentu. Seperti tes benar-salah atau tes bentuk melengkapi, cenderung untuk mengungkapkan aspek pengetahuan atau ingatan saja. Berbeda dengan tes esei yang menuntut testee lebih banyak menuntut proses mental yang lebih tinggi dari pada sekedar ingatan saja.

Perlu pula diketahui bahwa tes-tes objektif mempunyai kecenderungan mengarahkan kepada siswa untuk menerka jawabannya. Untuk itu dalam proses pemberian skor terhadap tes-tes bentuk objektif dapat dilakukan dengan formula pembetulan (*correction for chance*). Sebenarnya formula ini tidak untuk mengoreksi terhadap penerkaan jawaban melainkan untuk mengoreksi butir-butir soal yang tidak dikerjakan oleh testee.

Formula yang dimaksud adalah:

$$S = b - \frac{s}{n - 1}$$

Keterangan:

- S = Skor yang telah dikoreksi terhadap terkaan
- b = Jumlah jawaban benar
- s = Jumlah jawaban salah
- n = Jumlah option/alternatif jawaban yang disediakan
- 1 = bilangan konstan.

Bila tes bentuk benar-salah yang berarti memiliki option jawaban 2, maka formulanya sebagai berikut:

$$S = b - \frac{s}{2 - 1}$$

Untuk tes alternatif pilihan jawaban dengan option 3, maka formulanya, sebagai berikut:

$$S = b - \frac{s}{3 - 1}, \text{ begitu seterusnya.}$$

Dalam proses pemberian skor, guru harus telah mempersiapkan kunci jawabannya, baik itu bentuk tes objektif maupun tes esei. Begitu pula, apakah ada pembobotan yang berbeda untuk setiap butir soal atau jenis soal, ini juga harus telah ditetapkan terlebih dahulu sebelum proses koreksi/pemberian skor.

Peraturan-peraturan lain seperti apakah dalam mengoreksi dipertimbangkan pula terhadap ejaan yang salah tetapi maksudnya benar, baik - buruknya tulisan atau keruntutan dalam penyusunan kalimat. Seperti misalnya dalam tes matematika bila ada hasil perhitungan yang pecahan apakah harus "dikecilkan" bilangannya, atau apakah diperhitungkan bila suatu soal dikerjakan salah tetapi prosesnya benar. Hal-hal

seperti tersebut di atas perlu ditetapkan terlebih dahulu sebelum proses pemberian skor dilaksanakan. Mengenai bagaimana cara melaksanakan proses memberikan skor dan nilai hasil belajar akan dijelaskan lagi pada modul selanjutnya.

MENILAI TES HASIL BELAJAR

Langkah menilai tes hasil belajar adalah langkah akhir dalam penyusunan instrumen tes hasil belajar.

Telah dijelaskan di bagian awal modul ini bahwa instrumen penilaian hasil belajar harus memenuhi persyaratan: valid, reliabel, memiliki tingkat kesukaran yang memadai dan memiliki kemampuan diskriminasi yang baik. Keempat persyaratan tersebut konsep-konsepnya telah dijelaskan. Pada bagian ini akan dilanjutkan dengan "bagaimana mencari atau menghitung keempat persyaratan tersebut bagi seperangkat tes hasil belajar". Dengan diketahuinya hasil perhitungan tersebut kita dapat memutuskan apakah suatu instrumen penilaian hasil belajar itu, layak/memenuhi syarat atau tidak sebagai alat ukur.

MENGHITUNG VALIDITAS TES HASIL BELAJAR

Meskipun tes hasil belajar persyaratan utamanya harus memiliki validitas isi/kurikulum, namun perlu pula diketahui validitas empiris atau validitas statistiknya, sebab validitas empiris atau validitas statistik ini juga diperlukan. Misalnya tes hasil belajar yang digunakan untuk meramal keberhasilan belajar siswa di kemudian hari.

Jenis-jenis tes yang dibuat paralel juga dipersyaratkan memiliki jenis validitas statistik. Disebut dengan validitas statistik karena dalam proses penghitungan atau interpretasinya menggunakan teknik statistik.

Dalam hal ini ada dua cara dalam pengujian atau penghitungan validitas statistik ini, yaitu validitas internal dan validitas eksternal.

- a. *Validitas internal*, dipenuhi jika ada kesesuaian antara bagian-bagian instrumen (tes) dengan tes keseluruhan. Dengan kata lain tes dikatakan valid jika bagian-bagian tes tersebut mendukung tujuan tes keseluruhan.

Dimaksudkan dengan bagian-bagian tes dapat berupa butir-butir tes atau pokok-pokok materi (faktor-faktor) butir tes. Jika pengujian dilakukan dengan mengkorelasikan antara pokok-pokok (faktor-faktor) butir tes dengan butir tes keseluruhan akan menghasilkan jenis validitas faktor. Jika yang dikorelasikan adalah setiap butir tes dengan butir tes keseluruhan maka hasilnya disebut dengan validitas butir tes.

Oleh karena itu dapat dikatakan suatu tes memiliki validitas yang tinggi jika butir-butir tes yang membentuk tes keseluruhan tersebut tidak menyimpang dari fungsi dan tujuan tes tersebut. Juga dapat dikatakan bahwa suatu tes itu valid jika faktor-faktor (butir-butir tes faktor) yang merupakan bagian-bagian dari butir tes secara keseluruhan tidak menyimpang dari fungsi dan tujuan tes tersebut.

Menyimpang atau tidaknya fungsi butir-butir tes atau faktor-faktor butir-butir tes dapat diketahui dari hasil perhitungan korelasinya dengan butir tes secara keseluruhan.

Agar lebih jelas tentang bagaimana menghitung validitas butir soal dan validitas faktor berikut ini akan diberikan ilustrasi atau contohnya.

Bila kita ingin mengetahui seberapa tingkat penguasaan mahasiswa terhadap modul ini, maka perlu diadakan pengukuran (tes) terhadap mahasiswa tersebut. Selanjutnya, bagaimana membuat tes modul ini supaya memiliki validitas internal (butir soal; faktor).

Kita ketahui bahwa modul ini terdiri dari 3 pokok bahasan/penggalan, yaitu: (1) persyaratan instrumen penilaian hasil belajar; (2) cara menyusun instrumen tes penilaian hasil belajar; dan (3) menilai instrumen penilaian hasil belajar. Masing-masing pokok bahasan tersebut selanjutnya kita sebut dengan faktor I, faktor II, dan faktor III. Jika masing-masing faktor tersebut dibuat 10 butir tes berarti untuk tes modul ini terdiri dari 30 butir tes. Jika ditetapkan untuk setiap jawaban benar diberi bobot skor 1 (satu) maka skor maksimalnya adalah 30. Jika 30 butir tes tersebut disusun berdasarkan pada pokok-pokok materi modul ini berarti tes tersebut telah memiliki validitas isi. Untuk mengetahui apakah 30 butir tes

atau butir-butir tes pada masing-masing faktor tersebut valid atau tidak perlu dihitung validitas internalnya.

Misalkan, tes yang terdiri dari 30 butir soal tersebut telah diujicobakan kepada 10 mahasiswa yang mengikuti kuliah modul ini dan hasilnya seperti pada tabel 1 berikut ini:

Tabel 1 : Data Skor Tes Modul Penilaian Hasil Belajar

No. Subyek	Faktor I					Jumlah skor	Faktor II					Jumlah skor	Faktor III					Jumlah skor	Skor Total			
	Item Nomor						Item Nomor						Item Nomor									
	1	2	3	4	s/d	10	1	2	3	4	s/d	10	1	2	3	4	s/d	10				
1	1	1	-	1	-	1	8	1	1	1	1	-	1	8	1	1	1	1	-	1	7	23
2	1	1	-	1	-	-	7	1	1	-	1	-	-	6	1	1	1	-	-	1	8	21
3	1	1	1	-	-	-	7	-	1	1	1	-	1	7	1	-	1	1	-	1	7	21
4	-	1	-	1	-	1	6	1	1	1	1	-	-	7	1	-	-	1	-	1	7	20
5	-	-	1	1	-	-	5	1	1	-	-	-	-	4	1	-	1	-	-	-	6	15
6	1	1	1	-	-	1	6	1	-	1	1	-	1	7	-	-	1	1	-	1	6	19
7	1	1	1	-	-	-	6	-	-	1	1	-	-	7	1	1	-	1	-	1	8	21
8	-	1	-	1	-	1	7	1	-	1	1	-	1	8	1	-	1	-	-	1	7	23
9	1	1	1	1	-	1	9	1	1	1	-	-	1	8	1	1	1	1	-	1	9	26
10	1	1	1	1	-	1	10	1	1	1	1	-	-	7	1	1	-	1	-	1	9	26

Berdasarkan tabel di atas dapat dihitung baik validitas faktor maupun validitas butir soal. Analisis validitas butir faktor dilakukan dengan cara mengkorelasikan jumlah skor masing-masing faktor. Kemudian dilanjutkan dengan mengkorelasikan jumlah skor faktor I, II, dan III dengan skor total. Hal ini dapat dituliskan sebagai berikut:

- korelasi antara skor faktor I dengan II;
- korelasi antara skor faktor I dengan III;
- korelasi antara skor faktor II dengan III; dilanjutkan dengan ...
- korelasi antara skor faktor I dengan skor total;
- korelasi antara skor faktor III dengan skor total; dan
- korelasi antara skor faktor III dengan skor total.

Formula untuk menghitung besarnya angka koefisien korelasi, digunakan rumus korelasi product moment sebagai berikut:

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

Keterangan:

- r_{xy} = koefisien korelasi antara X dengan Y (misal antara skor faktor I dengan faktor II)
- x = penyimpangan tiap skor X dengan mean ($X - \bar{X}$)
- y = penyimpangan tiap skor Y dengan mean ($Y - \bar{Y}$)

\bar{X} = mean X

\bar{Y} = mean Y

Contoh: Korelasi antara skor faktor I dengan skor total

Tabel 2: Tabulasi Data Skor Faktor I (X) dan Skor Tabel (Y) yang dikorelasikan

No.	X	x	x ²	Y	y	y ²	xy
1	8	0,9	0,81	23	1,5	2,25	1,35
2	7	0,1	0,01	21	0,5	0,25	0,05
3	7	0,1	0,01	21	0,5	0,25	0,05
4	6	1,1	1,21	20	1,5	2,25	1,65
5	5	2,1	4,41	15	6,5	42,25	13,35
6	6	1,1	1,21	19	2,5	6,25	2,75
7	6	1,1	1,21	21	0,5	0,25	0,55
8	7	0,1	0,01	23	1,5	2,25	0,15
9	9	1,9	3,61	26	4,5	20,25	8,55
10	10	2,9	8,41	26	4,5	20,25	13,05
Σ	71		19,69	215		96,5	41,8
\bar{X}	7,1			$\bar{Y} = 21,5$			

Bila data tersebut dimasukkan ke dalam formula/rumus korelasi maka hasilnya adalah sebagai berikut:

$$\begin{aligned} r_{xy} &= \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} \\ &= \frac{41,8}{\sqrt{1900,085}} \\ &= \frac{41,8}{43,589964} \\ &= 0,9589363 \approx \underline{\underline{0,96}} \end{aligned}$$

Angka koefisien korelasi antara skor faktor I dengan skor total adalah sangat tinggi (0,96), hampir mendekati korelasi sempurna (1). Berarti item tes faktor I sangat mendukung tes secara keseluruhan, yang berarti pula bahwa item-item tes faktor I memiliki validitas faktor yang memadai.

Selanjutnya untuk menghitung besarnya korelasi antara faktor II dan III dengan skor total atau korelasi yang lainnya (antara masing-masing butir tes dengan tes keseluruhan atau dengan skor faktor masing-masing) dapat dilakukan dengan cara yang sama.

Pedoman untuk menyimpulkan dan menginterpretasi hasil perhitungan besarnya koefisien korelasi dapat digunakan patokan sebagai berikut:

Tabel Interpretasi nilai r *)

Besarnya nilai r	Interpretasi
0,80 - 1,00	Tinggi
0,60 - 0,80	Cukup
0,40 - 0,60	Agak rendah
0,20 - 0,40	Rendah
0,00 - 0,20	Sangat rendah (tak berkorelasi)

Selain itu Anda dapat mengetahui tingkat signifikansi nilai r dengan konsultasi pada tabel r *product moment*.

b) Validitas eksternal

Instrumen tes hasil belajar dianggap memiliki validitas eksternal jika skor yang dihasilkan dari instrumen tes tersebut sesuai dengan skor hasil tes lain yang sesuai dan yang telah valid. Misalnya kita ingin mengetahui validitas eksternal tes modul ini, maka dapat dilakukan dengan mengkorelasikan skor tes tersebut dengan skor hasil tes modul ini dengan hasil tes penilaian hasil belajar lainnya yang telah valid. Cara lain yaitu dengan mengkorelasikan skor atau nilai rapor atau skor hasil dari tes inteligensi.

, Karena kriteria pembandingan berada di luar alat tes itu sendiri maka yang dihasilkan adalah validitas eksternal.

Cara untuk menghitung validitas eksternal ini sama dengan cara menghitung validitas internal. Misalkan skor tes modul ini (X) hendak dikorelasikan dengan hasil tes inteligensi siswa (Y). Maka dengan formula/ rumus korelasi *product moment*, kedua skor tersebut X dengan Y dapat dicari besarnya koefisien korelasinya.

Formula untuk menghitung validitas (rumus korelasi *product moment*) tersebut hanyalah satu di antara banyak rumus untuk menghitung validitas eksternal dan internal.

Formula/rumus yang lain misalnya:

- (1) teknik korelasi Phi (ϕ);
- (2) teknik korelasi tetra-chorik;
- (3) teknik korelasi Biserial (λ -Bis).

Sebenarnya keempat teknik korelasi di atas merupakan pengembangan dari teknik korelasi *product moment*, disesuaikan dengan tujuan dan jenis data yang ada. Agar dalam hal ini kita mendapat gambaran yang agak konkret, masing teknik korelasi tersebut akan disinggung sedikit, terutama berkenaan dengan "bilamana digunakan".

(1) Teknik koelasi phi (ϕ)

Teknik korelasi phi ini digunakan jika data yang hendak dikorelasikan baik variabel "Y" maupun variabel "X" adalah data nominal atau bersifat diskrit. Persyaratan lain yaitu bila subjek (sampel) yang dianalisis lebih dari 200 orang.

Bentuk tes yang biasa dianalisis dengan teknik korelasi ini adalah tes benar-salah atau yang membedakan mereka yang *memenuhi syarat* dan yang *tidak memenuhi syarat*.

*) Tabel interpretasi nilai r di atas diambil dari Sutrisno Hadi: *Metodologi Research -3*, Yogyakarta, Yayasan Penerbit Fakultas Psikologi UGM, 1979, P.310.

Jika hendak menentukan validitas butir tes maka distribusi skor yang diperoleh para siswa harus diubah menjadi data diskrit/dikhotomi (kelompok atas dan kelompok bawah; atau memenuhi syarat dan tidak memenuhi syarat). Cara yang digunakan untuk menjadikan data asli ke data diskrit ini dapat menggunakan mean, median atau cara lain sebagai titik batas. Jika cara yang digunakan adalah menggunakan median sebagai titik pembagi dua kelompok tersebut maka jumlah subjek yang dibagi dapat sama. Tetapi jika mean yang digunakan mungkin jumlah subjek tersebut bisa tidak sama.

Langkah selanjutnya adalah memberikan skor 1 untuk jawaban benar dan 0 untuk jawaban salah (bila bentuk tes yang dianalisis adalah tes benar-salah).

Selanjutnya dapat diberikan penjelasan dalam bentuk bagan analisis kontingensi sebagai berikut:

Subjek	Jawaban terhadap item		Jumlah total
	Benar (Skor 1)	Salah (Skor 0)	
Skor total lebih besar dari mean/median *) (Skor 1)	a	b	a + b
Skor total lebih kecil dari mean/median *)	c	d	c + d
Jumlah total	a + c	a + c	a + b + c + d

Keterangan :

- a = banyak subjek yang skornya 1 untuk item tes dan skor 1 untuk data dikhotomi
- b = banyak subjek yang skornya 0 untuk item tes dan skornya 1 untuk data dikhotomi
- c = banyak subjek yang skornya-1 untuk item tes dan 0 untuk data dikhotomi
- d = banyak subjek yang skornya baik untuk item tes maupun data dikhotomi adalah 0.

Berdasarkan tabel kontingensi tersebut dapat diketahui besarnya angka korelasi phi ($r\phi$)-nya, dengan rumus:

$$r\phi = \frac{ad - bc}{\sqrt{(a+b)(b+d)(c+d)(a+c)}}$$

Bila data yang didikhotomikan berjumlah sama, maka digunakan rumus

$$r\phi = \frac{Pa - Pb}{2\sqrt{Pq}}$$

Keterangan:

- Pa = Proporsi golongan atas yang menjawab tes dengan benar
- Pb = Proporsi golongan bawah yang menjawab tes dengan benar
- P = Proporsi seluruh subjek yang dapat menjawab tes dengan benar.

Bila telah diketahui besarnya $r\phi$, untuk mengetahui signifikansi tes digunakan tabel phi kuadrat dengan cara:

$$x^2 = \frac{\sum x^2}{N} \text{ atau } = \sqrt{\frac{\sum x^2}{N}} \text{ dengan db} = 1$$

(2) Teknik korelasi *tetrachorik*

Teknik korelasi *tetrachorik* digunakan untuk menghitung besarnya koefisien korelasi antara dua variabel diskrit (*artificial dichotomies*). Sebenarnya data yang dichotomi tersebut merupakan data kontinum yang berbentuk kurva normal yang selanjutnya dibagi menjadi dua kelompok. Pembagian secara *dichotomis* tersebut misalnya menjadi *kelompok yang berhasil* dan *kelompok yang gagal*.

Persyaratan yang dituntut di dalam penggunaan teknik korelasi *tetrachorik* yaitu jumlah sample harus di atas 400 orang.

Rumus atau formula yang digunakan dalam teknik korelasi *tetrachorik* adalah:

a) $r_{kt} = \sin^{-1} (\sqrt{\frac{ad}{bc}})$; dan

b) $r_{kt} = \frac{ad - bc}{ad + bc}$ atau $\frac{bc - ad}{bc + ad}$

Teknik korelasi *tetrachorik* ini jarang digunakan karena selain rumit dalam penghitungannya, juga standar errornya lebih besar. Meskipun demikian bila ingin mendalami teknik ini disarankan untuk mempelajari buku-buku statistik pendidikan atau psikologi.

(3) Teknik korelasi *Biserial*

Teknik korelasi *point - biserial* digunakan untuk mengetahui suatu validitas instrumen, bila yang dikorelasikan adalah dua variabel. Variabel pertama bersifat dichotomis dan variabel keduanya adalah data kontinum.

Teknik korelasi *point biserial* ini cocok untuk analisis validitas tes bentuk benar-salah atau instrumen yang hanya memerlukan jawaban ya dan tidak. Untuk jawaban benar diberi skor 1 dan nol untuk jawaban yang salah.

Misalnya kita akan menganalisis butir soal benar-salah no: 7; dan jumlah butir soal keseluruhan 50 soal. Berdasarkan hasil tes tersebut dapat ditabulasi sebagai berikut:

Korelasi R. Bis
Soal nomor 7: Tes Penilalan Hasil Belajar

S k o r	Jawaban yang		frekuensi
	Benar	Salah	
46 - 50	3	-	3
41 - 45	4	2	6
36 - 40	6	4	10
31 - 35	5	8	13
26 - 30	9	7	18
19 - 25	7	6	13
14 - 20	6	6	14
11 - 15	4	2	6
	-	1	1
Total	44	36	80

Berdasarkan tabulasi tersebut, misalnya telah diketahui data sebagai berikut:

- $X_t = 25,9$ (rata-rata prestasi testee)
- $SD_t = 6,7$ (standar deviasi)
- $X_b = 27,4$ (rata-rata prestasi untuk kelompok yang menjawab benar)
- $X_s = 23,8$ (rata-rata prestasi untuk kelompok yang menjawab salah)
- $n_b = 44$ (banyaknya testee yang menjawab soal benar, proporsinya (P) = 0,55)
- $n_s = 36$ (banyaknya testee yang menjawab soal salah, proporsinya (q) = 0,45)
- $y = 0,3958$ (yaitu tinggi ordinat kurve normal yang memisahkan proporsi 0,55 dengan 0,40)*

Berdasarkan pada hasil di atas maka langkah berikutnya adalah memasukkan data tersebut ke dalam rumus r_{bis} , sebagai berikut:

$$\begin{aligned}
 r_{bis} &= \frac{X_b - X_t}{SD_t} \times \frac{P \cdot q}{y} \text{ atau } r_{bis} = \frac{X_b - X_t}{SD_t} \times P/y \\
 &= \frac{27,4 - 25,9}{6,7} \times \frac{0,55 \times 0,45}{0,3958} \\
 &= 0,537134 \times 0,6253158 \\
 &= 0,3359905 \approx 0,336
 \end{aligned}$$

Bila kita gunakan rumus yang kedua, maka

$$\begin{aligned}
 r_{bis} &= \frac{27,4 - 23,8}{SD_t} \times \frac{0,55}{0,388} \\
 &= 0,2238805 \times 1,3895907 \\
 &= 0,3111022 \approx \underline{\underline{0,311}}
 \end{aligned}$$

Perbedaan hasil perhitungan antara rumus pertama dan kedua disebabkan karena data di atas fiktif.

MENGHITUNG RELIABILITAS TES HASIL BELAJAR

Telah dijelaskan pada bagian sebelumnya bahwa reliabilitas tes hasil belajar berarti kemantapan atau kestabilan tes tersebut jika digunakan. Suatu alat tes disebut reliabel jika dapat menghasilkan data yang stabil/konsisten jika digunakan untuk mengukur gejala yang sama dalam waktu yang berbeda. Jika suatu luas tanah diukur dengan alat meteran dan pengukuran dengan alat tersebut dilakukan berulang-ulang mendapatkan hasil yang sama berarti meteran tersebut reliabel. Barangkali jika meteran tersebut terbuat dari bahan karet yang lentur akan terjadi hasil pengukuran yang dilakukan pertama kali dan yang berikutnya dapat menghasilkan data yang berbeda. Maka bila hal ini yang terjadi berarti alat ukur yang terbuat dari karet tersebut tidak reliabel.

Alat-alat ukur fisik seperti meteran, timbangan cenderung mempunyai reliabilitas yang baik. Tetapi alat ukur yang digunakan untuk hal-hal non fisik (psikososial) sangatlah sulit ditemui alat ukur yang benar-benar reliabel seperti pengukuran terhadap sikap, persepsi, kemampuan hasil belajar hampir tidak pernah terjadi hasil pengukuran ulang yang sama persis dengan hasil pengukuran sebelumnya.

Reliabilitas alat ukur non-fisik seperti gejala tersebut di atas biasanya ditunjukkan dengan angka koefisien korelasi (indeks λ). Indeks λ ini menunjukkan seberapa besar hasil pengukuran pertama dan kedua mempunyai korelasi.

Perhitungan reliabilitas item tes harus dilakukan hanya pada item-item tes yang telah valid, dengan kata lain untuk menghitung apakah suatu item tes telah reliabel didahului dengan perhitungan validitasnya.

Terdapat beberapa teknik untuk menghitung reliabilitas tes yaitu dengan teknik test-retest; teknik belah dua, dan teknik bentuk paralel.

a) Teknik test-retest

Untuk mengetahui reliabilitas tes dengan teknik test-retest, kita harus menggunakan tes tersebut dua kali terhadap subjek yang sama. Tenggang waktu antara pemberian tes pertama dan kedua diperkirakan berjarak antara 15 - 30 hari. Jika tenggang waktu terlalu dekat subjek cenderung masih ingat apa isi tes yang telah dia kerjakan, sedangkan jika tenggang waktu terlalu lama mungkin telah terjadi perubahan pada kondisi pengetahuan atau pengalaman subjek tersebut.

Hasil pengukuran pertama dengan yang kedua selanjutnya kita korelasikan (dengan teknik korelasi product moment) dengan rumus angka kasar sebagai berikut:

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

Misalkan sebuah alat tes kita berikan kepada subjek yang sama dalam waktu yang berlainan, hasilnya sebagai berikut:

Tes I : 8 ; 7 ; 7 ; 6 ; 5 ; 6 ; 6 ; 7 ; 9 ; 10
 Tes II : 8 ; 6 ; 7 ; 7 ; 4 ; 7 ; 7 ; 8 ; 8 ; 7

Bila data tersebut kita susun dalam tabel kerja-r, hasilnya menjadi seperti berikut.

Tabel: Korelasi antara hasil tes II dengan tes I

No.	X	Y	X ²	Y ²	XY
1	8	8	64	64	64
2	7	6	49	36	42
3	7	7	49	49	49
4	6	7	36	49	42
5	5	4	25	16	20
6	6	7	36	49	42
7	6	7	36	49	42
8	7	8	49	64	56
9	9	8	81	64	72
10	10	7	100	49	70
Σ X = 71 X̄ = 7,1		Σ Y = 69 Ȳ = 6,9	Σ X ² = 525	Σ Y ² = 489	Σ XY = 499

$$\begin{aligned}
 r_{tt} &= \frac{10(499) - (71.69)}{\sqrt{\{10.525 - (71)^2\} \{10.489 - (69)^2\}}} \\
 &= \frac{4990 - 4899}{\sqrt{(5250 - 5041)(4890 - 4761)}} \\
 &= \frac{91}{\sqrt{(209)(129)}} = \frac{91}{\sqrt{26961}} \\
 &= \frac{91}{164,19805} = 0,554
 \end{aligned}$$

b) Teknik belah-dua

Untuk dapat menggunakan teknik belah-dua dalam menghitung besarnya reliabilitas tes, alat tes tersebut harus memiliki jumlah item yang cukup banyak < 60 butir item. Semakin banyak jumlah item, reliabilitas yang diperoleh semakin baik.

Cara/prosedur melakukan uji reliabilitas tes adalah:

- (1) Mengumpulkan item-item tes yang telah valid \pm 60 butir.
- (2) Membagi item-item yang telah valid tersebut menjadi dua bagian, misalnya 30 butir item yang bernomor *genap* dan 30 butir yang lainnya bernomor *ganjil*. Cara lain membagi dua bagian ini yaitu dengan random (acak).
- (3) Seperti halnya cara mencari besarnya korelasi yang lain, kedua kelompok tersebut dihitung, dengan rumus korelasi product moment.
- (4) Mengingat besarnya angka korelasi yang diperoleh dengan teknik ini adalah dari alat ukur yang dibelah dua, maka angka korelasinya lebih rendah dari pada jika tidak dibelah. Untuk itu perlu adanya koreksi, dengan memasukkan ke dalam rumus berikut:

$$r_{tot} = \frac{2(r_{tt})}{1 + r_{tt}}$$

Andaikata suatu angka korelasi yang diperoleh dari penghitungan dengan teknik belah-dua adalah 0,554. Maka setelah angka korelasi tersebut dikoreksi dengan rumus di atas hasilnya sebagai berikut:

$$r_{tot} = \frac{2 \times 0,554}{1 + 0,554} = \frac{1,108}{1,544}$$

$$r_{tot} = 0,7129 \approx \underline{\underline{0,713}}$$

c) Teknik bentuk paralel

Teknik bentuk paralel disebut juga dengan *equivalen form* atau *alternatif form*. Cara yang dilakukan dengan teknik ini yaitu dengan membuat 2 jenis perangkat tes yang serupa. Kedua jenis perangkat ini setelah dicari validitasnya, selanjutnya dihitung reliabilitasnya.

Untuk menghitung reliabilitas tes ini yaitu hasil pengukuran dengan alat tes yang pertama (A) dikorelasikan dengan hasil pengukuran dengan alat tes yang kedua (B). Teknik atau rumus untuk menghitung reliabilitas tes ini sama dengan cara-cara sebelumnya, yaitu dengan teknik korelasi product moment. Hasil yang diperoleh dari teknik ini adalah angka korelasi yang untuk tidak perlu lagi dicek dengan rumus koreksi.

Teknik korelasi product moment tersebut adalah salah satu dari rumus yang dapat digunakan untuk mencari reliabilitas tes atau non-tes. Teknik statistik yang lainnya misalnya dengan rumus:

(1) Flanagan:

$$r_{11} = 2 \left(1 - \frac{V_1 + V_2}{V_t} \right)$$

r_{11} = reliabilitas instrumen

V_1/V_2 = varian (SD^2) belahan pertama/kedua

V_t = varian total

Untuk mencari varian digunakan rumus:

$$V = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}$$

(2) Rulon:

$$r_{11} = 1 - \frac{V_d}{V_t}$$

r_{11} = reliabilitas instrumen

V_d = varian beda

d = skor belahan awal dikurangi skor belahan akhir

V_t = varians total

(3) K - R 20:

$$r_{11} = \left(\frac{K}{K-1} \right) \left(\frac{V_t - \sum Pq}{V_t} \right)$$

r_{11} = reliabilitas instrumen

k = banyaknya butir item

V_t = varians total

P = Proporsi subjek yang menjawab benar/skor 1

q = Proporsi subjek yang menjawab salah/skor 0 ($q = P - 1$)

Sebenarnya masih banyak lagi rumus yang dapat digunakan untuk mencari reliabilitas instrumen seperti **UNIVERSITAS MEDAN AREA** sebagainya. Anda dapat mempelajarinya atau memperdalam dengan

menggunakan buku-buku/literatur lain. Tetapi untuk keperluan praktek di lapangan cukup bagi Anda jika melakukan penghitungan reliabilitas tes dengan cara yang dicontohkan di muka.

MENGHITUNG TINGKAT KESUKARAN DAN KEMAMPUAN DISKRIMINASI BUTIR SOAL TES

Telah dijelaskan di muka bahwa suatu tes yang baik selain valid dan reliabel juga harus memiliki tingkat kesukaran dan kemampuan diskriminasi. Untuk memperoleh tentang tingkat kesukaran dan kemampuan diskriminasi harus dilakukan analisis item. Dari analisis item ini akan kita ketahui apakah suatu item tes dapat berfungsi baik atau tidak.

TINGKAT KESUKARAN (K)

Bagaimana mencari dan menafsirkan bahwa suatu tes itu termasuk *sukar*, *sedang* atau *mudah*. Untuk jenis tes esei atau lisan cukup sulit menentukan tingkat kesukarannya. Sebab dalam proses pemberian skor terhadap jawaban tes esei atau lisan banyak terjadi kenisbian. Suatu jawaban tes esei atau lisan biasanya jarang yang diskor 0 (salah mutlak) atau bentuk mutlak. Rentangan skor untuk setiap tes esei dapat berkisar skor 1 - 4. Atas dasar ini korektor dapat menetapkan bahwa suatu tes dikerjakan benar jika setiap soalnya mendapat skor 3 dan jika kurang dari 3 dikategorikan jawaban salah.

Selanjutnya klasifikasi tingkat kesukaran dapat ditetapkan sebagai berikut:

- testee yang menjawab benar hanya sampai 27% termasuk soal tes yang *sukar*;
- testee yang menjawab benar antara 28% - 72% termasuk soal tes yang *sedang*; dan
- testee yang menjawab benar di atas 73% termasuk soal tes yang *mudah*.

Untuk tes bentuk objektif dapat dilakukan analisis butir soal dengan pedoman tabel berikut ini:

Tabel Formulas for Finding (WL + WH) at three Difficulty levels *

Percentase of Testee who "Do Not Know" the Correct Answer to the item	Number of options Each item Has			
	2	3	4	5
16	0,160 n	0,213 n	0,240 n	0,256 n
50	0,500 n	0,667 n	0,750 n	0,800 n
84	0,840 n	1,120 n	1,260 n	1,344 n

Contoh penggunaan tabel di atas

Misalnya untuk tes Matematika jumlah testee 80 orang (N). Berarti 27% x N adalah 21,6 = 22 orang. Jika bentuk tes yang kita berikan adalah benar-salah berarti optionnya adalah dua. Misalkan kelompok atas (22 orang) menjawab benar 20 orang dan kelompok bawah (22 orang) menjawab benar suatu item tes hanya 8 orang maka bila dijumlahkan sama dengan 28 orang testee yang menjawab benar. Langkah selanjutnya kita masukkan ke dalam pedoman untuk tes yang beroption 2, hasilnya adalah:

$$0,160 n = 0,160 \times 28 = 4,48 = \text{mudah}$$

$$0,500 n = 0,500 \times 28 = 14,00 = \text{sedang}$$

$$0,840 n = 0,840 \times 28 = 23,52 = \text{sukar}$$

UNIVERSITAS MEDAN AREA
 Karena kelompok bawah yang menjawab benar item tersebut 28 dan ini berada pada 23,52 ke atas berarti item tersebut termasuk *sukar*.

Cara menghitung tingkat kesukaran item tes yang kedua yaitu dengan menggunakan rumus sebagai berikut:

$$K = \frac{B - \frac{S}{n-1}}{N}$$

Keterangan :

- K = tingkat kesukaran
- B = banyak testee yang menjawab benar
- S = banyak testee yang menjawab salah
- n = jumlah option/alternatif jawaban
- N = jumlah testee keseluruhan.

Misalnya dalam sebuah tes IPA untuk item no. 7 dijawab benar oleh 50 orang, dijawab salah oleh 20 orang, dan tes IPA tersebut adalah tes objektif dengan 5 *options*, maka tingkat kesukaran soal tersebut dapat dicari sebagai berikut:

$$K = \frac{50 - \frac{20}{5-1}}{70}$$
$$= \underline{\underline{0,64}}$$

Di atas telah dijelaskan berbagai cara mencari tingkat kesukaran item tes. Selanjutnya, bagaimana menyimpulkan bahwa suatu item tes itu baik atau tidak bila yang digunakan dasar adalah tingkat kesukarannya. Ada suatu ketentuan umum, bahwa item tes itu baik jika tingkat kesukarannya berkisar pada 0,50. Apabila suatu tes memiliki tingkat kesukaran di bawah 0,10 atau di atas 0,90 maka tes itu kurang dapat berfungsi dengan baik, dan sebaiknya tes seperti itu direvisi atau tidak digunakan lagi.

KEMAMPUAN DISKRIMINASI (D)

Suatu item tes yang baik salah satunya harus memiliki kemampuan diskriminasi (daya membedakan). Suatu kemampuan diskriminasi yang benar pada suatu item tes jika item tes tersebut lebih banyak dijawab benar oleh kelompok atas dibandingkan dengan kelompok bawah. Dikatakan benar mutlak suatu kemampuan diskriminasi item tes jika dijawab benar oleh seluruh testee kelompok atas dan dijawab salah oleh seluruh testee kelompok bawah. Jika terjadi sebaliknya, yaitu kelompok testee bawah lebih banyak menjawab benar suatu item tes dibandingkan dengan kelompok testee atas, berarti tes tersebut mempunyai kemampuan diskriminasi yang salah. Dengan kata lain tes tersebut tidak baik atau perlu diganti.

Cara yang paling sederhana untuk menghitung kemampuan diskriminasi dengan rumus sebagai berikut:

$$D = \frac{B_a - B_b}{n_a \text{ atau } n_b} \times 100\%$$

- B_1 = kelompok atas yang menjawab benar
- B_2 = kelompok bawah yang menjawab benar
- n_1 = jumlah kelompok atas
- n_2 = jumlah kelompok bawah

Makin besar persentase yang diperoleh dari perhitungan di atas dengan hasil positif berarti semakin besar kemampuan diskriminasinya. Misalkan suatu butir tes dikerjakan benar 15 testee, kelompok atas dan 5 testee untuk kelompok bawah, dan jika jumlah $n_1/n_2 = 20$ maka kemampuan diskriminasi tes tersebut adalah

$$D = \frac{15 - 5}{20} \times 100\% = 50\%$$

Ebes (1965) memberikan petunjuk untuk menaksirkan nilai kemampuan diskriminasi (D) sebagai berikut:

- $D \geq 0,40$: butir tes berfungsi sangat memuaskan
- $0,30 \leq D \leq 0,39$: butir tes perlu direvisi sedikit/tidak direvisi sama sekali
- $0,20 \leq D \leq 0,29$: butir tes harus direvisi sebagian
- $D \leq 0,19$: butir tes harus dibuang atau direvisi total.

Selain tingkat kesukaran ataupun kemampuan diskriminasi di atas, bagi tes objektif juga perlu dipertimbangkan tentang kemampuan *options* pilihan jawaban sebagai distraktor daripada kuncinya. Suatu *option* jawaban (bukan kunci jawaban) akan berfungsi sebagai distraktor jika testee ada yang memilihnya paling tidak 5%. *Option* jawaban yang tidak ada pemilihannya berarti kurang baik atau kurang berfungsi sebagai distraktor.

Demikian penjelasan mengenai menilai instrumen tes hasil belajar. Selanjutnya agar pemahaman Anda terhadap penjelasan-penjelasan yang diberikan menjadi lebih mantap, silakan kerjakan latihan berikut ini:

- 1) Apakah yang hendak diketahui dari hasil uji coba suatu tes hasil belajar?
- 2) Hitunglah dan simpulkanlah hasil analisis validitas internal dari data tes berikut!

Data: Skor Item-Item Faktor (X) dan Skor Total (Y)

Subyek/Testee	X	Y
A	5	45
B	4	45
C	3	39
D	5	38
E	3	34
F	2	21
G	2	26
H	1	16
I	4	40
J	3	24

Keterangan:

X = Skor item faktor I

Y = Skor total

- 3) Bila kita memiliki data hasil tes pelajaran IPS yang telah dibelah menjadi belahan I dan II sebagai berikut:

Data/Skor Belahan I

45
45
39
38
34
21
26
16
40
24

Data/Skor Belahan II

45
42
40
38
32
20
24
17
41
24

Hitunglah besarnya angka reliabilitasnya dengan rumus Fianagon!

$$r_{11} = 2 \left(1 - \frac{V_1 + V_2}{V_t} \right)$$

- 4) Jika suatu item tes dengan option 4 dikerjakan dengan benar oleh 35 orang testee kelompok atas dan 10 orang testee kelompok bawah, bagaimana tingkat kesukaran item tes tersebut?
- a) Gunakan tabel untuk menghitung tingkat kesukaran soal tersebut; dan
b) Gunakan rumus:

$$K = \frac{B - \frac{S}{n-1}}{N}$$

- 5) Berdasarkan analisis suatu item tes diperoleh data sebagai berikut:

Pilihan Jawaban	a	b *)	c	D	E	Jumlah
Kelompok atas	5	17	7	4	2	35
Kelompok bawah	8	8	9	5	5	35
Jumlah	13	25	16	9	7	70

*) b ada Kunci Jawaban

- a) berapa tingkat kesukaran soal tersebut?
b) berapa kemampuan diskriminasi soal tersebut?
c) bagaimana fungsi options sebagai distraktornya?

Untuk menjawab/mengerjakan latihan-latihan tersebut, berikut ini diberikan petunjuk sebagai arahan cara mengerjakannya.

- 1). Untuk mengetahui suatu perangkat tes yang baik tidak mungkin bila tanpa data hasil uji coba dari tes tersebut. Data hasil uji coba suatu tes dapat dianalisis apakah tes itu memenuhi syarat atau tidak dari berbagai tuntutan tes yang baik.

- 2) Untuk menjawab latihan nomor 2 Anda dapat melihat pada contoh pada uraian tentang validitas internal.
- 3) Untuk dapat menjawab latihan no. 3 Anda harus dapat menghitung standar deviasi. Varian adalah kuadrat dari standar deviasi. Buku-buku statistik dasar pasti membahas soal standar deviasi, coba Anda kaji, untuk selanjutnya dapat Anda gunakan untuk menjawab latihan 3 ini.
- 4) Latihan no 4, tabel dicontohkan pada bagian uraian. Jadi Anda hanya memasukkan data ke dalam rumus yang tersedia.
- 5) Dengan menggunakan salah satu rumus untuk mencari tingkat kesukaran dan kemampuan diskriminasi soal Anda akan dapat menjawab latihan no. 5 ini. Selanjutnya untuk kemampuan options sebagai distraktor dapat digunakan ketentuan minimal dipilih oleh 5% testee.

RANGKUMAN

Sebagai seorang guru di sekolah menyusun instrumen penilaian hasil belajar baik yang berwujud tes maupun non-tes mutlak harus dilakukan. Untuk dapat menyusun instrumen penilaian hasil belajar yang dapat berfungsi dan mencapai tujuan yang diharapkan perlulah dipersiapkan dengan matang.

Uji coba instrumen sebelum digunakan merupakan usaha yang baik untuk mendapatkan seperangkat instrumen yang baik. Dengan uji coba instrumen bentuk tes khususnya, akan diketahui validitas, reliabilitas, dan tingkat kesukaran maupun kemampuan diskriminasinya. Dengan diketahui hal-hal tersebut dapat kita nilai kualitas dari setiap butir soal. Atas dasar itu pula kita dapat menetapkan mana butir soal yang baik dan yang tidak baik untuk dipergunakan selanjutnya. Berbagai perhitungan dengan teknik statistik dapat digunakan untuk mencari validitas, reliabilitas, tingkat kesukaran maupun diskriminasi soal tes.

TES FORMATIF 3

Petunjuk: Pilihlah salah satu pilihan jawaban yang Anda anggap paling benar dengan memberi tanda silang (X).

- 1) Kondisi penyelenggaraan tes yang dapat mempengaruhi skor hasil tes siswa adalah sebagai berikut, *kecuali:*
 - A. tempat penyelenggaraan tes
 - B. waktu yang disediakan untuk mengerjakan tes
 - C. subjektivitas guru dalam mengoreksi hasil tes
 - D. jam/pukul berapa penyelenggaraan tes dilaksanakan.
- 2) Maksud pemberian waktu yang cukup lama dalam uji coba tes adalah agar para siswa (testee) dapat mencoba mengerjakan seluruh soal tes. Hal ini untuk tujuan sebagai berikut:
 - A. analisis tingkat kesukaran setiap butir soal
 - B. mengetahui validitas-isi tiap butir soal
 - C. mengetahui reliabilitas tes

UNIVERSITAS MEDAN AREA masing-masing testee dalam mengerjakan soal tersebut.

- 3) Jika 100 butir soal tes objektif dengan 5 alternatif pilihan jawaban (*options*), seorang siswa dapat menjawab 75 butir dengan benar sedangkan 25 butir lainnya salah. Berapakah skor siswa tersebut jika dikoreksi dengan *Formula Pembetulan* hasilnya adalah
- 69
 - 66
 - 62
 - 50.
- 4) Suatu alat tes dikatakan memiliki validitas internal jika memiliki ciri seperti berikut ini *kecuali*
- setiap skor butir tes memiliki angka yang signifikan jika dikorelasikan dengan skor total tes itu sendiri
 - skor total tes memiliki angka yang signifikan jika dikorelasikan dengan skor total tes yang lain yang terstandar
 - setiap skor dari faktor-faktor butir tes memiliki angka yang signifikan jika dikorelasikan dengan skor total tes itu sendiri
 - setiap butir tes mendukung keseluruhan butir tes itu sendiri.
- 5) Suatu alat tes akan memiliki validitas eksternal jika memiliki ciri berikut, *kecuali*
- skor tes tersebut memiliki angka yang signifikan bila dikorelasikan dengan tes yang sama yang telah terstandar
 - skor tes tersebut memiliki angka yang signifikan jika dikorelasikan dengan skor tes inteligensi
 - skor tes tersebut memiliki angka yang signifikan jika dikorelasikan dengan tes yang paralel
 - skor tes tersebut memiliki angka yang signifikan jika dikorelasikan dengan nilai rata-rata rapor siswa sejak kelas I hingga kelas III SMA.
- 6) Jika suatu perangkat tes yang telah dibelah menjadi dua (belahan A dan belahan B), diketahui:
- standar deviasi belahan A = 7,3
 - standar deviasi belahan B = 7,8
 - standar deviasi total = 14,4
 - tabel = 0,485.
- Bagaimanakah reliabilitas tes tersebut bila dihitung dengan rumus Flanagan?
- tes tersebut reliabel dengan $r_{hit} > r_{tabel}$
 - tes tersebut tidak reliabel dengan $r_{hit} < r_{tabel}$
 - tes tersebut reliabel dengan $r_{hit} = r_{tabel}$
 - tes tersebut tidak reliabel dengan $r_{hit} \leq r_{tabel}$.
- 7) Bila suatu perangkat tes diketahui angka koefisien korelasi belahan pertama dengan kedua adalah 0,80. Berapakah angka koefisien korelasi yang sebenarnya?
- 0,87
 - 0,88
 - 0,89
 - 0,90.
- 8) Bila suatu butir tes dengan 5 options jawaban yang dikerjakan oleh 100 testee, diketahui hal-hal berikut:
- kelompok atas menjawab benar 23 testee
 - kelompok bawah menjawab benar 11 testee
- Bagaimanakah tingkat kesukaran butir tes tersebut?
- tingkat kesukaran butir tes tersebut ada pada klasifikasi sangat sukar
 - tingkat kesukaran butir tes tersebut ada pada klasifikasi sukar
 - tingkat kesukaran butir tes tersebut ada pada klasifikasi cukup sukar
 - tingkat kesukaran butir tes tersebut ada pada klasifikasi mudah.

- 9) Saran apa yang dapat Anda buat, dari butir tes yang diketahui:
- jumlah kelompok atas atau bawah = 50 orang;
 - kelompok atas yang menjawab benar = 26 orang; dan
 - kelompok bawah yang menjawab benar = 28 orang?
- A. tes tersebut dapat digunakan karena baik kelompok atas maupun bawah yang dapat mengerjakan benar mendekati 50 %
- B. tes tersebut dapat dipergunakan dengan revisi sebagian, karena kelompok bawah lebih banyak yang menjawab benar dibandingkan dengan kelompok atas
- C. tes tersebut tidak dapat digunakan karena tingkat kesukarannya rendah.
- D. tes tersebut tidak dapat dipergunakan karena tidak memiliki kemampuan diskriminasi.
- 10) Suatu options jawaban yang bukan kunci jawaban suatu tes objektif dipilih sebagai jawabannya oleh 90% testee, artinya
- A. stem tes tersebut harus direvisi
 - B. options tes tersebut harus direvisi
 - C. baik options maupun stem tes tersebut harus direvisi
 - D. tes tersebut lebih baik tetap dipergunakan tanpa revisi.

Bila Anda telah mengerjakan ke 10 butir tes objektif di atas, selanjutnya cocokkan dengan kunci jawabannya yang tersedia di akhir modul ini. Kemudian gunakan rumus di bawah ini untuk mengetahui tingkat penguasaan Anda terhadap materi yang telah dipelajari.

Rumus:

$$\text{Tingkat Penguasaan} = \frac{\text{Jumlah jawaban benar}}{10} \times 100\%$$

Makna klasifikasi tingkat penguasaan materi:

90%	-	100%	=	baik sekali
80%	-	89%	=	baik
70%	-	79%	=	sedang/cukup
<		70%	=	kurang.

Jika tingkat penguasaan Anda lebih dari 80%, silakan Anda pelajari modul selanjutnya. Jika kurang dari 80%, Anda harus mengulang mempelajari modul ini, khususnya pada bagian yang belum Anda kuasai.

TEKNIK DAN PROSEDUR PEMBERIAN SKOR DAN NILAI HASIL BELAJAR

PENDAHULUAN

Modul ini terdiri atas dua pokok bahasan. Pokok bahasan *pertama* adalah "Cara Memberi Skor Tes Hasil Belajar" dan pokok bahasan *kedua* adalah "Cara Memberi Nilai Hasil Belajar". Pokok bahasan pertama meliputi:

1. Cara memberi skor hasil tes objektif.
2. Cara memberi skor hasil tes esei atau mengarang.

Pokok bahasan kedua meliputi:

1. Pendekatan dalam penilaian hasil belajar.
2. Teknik statistik yang berkaitan dengan penilaian hasil belajar.
3. Transformasi skor tes ke dalam skala nilai.

Dengan mempelajari modul ini Anda akan memperoleh pengetahuan dan keterampilan dalam proses memberi skor dan nilai tes hasil belajar. Dengan pengetahuan dan keterampilan baru ini Anda dapat mencoba dan mengembangkannya melalui kegiatan belajar-mengajar di kelas, di mana Anda bertugas sebagai guru.

Secara umum tujuan yang hendak dicapai setelah mempelajari modul ini adalah Anda dapat:

1. Melakukan proses pemberian skor tes hasil belajar dari alat tes objektif maupun tes esei atau karangan.
2. Melakukan proses pemberian nilai hasil belajar dengan menggunakan pendekatan PAN dan PAP.
3. Menerapkan teknik statistik untuk keperluan penilaian hasil belajar.

Secara khusus tujuan yang hendak dicapai setelah Anda mempelajari modul ini adalah agar Anda dapat:

- a. membedakan arti skor dan nilai hasil belajar;
- b. mengemukakan prosedur dalam memberi skor hasil tes objektif;
- c. mengemukakan prosedur dalam memberi skor hasil tes esei atau karangan;
- d. memberikan contoh cara memberikan skor hasil tes objektif;
- e. memberi contoh cara memberi skor hasil tes esei atau karangan;
- f. menjelaskan pendekatan PAN dalam penilaian hasil belajar;
- g. menjelaskan pendekatan PAP dalam penilaian hasil belajar;
- h. memberikan alasan-alasan pemilihan pendekatan PAN atau PAP untuk suatu penilaian hasil belajar;
- i. menjelaskan pengertian beberapa ukuran tendensi sentral;
- j. menghitung besarnya mean, median, dan mode dari suatu distribusi skor hasil tes;
- k. meranking suatu hasil distribusi skor hasil tes;
- l. menghitung besarnya variabilitas: (a) kuartil, (b) desil, (c) persentil, (d) standar deviasi, dari suatu distribusi skor tes hasil belajar;
- m. mentransformasi skor tes hasil belajar ke dalam skala 5 (sta - five);
- n. mentransformasi skor tes hasil belajar ke dalam skala 10 (sta - ten/C - Scale);
- o. mentransformasi skor tes hasil belajar ke dalam skala 100 (persentil);
- p. mentransformasi skor tes hasil belajar ke dalam skala 11 (sta - eleven).

1. CARA MEMBERI SKOR TES HASIL BELAJAR

Hendaknya dibedakan antara pemberian skor dengan pemberian nilai tes hasil belajar. Pemberian skor adalah proses yang objektif dalam perhitungan jawaban yang benar dan yang salah terhadap hasil tes setiap siswa. Data yang dihasilkan dari proses pemberian skor berupa data kuantitatif, belum bermakna dan biasa disebut dengan nama skor mentah atau skor saja.

Pemberian nilai merupakan langkah selanjutnya setelah pemberian skor. Proses dalam pemberian nilai merupakan interpretasi terhadap skor dengan menggunakan pedoman kriteria tertentu. Untuk hasil pekerjaan tes esei kadang tidak melalui proses pemberian skor terlebih dahulu. Cara seperti ini biasa disebut dengan cara *Global*, sedangkan bila melalui proses pemberian skor terlebih dahulu disebut dengan cara *Analisis*. Dengan cara global hasil yang didapatkan terlalu kasar. Sebab pemeriksa langsung memberikan nilai pada pekerjaan siswa setelah secara global pekerjaan tersebut dikoreksi/diperiksa. Kata-kata atau simbol-simbol yang biasa digunakan dalam pemberian nilai secara global misalnya: 'Baik (B)' - 'Cukup (C)' - 'Kurang (K)'. Penilaian cara global ini tidak tepat untuk pekerjaan tes sumatif di samping siswa tidak senang/tidak puas jika pekerjaannya dikoreksi dengan cara global ini.

Cara penilaian global ini tidak dibicarakan dalam uraian ini, melainkan hanya penilaian dengan cara analisis saja. Artinya dalam memberikan nilai terhadap hasil pekerjaan siswa dari tes esei atau karangan didahului dengan proses pemberian skor. Berikut ini akan dijelaskan tentang cara pemberian skor hasil tes objektif dan tes esei atau karangan.

CARA MEMBERI SKOR HASIL TES OBJEKTIF

Memberi skor untuk pekerjaan tes-tes objektif lebih mudah dibandingkan dengan pekerjaan tes esei atau karangan.

Pemberian skor untuk pekerjaan tes-tes objektif dapat dilakukan dengan beberapa cara. *Pertama* dengan memberi bobot angka (skor) 1 (satu) untuk setiap butir soal yang dijawab dengan benar dan 0 (nol) untuk jawaban salah. Skor siswa dari suatu tes berarti jumlah semua jawaban yang benar. Cara yang pertama ini tidak perlu mempertimbangkan tingkat kesukaran/tingkat kognitif yang diungkap maupun tipe-tipe tes objektif yang digunakan. Kesemua butir tes diperlakukan sama dalam pembobotannya.

Kedua adalah memberi bobot skor yang berbeda-beda untuk setiap tipe soal tes. Misalkan tipe tes benar-salah (*true - false*) setiap butir soal yang dijawab benar diberi bobot skor 1, tipe memasangkan (*matching*) diberi bobot skor 1,5, dan tipe variasi pilihan (*multiple choices*) diberi bobot skor 2. Rasional cara pemberian skor ini adalah bahwa setiap tipe soal tes objektif memiliki tingkat kesukaran yang berbeda. Maka setiap tipe soal tes perlu diberi bobot yang berbeda. Contoh di atas dapat dikatakan bahwa tes benar-salah lebih mudah dari pada tipe tes memasangkan, namun keduanya lebih mudah dibandingkan dengan tipe variasi pilihan.

Ketiga adalah dalam pemberian skor mempertimbangkan selain tipe soal tes juga tingkat aspek kognitif yang diukur oleh setiap butir soal tes. Untuk tingkat aspek kognitif ini dapat diklasifikasikan menjadi dua, yaitu tingkatan kognitif rendah dan tinggi. Yang termasuk dalam tingkatan kognitif rendah yaitu Pengertian (C_1), Pemahaman (C_2), dan Aplikasi (C_3), sedangkan yang termasuk tingkatan kognitif tinggi adalah Analisis (C_4), Sintesis (C_5), dan Evaluasi (C_6).

Bobot skor untuk butir tes yang mengungkap aspek kognitif rendah lebih kecil dibandingkan dengan yang mengungkap kognitif tinggi. Dengan kata lain pemberian skor cara ketiga ini selain pembobotan dilakukan untuk tipe-tipe soal tes yang berbeda, juga pada tingkatan kognitif yang diukur oleh butir-butir soal tes dalam tipe yang sama.

Cara ketiga dalam pemberian bobot skor ini kelihatan lebih proporsional, namun terlalu rumit, sehingga tidak umum dilakukan. Cara pemberian skor pertama dan kedua biasa dilakukan dalam praktek di lapangan. Cara pemberian skor yang terakhir yaitu dengan rumus koreksi dan cara ini telah

CARA MEMBERI SKOR HASIL TES ESEI ATAU KARANGAN

Telah dijelaskan di atas bahwa proses pemberian skor hasil tes objektif relatif lebih sederhana, lebih mudah dan hasilnya lebih objektif. Hal ini dapat dimengerti karena untuk mengoreksi jawaban yang diberikan siswa cukup bisa dilakukan dengan kunci jawaban yang pasti. Memberi skor untuk pekerjaan tes-tes objektif dapat dilakukan oleh orang lain, meskipun orang tersebut tidak mengetahui sama sekali materi tes. Hal ini sangat tidak mungkin dapat dilakukan untuk pekerjaan hasil tes esei atau karangan. Sebab jawaban yang diberikan oleh masing-masing siswa mempunyai variasi yang lebih luas dibandingkan dengan tes objektif.

Ketidakefektifan dalam memberi skor untuk pekerjaan tes esei atau karangan sangat mungkin terjadi. Apalagi jika seorang pemberi skor (korektor) sadar atau tidak sadar memperhatikan faktor tulisan dan bahasa yang digunakan siswa. Seharusnya kedua faktor tersebut dikesampingkan dalam proses pemberian skor.

Julian C. Stanley dalam bukunya yang berjudul *Measurement in Today's Schools* memberikan saran untuk mempermudah dalam pemberian skor hasil tes sebagai berikut:

- 1) Instruksi tes hendaknya jelas bagi siswa, yaitu apa yang harus dikerjakan siswa. Demikian pula pertanyaan-pertanyaan yang diajukan hendaknya jelas.
- 2) Sebaiknya tidak disediakan soal-soal yang dapat dipilih oleh siswa. Mereka semua harus mengerjakan soal tes yang sama. Sebab dalam penilaian jawaban merekalah yang diperbandingkan.
- 3) Sebaiknya nama siswa ditulis di bawah pekerjaan atau di akhir pekerjaannya (lembar jawabnya). Hal ini untuk menghindari adanya subjektivitas pemeriksa, karena dapat melihat seketika itu nama siswa yang dikoreksi.

Selanjutnya dalam proses pemberian skor hasil tes esei atau karangan beberapa ahli dalam bidang tes menyarankan atau membuat ketentuan untuk menjadi pedoman.

Cochran dan Weidemann (1934) memberi saran untuk menilai (memberi skor) untuk tes-tes esei sebagai berikut:

- 1) Bacalah terlebih dahulu pekerjaan-pekerjaan siswa untuk mendapatkan gambaran umum mengenai keadaan jawaban mereka, untuk selanjutnya memperkirakan skor yang diharapkan.
- 2) Periksa dan berilah skor untuk sebuah pertanyaan bagi semua pekerjaan siswa, sebelum memeriksa nomor atau soal berikutnya. Keuntungan dari cara yang demikian adalah:
 - a) Dengan membandingkan jawaban-jawaban masing-masing siswa akan sebuah soal, ternyata dalam pemberian skornya lebih eksak dan adil.
 - b) Karena pada setiap kali hanya dipikirkan pokok-pokok jawaban yang berkaitan dengan sebuah soal tertentu saja, maka ketelitian pemeriksa meningkat. Akibat berikutnya waktu memeriksa dapat lebih singkat.

Saran butir 2) ini dapat digambarkan sebagai berikut:

Nama Siswa	A	B	C	D	E	dst
Pertanyaan/ soal						
Nomor 1						
Nomor 2						
Nomor 3						
dst						

- 3) Sebelum memeriksa, hendaknya dibaca terlebih dahulu sumber bacaan atau materi serta catatan-catatan yang dimiliki siswa yang dipakai sebagai bahan pertanyaan-pertanyaan.
- 4) Buatlah daftar pokok-pokok terpenting yang harus dicantumkan untuk menjawab sesuatu pertanyaan. Tiap-tiap pokok jawaban tersebut berilah bobot dan skor tertentu. Skor yang diberikan kepada pokok-pokok yang harus dicantumkan tersebut merupakan skor minimum. Jika ada seorang siswa memberikan jawaban melebihi pokok-pokok jawaban dan tidak menyimpang dari pertanyaan, dapat diberi tambahan skor (skor ekstra). Skor ekstra ini harus dibatasi dalam jumlah maksimum tertentu.
- 5) Setelah pokok-pokok dipertimbangkan selanjutnya proses pemberian skor terhadap pekerjaan siswa dapat dimulai. Caranya bacalah satu kali baru selanjutnya diteliti kembali untuk melihat adanya fakta-fakta atau pokok-pokok yang bersifat detail. Berilah tanda-tanda bila menjumpai adanya kesalahan dan tuliskan pembedulannya secara singkat. Pokok-pokok jawaban yang semestinya dicantumkan tetapi tidak ditulis oleh siswa berarti tidak ada skor untuk pokok tersebut. Skor akhir untuk masing-masing pertanyaan merupakan penjumlahan skor minimum dari pokok-pokok jawaban yang diberikan siswa ditambah skor ekstra jika ada. Begitu seterusnya untuk memberikan skor pada masing-masing butir pertanyaan.

Ahli lain, Sims dalam tulisannya yang berjudul *The Objectivity, Reliability, and Validity of on Essay Examination Graded by Rating* memberi saran untuk memeriksa hasil tes esei atau karangan sebagai berikut:

- 1) Bacalah sepiantas lalu/secara cepat semua pekerjaan siswa. Sesuai dengan pendapat pemeriksa mengenai nilai masing-masing pekerjaan tersebut pilah-pilahkanlah (klasifikasikanlah) menjadi 5 kelompok. Kelompok I adalah pekerjaan-pekerjaan siswa yang *sangat baik*, kelompok II termasuk kriteria *baik*, kelompok III kriteria *sedang*, kelompok IV kriteria *kurang*, dan kelompok V kriteria *sangat kurang*.
- 2) Kertas-kertas pekerjaan dalam tiap-tiap kelompok dibaca kembali untuk menghindari penempatan yang salah.

Pemilihan dan penempatan kertas-kertas pekerjaan siswa ke dalam kelompok-kelompok di atas, yaitu yang lebih kurang setara nilainya, dapat menghindari kemungkinan timbulnya bias dalam pemberian skor. Seperti dikemukakan *Stalnaker*, bahwa skor-skor/ nilai-nilai yang diberikan terhadap sebuah pekerjaan seringkali terpengaruh oleh pekerjaan yang dibaca sebelumnya. Dalam hal ini memang lebih baik lagi jika pengelompokan tersebut dilakukan setiap kali setelah membaca sebuah jawaban tertentu. Dengan kata lain setiap jawaban butir soal tersusun kelompok pekerjaan siswa-siswa yang berlainan. Meskipun demikian yang terakhir ini memang cukup berat bagi pemeriksa tetapi baik untuk dilakukan. Hal yang paling sulit dihindarkan oleh pemeriksa tes esei adalah pengaruh nama-nama siswa.

Berikut ini diberikan sebuah contoh soal tes esei dan sebuah lagi tugas mengarang.

- (1) Jelaskanlah mengapa tes esei tidak tepat digunakan untuk tes sumatif!
- (2) Buatlah *pendahuluan* suatu makalah dengan judul *Pacaran dan Kumpul Kebo di Kalangan Pelajar*.

Catatan: - Pendahuluan makalah tersebut tidak boleh lebih dari satu setengah halaman folio!
- Jumlah paragraf antara 3 s/d 4.

Memperhatikan saran-saran cara memeriksa dan memberi skor tes esei atau karangan di atas maka 2 butir soal tersebut dapat dilakukan proses pemeriksaan pekerjaan siswa sebagai berikut:

Pertama membuat daftar pokok-pokok jawaban yang dianggap memadai bagi tujuan butir tes. Bila kita perhatikan soal No. (1), soal tersebut bertujuan mengungkap kemampuan evaluasi siswa terhadap kegunaan tes esei dalam fungsi sumatif. Pokok-pokok jawaban soal No. (1) tersebut dapat dibuat

(1) Tes esei tidak tepat digunakan untuk tes sumatif karena:

- (a) Validitas tes esei menjadi rendah karena dalam waktu relatif singkat hanya mengukur sebagian kecil materi yang diajarkan. Padahal dalam tes sumatif dituntut dapat mengungkap tujuan atau materi yang diajarkan secara representatif.
- (b) Secara umum reliabilitas tes esei rendah, apalagi jika pembuatan tes esei tersebut tidak direncanakan dengan seksama. Misalnya: (a) aspek apa yang diungkap, (b) apa yang seharusnya dikerjakan siswa dalam menjawab tes tersebut, (c) bagaimana cara penskorannya dan sebagainya. Bila hal ini tidak mendapat perhatian oleh pembuat tes jelas reliabilitas tes tersebut rendah.
- (c) Siswa yang sebenarnya secara keseluruhan materi lebih mampu/lebih menguasai, dapat terjadi memperoleh nilai jelek gara-gara soal yang diajukan belum terkuasai dengan baik. Sebaliknya siswa yang kebetulan hanya belajar materi tertentu saja dan kebetulan menjadi soal yang diajukan dalam tes sumatif jelas akan mendapat nilai baik meskipun sebenarnya secara keseluruhan dia tidak menguasai.

(2) Pendahuluan makalah dengan judul *Pacaran dan Kumpul Kebo*

Dalam tugas mengarang ini telah dibatasi oleh ketentuan-ketentuan: (a) tidak boleh lebih dari satu setengah halaman folio, dan (b) jumlah paragraf antara 3 s/d 4. Ketentuan seperti ini sangat diperlukan untuk mempermudah dalam pemberian skor. Siswa harus berpikir tentang ide pokok-pokok yang harus dikemukakan dalam setiap paragraf, sebelum dia menuliskannya. Dengan ketentuan tersebut juga memaksa siswa untuk menggunakan kalimat yang seefektif (lugas) dan penuh isi, karena hanya dibatasi paling banyak satu setengah halaman.

Atas dasar ketentuan tersebut pembuat soal ini dapat membuat rambu-rambu/pokok-pokok jawaban dengan mendasarkan pada konsep tentang pendahuluan suatu makalah dan ketentuan-ketentuan yang ditetapkan.

Pokok-pokok jawaban soal No. (2) ini dapat disusun sebagai berikut:

- (a) Harus dicantumkan tentang konsep pacaran dan kumpul kebo.
- (b) Harus dicantumkan latar belakang sosio-kultur/teknologi dan latar belakang kondisi psikologis anak usia remaja (pelajar).
- (c) Harus dicantumkan pokok-pokok persoalan yang hendak dibahas atau dijelaskan pada bagian selanjutnya, misalnya:
 - Pacaran sebagai kebutuhan yang wajar bagi remaja/pelajar.
 - Pacaran yang tidak bertanggung jawab dan dampaknya
 - Pacaran yang bertanggung jawab.
 - dan lain sebagainya.
- (d) Setiap paragraf harus berisi satu ide pokok.

Langkah kedua menetapkan bobot skor. Perlu diingat apakah untuk setiap butir soal dalam suatu tes akan diberi bobot skor yang sama atau tidak. Hal ini tergantung pada pemeriksa atau pembuat tes. Ada orang yang memberi bobot skor berbeda untuk butir tes yang mengungkap aspek kognitif yang berbeda, tetapi juga banyak orang yang tidak membedakan bobot skor untuk setiap butir tes.

Untuk 2 soal di atas misalnya masing-masing diberi bobot skor maksimal adalah 5, maka dalam proses pemberian skor dapat dilakukan sebagai berikut:

Soal No.1

Pokok jawaban (1a) dapat diberi bobot skor minimal 3, butir (1b) dan (1c) diberi bobot skor masing-masing 1. Hal ini dilakukan karena pokok jawaban butir (1a)-lah yang paling penting/harus ada pada jawaban yang dibuat siswa. Bila ada siswa yang menjawab lebih dari pokok-pokok jawaban tersebut dan tidak menyimpang dari persoalan dapat diberi skor ekstra 1.

Soal No. 2

Pokok jawaban butir (2a) dan (2b) masing-masing diberi bobot skor 1, sedangkan butir (2c) dan (2d) diberi bobot skor masing-masing 1,5. Satu skor ekstra dapat diberikan kepada siswa yang dalam menjawab benar dan benar. Dapat pula dilakukan pengurangan satu skor, jika siswa melanggar ketentuan yang ditetapkan.

Langkah ketiga adalah pengelompokan kualitas pekerjaan siswa untuk setiap nomor/butir soal. Karena yang dicontohkan di sini 2 soal maka pengelompokan pekerjaan siswa ini dengan membaca sekilas tentang jawaban siswa untuk setiap butir soal. Klasifikasi/pengelompokan dibagi menjadi kelompok sangat baik, baik, sedang, kurang, dan kurang sekali.

Langkah keempat adalah pemeriksaan dan pemberian skor terhadap pekerjaan siswa. Bila semua pekerjaan telah diberi skor untuk setiap butir atau pokok-pokok jawaban yang diberikan siswa, proses terakhir adalah menjumlahkan skor-skor masing-masing siswa tersebut. Jumlah dari skor-skor yang ada pada pekerjaan setiap siswa itu adalah skor total yang dimiliki oleh siswa tersebut.

Agar pemahaman dan keterampilan Anda terhadap pokok bahasan di atas mantap, cobalah kerjakan latihan berikut ini:

- 1) Bilamana pemeriksaan secara global terhadap hasil pekerjaan siswa dilakukan? Jelaskan!
- 2) Apakah yang dimaksud dengan nilai dan skor hasil belajar?
- 3) Apa kelebihan pemeriksaan atau pemberian skor tes dengan cara memeriksa sebuah persoalan atau butir tes terhadap seluruh siswa yang dites?
- 4) Buatlah sebuah tes esei dan buatlah pula pokok-pokok jawabannya serta bobot skornya!
- 5) Tetapkanlah apa saja yang harus dinilai/diberi skor untuk tes/tugas mengarang.

Bila Anda telah mengerjakan latihan-latihan di atas, selanjutnya cocokkan pekerjaan Anda dengan petunjuk-petunjuk berikut ini.

- 1) Hasil tes siswa dapat dikoreksi atau diperiksa secara global jika maksud diadakannya tes tersebut tidak untuk fungsi sumatif. Pekerjaan siswa dapat dikoreksi secara global, jika maksud diadakannya tes tersebut hanya untuk mengetahui tingkat penguasaan siswa terhadap suatu materi tertentu. Dengan kata lain tes tersebut untuk diagnosis atau menjalankan fungsi formatif.
- 2) Nilai merupakan hasil proses evaluasi yang berupa data kualitatif, sedangkan skor merupakan hasil dari proses pengukuran yang datanya bersifat kuantitatif. Dalam bidang pendidikan di sekolah biasanya nilai diperoleh dari hasil pemrosesan skor dengan menggunakan kriteria tertentu.
- 3) Kelebihan dengan cara memeriksa satu buah butir tes terhadap keseluruhan siswa yang dites adalah:
 - (a) Waktu pemeriksaan lebih dapat dipersingkat karena pemahaman pemeriksa terhadap jawaban-jawaban siswa makin lama menjadi makin jeli dan semakin hafal terhadap pokok-pokok jawaban yang benar ataupun yang salah.
 - (b) Dalam proses pemberian skor akan lebih adil dan realistis serta lebih objektif. Apalagi jika didahului dengan pengelompokan pekerjaan siswa sesuai dengan kualitasnya, maka objektivitas tersebut akan lebih terjamin.
- 4) Untuk menjawab latihan butir 4 ini Anda dapat mendiskusikan dengan teman Anda dengan acuan contoh-contoh yang ada pada uraian pokok bahasan modul ini.
- 5) Untuk menetapkan pokok-pokok jawaban tugas mengarang antara lain yang harus dikemukakan oleh siswa dalam karangannya, antara lain:
 - (a) Kesesuaian antara judul dengan isi keseluruhan.
 - (b) Sumber kepustakaan yang relevan atau tidak dengan permasalahan yang dibahas.
 - (c) Adanya data empiris/kuantitatif yang mendukung isi karangan.
 - (d) Susunan paragraf, apalagi setiap paragraf berisi satu ide pokok saja, apakah alur ceritanya antara paragraf satu dengan paragraf berikutnya menunjukkan hubungan yang runtut.
 - (e) Apakah dalam penyusunan kalimat sudah menggunakan Bahasa Indonesia yang baik dan benar.
 - (f) Apakah secara keseluruhan isi karangan itu menarik (aktual, problematik, orisinal, memberikan manfaat bagi pihak yang berkepentingan).
 - (g) Apakah tulisan karangan tersebut baik, rapi, dan mudah dibaca.

Setelah pokok-pokok jawaban tersebut dibuat, selanjutnya adalah pemberian bobot skor terhadap pokok-pokok jawaban tersebut. Untuk pokok jawaban butir a, b, c, d, e, dan f masing-masing diberi skor 1,5, sedangkan butir g skor 1. Satu skor ekstra dapat diberikan pada pekerjaan siswa yang melampaui dan menentang hal di luar pokok jawaban tersebut namun masih relevan dengan judul karangan.

RANGKUMAN

Pada prinsipnya proses pemberian skor harus dibedakan dengan proses pemberian nilai hasil belajar. Dalam bidang pendidikan di sekolah proses pemberian nilai biasa dilakukan setelah proses pemberian skor.

Dalam pemberian skor untuk tes-tes objektif relatif sederhana dan mudah dilakukan. Hasil dari pemberian skor tes-tes objektif ini pun menunjukkan objektivitas yang cenderung sempurna.

Pemberian skor untuk tes objektif berbeda dengan tes esei atau karangan. Hasil dari pemberian skor tes esei cenderung kurang objektif. Untuk meningkatkan objektivitas ini para ahli memberikan beberapa saran. Secara umum saran-saran untuk memeriksa atau memberikan skor terhadap pekerjaan tes esei adalah:

- 1) Siapkan terlebih dahulu daftar pokok-pokok jawaban yang memadai bagi tujuan setiap butir tes.
- 2) Berikan skor tertentu pada pokok-pokok jawaban yang esensial.
- 3) Berikan skor terhadap sebuah pertanyaan pada semua kertas pekerjaan siswa sebelum dilanjutkan pada pertanyaan lain.

TES FORMATIF 1

Petunjuk: Silanglah salah satu option jawaban A, B, C, atau D yang Anda anggap paling benar.

- 1) Yang merupakan hasil proses pengukuran adalah
 - A. Dari 100 soal yang diajukan, Ali mendapat nilai sangat memuaskan
 - B. Dari 100 soal yang diajukan, Ali dapat mengerjakan 95 soal dengan benar
 - C. Prestasi Ali dalam mengerjakan 100 soal yang diajukan, berkedudukan pada rata-rata kelas
 - D. Karena skor Ali ada pada rata-rata kelas berarti prestasi dia termasuk sedang di kelompoknya.
- 2) Yang merupakan hasil proses penilaian adalah
 - A. Hasil karangan Ali memenuhi kriteria yang disediakan
 - B. Karangan Ali mendapat pujian dari para guru
 - C. Karangan Ali mendapat juara I tingkat kabupaten.
 - D. Karangan Ali setiap pokok-pokok jawaban yang dituntut mendapat skor maksimal.
- 3) Manakah di antara pernyataan di bawah ini yang tergolong pemeriksaan secara analisis?
 - A. Pemeriksa memberi nilai B pada pekerjaan Toni sebagai simbol "baik"
 - B. Pemeriksa memberikan skor-skor pada setiap pokok jawaban dan menjumlahnya
 - C. Pemeriksa dengan membaca keseluruhan pekerjaan siswa kemudian memberi nilai terhadap pekerjaan tersebut
 - D. Pemeriksa mengelompokkan pekerjaan siswa menjadi 5 klasifikasi, yaitu A, B, C, D, dan E selanjutnya sebagai nilai masing-masing.
- 4) Berikut ini adalah maksud saran yang dikemukakan oleh Julian C. Stanley bahwa "Instruksi tes hendaknya jelas bagi siswa yaitu apa yang harus dikerjakan siswa" untuk menjawab persoalan yang diajukan, **KECUALI:**
 - A. untuk mempermudah pemeriksaan pekerjaan siswa
 - B. agar tidak terjadi salah penafsiran siswa terhadap soal tes

- C. agar siswa lebih terarah dalam mendekati persoalan yang ada pada tes
 D. agar siswa dapat dengan mudah menjawab persoalan yang ada pada tes.
- 5) Untuk meningkatkan objektivitas dalam memeriksa atau memberi skor pada tes esei dapat dilakukan hal-hal berikut, KECUALI:
 A. mengganti nama siswa dengan nomor kode
 B. memeriksa secara analisis
 C. memeriksa nomor soal yang sama terhadap keseluruhan siswa
 D. mempersiapkan daftar pokok-pokok jawaban untuk setiap soalnya.
- 6) Memeriksa tes bentuk objektif hasilnya lebih objektif daripada memeriksa tes bentuk esei, karena:
 A. kunci jawaban tes objektif lebih eksak daripada tes esei
 B. memeriksa tes objektif lebih mudah daripada memeriksa tes esei
 C. bentuk tes esei lebih sulit dibandingkan dengan tes objektif menurut siswa yang mengerjakannya
 D. tes bentuk objektif lebih memiliki validitas yang memadai daripada tes bentuk esei.
- 7) Sebuah tes esei berbunyi "Jelaskan bahwa tes bentuk objektif lebih valid daripada tes esei, untuk fungsi sumatif!" Maka yang menjadi pokok jawaban terpenting adalah
 A. tes objektif lebih dapat mengukur tujuan secara representatif daripada tes esei
 B. tes objektif dapat mengukur materi pelajaran lebih komprehensif daripada tes esei
 C. dalam waktu yang relatif singkat tes objektif lebih efektif daripada tes esei
 D. tes esei dapat digunakan untuk maksud menjalankan fungsi sumatif jika direncanakan dengan seksama.
- 8) Sebuah tes berbunyi "Bagaimana cara memberi skor pada tes esei?", maka yang menjadi pokok jawabannya adalah seperti berikut, KECUALI:
 A. mengelompokkan nama-nama siswa yang memiliki kualitas jawaban yang relatif sama terhadap suatu butir pertanyaan tertentu
 B. memberikan bobot skor terhadap pokok-pokok jawabannya
 C. menetapkan skor ekstra maksimum untuk setiap butir soal
 D. memberikan nilai setiap skor pokok-pokok jawaban semua butir soal kemudian dijumlah.
- 9) Tes bentuk esei lebih sering digunakan di sekolah-sekolah khususnya sebagai fungsi sumatif, karena alasan berikut, KECUALI:
 A. mudah dan cepat membuatnya
 B. dapat diperiksa secara global
 C. dapat mengungkap kemampuan kognitif yang kompleks
 D. mudah dalam pemeriksaannya.
- 10) Manakah di antara tes esei berikut ini yang paling kompleks dalam mempersiapkan daftar pokok-pokok jawabannya?
 A. Jelaskan beberapa alasan kegunaan tes esei untuk mengukur hasil belajar siswa!
 B. Buatlah definisi pengukuran menurut pendapat saudara sendiri!
 C. Jelaskan bahwa tes objektif lebih valid daripada tes esei untuk mengukur hasil belajar selama satu semester!
 D. Kemukakan langkah-langkah pemberian skor terhadap pekerjaan tes esei!

Setelah Anda mengerjakan soal tes formatif di atas, selanjutnya koreksilah dengan kunci jawaban pada bagian akhir modul ini. Hitunglah jawaban Anda yang benar dan gunakan rumus di bawah ini untuk mengetahui tingkat penguasaan Anda.

Rumus:

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah jawaban Anda yang benar}}{10} \times 100\%$$

Arti tingkat penguasaan yang Anda capai:

90% - 100% = baik sekali

80% - 89% = baik

70% - 79% = sedang

< 70% = kurang

Jika Anda mencapai tingkat penguasaan 80% ke atas, silakan meneruskan mempelajari pokok bahasan berikutnya. Bagus! Bila tingkat penguasaan Anda kurang dari 80%, Anda harus mengulangi materi ini, terutama bagian yang belum Anda kuasai.

2. CARA DAN PROSEDUR MEMBERI NILAI HASIL BELAJAR

Pada bagian pendahuluan telah dikemukakan bahwa dalam pokok bahasan kedua ini akan dibicarakan: (1) pemekatan dalam penilaian hasil belajar, (2) penggunaan teknik statistik dalam penilaian hasil belajar, dan (3) transformasi skor hasil tes ke dalam beberapa skala penilaian dan interpretasinya.

PENDEKATAN DALAM PENILAIAN HASIL BELAJAR

Penilaian hasil belajar diawali dengan proses pengukuran mempunyai tujuan untuk mengetahui seberapa jauh siswa dapat mencapai tujuan instruksional yang diharapkan. Hasil pengukuran yang berupa skor (angka) hasil belajar mempunyai rentangan yang relatif. Rentangan skor siswa ini perlu diolah lagi supaya lebih bermakna bagi siswa khususnya dan pihak lain yang berkepentingan. Proses pengolahan skor biasanya dilakukan dengan membandingkan skor tersebut dengan suatu kriteria atau pedoman tertentu. Proses inilah yang selanjutnya menghasilkan nilai hasil belajar.

Pedoman atau kriteria yang digunakan dalam proses penilaian pada pendidikan di sekolah antara lain:

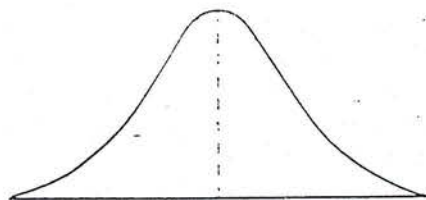
- (1) Kriteria yang berpedoman pada skor teman sekelas atau sekelompok usia sebaya.
- (2) Kriteria dengan berpedoman pada "batas minimal lulus" atau batas minimal pencapaian tujuan/tingkat penguasaan.
- (3) Kriteria yang berdasarkan pada prestasi atau hasil yang pernah diperoleh siswa di masa lalunya.
- (4) Kriteria yang menggunakan kemampuan dasar atau potensi anak itu sendiri. Hal ini biasanya berupa hasil tes kemampuan umum atau bakat dengan alat tes psikologis yang terstandar.

PENDEKATAN PENILAIAN ACUAN NORMA (PAN)

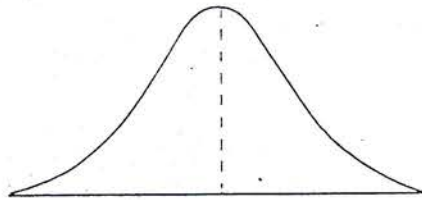
Menggunakan pendekatan PAN berarti dalam pemberian nilai didasarkan pada kondisi hasil tes "apadanya" atau pada fakta yang dihadapi. Fakta yang dimaksud adalah hasil atau skor-skor yang diperoleh sekelompok siswa yang dites.

Dalam menilai prestasi belajar dengan PAN kriteria yang digunakan adalah nilai rata-rata kelas (Mean) dan standar deviasinya. Oleh sebab itu penetapan ketentuan batas lulus dapat berubah-ubah sesuai dengan kondisi kemampuan kelompok siswa terhadap materi tes. Seorang siswa yang diuji di kelompok/kelas III^B mendapat nilai baik dapat terjadi menjadi nilai sedang jika dibandingkan dengan kelas III^A dalam tes yang sama. Dengan kata lain, hasil yang diperoleh tersebut relatif. Oleh karena itu pendekatan ini disebut *standar relatif*.

Dasar pikiran penggunaan pendekatan PAN adalah bahwa setiap gejala populasi yang heterogen terdapat kelompok baik, kelompok sedang, dan kelompok kurang. Hal ini dimulai dalam bidang pengukuran psikologis dengan pengukuran bakat, yang konon bakat ini dibawa sejak lahir. Gejala yang nampak dari pengukuran bakat ini berupa indeks kecerdasan atau *Intelligence Quotient* (IQ). Populasi IQ akan tergambar sebagai sebuah kurva normal. Bila anak-anak yang telah diukur IQ-nya tersebut belajar, maka prestasi atau hasil belajar mereka pun akan tergambar sebagai kurva normal.



Gambar 1
Kurva normal IQ



Gambar 2
Kurva normal prestasi belajar

Pendekatan PAN ini digunakan pertama kali pada tahun 1908 (Cureton, 1971), dengan landasan dasar bahwa tingkat pencapaian hasil belajar siswa akan tersebar menurut kurva normal. Dengan demikian penggunaan pendekatan PAN merupakan hal yang tidak dapat dibantah lagi (Suharsimi Arikunto, 1987).

Akhirnya dapat disimpulkan bahwa dalam penilaian hasil belajar dengan PAN pasti akan diperoleh distribusi sebagian siswa yang tidak berhasil/kurang, sebagian siswa termasuk berprestasi sedang, dan sebagian yang lain adalah kelompok siswa yang berhasil dengan baik.

PENDEKATAN PENILAIAN ACUAN PATOKAN (PAP)

PAP disebut juga dengan PAK (Penilaian Acuan Kriteria). Pada dasarnya pendekatan PAP dilakukan dengan cara membandingkan hasil belajar siswa dengan suatu patokan atau kriteria tertentu. Kriteria ini telah ditetapkan sebelum proses pengukuran/testing. Dengan demikian kriteria yang digunakan tidak perlu dicari di tempat lain.

Tentang kriteria yang menjadi batas lulus atau tingkat penguasaan minimal memang tidak ada ketentuan yang pasti. Hal ini sangat bergantung pada tujuan yang diharapkan dicapai oleh siswa menurut guru atau lembaga yang berkepentingan. Dalam sistem belajar tuntas (*mastery learning*) barangkali ditetapkan 70% atau 75% sebagai batas lulus, sedangkan untuk sistem belajar dengan modul di UT seperti Anda sekarang ini ditetapkan 80%. Ini berarti bahwa mereka yang telah mencapai tingkat penguasaan *sama* atau *lebih* dari kriteria yang ditetapkan diperbolehkan untuk mempelajari modul atau materi berikutnya. Bagi yang belum mencapai kriteria yang ditetapkan harus mengulang materi atau modul yang sedang dipelajarinya itu.

Adanya kriteria batas lulus yang ditetapkan menunjukkan bahwa pendekatan PAP memiliki ciri yang mutlak dan konsisten. Arti mutlak dan konsisten ini adalah bahwa hasil pengukuran dari satu periode ke periode berikutnya untuk semua siswa harus diperlakukan sama.

Penggunaan pendekatan PAP ini menuntut guru untuk membuat alat tes yang dapat mengukur tujuan instruksional secara representatif sebab kesimpulan yang dibuat dari hasil penilaian ini bermakna "persentase tingkat penguasaan terhadap materi yang telah dipelajari".

STATUS NILAI DILIHAT DARI PENGGUNAAN PENDEKATAN PAN DAN PAP

Skor dan nilai yang diperoleh siswa merupakan penghargaan atas usaha belajarnya. Namun skor dan nilai tersebut belum berarti dan bermakna jika kita tidak mengetahui pendekatan penilaian yang digunakannya. Bila pendekatan yang digunakan adalah PAP kita akan lebih mudah memberi maknanya, dibandingkan jika kita menggunakan PAN.

Jika PAP yang kita gunakan berarti skor ataupun nilai yang didapat oleh setiap siswa merupakan tingkat prosentase penguasaan terhadap materi yang diujikan. Kesimpulan seperti ini tidak berlaku

jika pendekatan yang kita gunakan adalah PAN sebab nilai yang diperoleh siswa dari hasil penilaian dengan pendekatan PAN merupakan hasil perbandingan skor individu siswa dengan kelompoknya. Misalnya: Seorang siswa dapat menjawab benar suatu tes objektif sebanyak 60 butir dari 120 butir soal tes. Skor 60 ini belum mempunyai makna, jika kita tidak mengolahnya dengan suatu pendekatan dalam penilaian. Jika dalam penilaian tersebut pendekatan yang digunakan adalah PAP, maka kita harus mengetahui ketentuan atau kriteria *batas lulus* yang ditetapkan. Jika kriteria batas lulusnya adalah 75% maka individu tersebut belum berhasil, sebab dia baru mencapai 50%. Bila pendekatan yang digunakan dalam penilaian adalah PAN *barangkali* dia dapat berhasil atau lulus, kalau skor rata-rata kelas yang mengikuti tes tersebut lebih rendah dari 60.

Baik pendekatan PAN maupun PAP erat kaitannya dengan kegiatan belajar-mengajar yang berorientasi pada TIK. Dengan kata lain, TIK dijadikan arahan dalam kegiatan belajar-mengajar juga merupakan acuan dalam pembuatan alat ukur tes. Pendekatan PAN tidak berpedoman pada tuntutan penguasaan terhadap TIK, sedangkan pendekatan PAP menggunakan tuntutan tersebut.

Manfaat yang bisa diperoleh dari penilaian dengan pendekatan PAN, adalah nilai yang diperoleh siswa dapat digunakan untuk meramalkan keberhasilan studi siswa. Misalnya siswa yang selalu memiliki skor tes pada posisi *Mean + 1,5 standar deviasi*, berarti siswa tersebut sangat potensial dalam kelompoknya. Atas dasar posisi yang tetap ini kita dapat meramalkan bahwa kesuksesan dia dalam belajar kelak sangat besar. Sedangkan pendekatan PAP memiliki manfaat untuk penyelenggaraan program remedial atau pengayaan pada siswa. Dari hasil penilaian pendekatan PAP akan diperoleh informasi tentang tingkat persentase penguasaan masing-masing siswa terhadap suatu unit materi ajar.

TEKNIK STATISTIK YANG BERKAITAN DENGAN PENILAIAN HASIL BELAJAR

Hasil pengukuran terhadap penguasaan bahan pelajaran biasanya berwujud angka atau skor. Agar angka atau skor ini dapat memberikan informasi yang bermakna perlu dianalisis dan ditafsirkan. Untuk menganalisis skor hasil belajar ini biasa digunakan teknik statistik. Dengan teknik statistik inilah akan diperoleh hasil perhitungan yang objektif dan dapat dipertanggungjawabkan secara ilmiah.

Teknik statistik yang berkaitan dengan analisis skor hasil belajar adalah ukuran *tendensi sentral* dan ukuran *variabilitas*. Yang termasuk ukuran tendensi sentral adalah *Mean*, *Median*, dan *Mode*, sedangkan yang termasuk dalam ukuran variabilitas adalah *Range*, *Kuartil*, *Desil*, *Persentil*, dan *Standar Deviasi*. Masing-masing ukuran variabilitas dan tendensi sentral tersebut akan dijelaskan di bawah ini.

BEBERAPA UKURAN TENDENSI SENTRAL UNTUK MENENTUKAN STATUS/ KEDUDUKAN SISWA

Dalam menentukan kedudukan skor hasil belajar siswa dalam kelas atau kelompoknya digunakan ukuran tendensi sentral. Pertama yaitu menggunakan teknik ranking. Dengan meranking skor-skor siswa akan dapat diketahui skor yang dapat dianggap sebagai tendensi sentralnya. Misalnya diketahui sejumlah siswa yang masing-masing memiliki skor hasil belajar sebagai berikut:

Kode

Nama siswa : - A B C D E F G H I J K L M N

Skor : 88 87 83 86 87 91 88 100 95 93 94 91 91 94

Bila skor-skor siswa tersebut diranking maka urutannya menjadi berikut ini:

UNIVERSITAS MEDAN AREA

Kode																
Nama siswa	:	C	D	B	E	A	G	F	L	M	J	N	K	I	H	
Skor	:	83	86	87	87	88	88	91	91	91	93	94	94	95	100	
								8	-7	-6		4	3			
								↓	↓	↓		↓	↓			
Ranking	:	14	13	13,5	11,5	9,5	9,5	7	7	7	5	3,5	3,5	2	1	

Bila diperhatikan subjek N dan K memiliki skor yang sama yaitu 94, sebenarnya N dan K tersebut memiliki ranking 3 dan 4 atau sebaliknya. Karena skor mereka sama maka cara merankingnya dengan menjumlahkan ranking kedua skor tersebut kemudian dibagi dua ($\frac{3+4}{2} = 3,5$). Begitu pula untuk siswa F, L, dan M yang memiliki skor sama "91", sehingga ranking masing-masing adalah $\frac{6+7+8}{3} = 7$

Dari ke-14 siswa tersebut bahwa ranking yang ada di tengah atau yang membagi separo bagian siswa berada di ranking bawah dan separonya lagi berada di ranking atas, adalah dua skor yaitu 91 dan 91 (siswa F dan L). Skor yang ada di tengah inilah yang merupakan skor tendensi sentral. Berdasarkan hasil ranking ini pula dapat diketahui kedudukan ranking masing-masing siswa terhadap penguasaan materi yang ditekankan.

Ukuran tendensi sentral yang kedua adalah *Mean* atau *Nilai Rata-rata*. Mean sebagai ukuran tendensi sentral lebih banyak digunakan dalam penilaian hasil belajar. Alasannya, karena Mean lebih memberikan gambaran yang cermat dari suatu distribusi skor.

Besarnya Mean dari suatu distribusi skor diperoleh dengan menjumlah seluruh skor kemudian membaginya dengan jumlah subjeknya.

Bila data di atas dihitung Mean-nya hasilnya adalah sebagai berikut:

$$\text{Mean} = \frac{88 + 87 + 83 + 86 + 87 + 91 + 88 + 100 + 95 + 93 + 94 + 91 + 91 + 94}{14}$$

$$= 90,571428 \text{ dibulatkan menjadi } 90,57$$

Bila angka Mean digunakan untuk menentukan status siswa, maka berdasarkan hasil perhitungan di atas dapat dikatakan bahwa ada 6 siswa yang termasuk klasifikasi gagal dan 8 siswa yang termasuk klasifikasi berhasil.

Dalam menghitung besarnya Mean untuk jumlah subjek (N) yang relatif sedikit seperti di atas dapat dilakukan seperti di atas. Tetapi jika N-nya lebih dari 50 cara di atas jelas tidak efektif, apalagi tanpa dibantu dengan mesin hitung. Untuk menghitung Mean yang N-nya lebih dari 50 dapat dilakukan dengan cara membuat tabel distribusi kelompok. Dengan menggunakan tabel distribusi kelompok ini besarnya Mean diperoleh dengan menggunakan rumus *angka kasar* atau rumus *Mean Terkaan*.

Agar lebih jelas di bawah ini dicontohkan cara menghitung besarnya Mean dengan kedua rumus di atas. Misalnya saja data berikut ini adalah skor hasil tes Bahasa Indonesia, Kelas V SD.

87	83	84	85	89	98	95	97	93	97	94
63	64	66	65	67	68	69	86	83	82	80
77	72	79	66	84	89	77	75	90	100	81
82	82	77	74	88	90	92	75	77	79	91
92	93	87	86	75	74	60	72	95	96	98

a) Menghitung Mean dengan Rumus Angka Kasar.

Tabel 1: Distribusi Frekuensi Skor Tes Bahasa Indonesia
Kelas V SD

Kelas Interval	f	X	fX
100 - 104	1	102	102
95 - 99	7	97	679
90 - 94	8	92	736
85 - 89	8	87	696
80 - 84	9	82	738
75 - 79	9	77	693
70 - 74	4	72	281
65 - 69	6	67	402
60 - 64	3	62	186
N = 55		$\sum fX = 4513$	

Rumus Mean angka kasar adalah:

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

di mana f = frekuensi

X = titik tengah masing-masing kelas interval

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

$$= \frac{4513}{55} = 82,54545 \text{ dibulatkan menjadi } 82,05$$

b) Menghitung Mean dengan Rumus Terkaan

Tabel 2: Distribusi Frekuensi Skor Tes Bahasa Indonesia
Kelas V SD

Kelas Interval	f	x'	fx'
100 - 104	1	4	4
95 - 99	7	3	21
90 - 94	8	2	16
85 - 89	8	1	8
80 - 84	9	0	0
75 - 79	9	-1	-9
70 - 74	4	-2	-8
65 - 69	6	-3	-18
60 - 64	3	-4	-12
N = 55		$\sum fx' = 2$	

Penentuan MT atau angka 0 pada kolom x' adalah pada kelas interval sembarang. Biasanya penentuan tersebut dipilih pada kelas interval yang berada di tengah, seperti contoh di atas dipilih pada kelas interval 80-84. Rumus Mean Terkaan adalah:

$$\text{Mean} = \text{MT} + \left[\frac{\sum fX'}{N} \right] i$$

MT = Mean terkaan
 i = interval kelas
 x' = penyimpangan skor dari Mean terkaan

$$\begin{aligned} \text{Mean} &= \text{MT} + \left[\frac{\sum fX'}{N} \right] i \\ &= 82 + \left[\frac{2}{55} \right] 5 = 82,1818 \text{ dibulatkan menjadi } 82,18 \end{aligned}$$

Berdasarkan hasil perhitungan Mean, baik dengan rumus angka kasar maupun dengan rumus Mean Terkaan hasilnya relatif sama.

Kesimpulan yang didapat dari hasil perhitungan tersebut adalah bahwa skor siswa yang berada di atas Mean (82) termasuk kelompok berhasil, dan di bawah skor 82 termasuk kelompok gagal. Karena skor Mean merupakan ukuran tendensi sentral yang memiliki stabilitas yang tinggi di antara ukuran tendensi sentral yang lain, maka skor Mean tersebut sering digunakan untuk dasar perhitungan statistik yang lain.

Tendensi sentral yang berikutnya adalah *Median*. Median adalah skor yang berada di tengah, yaitu yang membatasi 50% skor di atas dan 50% skor di bawah. Pada data yang tidak dikelompokkan, Median adalah skor yang berada di tengah suatu distribusi. Bila distribusi N adalah bilangan genap, maka jelas skor yang berada di tengah tidak dijumpai. Dalam hal ini kita mengambil dua skor yang berdekatan di tengah-tengah suatu distribusi. Misalkan 2 skor yang berada di tengah tersebut adalah skor 88 dan 89, maka mediannya adalah $(88 + 89) : 2 = 88,5$.

Untuk mencari Median terhadap data kelompok, digunakan rumus berikut ini:

$$\text{Median} = Bb_{ny} + \frac{\left[\frac{1}{2} N - Kfb \right]}{fd} \times i$$

Keterangan:

Bb_{ny} = Batas bawah nyata kelas interval yang mengandung Median

fd = Frekuensi pada kelas interval yang mengandung Median

Kfb = Kumulatif frekuensi di bawah kelas interval yang mengandung Median

i = Interval

N = jumlah subjek.

Berikut ini dicontohkan cara menghitung Median dari data kelompok.

UNIVERSITAS MEDAN AREA

Tabel 3: Distribusi Frekuensi Skor Tes Pelajaran IPS
Siswa-Kelas VI SD

Kelas interval	f	Kf
55 - 59	1	50
50 - 54	1	49
45 - 49	3	48
40 - 44	4	45
35 - 39	6	41
30 - 34	7	35
25 - 29	12 — fd	28 — Letak Median
20 - 24	6	16 — Kfb
15 - 19	8	10
10 - 14	2	2
	N = 50	

Berdasarkan tabel di atas maka besarnya Median adalah sebagai berikut:

$$\begin{aligned}
 \text{Median} &= Bb_{ny} + \frac{\left[\frac{1}{2} N - Kfb \right]}{fd} \times i \\
 &= 24,5 + \frac{(25 - 16)}{12} \times 5 \\
 &= 24,5 + 3,75 = 28,25
 \end{aligned}$$

Dari hasil perhitungan Median tersebut dapat disimpulkan bahwa skor-skor yang berada di atas Median (28,25) termasuk siswa yang berhasil, sedangkan yang berada di bawahnya termasuk skor-skor siswa yang gagal.

Tendensi sentral yang terakhir adalah *Mode*. Mode adalah skor yang memiliki frekuensi yang tertinggi. Untuk data berkelompok atau tunggal kita bisa langsung menetapkan skor berapa yang sebagai Mode-nya. Caranya dengan melihat frekuensi yang terbanyak, yang dimiliki oleh suatu skor atau kelas interval tertentu, maka skor tersebutlah sebagai Mode-nya. Untuk data berkelompok skor yang menjadi Mode-nya adalah titik tengah dari kelas interval tertentu. Sebagai misal Tabel 3 yang menjadi Mode-nya adalah 27 (titik tengah kelas interval 25 - 29), sebab interval kelas ini memiliki frekuensi tertinggi yaitu 12.

Cara mencari besarnya Mode di atas adalah cara singkat dan hasilnya relatif kasar. Untuk menghitung Mode yang sebenarnya digunakan rumus:

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean} \quad \text{atau}$$

$$\text{Mode} = Bbs + \left(\frac{fa}{fa + fb} \right) i$$

Keterangan :

Bbs = Batas bawah semua kelas interval yang mengandung Mode atau kelas interval yang memiliki frekuensi terbanyak.

fa = Frekuensi yang berada di atas frekuensi yang mengandung Mode.

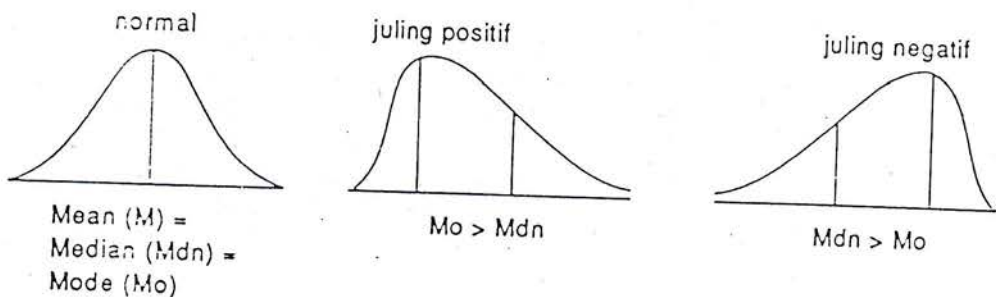
fb = Frekuensi yang berada di bawah frekuensi yang mengandung Mode.

Sebagai contoh dapat mempergunakan data Tabel 3. Dari Tabel 3 tersebut dapat kita lihat bahwa frekuensi tertinggi adalah 12, yang berada pada kelas interval 25 - 29. Dengan rumus di atas dapat dicari besarnya Mode sebagai berikut:

$$\begin{aligned} \text{Mode} &= \text{Bbs} + \left(\frac{f_a}{f_a + f_b} \right) 5 \\ &= 25 + \left(\frac{7}{7 + 6} \right) 5 \\ &= 27,69 \end{aligned}$$

Bila kita ingin menggunakan rumus $\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$ terlebih dahulu kita harus menghitung besarnya Mean.

Dengan diketahuinya Mean, Median, dan Mode suatu distribusi skor tes hasil belajar, dapat diketahui apakah distribusi tersebut *normal* atau *tidak*. Suatu distribusi dikatakan normal jika kedudukan Mean, Median, dan Mode berada pada satu titik/garis. Distribusi normal ini mengandung makna bahwa kemampuan kelompok terhadap suatu materi tes adalah homogen. Jika kedudukan *Mode* lebih dari *Median*, distribusi skor tes akan menunjukkan bentuk kurva yang juling positif. Makna dari bentuk kurva juling positif adalah bahwa sebagian besar siswa mempunyai skor atau kemampuan hasil belajar yang rendah. Sebaliknya, jika *Mode* lebih besar dari *Median* distribusinya akan berbentuk kurva juling negatif. Distribusi kurva juling negatif ini berarti bahwa sebagian besar siswa yang dites mendapatkan nilai yang baik. Interpretasi selanjutnya dapat diasumsikan bahwa materi uji mungkin terlalu mudah, sehingga sebagian besar mereka dapat mengerjakan dengan benar, begitu pula sebaliknya.



Gambar 3 Berbagai Distribusi Skor

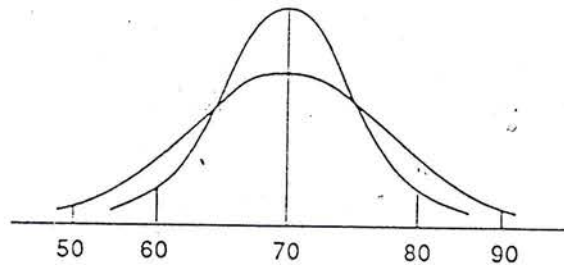
Dengan mempelajari kedudukan siswa atas dasar nilai tendensi sentral, kita dapat dengan cepat mengetahui kondisi pencapaian tujuan instruksional, meskipun kesimpulan tersebut masih bersifat kasar.

BEBERAPA UKURAN VARIABILITAS YANG BERKAITAN DENGAN PENILAIAN HASIL BELAJAR

Dengan mengetahui ukuran-ukuran tendensi sentral suatu distribusi skor tes kita dapat mengetahui atau membuat beberapa kesimpulan umum tentangnya. Namun kesimpulan tersebut belumlah cukup untuk

dimengerti, bila kita ingin mengetahui kondisi individu dalam keseluruhannya. Untuk mengatasi masalah ini kita harus tahu bagaimana pencaran (variabilitas) distribusi hasil tes itu.

Sebagai contoh kita mempunyai dua kelas paralel yaitu kelas V^A dan V^B . Kedua kelompok siswa tersebut setelah diberikan tes yang sama ternyata Mean keduanya sama yaitu 70. Tentu saja kesimpulan umum yang dapat ditarik dari Mean yang sama tersebut adalah bahwa prestasi belajar siswa V^A dan V^B adalah sama. Akan tetapi bila diberi informasi tambahan bahwa pencaran skor pada siswa V^A adalah 50 - 90, dan siswa kelas V^B adalah 60 - 80, maka kita segera menyadari bahwa kedua kelas tersebut memiliki kemampuan yang berbeda. Hal ini akan terlihat jelas pada gambar kurva di bawah ini.



Gambar 4 Kurva Perbandingan Distribusi Skor

- Mean Skor kelas V^A = Mean skor kelas V^B
- Pencaran skor kelas V^A lebih luas dari pencaran skor kelas V^B
- Jadi, meskipun besarnya Mean kedua kelas adalah sama, karena pencarannya berbeda maka kondisi kelas tersebut berbeda.

Ukuran variabilitas dalam teknik statistik meliputi: *Range*, *Kuartil*, *Desil*, *Persentil*, dan *Deviasi*. Range merupakan pencaran atau variabilitas yang paling kasar yaitu berupa jarak antara skor yang tertinggi dengan yang terendah dari suatu distribusi. Kuartil, Desil, dan Persentil masing-masing suatu distribusi skor yang membagi skor menjadi 4 bagian, 10 bagian, dan 100 bagian. Secara lengkap pencaran skor-skor standar dapat dilihat pada Tabel 4.

Berikut ini hendak dijelaskan beberapa ukuran variabilitas atau pencaran skor dari suatu distribusi, khususnya yang berkaitan dengan penilaian hasil belajar.

(1) *Kuartil*

Kuartil adalah titik-titik yang membagi suatu distribusi data menjadi empat bagian sama besar. Ada tiga titik kuartil yaitu Kuartil 1 (K_1), K_2 , dan K_3 . Pembagian distribusi menurut klasifikasi kuartil dapat dilihat pada Tabel 5 berikut ini.

Tabel 5 Pembagian Distribusi Skor Dengan Kuartil

Skor	Frekwensi	Kedudukan K	Persentase
-----	-----	K_3 } K_2 } K_1 }	25%
-----	-----		25%
-----	-----		25%
-----	-----		25%

Cara menghitung besarnya angka Kuartil sama dengan bilamana Anda menghitung Median, bedanya hanya pada unsur N-nya saja. Dalam Median hanya ada satu N yang merupakan titik yang membagi 50% skor di atas dan 50% lainnya di bawah.

Kuartil memiliki tiga titik yaitu titik K_1 yaitu 25% dari N; $K_2 = 50%$ dari N = Median; dan $K_3 = 75%$ dari N.

Rumus untuk menghitung besarnya kuartil adalah

$$K_n (K_1, K_2 \text{ atau } K_3) = Bb_{ny} + \frac{\left[\frac{n}{4} N - Kfb \right]}{fi} \times i$$

Keterangan:

K_n = Kuartil ke n

N = Jumlah subjek/frekuensi

Kfb = Frekuensi kumulatif yang berada di bawah kelas interval yang ditempati kuartil

fi = Frekuensi di mana kelas interval kuartil berada.

Untuk memperjelas uraian di atas, di bawah ini akan dikemukakan contoh perhitungan nilai kuartil.

Tabel 6: Distribusi Frekuensi Skor Siswa Kelas VI SD dalam Tes Bahasa Indonesia

Skor	f	Kf
95	2	80
90	4	78
85	5	74
80	7	69
75	10	62
70	12	52
65	15	40
60	9	25
55	6	16
50	7	10
45	2	3
40	1	1
N = 80		

Dengan menggunakan rumus yang ada besarnya K_1 , K_2 , dan K_3 dapat dicari sebagai berikut.

$$K_1 = Bb_{ny} + \frac{\frac{1}{4} N - Kfb}{fi} \times i$$

$$= 59,5 + \frac{20 - 16}{9} \times 1 = 59,94$$

$$K_2 = Bb_{ny} + \frac{\frac{2}{4} N - Kfb}{fi} \times i$$

$$= 64,5 + \frac{40 - 25}{15} \times 1 = 65,5$$

$$\begin{aligned}
 K_3 &= Bb_{ny} + \frac{\frac{3}{4} N - Kfb}{fi} \times i \\
 &= 74,5 + \frac{60 - 52}{10} \times 1 \\
 &= 75,3
 \end{aligned}$$

Untuk data kelompok Anda hanya mengubah bilangan i (interval kelas) sesuai dengan interval kelas yang ada pada data kelompok yang bersangkutan.

Atas dasar hasil perhitungan Kuartil 1 (K_1), K_2 , dan K_3 maka dapat dibuat kesimpulan sebagai berikut:

- Siswa yang memiliki skor di bawah skor K_1 ($59,94 \leq 60$) perlu mendapat bimbingan belajar atau mengulang materi yang diujikan.
- Siswa yang memiliki skor di atas skor K_3 ($75,3 \leq 75,0$) perlu mendapat program pengayaan atau diperbolehkan melanjutkan pada materi selanjutnya.
- Siswa yang memiliki skor di atas K_1 dan di bawah K_3 adalah mereka yang berprestasi cukup/ sedang.

(2) Desil dan Persentil

Pengertian-pengertian yang telah Anda ketahui tentang Median dan Kuartil dapat dijadikan dasar untuk mempelajari Desil maupun Persentil. Bila Median hanya ada satu titik dalam suatu distribusi, Kuartil terdiri atas tiga titik, maka Desil dan persentil masing-masing memiliki 9 dan 99 titik. Atau dapat dikatakan bahwa Desil membagi distribusi data menjadi 10 bagian sama besar, sedangkan persentil membagi distribusi data menjadi 100 bagian sama besar.

Rumus untuk menghitung nilai Desil maupun Persentil tertentu pada prinsipnya tidak berbeda dengan rumus yang digunakan dalam mencari Median atau Kuartil. Rumus untuk menghitung Desil (D) dan Persentil (P) dapat dituliskan sebagai berikut:

Rumus Desil:

$$D_n = Bb_{ny} + \left[\frac{\frac{n}{10} N - Kfb}{fi} \right] \times i$$

Rumus Persentil:

$$P_n = Bb_{ny} + \frac{\frac{n}{100} N - Kfb}{fi} \times i$$

Berikut ini dicontohkan cara menghitung Desil dan Persentil dari suatu distribusi skor tes hasil belajar

Tabel 7: Distribusi Frekuensi Skor Tes Hasil Belajar
Bahasa Indonesia

Kelas Interval	f	Kf
100 - 104	1	55
95 - 99	7	54
90 - 94	8	47
85 - 89	8	39
80 - 84	9	31
75 - 79	9	22
70 - 74	4	13
65 - 69	6	9
60 - 64	3	3
	N = 55	

Sebagai misal kita ingin mencari kedudukan Desil (D_3 , D_7), serta Persentil (P_{35} dan P_{80}) dari distribusi tabel 7 di atas.

$$\begin{aligned} \text{a. } D_3 &= 74,5 + \frac{\frac{3}{10} \cdot 55 - 13}{9} \times 5 \\ &= 74,5 + 1,94 \\ &= 76,444 \end{aligned}$$

$$\begin{aligned} \text{b. } D_7 &= 84,5 + \frac{\frac{7}{10} \cdot 55 - 31}{8} \times 5 \\ &= 84,5 + 4,6875 \\ &= 89,1875 \end{aligned}$$

$$\begin{aligned} \text{c. } P_{35} &= 74,5 + \frac{\left(\frac{35}{100} \times 55\right) - 13}{9} \times 5 \\ &= 74,5 + 3,472 \\ &= 77,972 \end{aligned}$$

$$\begin{aligned} \text{d. } P_{80} &= 89,5 + \frac{\left(\frac{80}{100} \times 55\right) - 39}{8} \times 5 \\ &= 89,5 + 3,125 \\ &= 92,626 \end{aligned}$$

Cara menghitung Desil dan Persentil lainnya dapat dilakukan dengan cara yang sama seperti di atas.

DEVIASI ATAU PENYIMPANGAN

Deviasi adalah suatu nilai yang menyimpang dari Mean. Terdapat dua macam deviasi yaitu deviasi rata-rata/Mean Deviasi (*Average Deviation*) dan deviasi standar (*Standard Deviation*).

Deviasi rata-rata (AD) adalah nilai yang menunjukkan rata-rata penyimpangan baik penyimpangan yang positif maupun yang negatif. Sebab hakikat dari nilai deviasi, baik yang positif maupun yang negatif merupakan nilai penyimpangan dari Mean.

Cara menghitung nilai AD ialah dengan menjumlahkan keseluruhan penyimpangan skor dari Mean kemudian dibagi dengan jumlah subjek (N). Cara menghitung besarnya nilai AD menggunakan rumus sebagai berikut:

Sebagai contoh cara menghitung AD dari data yang disajikan berikut ini.

Tabel 8: Distribusi Frekuensi Skor Tes Hasil Belajar

Skor (X)	f	fX	x	fx
9	3	27	2,2	6,6
8	5	40	1,2	6,0
7	22	154	0,2	4,4
6	4	24	-0,8	-3,2
5	3	15	-1,8	-5,4
4	3	12	-2,8	-8,4
	N = 40	∑ fX = 272		∑ fx = 34

Untuk menghitung besarnya AD ditempuh langkah-langkah sebagai berikut:

- Menghitung besarnya Mean dari Tabel 8 di atas, besarnya Mean adalah:

$$\begin{aligned} \text{Mean} &= \frac{\sum fX}{N} \\ &= \frac{272}{40} = 6,8 \end{aligned}$$

- Menghitung besarnya penyimpangan (x) masing-masing skor dari Mean. Caranya adalah setiap skor (X) dikurangi Mean. Sebagai contoh besarnya penyimpangan untuk skor 9 adalah 2,2 (9-6,8).
- Mengalikan masing-masing frekuensi (f) dengan penyimpangannya (x). Misalnya untuk skor 9 frekuensinya 3 dan penyimpangannya 2,2. Maka nilai fx adalah 3 x 2,2 = 6,6.
- Menjumlahkan keseluruhan nilai fx tanpa memperhatikan tanda negatif. ∑ fx untuk data pada tabel 8 di atas adalah 34.

Memasukkan data dari tabel 8 ke dalam rumus AD.

$$AD = \frac{\left(\sum fX\right)}{N}$$

$$= \frac{34}{40} = 0,85$$

Karena nilai AD tidak mempertimbangkan nilai negatif, maka kurang dapat memberikan informasi yang bermakna. Untuk itu dalam mengatasi hal ini digunakan nilai penyimpangan yang lain yaitu Standar Deviasi (SD).

Standar Deviasi (SD) merupakan ukuran variabilitas yang banyak bermanfaat bagi dasar perhitungan statistik yang lain. Dalam kaitannya dengan penilaian hasil belajar dengan mengetahui besarnya SD kita akan mengetahui heterogenitas atau homogenitas prestasi siswa dalam suatu tes. Di samping itu SD erat kaitannya dengan kurva normal. Daerah kurva normal dalam ilmu-ilmu sosial biasanya dibagi menjadi 6 daerah SD. Atas dasar nilai SD ini prestasi siswa dapat dikelompokkan. Contoh klasifikasi yang menggunakan SD bersama-sama dengan Mean dikemukakan oleh Rummers Gage dan kawan-kawan (1959, 274 - 275) sebagai berikut:

Grade A = Mean + 1,5 SD ke atas

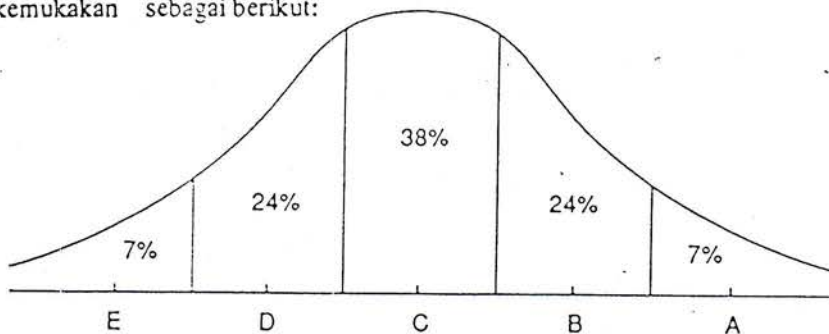
Grade B = Mean + 0,5 SD sampai dengan Mean + 1,49 SD

Grade C = Mean - 0,5 SD sampai dengan Mean + 0,49 SD

Grade D = Mean - 1,5 SD sampai dengan Mean - 0,49 SD

Grade E = Mean - 1,51 SD ke bawah

Pembagian 'Grade' A, B, C, D dan E bila digambarkan dalam Kurva Normal secara persentase dapat dikemukakan sebagai berikut:



Cara menghitung besarnya SD dapat menggunakan berbagai rumus:

(a) Rumus dasar SD:

$$SD = \sqrt{\frac{\sum fx^2}{N}}$$

Sebagai contoh perhitungan SD digunakan data pada tabel 8 di atas.

Tabel.9: Perhitungan Standar Deviasi

X	f	x	x ²	fx ²
9	3	2,2	4,48	14,52
8	5	1,2	1,44	7,2
7	22	0,2	0,04	0,88
6	4	-0,8	0,64	2,56
5	3	-1,8	3,24	9,72
4	3	-2,8	7,84	23,52
N = 40				∑ fx ² = 34

Berdasarkan tabel di atas dengan rumus dasar SD, dapat diketahui besarnya standar deviasi yaitu

$$\begin{aligned}
 SD &= \sqrt{\frac{\sum fx^2}{N}} \\
 &= \sqrt{\frac{58,4}{40}} = 1,2083
 \end{aligned}$$

(b) Untuk menghitung SD data kelompok digunakan rumus:

$$\text{b.1. } SD = \sqrt{\frac{\sum fX^2}{N} - \left[\frac{\sum fX}{N} \right]^2} \quad \text{atau}$$

$$\text{b.2. } SD = i \cdot \sqrt{\frac{\sum fx'^2}{N} - \left[\frac{\sum fx'}{N} \right]^2}$$

Sebagai contoh penggunaan kedua rumus di atas berikut ini disajikan tabel data kelompok.

Tabel 10: Perhitungan Standar Deviasi Data Kelompok

Kelas Interval	f	X ['])	f(X)	fX ²
109 - 117	3	113	339	38307
100 - 108	3	104	312	32448
91 - 99	3	95	285	27075
82 - 90	4	86	344	29584
73 - 81	5	77	385	29645
64 - 72	10	68	680	46240
55 - 63	9	59	531	31329
46 - 54	5	50	250	12500
37 - 45	6	41	246	10086
Σ	48		3372	257214

Keterangan:

*) X = titik tengah suatu kelas interval.

Dengan menggunakan rumus (b.1) besarnya SD adalah:

$$\begin{aligned}
 SD &= \sqrt{\frac{\sum fX^2}{N} - \left[\frac{\sum fX}{N} \right]^2} \\
 &= \sqrt{\frac{257214}{48} - \left[\frac{3372}{48} \right]^2} = \sqrt{5358,6250 - 4950,0625} \\
 &= \sqrt{423,5625} = 20,58
 \end{aligned}$$

Bila kita tidak ingin menggunakan bilangan/angka besar seperti di atas, kita dapat menggunakan rumus SD (b.2). Rumus (b.2) ini dapat dipilih atau digunakan jika tersedia mesin hitung. Untuk dapat menggunakan rumus (b.2) tabel di atas harus dimodifikasi sebagai berikut.

Tabel 11: Perhitungan Standar Deviasi Data Kelompok

Kelas Interval	f	x'	fx'	fx'. ²
109 - 117	3	5	15	75
100 - 108	3	4	12	48
91 - 99	3	3	9	27
82 - 90	4	2	8	16
73 - 81	5	1	5	5
64 - 72	10	0	0	0
55 - 63	9	-1	-9	9
46 - 54	5	-2	-10	20
37 - 45	6	-3	-18	54
Σ	48		12	254

*j) X = titik tengah suatu kelas interval.

Dengan menggunakan rumus (b.2) besarnya SD adalah:

$$\begin{aligned}
 SD &= i \sqrt{\frac{\sum fx'.^2}{N} - \left[\frac{\sum fx'}{N} \right]^2} \\
 &= 9 \sqrt{\frac{254}{48} - \left[\frac{12}{48} \right]^2} = 9 \sqrt{5,291666 - 0,0625} \\
 &= 9 \sqrt{5,229166} = 20,58
 \end{aligned}$$

Berdasarkan hasil perhitungan di atas baik dengan menggunakan rumus (b.1) maupun (b.2) hasil yang diperoleh relatif sama yaitu 20,58.

b. Transformasi Skor Tes ke dalam Beberapa Skala Penilaian

Mentransformasikan skor ke dalam nilai standar dapat dilakukan dengan pendekatan PAP ataupun PAN. Bila kita dalam mentransformasikan skor ke dalam standar nilai menggunakan PAP berarti mengklasifikasi tingkat penguasaan terhadap materi yang diteskan. Misalkan suatu unit pelajaran telah disusun alat ukurnya dengan tes objektif. Tes objektif ini terdiri atas 100 butir soal tes yang secara representatif mengukur TIK unit pelajaran tersebut.

Selanjutnya dengan pembagian secara persentase keadaan suatu skala tertentu nilai siswa dapat ditempatkan. Misalkan dalam penilaian skala - 5 dengan PAP dapat diklasifikasikan sebagai berikut:

Tabel 12: Transformasi Skala - 5 dengan PAP

Persentase Tingkat penguasaan	Nilai	Makna Nilai
81 - 100%	A 4	Baik sekali
61 - 80%	B 3	Baik
41 - 60%	C 2	Cukup/sedang
21 - 40%	D 1	Kurang
0 - 20%	E 0	Kurang sekali

1) Z-Skor

Untuk mentransformasikan skor ke dalam skala yang lain dengan pendekatan PAP dapat dilakukan dengan klasifikasi/pembagian dalam kurva normal (Lihat Pencaran Kurva Normal oleh Stanley, 1964). Tabel 13: 1). Transformasi skor ke dalam skala nilai standar biasanya menggunakan pendekatan PAN. Dalam hal ini sebagai pedoman adalah nilai *Mean* dan *Standar Deviasi*, dan nilai yang didapat adalah nilai Z atau Z Skor. Rumus untuk mencari Z-Skor adalah:

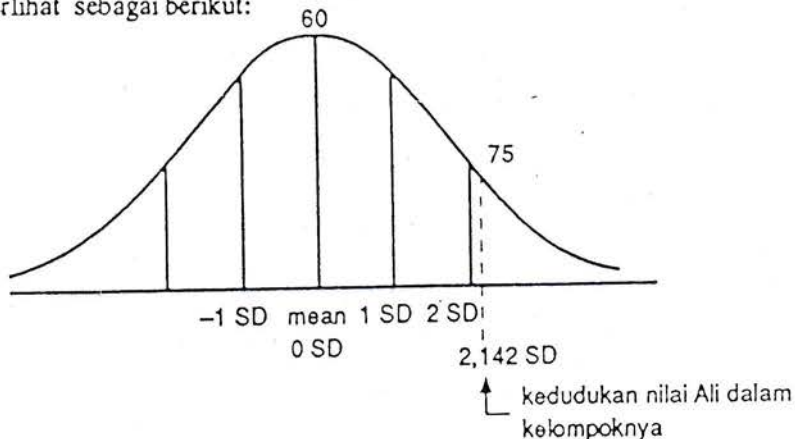
$$Z - \text{Skor} = \frac{X - \text{Mean}}{SD}$$

Sebagai Contoh: Ali mendapat Skor (X) 75 dalam tes PMP ; Mean dari hasil tes tersebut adalah 60; dan SD-nya adalah 7. Maka besarnya Z-Skor 75 adalah:

$$Z - \text{Skor}_{75} = \frac{75 - 60}{7} = 2,142$$

(dibulatkan menjadi 2,143)

Berarti skor 75 yang dimiliki Ali berkedudukan pada 2,143 SD dari Mean. Bila digambarkan dalam kurva normal maka terlihat sebagai berikut:



Gambar 5
Kedudukan Nilai dalam Kurva Normal

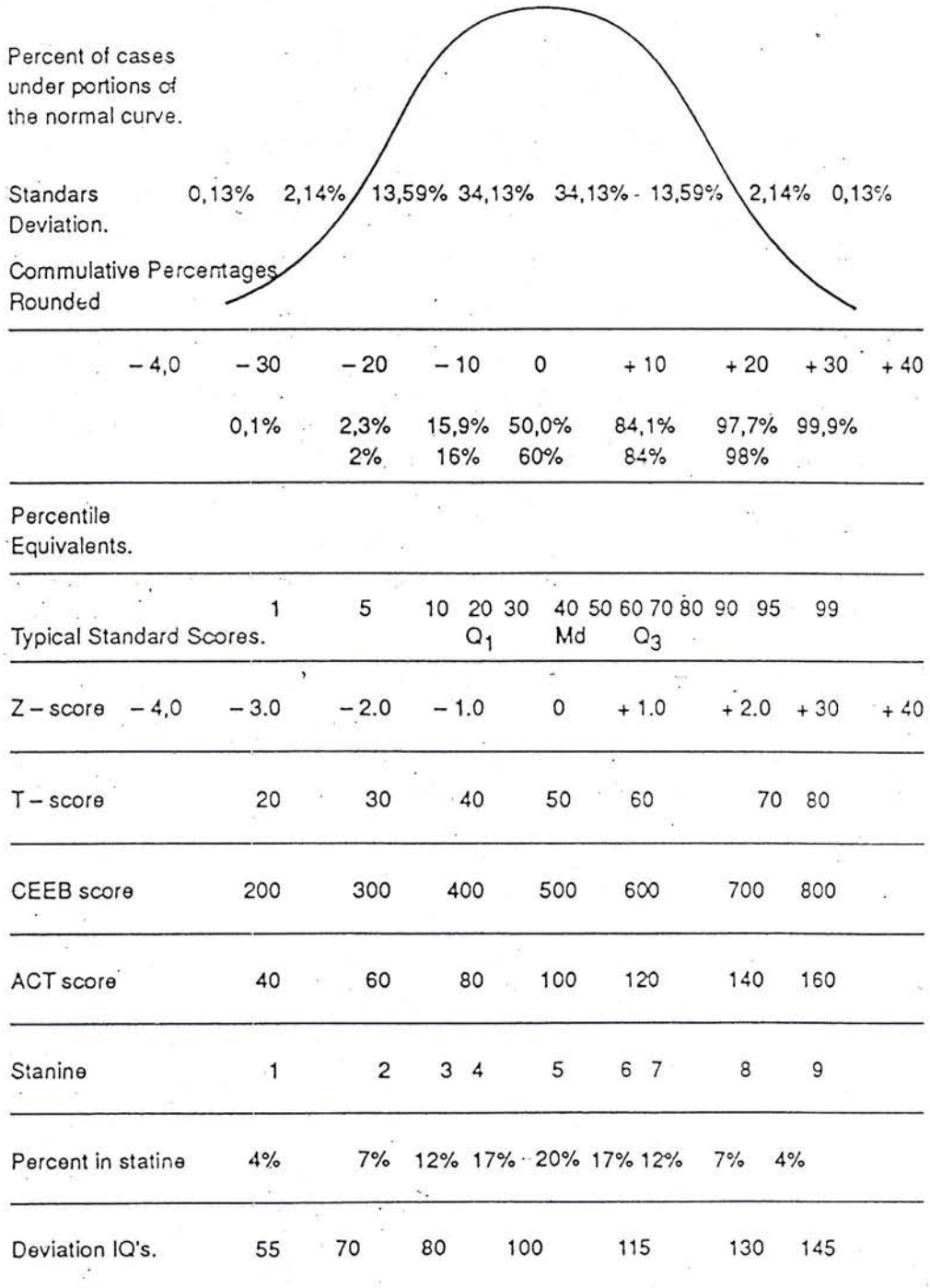
Tabel 13

PENCARAN SKOR-SKOR STANDARD
DALAM KERANGKA KURVA NORMAL *)

Percent of cases
under portions of
the normal curve.

Standars
Deviation.

Commulative Percentages
Rounded



*) dari Stanley, 1964 : halaman 74

Dengan mengetahui nilai Z dari keseluruhan skor siswa kita dapat menentukan kedudukan masing-masing siswa. Selanjutnya nilai Z tersebut dapat digunakan untuk menentukan nilai yang diperoleh masing-masing siswa. Hal ini dapat dilakukan dengan memasukkan nilai Z ke dalam klasifikasi skala nilai yang diminta.

Berikut ini beberapa Skala Nilai Standar dengan pendekatan PAN.

a) Tabel 14: Skala - 5 (Sta - five)

Nilai	Z - Skor
A/4	2 SD ————— ke atas
B/3	1 SD ————— 2 SD
C/2	- 1 SD ————— 1 SD
D/1	- 2 SD ————— - 1 SD
E/0	di bawah ————— - 2 SD

b) Tabel 15: Skala - 9 (Sta - nine)

Nilai	Z - Skor
9	1,75 SD ————— ke atas
8	1,25 SD ————— 1,75 SD
7	0,75 SD ————— 1,25 SD
6	0,25 SD ————— 0,75 SD
5	- 0,25 SD ————— 0,25 SD
4	- 0,75 SD ————— - 0,25 SD
3	- 1,25 SD ————— - 0,75 SD
2	- 1,75 SD ————— - 1,25 SD
1	di bawah ————— - 1,75 SD

c) Tabel 16: Skala 10 (Sta - Ten/C - Skor)

Nilai	Z - Skor
10	2,0 SD ————— ke atas
9	1,5 SD ————— 2,0 SD
8	1,0 SD ————— 1,5 SD
7	0,5 SD ————— 1,0 SD
6	0 SD (Mean) ————— 0,5 SD
5	- 0,5 SD ————— 0 SD
4	- 1,0 SD ————— - 0,5 SD
3	- 1,5 SD ————— - 1,0 SD
2	- 2,0 SD ————— - 1,5 SD
1	di bawah ————— - 2,0 SD

d) Tabel 17: Skala - 11 (Sta - Eleven) *)

Nilai	Z - Skor
10	2,25 SD ————— ke atas
9	1,75 SD ————— 2,25 SD
8	1,25 SD ————— 1,75 SD
7	0,75 SD ————— 1,25 SD
6	0,25 SD ————— 0,75 SD
5	- 0,25 SD ————— 0,25 SD
4	- 0,75 SD ————— - 0,25 SD
3	- 1,25 SD ————— - 0,75 SD
2	- 1,75 SD ————— - 1,25 SD
1	- 2,25 SD ————— - 1,75 SD
0	di bawah ————— - 2,25 SD

*) Skala - 11 sebenarnya sama dengan Skala - 9, perbedaannya pada skala - 11 terdapat nilai 0 dan 10, sedangkan Skala - 9 tidak.

Dalam penggunaan tabel skala nilai standar di atas dapat diberi contoh sebagai berikut. Luthfi dalam tes bidang pelajaran IPA mendapat skor 85, sedangkan mean/rata-rata skor siswa di kelas Luthfi dalam tes tersebut adalah 70 dengan standar deviasi 13. Bila skor Luthfi ditransformasi ke dalam Skala - 5, Skala - 9/11 dan Skala - 10 hasilnya adalah:

$$Z_{85} = \frac{85 - 70}{13} = 1,15 \text{ SD}$$

Nilai Z yang dimiliki Luthfi adalah 1,15 SD. Untuk mentransformasi ke skala-skala di atas kita lihat kedudukan Z (1,15 SD) tersebut di dalam skala standar.

Pada Skala - 5, 1,15 SD berkedudukan dalam kelas interval 1 SD - 2 SD yang berarti mendapat nilai B/3 atau dapat dikatakan nilai Luthfi dalam tes IPA adalah baik. Pada Skala - 9/11; 1,15 SD berkedudukan dalam kelas interval Z-Skor 0,75 SD - 1,125 SD atau mendapat nilai 7 atau nilai baik. Pada Skala - 10; 1,15 SD berkedudukan dalam kelas interval Z-Skor 1,0 SD - 1,5 SD atau pada nilai 8.

Bila disimpulkan prestasi Luthfi dalam tes IPA baik ditransformasi ke skala - 10 relatif sama yaitu ada pada klasifikasi *baik*.

2) T - Skor

Selain cara mentransformasikan skor ke dalam skala nilai standar di atas, juga masih ada cara lain yang disebut dengan T-Skor. Rasional dari T-Skor ialah mendasarkan pada nilai Mean = 50 dan SD = 10 untuk rentangan skala 1 sampai 100.

Penyebaran skor mentah ke dalam T-skor dimaksudkan selain untuk mengetahui kedudukan prestasi seseorang dalam kelompoknya, juga untuk mengetahui kedudukan prestasi seseorang dalam beberapa bidang studi yang masing-masing memiliki Mean dan SD yang berbeda. Beberapa hasil tes suatu tim panitia ujian sekolah akan dengan mudah menentukan batas lulus dari berbagai mata pelajaran/ bidang studi yang menggunakan nilai yang sama (T-Skor).

Rumus untuk menghitung besarnya T-Skor adalah sebagai berikut:

$$\begin{aligned} T - \text{Skor} &= \left[\frac{X - \text{Mean}}{SD} \right] 10 + 50 \quad \text{atau} \\ &= 10 Z + 50 \end{aligned}$$

Misalnya: Rudi dalam mata pelajaran Bahasa Indonesia mendapat skor 75 dengan mean kelas = 70 dan SD-nya = 5. Dalam mata pelajaran Matematika skornya = 55, dengan mean kelas 45, SD-nya = 4, dan dalam mata pelajaran PMP skornya 65 dengan mean kelas = 60 dan SD-nya 7.

Bila skor-skor Rudi tersebut diubah ke dalam T-Skor hasilnya akan diperoleh:

$$\begin{aligned} T - \text{Skor Bahasa Indonesia} &= \left[\frac{75 - 70}{5} \right] 10 + 50 \\ &= 10 + 50 = 60 \end{aligned}$$

$$\begin{aligned} T - \text{Skor Matematika} &= \left[\frac{55 - 45}{4} \right] 10 + 50 \\ &= 25 + 50 = 75 \end{aligned}$$

$$\begin{aligned} T - \text{Skor PMP} &= \left[\frac{65 - 60}{7} \right] 10 + 50 \\ &= 7,1428 + 50 = 57,1428 \approx 57 \end{aligned}$$

Berdasarkan hasil perhitungan di atas ternyata nilai Rudi dalam bidang studi/mata pelajaran Matematika lebih tinggi dibandingkan nilainya dalam mata pelajaran Bahasa Indonesia dan PMP. Meskipun bila hanya dilihat skor tes ketiga mata pelajaran tersebut, skor tes Matematikanya yang paling rendah.

T-Skor di atas sejak Kurikulum 75 sudah tidak lagi diberlakukan/diterapkan untuk mentransformasi skor hasil tes, melainkan menggunakan Jenjang Persentil (JP).

3) Jenjang Persentil (JP)

Jenjang Persentil dalam literatur statistik sering disebut dengan *Percentile Rank*. JP berbeda dengan jenjang/kedudukan menurut angka. Perbedaan *pertama* adalah jika kedudukan menurut ranking angka, yang terendah berada pada distribusi bagian atas, sedangkan untuk JP angka yang terendah berada pada distribusi yang paling bawah. *Kedua*, JP adalah kedudukan yang perhitungannya didasarkan atas jenjang 100 angka, sedangkan kedudukan ranking angka bergantung pada jumlah angka yang ada pada distribusi. Untuk jelasnya dapat dilihat Tabel 18 berikut.

Tabel 18: Perbedaan Kedudukan Rangkaing Angka dengan Jenjang Persentil

Nilai	Kedudukan Rangkaing Angka	Kedudukan Jenjang Persentil
10	1	100
9	2	90
8	3	80
7	4	70
6	5	60
5	6	50
4	7	40
3	8	30
2	9	20
1	10	10
0	11	0

Rumus untuk menghitung JP adalah:

$$JP = \left[\frac{X - Bb_{ny}}{i} \right] fd + Kfb \times \frac{100}{N}$$

Keterangan:

JP = Jenjang Persentil

X = Skor yang diketahui dari hasil suatu tes

Bb_{ny} = Batas bawah nyata interval yang mengandung X

i = Interval kelas

fd = Frekuensi kelas interval yang mengandung X

Kfb = Kumulatif frekuensi di bawah kelas interval yang mengandung X

N = Jumlah subjek/frekuensi

Sebagai contoh, bila suatu distribusi skor tes setelah dikelompokkan seperti tabel 19 berikut, berapakah kedudukan JP dari skor 69?

Tabel 19: Distribusi Skor Hasil Tes

Kelas Interval	f	Kf
109 - 117	3	48
100 - 108	3	45
91 - 99	3	42
82 - 90	4	39
73 - 81	5	35
64 - 72	10	30
55 - 63	9	20
46 - 54	5	11
37 - 45	6	6
Σ	48	

$$\begin{aligned}
 JP_{69} &= \left[\frac{69 - 63,5}{9} \right] 10 + 20 \times \frac{100}{48} \\
 &= 61,1 + 20 (2,0833) \\
 &= 6,1 + 41,66 \\
 &= 47,766
 \end{aligned}$$

Jadi kedudukan skor 69 berada pada JP = 47,766. Ini berarti 47,766% frekuensi atau siswa yang dites skornya di bawah 42.

Untuk memantapkan pemahaman Anda terhadap pokok bahasan ini coba kerjakan latihan-latihan di bawah ini.

- 1) Apa yang mendasari digunakannya PAN dalam penilaian hasil belajar? Jelaskan!
- 2) Dilihat dari segi tingkat pencapaian TIK pendekatan penilaian mana yang lebih tepat? Jelaskan!.
- 3) Buatlah ranking dari skor hasil tes siswa berikut ini!

74,	81,	71,	72,	93,	74,	70,	74,	83,	93
85,	86,	92,	90,	85,	76,	89,	81,	74,	87
- 4) Hitunglah data pada latihan 3 di atas besarnya:

a) Mean	b) Median	c) Mode
d) K_3	e) D_7	f) P_{33}
- 5) Transformasikan data skor tes di bawah ini ke dalam skala 5, 10 dan 11. Data skor tes tersebut adalah:

74	89	80	93	64	67	72	70	66	85	89	81
81	71	74	82	85	63	72	81	81	95	84	81
80	70	69	66	60	83	85	98	84	68	90	82
69	72	87	88	79	70	81	75	74	78	80	81
- 6) Berdasarkan data latihan 5, hitunglah kedudukan Jenjang Persentil (JP) skor 89? dan transformasikan skor 81 ke dalam T-Skor.

Dalam mengerjakan latihan-latihan di atas silakan mengikuti petunjuk sebagai arahan di bawah ini.

Latihan: 1

Penilaian dengan pendekatan PAN didasarkan pada distribusi populasi gejala-gejala alam sekitar yang menunjukkan bentuk kurva normal. Begitu pula tentang kemampuan siswa dalam satu kelas ada yang menjadi kelompok *pandai/baik, sedang*, dan *kurang* Asumsi tersebutlah yang mendasari digunakannya PAN dalam penilaian hasil belajar.

Latihan: 2

Pendekatan penilaian PAP yang lebih tepat untuk mengetahui tingkat penguasaan atau tingkat pencapaian TIK. Dengan pendekatan PAP secara langsung guru harus menyusun alat evaluasi (TES) yang harus representatif dapat mengukur TIK yang akan dicapai. Alat tes yang representatif tersebut jika dikerjakan dengan betul semua berarti 100% TIK telah dapat dikuasai. Dengan PAP biasanya telah ditetapkan suatu kriteria keberhasilan minimal dalam wujud tingkat persentase tertentu (seperti 60%, 70%, atau 80%). Dengan diketahuinya rata-rata persentase siswa yang dites akan diketahui tingkat pencapaian TIK-nya.

Latihan: 3

Dari data yang diketahui dapat diurut sebagai berikut:

70 71 72 74 74 74 76 81 81 83
85 85 86 87 89 90 92 93 93

Berdasarkan urutan skor-skor tersebut maka ranking skor masing-masing adalah:

Skor (X)	f	Rangking
93	2	1,5
92	1	3
90	1	4
89	1	5
87	1	6
86	1	7
85	2	8,5
83	1	10
81	2	11,5
76	1	13
74	4	15,5
72	1	18
71	1	19
70	1	20

Latihan: 4

Berdasarkan data pada latihan 3, besarnya:

a) Mean = $\frac{\sum fX}{N}$, silakan hitung!

b) Median adalah skor yang membagi 50% frekuensi di atas dan 50% frekuensi di bawah. Hitunglah dengan rumus:

UNIVERSITAS MEDAN AREA

$$\text{Median} = \frac{Bb}{ny} + \frac{\left[\frac{1}{2} N - Kfb \right]}{fd} \times i$$

- c) Mode adalah skor yang paling banyak frekuensinya. Selanjutnya gunakan rumus berikut ini untuk menghitungnya.

$$\text{Mode} = Bbs + \left(\frac{fa}{fa + fb} \right) \times i$$

- d) Kuartil 3 (K_3) adalah kedudukan skor yang mengandung K_3 , yaitu yang berada pada $3/4$ frekuensi keseluruhan dari bawah. Gunakan rumus berikut ini untuk mengetahui besarnya skor K_3 .

$$K_3 = Bb_{ny} + \left(\frac{\frac{3}{4} N - Kfb}{fi} \right) \times i$$

- e) Desil 7 (D_7) yaitu kedudukan skor yang mengandung D_7 , yaitu $0,7$ dari N /frekuensi keseluruhan dari bawah. Gunakan rumus berikut ini untuk mengetahui besarnya skor D_7 .

$$D_7 = Bb_{ny} + \left(\frac{\frac{7}{10} N - Kfb}{fi} \right) \times i$$

- f) Persentil 55 (P_{55}), adalah kedudukan skor yang mengandung P_{55} , yaitu $0,55$ dari N /frekuensi keseluruhan dari bawah. Gunakan rumus berikut ini untuk mengetahui besarnya P_{55} .

$$P_{55} = Bb_{ny} + \left(\frac{\frac{55}{100} N - Kfb}{fi} \right) \times i$$

Latihan: 5

Untuk dapat mentransformasi data skor, Anda harus mencari besarnya Mean dan SD-nya. Atas dasar Mean dan SD inilah Anda bisa mengubah skor-skor yang ada ke dalam Skala - 5, Skala - 10, dan Skala - 11, yaitu dengan memasukkan ke dalam tabel Skala - 5, Skala - 10, dan Skala - 11.

Latihan: 6

Jika Mean dan SD skor pada latihan 5 sudah Anda peroleh, serta Anda telah membuat tabel distribusi frekuensi kelompok maka JP_{89} , maupun T-Skor 81 dapat Anda hitung dengan rumus yang ada dalam modul ini.

RANGKUMAN

1. Dalam penilaian hasil belajar di sekolah ada dua pendekatan yang digunakan yaitu PAN dan PAP. Pendekatan PAN didasarkan pada asumsi gejala populasi yang cenderung berbentuk kurva normal. PAN dalam penilaian hasil belajar berorientasi pada prestasi siswa apa adanya; nilai seorang siswa dibandingkan dengan nilai dalam kelompoknya. Pada PAP penilaian dilakukan dengan dasar tuntutan penguasaan terhadap materi yang diujikan. Tuntutan tersebut mengacu pada kriteria persentase tertentu sebagai tingkat penguasaan yang diharapkan.
2. Dalam mengolah skor hasil tes digunakan analisis statistik. Teknik statistik yang berkaitan dengan penilaian hasil belajar adalah beberapa ukuran tendensi sentral dan variabilitas. Ukuran tendensi sentral yang dimaksud adalah Mean, Median, dan Mode. Sedangkan ukuran variabilitas adalah Range, Kuartil, Desil, Persentil, dan Deviasi. Di antara ukuran tendensi sentral dan variabilitas tersebut Mean dan Standar Deviasilah yang sering digunakan untuk hasil belajar ke dalam skor nilai standar.

3. Skor hasil pengukuran dalam penilaian hasil belajar belum memiliki makna yang berarti. Agar skor hasil pengukuran tersebut dapat bermakna haruslah ditransformasi ke dalam skala nilai standar.

Ada beberapa skala nilai standar, yaitu:

Skala - 5, Skala - 9, Skala - 10, Skala - 11, T - Skor, dan Jenjang Persentil. Di dalam mentransformasikan skor ke dalam Skala 5, 9, 10, dan 11 berdasarkan pada mean kelas dan standar deviasinya. Sedangkan untuk mentransformasi skor ke dalam T-Skor berdasarkan pada pedoman Mean = 50 dan SD = 10 serta skor total 100. Untuk jenjang persentil, penilaian dilakukan dengan berdasarkan kedudukan setiap skor siswa pada urutan atau ranking dari 0 hingga 100.

TES FORMATIF 2

Petunjuk: Silanglah satu option jawaban pada setiap soal di bawah ini yang Anda anggap paling benar.

- 1) Untuk mengetahui prestasi seorang siswa dibandingkan dengan teman sekelasnya, maka dalam penilaian digunakan pedoman:
 - A. PAN
 - B. PAP
 - C. Prestasi siswa sebelumnya
 - D. Prestasi rata-rata kelas.
- 2) Untuk keperluan tes formatif sebaiknya digunakan penilaian dengan pendekatan PAP sebab:
 - A. tes formatif bertujuan untuk mengetahui tingkat penguasaan siswa terhadap unit pelajaran, sehingga PAP yang sesuai
 - B. tes formatif bertujuan untuk program diagnosis, sehingga PAP yang sesuai
 - C. tes formatif bertujuan untuk maksud perbaikan program pengajaran, sehingga perlu penilaian dengan PAP
 - D. tes formatif harus dinilai dengan pedoman PAP, karena dapat diketahui tingkat penguasaan siswa terhadap materi yang diujikan.
- 3) Teknik ranking dalam evaluasi hasil belajar dapat sebagai informasi tentang:
 - A. tingkat kecerdasan siswa
 - B. tingkat penguasaan materi yang diajarkan
 - C. tingkat/kedudukan siswa dalam kelas
 - D. tingkat keberhasilan/kegagalan guru dalam mengajar.
- 4) Skor tendensi sentral *Mean* dapat merupakan informasi seperti berikut, kecuali:
 - A. kemampuan rata-rata siswa dalam kelas
 - B. batas lulus dan tidak lulus
 - C. tingkat penguasaan materi pada suatu kelas
 - D. tingkat penguasaan materi seorang siswa.
- 5) Mode mempunyai kelebihan sebagai alat informasi hasil belajar siswa, karena:
 - A. kecermatan dalam perhitungan
 - B. cepatnya diketahui kemampuan rata-rata siswa di suatu kelas
 - C. segera diketahui 50% siswa yang pandai dan 50% siswa yang kurang pandai
 - D. sangat bermanfaat untuk analisis selanjutnya.

6) Berikut ini adalah skor tes Bahasa dari 25 siswa kelas II SMA:

35 35 37 39 40 40 41 43 49 50 55 56 57
57 58 60 61 63 65 65 66 70 71 72 74

Besarnya Mean dari skor tersebut adalah:

- A. 54,4
- B. 54,5
- C. 55,4
- D. 55,5

7) Kelebihan MEAN sebagai tendensi sentral adalah sebagai berikut, kecuali:

- A. Merupakan informasi yang paling cermat dibandingkan dengan ukuran tendensi sentral lainnya
- B. Cara penghitungannya lebih mudah dibandingkan dengan cara menghitung Mode untuk kelompok siswa yang jumlahnya kurang dari 20 orang
- C. Banyak digunakan dalam analisis statistik pendidikan
- D. Dapat untuk pedoman klasifikasi nilai siswa.

8) Luthfi siswa kelas II SMA dalam THB mendapat skor:

- 90 dengan Mean 69 dan SD 10 untuk bidang studi Matematika
- 100 dengan Mean 90 dan SD 5 untuk bidang studi IPS

Bila kedua skor tersebut dinilai dengan menggunakan pendekatan PAN, maka:

- A. Prestasi belajar Luthfi dalam bidang studi Matematika lebih baik dari pada di IPS di kelasnya
- B. Prestasi belajar Luthfi dalam bidang studi IPS lebih baik dari pada Matematika di kelasnya
- C. Prestasi belajar Luthfi untuk kedua bidang studi tersebut sama
- D. Kedua skor tersebut belum mempunyai arti apa-apa.

9) Anita dalam pelajaran Bahasa Indonesia memperoleh skor 45 dengan mean kelas 25 dan SD 9, sedangkan dalam pelajaran PMP memperoleh skor 75 dengan mean kelas 65 dan SD 7. Maka kesimpulan yang dapat ditarik dengan berdasarkan pada perhitungan T-Skor adalah:

- A. Nilai Anita dalam pelajaran Bahasa Indonesia lebih baik daripada pelajaran PMP
- B. Nilai Anita dalam pelajaran PMP lebih baik daripada pelajaran Bahasa Indonesia
- C. Kedua nilai tersebut belum dapat diartikan
- D. Anita lebih menguasai bahan pelajaran Bahasa Indonesia daripada PMP.

10) Untuk melaporkan hasil belajar siswa lebih baik mempergunakan Skala - 5 karena:

- A. banyak disenangi para guru
- B. cermat hasilnya
- C. jelas maknanya
- D. mudah analisisnya.

11) Kelebihan transformasi skor ke dalam T-Skor dibandingkan dengan cara transformasi yang lain adalah:

- A. mudah analisisnya
- B. cermat hasilnya
- C. dapat membedakan kemampuan antar-siswa
- D. lebih mempunyai arti/makna bagi siswa.

12) Rachman dalam pelajaran PMP mendapat nilai 7 (tujuh) berarti Rachman:

- A. mempunyai prestasi baik dalam pelajaran PMP
- B. berkedudukan pada ranking ketujuh

UNIVERSITAS MEDAN AREA 70 persen

- D. mempunyai prestasi baik dibandingkan dengan teman-teman sekelasnya.

Petunjuk: Kerjakan soal tes esai berikut ini:

13) Hitunglah Mean dan SD dari data skor tes di bawah ini:

Skor (X)	f	Kf
95 - 99	2	63
90 - 94	3	61
85 - 89	3	58
80 - 84	4	55
75 - 79	6	51
70 - 74	9	45
65 - 69	13	36
60 - 64	8	23
55 - 59	7	15
50 - 54	3	8
45 - 49	4	5
40 - 44	1	1

14) Transformasikanlah data skor tes (soal no. 13) di atas ke dalam skala-5.

15) Transformasikanlah skor 64 dan 87 data skor tes (soal no. 13) di atas ke dalam T-Skor.

16) Transformasikanlah skor 59 dan 75 data skor tes (Soal No. 13) di atas ke dalam jenjang persentil.

Setelah Anda mengerjakan soal tes formatif di atas selanjutnya koreksilah dengan kunci jawaban pada bagian akhir modul ini. Hitunglah jawaban Anda yang benar dan gunakan rumus di bawah ini untuk mengetahui tingkat penguasaan Anda.

Rumus:

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah jawaban Anda yang benar}}{16} \times 100\%$$

Arti tingkat penguasaan yang Anda capai:

- 90 - 100% = Baik sekali
- 80 - 89% = Baik
- 70 - 79% = Sedang
- < 70% = Kurang

Jika Anda mencapai tingkat penguasaan 80% ke atas, silakan meneruskan mempelajari materi berikutnya. Bila tingkat penguasaan Anda kurang dari 80%, Anda harus mengulangi materi ini, terutama bagian yang belum Anda kuasai: