

TEKNIK PENGUKURAN DAN PENILAIAN HASIL BELAJAR

Dr. Sumardi, M.Hum

**TEKNIK PENGUKURAN DAN PENILAIAN
HASIL BELAJAR**

Kata Pengantar

Ada paling tidak empat aktivitas yang harus dilakukan oleh guru untuk mengetahui hasil belajar peserta didik, sekaligus untuk mengidentifikasi efektivitas proses pembelajaran dalam rangka mencapai tujuan yang telah ditentukan. Aktivitas itu meliputi menguji (memberikan berbagai pertanyaan yang harus dijawab oleh peserta didik); mengukur (menentukan besaran angka yang merefleksikan seberapa besar kemampuan peserta didik dalam menjawab pertanyaan-pertanyaan yang diberikan oleh guru); menilai (menginterpretasikan angka-angka hasil pengukuran); dan mengevaluasi (memutuskan tingkat keberhasilan belajar peserta didik dan juga keberhasilan proses pembelajaran yang telah dilakukan oleh seorang guru). Karakteristik keempat aktivitas tersebut tentunya berbeda satu sama lain, dimulai dari aktivitas yang paling sederhana menuju aktivitas yang lebih kompleks. Untuk menguji peserta didik, guru memerlukan tes yang baik. Dari tes yang diberikan kepada peserta didik itu, guru kemudian dapat memberikan skor (dalam bentuk angka) berdasarkan jawaban yang diberikan oleh peserta didik. Selanjutnya angka itu diinterpretasikan agar menjadi informasi yang lebih bermakna berkaitan dengan perkembangan hasil belajar peserta didik. Informasi itu tentu saja sangat berguna untuk kepentingan membuat keputusan (evaluasi) terhadap proses pembelajaran yang telah dilakukan.

Secara spesifik, pengukuran dan penilaian merupakan bagian integral dari keseluruhan proses pembelajaran karena hasil dari kedua aktivitas tersebut akan mempengaruhi keputusan yang dibuat oleh guru berkaitan dengan hasil belajar peserta didik, kebutuhan yang menunjang proses pembelajaran, pemenuhan kurikulum, dan sebagainya. Aktivitas pengukuran dan penilaian mestinya tidak hanya diarahkan untuk sekedar mengetahui sejauhmana peserta didik mencapai pengetahuan tertentu, tetapi juga perlu diarahkan untuk mengidentifikasi keterampilan dan sikap yang harus dimiliki oleh peserta didik agar mereka memiliki kesiapan dalam menghadapi dunia yang terus berubah. Bahkan karena tuntutan dunia yang terus berubah itu, pengukuran dan penilaian hasil belajar perlu diarahkan untuk mendapatkan informasi berkaitan dengan kemampuan peserta didik

untuk berpikir kritis, melakukan analisis terhadap suatu permasalahan, dan membuat suatu inferensi dari berbagai kasus faktual. Guna mencapai tujuan tersebut, guru harus memiliki pemahaman dan keterampilan yang baik berkaitan dengan teknik pengukuran dan penilaian hasil belajar yang telah dicapai oleh peserta didik sehingga informasi yang diperoleh dari kedua aktivitas tersebut merupakan informasi yang akurat dan tidak bias makna. Oleh karena itu, perlu adanya panduan berupa buku yang dapat digunakan oleh guru dan calon guru untuk mengembangkan pengetahuan dan pemahaman terhadap kedua hal tersebut.

Buku ini pada dasarnya ditulis sebagai buku ajar mata kuliah Evaluasi Pembelajaran yang dimaksudkan untuk membekali calon guru berkaitan dengan kompetensi pedagogis, terutama kompetensi untuk melakukan evaluasi hasil belajar. Di dalam buku ini, dibahas berbagai hal yang berkaitan dengan konsep asesmen (penilaian) pembelajaran. Pembahasan meliputi pengertian berbagai istilah yang saling berkaitan, yaitu tes, pengukuran, penilaian, dan evaluasi pembelajaran. Banyak orang awam, bahkan guru dan penentu kebijakan di bidang pendidikan menggunakan istilah-istilah tersebut secara bergantian untuk mengacu hal yang sama. Padahal istilah-istilah tersebut berbeda satu sama lain dan mengacu pada aktivitas hirarkis yang perlu dilakukan untuk mengetahui hasil belajar (*achievement*) peserta didik, sekaligus untuk mengetahui efektivitas proses pembelajaran. Dalam buku ini juga dibahas berbagai alasan mengapa guru perlu melakukan asesmen pembelajaran.

Pembahasan berkaitan dengan tes, pengukuran, asesmen, dan evaluasi dalam buku ini disajikan secara ringkas dan praktis sehingga diharapkan guru dan calon guru dapat lebih mudah untuk memahami berbagai terminologi tersebut dan pada akhirnya mampu menerapkannya dalam rangka mengidentifikasi ketercapaian tujuan pembelajaran yang telah ditentukan. Buku ini diawali dengan membahas berbagai konsep yang berkaitan dengan perbedaan tes, pengukuran, penilaian, dan evaluasi. Selanjutnya secara berturut-turut, buku ini membahas tes prestasi dan teknik pengembangannya; variasi bentuk tes; karakteristik tes yang baik; tes standard dan tes buatan guru; pelaksanaan tes (*test administration*); teknik penskoran; konversi skor mentah menjadi skor standar; penilaian

otentik; penilaian berbasis HOTS (*higher order thinking skills*); pelaporan hasil penilaian; dan evaluasi program pembelajaran.

Penulis menyadari bahwa buku ini mungkin masih terdapat berbagai kekurangan, baik dilihat dari konten atau materi, penyajian materi, penjelasan materi, pengutipan pendapat ahli, dan sebagainya. Seperti kata peribahasa “Tak ada gading yang tak retak” yang bermakna bahwa tidak ada kesempurnaan yang ada pada manusia. Oleh karena itu, penulis mengharapkan adanya saran dan masukan konstruktif dari berbagai pihak dalam rangka perbaikan dan penyempurnaan buku ini. Selanjutnya, penulis berharap buku ini akan menjadi kontribusi yang bermakna bagi guru, calon guru, pemerhati pendidikan, dan berbagai pihak yang konsen dalam dunia pendidikan.

Surakarta, November 2020

Penulis

DAFTAR ISI

KATA PENGANTAR

DAFTAR ISI

Bab 1 Tes, Pengukuran, Penilaian dan Evaluasi

A. Tes

1. Tes Kemajuan (*Progress Test*)
2. Tes Sumatif (*Sumative Test*)
3. Tes formatif (*Formative Test*)
4. Tes Penempatan (*Placement Test*)
5. Tes Diagnostik (*Diagnostic Test*)
6. Tes Kecakapan (*Proficiency Test*)
7. Tes Prestasi (*Achievement Test*)
8. Tes objektif dan tes subjektif

B. Pengukuran

C. Asesmen / Penilaian

1. Asesmen formal
2. Asesmen Informal
3. Penilaian Diri

D. Evaluasi

E. Hubungan antara Tes, Pengukuran, Asesmen dan Evaluasi

Bab 2 Teknik Pengembangan Tes Prestasi

A. Tes Prestasi

B. Tahapan Pengembangan Tes Prestasi

1. Menulis *outline*
2. Menulis Kisi-kisi Soal
3. Menuliskan Draft Tes
4. Mengurutkan Butir-butir Tes
5. Uji-coba Tes
6. Menganalisis Soal

Bab 3 Variasi Bentuk Tes Prestasi

A. Tes Objektif

1. Keunggulan dan kelemahan tes objektif
2. Jenis-jenis tes objektif

B. Tes Subjektif

1. Jenis-jenis tes subjektif
2. Kelebihan dan kekurangan tes subjektif
3. Teknik meningkatkan kualitas tes subjektif

Bab 4 Karakteristik Tes yang Baik

- A. Validitas Tes
 - 1. Jenis-jenis validitas tes
 - 2. Uji validitas
- B. Reliabilitas Tes
 - 1. Faktor-faktor yang mempengaruhi reliabilitas tes
 - 2. Metode pengujian reliabilitas
 - 3. Uji reliabilitas tes
- C. Kepraktisan Tes
- D. Dampak Tes
- E. Tingkat Kesulitan Tes
- F. Daya Beda Tes

Bab 5 Tes Standard dan Tes Buatan Guru

- A. Tes Standar
 - 1. Pengertian Tes Standar
 - 2. Kriteria Tes Standar
 - 3. Karakteristik Tes Standar
 - 4. Kegunaan Tes Standar
- B. Tes Buatan Guru
 - 1. Pengertian Tes Buatan Guru
 - 2. Karakteristik Tes Buatan Guru
 - 3. Kegunaan Tes Buatan Guru

Bab 6 Pelaksanaan Tes

- A. Pengertian Pelaksanaan Tes
- B. Prinsip-prinsip Pelaksanaan Tes
- C. Tahapan Pelaksanaan Tes
 - 1. Tahap sebelum tes dilaksanakan
 - 2. Tahap pada saat tes dilaksanakan
 - 3. Tahap setelah tes dilaksanakan

Bab 7 Teknik Penskoran

- A. Teknik Penskoran Domain Kognitif
 - 1. Teknik Penskoran Soal pilihan Ganda
 - 2. Teknik Penskoran Soal Benar-salah
 - 3. Teknik Penskoran Soal Bentuk Menjodohkan
 - 4. Teknik Penskoran tes Bentuk Uraian
 - 5. Penskoran Bentuk Tes campuran
 - 6. Penskoran Tugas
- B. Teknik Penskoran Domain Afektif
- C. Teknik Penskoran Domain Psikomotor

Bab 8 Konversi Skor

- A. Penilaian Acuan Patokan
 - 1. Kegunaan Penilaian Acuan Patokan
 - 2. Konversi Skor dengan Pendekatan Penilaian Acuan Patokan
- B. Penilaian Acuan Norma
 - 1. Kegunaan Penilaian Acuan Norma
 - 2. Konversi Skor dengan Pendekatan Penilaian Acuan Norma

Bab 9 Penilaian Autentik

- A. Penilaian Non-autentik dan Penilaian Autentik
- B. Kelebihan dan Kekurangan Penilaian Non-autentik dan Autentik
- C. Jenis-jenis Penilaian Autentik
 - 1. Penilaian Berbasis Kinerja (*Performance-based Assessments*)
 - 2. Penilaian Berbasis Proyek (*Project-based Assessments*)
 - 3. Penilaian Berbasis Portofolio (*Portfolio-based Assessments*)

Bab 10 Penilaian Berbasis HOTS (Higher Order Thinking Skills)

- A. Pembelajaran Berbasis HOTS
 - 1. Strategi Pembelajaran HOTS
 - 2. Model-model Pembelajaran HOTS
- B. Penilaian Berbasis HOTS
 - 1. Level Kognitif
 - 2. Langkah-langkah Pengembangan Tes Berbasis HOTS

Bab 11 Pelaporan Hasil Penilaian

- A. Hakikat dan Tujuan Pelaporan Hasil Penilaian
- B. Manfaat Pelaporan Hasil Penilaian
 - 1. Manfaat Pelaporan Hasil Penilaian untuk Peserta Didik
 - 2. Manfaat Pelaporan Hasil Penilaian untuk Orang Tua
 - 3. Manfaat Pelaporan Hasil Penilaian untuk Guru dan Kepala Sekolah
 - 4. Manfaat Pelaporan Hasil Penilaian untuk Masyarakat
- C. Model Laporan Hasil Penilaian
 - 1. Identitas Peserta Didik
 - 2. Format Nilai Hasil Belajar Peserta Didik
 - 3. Format ketercapaian kompetensi peserta didik
 - 4. Pengembangan diri
 - 5. Akhlak mulia dan kepribadian
 - 6. Ketidakhadiran
 - 7. Catatan wali kelas

8. Catatan prestasi peserta didik
9. Keterangan pindah sekolah

Bab 12 Evaluasi Program Pembelajaran

- A. Hakaikat Evaluasi
- B. Evaluasi Program Pembelajaran
- C. Model Evaluasi Program Pembelajaran
 1. Model Evaluasi Kirkpatrick
 2. Model Evaluasi Stake
 3. Model Evaluasi Alkin
 4. Model CIPP Stufflebeam
- D. Pentingnya Evaluasi Program Pembelajaran

Daftar Pustaka

Lampiran 1 Tabel r Product Moment Pada Sig.0,05 (Two Tail)

Lampiran 2 Tabel Uji-t

Bab 1

Tes, Pengukuran, Penilaian dan Evaluasi

Asesmen pembelajaran atau biasa juga disebut dengan penilaian hasil belajar adalah aktivitas yang sangat penting dilakukan oleh guru untuk mengetahui tingkat perkembangan hasil belajar peserta didik. Tetapi sebagian guru belum memiliki pemahaman konsep yang baik terhadap istilah tersebut. Dalam bagian ini, dibahas berbagai hal yang berkaitan dengan konsep asesmen pembelajaran. Pembahasan meliputi pengertian berbagai istilah yang saling berkaitan, yaitu tes, pengukuran, asesmen, dan evaluasi pembelajaran. Banyak orang awam, bahkan guru dan penentu kebijakan di bidang pendidikan menggunakan istilah-istilah tersebut secara bergantian untuk mengacu hal yang sama. Padahal istilah-istilah tersebut berbeda satu sama lain dan mengacu pada aktivitas hirarkis yang perlu dilakukan untuk mengetahui hasil belajar (*achievement*) peserta didik sekaligus untuk mengetahui efektivitas proses pembelajaran. Pada bagian ini juga dibahas berbagai alasan mengapa guru perlu melakukan asesmen pembelajaran.

Ada paling tidak empat kegiatan yang harus dilakukan oleh guru untuk mengetahui hasil belajar peserta didik sekaligus untuk mengukur efektivitas proses pembelajaran dalam rangka mencapai tujuan yang telah ditentukan. Aktivitas itu meliputi menguji (memberikan berbagai pertanyaan yang harus dijawab oleh peserta didik); mengukur (menentukan besaran angka yang merefleksikan seberapa besar kemampuan peserta didik dalam menjawab pertanyaan-pertanyaan yang diberikan); menilai (menginterpretasikan angka hasil pengukuran); dan mengevaluasi (memutuskan tingkat keberhasilan belajar peserta didik dan juga keberhasilan proses pembelajaran yang telah dilakukan). Karakteristik berbagai aktivitas tersebut tentunya berbeda satu sama lain dimulai dari aktivitas yang paling sederhana menuju aktivitas yang lebih kompleks. Untuk menguji peserta didik, guru memerlukan tes yang baik. Dari tes yang diberikan kepada peserta didik itu, guru kemudian dapat memberikan skor (dalam bentuk angka) berdasarkan jawaban yang diberikan oleh peserta didik. Selanjutnya angka itu diinterpretasikan agar menjadi informasi yang bermakna berkaitan dengan perkembangan hasil belajar peserta didik.

Informasi itu tentu saja sangat berguna untuk kepentingan evaluasi terhadap proses pembelajaran yang telah dilakukan. Lalu apa yang dimaksud dengan tes, pengukuran, asesmen, dan evaluasi itu? Apa pula kaitan satu sama lain dalam konteks pembelajaran?

A. Tes

Allen & Yan (1979) menyatakan bahwa tes merupakan alat untuk menentukan sampel dari perilaku peserta didik. Sampel mengacu pada sebagian perilaku peserta didik yang ingin diamati oleh guru, karena pada dasarnya sulit bagi guru untuk mengamati keseluruhan perilaku yang dimiliki oleh siswa. Pengertian tersebut cenderung berkaitan dengan aspek-aspek psikologis yang dimiliki oleh peserta didik dan belum menyentuh aspek-aspek kognitif sebagai hasil belajar siswa. Pengertian berbeda diberikan oleh Brown (2004) bahwa tes merupakan metode untuk mengukur kemampuan, pengetahuan, dan kinerja seseorang dalam domain tertentu. Pengertian ini tampaknya lebih mengutamakan fungsi tes sebagai instrumen sebagai alat untuk mengukur kemampuan, pengetahuan, dan kinerja (*performance*) sebagai hasil dari proses belajar peserta didik. Secara faktual dapat dipahami bahwa tes diperlukan setelah proses pembelajaran dilakukan dan dimaksudkan untuk mengukur tingkat ketercapaian hasil belajar peserta didik pada materi pelajaran tertentu.

Pengertian yang lebih spesifik diberikan oleh Mardapi (2008) yang menyatakan bahwa tes merupakan seperangkat pertanyaan yang memiliki jawaban yang mempunyai atribut benar dan salah. Pengertian ini menekankan bahwa pertanyaan-pertanyaan yang menjadi fondasi dari tes harus memiliki atribut benar dan salah. Dengan kata lain, jawaban-jawaban yang diberikan oleh peserta didik dalam merespon pertanyaan-pertanyaan itu dapat diklasifikasikan menjadi jawaban benar dan jawaban salah. Pengertian ini penting untuk dipahami guna membedakan respon yang diberikan oleh responden (mungkin juga oleh peserta didik) terhadap pertanyaan atau pernyataan dalam kuesioner tertentu. Jawaban sebagai respon terhadap pertanyaan atau pernyataan dari kuesioner itu tidak dapat diklasifikasikan kedalam jawaban benar dan jawaban salah, karena jawaban-jawaban itu biasanya bersifat kontinum. Oleh karena itu, kuesioner dikategorikan sebagai instrumen non-tes.

Berbagai pengertian tersebut tentunya memberikan wawasan yang lebih baik tentang tes. Tes tidak hanya digunakan untuk mengukur perilaku dalam bidang psikologi, tetapi tes juga banyak digunakan dalam bidang pendidikan. Dengan demikian dapat dipahami bahwa tes merupakan seperangkat atau sejumlah pertanyaan yang memerlukan jawaban dengan maksud untuk mengukur tingkat keberhasilan belajar seseorang (peserta

didik) atau mengungkapkan aspek-aspek tertentu dari orang yang dikenai tes itu. Hasil dari tes merupakan informasi yang berkaitan dengan karakteristik seseorang, baik secara individu atau kelompok. Karakteristik itu dapat berupa kemampuan dan keterampilan yang dimiliki seseorang atau sekelompok orang.

Pada dasarnya tes dapat digunakan untuk berbagai macam keperluan. Misalnya tes digunakan untuk seleksi. Ketika jumlah calon peserta didik yang akan diterima di sebuah sekolah tertentu melebihi kuota yang ditentukan, maka tes dapat digunakan untuk menyeleksi calon peserta didik itu sesuai dengan kriteria yang telah ditentukan. Demikian pula perguruan tinggi dapat menyeleksi calon mahasiswa sesuai dengan kriteria tertentu dengan menggunakan tes. Perusahaan yang membutuhkan karyawan dengan kemampuan dan keterampilan tertentu pun dapat menggunakan tes sebagai instrumen untuk menyeleksi calon karyawan yang berkualitas sesuai dengan standar perusahaan.

Tes juga dapat digunakan untuk keperluan klasifikasi. Penggunaan tes untuk klasifikasi sangat populer di bidang pendidikan bahasa sebagai tes penempatan (*placement test*). Dengan menggunakan tes, kelas bahasa dapat menempatkan peserta didik sesuai dengan kemampuan awal yang dimiliki oleh peserta didik itu, misalnya kemampuan dasar (*elementary level*), kemampuan menengah (*intermediate level*), dan kemampuan lanjut (*advanced level*). Klasifikasi ini tentu sangat berguna bagi guru bahasa, misalnya guru bahasa Inggris, untuk menentukan metode, media, teknik, dan bahan ajar yang tepat sesuai dengan tingkat kemampuan peserta didik.

Tes juga dapat digunakan untuk keperluan evaluasi. Hasil dari evaluasi dapat digunakan, misalnya untuk menentukan apakah siswa naik atau tinggal kelas; untuk menentukan apakah seseorang layak membuka praktik kedokteran atau tidak; dan juga untuk menentukan efektivitas program pembelajaran. Selain itu, tes dapat pula digunakan untuk keperluan konseling. Misalnya *interest inventory* dapat digunakan untuk menentukan peluang sukses dalam karir seseorang atau dengan *personality test* untuk membantu pasangan calon pengantin dalam mempersiapkan mental sebelum pernikahannya. Tes juga sangat berguna untuk kepentingan penelitian. Dalam penelitian eksperimen, peneliti dapat menggunakan tes untuk mengklasifikasikan atau mengelompokkan subjek penelitian yang memiliki kreativitas tinggi dan rendah dengan tes kreativitas. Peneliti juga dapat menggunakan tes untuk mengukur perilaku seseorang dan kemudian menguji tata hubungan antarvariabel. Begitu besar manfaat tes dalam berbagai bidang kehidupan. Untuk mengevaluasi tingkat kebermanfaatan tes tersebut diperlukan sistem pengukuran yang tepat sehingga tes tersebut akan lebih bermakna.

Berkaitan dengan tujuannya, tes dapat dibedakan menjadi beberapa jenis tes. Berbagai jenis tes itu ada yang diberikan kepada *testee* (orang atau peserta didik yang melaksanakan tes) sebelum, pada saat, dan setelah program pembelajaran berlangsung. Sebagaimana diketahui bahwa tes dilaksanakan karena berbagai macam tujuan dan digunakan dalam berbagai tahapan proses pembelajaran sebagai sarana untuk mengumpulkan informasi berkaitan dengan kondisi peserta didik, baik sebelum dan setelah proses pembelajaran. Sebagai guru tentu saja harus mampu menentukan jenis tes mana yang harus digunakan untuk mengumpulkan informasi tentang kompetensi peserta didik itu. Pada bagian berikut, diuraikan berbagai jenis tes yang biasa digunakan dalam bidang pendidikan.

1. Tes Kemajuan (*Progress Test*)

Seperti namanya, tes kemajuan, tes ini dimaksudkan untuk mengetahui tingkat kemajuan belajar peserta didik setelah mengikuti proses pembelajaran dalam jangka waktu tertentu. Tes ini dikembangkan sendiri oleh guru berdasarkan buku teks atau materi yang pernah diajarkan. Tes ini biasanya dilaksanakan setelah proses pembelajaran untuk beberapa bab / unit / topik / pokok bahasan diselesaikan. Tujuan utama dari tes ini adalah untuk mendapatkan informasi mengenai sejauhmana atau sebaik apa peserta didik, baik secara individu atau kelompok, telah menguasai materi ajar dan mencapai tujuan pembelajaran yang telah ditentukan. Guru biasanya memberikan tes kemajuan dalam bentuk ulangan harian. Agar tes jenis ini dapat memberikan informasi yang akurat tentang kemajuan belajar peserta didik, guru harus menentukan sampel yang representatif dari keseluruhan materi yang telah diajarkan sebelumnya. Dengan kata lain, materi yang diujikan harus mewakili keseluruhan materi yang telah diajarkan oleh guru. Sebenarnya tes ini tidak hanya berguna untuk mengetahui kemajuan belajar peserta didik, tetapi juga dapat digunakan oleh guru untuk mendapatkan informasi terhadap efektivitas metode dan teknik pembelajaran yang digunakan. Apabila metode dan teknik pembelajaran itu dipandang tidak berfungsi dengan baik untuk membantu kemajuan belajar peserta didik, guru dapat merevisi dan memodifikasi metode dan teknik itu agar lebih efektif atau bahkan menggantinya dengan metode atau teknik pembelajaran yang sama sekali baru.

Ada beberapa hal yang harus dipertimbangkan oleh guru ketika akan mengembangkan tes kemajuan (Dorobot, 2007), yaitu:

- a. Materi tes harus berdasarkan program pembelajaran dan kurikulum (buku teks dan buku-buku lain) yang digunakan selama program pembelajaran itu berlangsung.

- b. Tes harus digunakan untuk menguji efektivitas proses pembelajaran dan hasil belajar peserta didik.
- c. Tes dimaksudkan untuk membantu guru dalam menentukan tingkat kemajuan belajar peserta didik, baik secara individu atau kelompok.
- d. Tes harus memiliki dampak positif (*positive washback effect*) terhadap cara belajar peserta didik.
- e. Tes harus memperkuat pemahaman peserta didik berkaitan dengan apa yang telah diajarkan oleh guru.
- f. Tes harus mengarahkan peserta didik untuk dapat menunjukkan penguasaannya terhadap pengetahuan dan keterampilan yang telah dipelajari.
- g. Tes harus menghasilkan skor yang tinggi karena tes ini dimaksudkan untuk mengidentifikasi kemajuan belajar peserta didik.

Tes kemajuan mestinya dibuat sendiri oleh guru yang mengajar peserta didik itu karena tes ini digunakan untuk mengukur apa yang telah diajarkan dan dipelajari sebelumnya. Selain itu, guru adalah orang yang paling memahami mengenai apa dan bagaimana mengevaluasi hasil belajar peserta didik yang diajar.

2. Tes Sumatif (*Sumative Test*)

Tes ini biasa diberikan diakhir semester dan bertujuan untuk mengetahui apakah peserta didik telah mencapai tujuan pembelajaran seperti yang dinyatakan dalam silabus untuk mata pelajaran tertentu. Tes ini banyak dimanfaatkan untuk menentukan apakah seorang peserta didik dapat naik kelas atau tidak. Oleh karena itu, tes ini banyak diterapkan pada jenjang sekolah dasar hingga sekolah menengah. Kelemahan dari jenis tes ini adalah guru tidak dapat memberikan *feedback* terhadap proses pembelajaran yang telah dilakukan karena peserta didik mungkin telah menempati kelas yang berbeda setelah dinyatakan naik kelas. Selain itu, tes ini banyak mengakibatkan stress atau beban psikologis, baik bagi peserta didik maupun guru. Mengapa hal itu terjadi? Karena cakupan bahan atau materi ajar yang harus diujikan lebih banyak sehingga beban belajar peserta didik juga semakin banyak. Juga ada tuntutan dari setiap peserta didik untuk dapat naik kelas. Selain itu, tes ini dapat juga digunakan sebagai gambaran dari tingkat keberhasilan guru dalam melaksanakan proses pembelajaran yang telah dilakukan. Apabila banyak peserta didik yang gagal dalam tes sumatif, tentu saja beban psikologis guru juga semakin berat.

3. Tes formatif (*Formative Test*)

Bebeda dengan tes sumatif, tes formatif diberikan dengan maksud untuk memberikan informasi berkaitan dengan efektivitas dan efisiensi proses pembelajaran. Tes ini diberikan pada saat atau segera setelah pembelajaran berlangsung. Hasil dari tes ini digunakan sebagai dasar untuk memberikan *feedback* terhadap proses belajar peserta didik dan juga *feedback* terhadap efektivitas metode dan teknik pembelajaran yang digunakan oleh guru. Tes ini hanya mengukur penguasaan peserta didik berkaitan dengan apa yang telah diajarkan di kelas. Secara praktis, tes ini dapat dimanfaatkan oleh guru untuk memperbaiki kualitas pembelajaran yang dilakukan dan juga dapat digunakan untuk mendorong peserta didik agar belajar lebih baik.

4. Tes Penempatan (*Placement Test*)

Tes jenis ini juga biasa disebut sebagai tes masuk (*entry test*) dan bertujuan untuk menempatkan atau mengklasifikasi peserta didik pada kelompok-kelompok yang lebih homogen. Homogenitas ini mengacu pada kemampuan atau keterampilan yang relatif sama yang dimiliki oleh setiap peserta didik. Penempatan setiap peserta didik yang memiliki kemampuan dan keterampilan yang relatif sama ini akan memudahkan bagi guru untuk dapat memberikan layanan pembelajaran yang tepat sehingga efektivitas pembelajaran dapat dicapai.

5. Tes Diagnostik (*Diagnostic Test*)

Tidak seperti jenis tes lain yang dimaksudkan untuk mengetahui tingkat kesuksesan belajar peserta didik, tes ini sebaliknya dimaksudkan untuk mengetahui berbagai kendala belajar yang dihadapi oleh peserta didik. Dengan menggunakan tes ini, guru ingin mengetahui pada bagian mana peserta didik, baik secara individu atau kelompok, memiliki kendala atau masalah untuk dapat menguasai atau mencapai tujuan pembelajaran yang telah ditentukan. Jadi, tes jenis ini lebih dimaksudkan untuk mengetahui penyebab kegagalan belajar peserta didik, bukan untuk mengetahui berapa banyak pertanyaan yang dijawab benar oleh peserta didik. Informasi diagnostik yang diperoleh dari hasil tes diagnostik ini sangat berguna bagi guru untuk mendesain proses pembelajaran yang efektif. Selain itu, juga dapat digunakan oleh guru untuk mendesain aktivitas remedial guna membantu keberhasilan belajar peserta didik. Peserta didik juga dapat memanfaatkan informasi diagnostik untuk menganalisis sendiri masalah-masalah belajar yang mereka hadapi.

6. Tes Kecakapan (*Proficiency Test*)

Jenis tes ini bertujuan untuk mendeskripsikan kemampuan peserta didik dalam menguasai bahasa asing, misalnya bahasa Inggris. Tes ini merupakan tes standar yang didesain dan dikembangkan oleh lembaga tertentu dan tidak dikembangkan sendiri oleh guru. Berbeda dengan tes prestasi (*achievement test*) yang dikembangkan berdasarkan pada kurikulum dan silabus tertentu, *proficiency test* tidak dikembangkan berdasarkan kurikulum dan silabus tertentu. Tes jenis ini dapat digunakan oleh peserta didik sebagai bukti akan kemampuannya dalam menguasai bahasa asing. Dalam konteks pembelajaran bahasa, contoh *proficiency test* adalah *Test of English as Foreign Language (TOEFL)*, *Tests of English as International Communication (TOEIC)*, *English for Academic Purposes (EAP)*, *International English Language Testing System (IELTS)* dan sebagainya.

7. Tes Prestasi (*Achievement Test*)

Tes prestasi bertujuan untuk mengetahui sejauhmana penguasaan peserta didik terhadap bahan ajar yang telah dipelajari dengan mengacu pada mata pelajaran tertentu dan tujuan program pembelajaran yang telah dinyatakan secara eksplisit. Tes ini berbeda dengan tes kecakapan yang tidak terkait dengan silabus atau program pembelajaran. Kedua istilah ini (tes prestasi dan tes kecakapan) kadang-kadang digunakan secara bergantian. Mereka yang membedakan antara dua istilah ini menekankan fakta bahwa tes prestasi didasarkan pada pembelajaran yang telah dilaksanakan atau berdasarkan buku yang menjadi pegangan pada saat pembelajaran berlangsung. Sementara tes kecakapan dikembangkan terlepas dari pembelajaran dan buku yang digunakan. Perbedaan mencolok antara dua jenis tes ini berkaitan dengan tujuan masing-masing tes itu. Tes kecakapan (*proficiency test*) umumnya digunakan untuk sertifikasi kompetensi belajar. Sedangkan tes prestasi dapat digunakan sekaligus sebagai tes diagnostik untuk mengidentifikasi kendala belajar yang dimiliki oleh peserta didik dan hasil dari tes prestasi itu dapat dimanfaatkan sebagai dasar untuk perbaikan proses pembelajaran berikutnya.

8. Tes objektif dan tes subjektif

Berdasarkan cara penskorannya, tes dapat dibedakan menjadi tes objektif dan tes subjektif. Kedua jenis tes ini tentu memiliki karakteristik yang berbeda satu dan lainnya. Selain itu, kedua tes ini memiliki kekuatan dan kelemahan masing-masing. Tes objektif adalah tes yang cara penskorannya tidak dipengaruhi oleh subjektivitas orang yang memberi skor (*tester/scorer/rater*). Pada tes objektif; seperti tes pilihan ganda, tes benar-salah, tes menjodohkan, dan tes yang menghendaki jawaban singkat,

pemeriksaan dan penskorannya dapat dilakukan oleh orang yang tidak memiliki pengetahuan khusus tentang materi yang diujikan. Siapapun dapat memeriksa dan memberi skor berdasarkan kunci jawaban yang telah disiapkan. Tes objektif sangat cocok untuk mengukur kemampuan peserta didik yang menuntut proses mental yang tidak begitu tinggi; seperti hafalan dan ingatan. Sebaliknya, tes subjektif adalah tes yang penskorannya sering dipengaruhi oleh subjektivitas orang yang memberi skor. Opini, anggapan, dan pertimbangan *tester* sangat mempengaruhi skor yang diperoleh oleh *testee*. *Tester* harus memiliki pengetahuan yang memadai berkaitan dengan materi ajar yang diujikan. Pada umumnya tes subjektif berbentuk uraian atau esai. Dalam konteks pembelajaran bahasa, tes kemampuan berbicara dan menulis juga termasuk tes subjektif. Paling tidak ada dua cara yang dapat dilakukan untuk mengurangi subjektivitas dalam memberikan skor. Pertama, guru harus menyiapkan rubrik penskoran yang memuat berbagai aspek yang menjadi fokus penskoran. Kedua, proses penskoran dilakukan oleh dua orang atau lebih (*inter-rater*). Penskoran yang dilakukan oleh lebih dari dua orang ini juga dimaksudkan untuk mencapai tingkat reliabilitas hasil penskoran yang lebih baik.

B. Pengukuran

Banyak aktivitas dalam kehidupan manusia yang memerlukan pengukuran sebagai cara untuk melukiskan karakteristik orang atau benda tertentu. Tukang kayu yang mengukur panjang sebuah meja kemudian tahu bahwa panjang meja itu adalah 2 meter. Pekerja bangunan mengukur luas rumah yang akan dibangun kemudian dia tahu lebar rumah itu 12 meter dan panjangnya 20 meter. Seorang tukang jahit juga tidak lepas dari kegiatan mengukur kain yang akan dibuat baju atau celana. Penjual buah ingin mengetahui berat jeruk yang dibeli oleh seorang konsumen juga tidak lepas dari aktivitas pengukuran. Semua aktivitas yang dimaksudkan untuk menentukan ukuran suatu benda dalam bentuk angka-angka adalah pengukuran. Termasuk seorang guru yang berusaha memberikan skor hasil ulangan peserta didiknya juga melakukan pengukuran, karena guru itu pada akhirnya setelah memeriksa jawaban ulangan peserta didiknya akan memberikan skor dalam bentuk angka tertentu, misalnya 85. Hanya saja instrumen yang digunakan oleh setiap profesi untuk melakukan pengukuran berbeda-beda. Tukang kayu dan tukang jahit melakukan pengukuran dengan menggunakan instrumen alat meter (meteran), penjual buah melakukan pengukuran dengan menggunakan timbangan, dan guru melakukan pengukuran hasil belajar peserta didik dengan menggunakan instrumen tes. Lalu apa yang dimaksud pengukuran itu?

Pengukuran merupakan prosedur sistematis yang digunakan untuk menentukan angka yang merepresentasikan karakteristik individu atau objek tertentu (Allen & Yan, 1979). Dalam konteks pembelajaran, angka-angka itu mengacu pada skor yang diperoleh oleh peserta didik setelah mengikuti ujian atau tes tertentu. Proses penentuan angka ini tidak boleh dilakukan secara sembarangan, tetapi harus dilakukan secara hati-hati berdasarkan prosedur yang telah ditentukan dan prosedur itu harus dapat diulang-ulang. Demikian itu yang dimaksud dengan prosedur sistematis dalam menentukan angka. Sebagai contoh, setiap peserta didik yang telah mengerjakan ulangan mata pelajaran matematika, maka mereka akan mendapat skor tertentu. Skor itu diperoleh karena peserta didik telah mengerjakan soal ulangan yang memuat instruksi, butir-butir pertanyaan, dan prosedur penskoran yang sama untuk setiap peserta didik. Skor tentu tidak akan bermakna apabila setiap peserta didik diberi instruksi, butir-butir pertanyaan dan prosedur penskoran yang berbeda-beda, karena perbedaan itu mencerminkan tindakan yang tidak sistematis dalam melakukan pengukuran.

Dalam melakukan pengukuran, orang atau guru sering melakukan kesalahan sehingga hasilnya kurang akurat. Tukang kayu melakukan kesalahan pengukuran karena alat ukur (meteran) yang digunakan sudah rusak dan satuan angka dalam meteran itu sudah tidak terlihat dengan jelas. Penjual buah melakukan kesalahan pengukuran karena timbangan yang dipakai sudah tidak standar. Guru melakukan kesalahan pengukuran hasil belajar peserta didik karena instrumen tes yang digunakan berkualitas buruk, misalnya tes yang digunakan tidak valid dan hasil tes tidak reliabel. Sejalan dengan ini, Mardapi (2008) menekankan bahwa dalam melakukan pengukuran untuk menentukan karakteristik individu harus sedapat mungkin mengandung kesalahan yang kecil. Kesalahan pengukuran dalam ilmu-ilmu alam lebih sederhana dibandingkan kesalahan pengukuran dalam ilmu-ilmu sosial. Kesalahan pengukuran pada ilmu-ilmu alam sering disebabkan oleh alat ukurnya, sedangkan kesalahan pengukuran pada ilmu sosial banyak disebabkan oleh alat ukur yang digunakan, cara mengukur, dan keadaan objek yang diukur.

Secara spesifik, agar pengukuran hasil belajar peserta didik akurat, alat ukur (tes) yang digunakan oleh guru untuk mengukur hasil belajar peserta didik pun harus baik, yaitu alat ukur harus valid dan reliabel. Selain itu, cara bagaimana pengukuran itu dilakukan juga harus baik. Pengukuran tidak boleh dipengaruhi oleh subjektivitas guru karena subjektivitas guru tentu saja akan mengakibatkan hasil pengukuran tidak reliabel dan akhirnya interpretasi hasil pengukuran akan bias makna. Selain itu, guru juga perlu memperhatikan keadaan objek (peserta didik) yang diukur. Kelompok peserta didik yang diberikan tes untuk mengukur kemampuannya pada jam

pertama proses pembelajaran dan peserta didik yang di beri tes pada jam terakhir proses pembelajaran tentu akan menghasilkan tingkat akurasi hasil pengukuran yang berbeda. Tingkat kesiapan peserta didik dalam mengikuti tes juga merupakan salah satu sumber kesalahan pengukuran.

Validitas (kesahihan) alat ukur yang tinggi merupakan cerminan dari sejauhmana alat ukur tersebut mampu mengukur kompetensi peserta didik yang seharusnya diukur. Misalnya, guru akan mengukur kompetensi berbicara peserta didik, tetapi guru tersebut menggunakan alat ukur berbentuk pilihan ganda. Tentu saja hasil pengukuran ini tidak mencerminkan kemampuan kompetensi peserta didik yang sebenarnya karena alat ukur yang digunakan tidak valid. Kesahihan alat ukur dapat dilihat paling tidak dari kisi-kisi yang digunakan sebagai dasar pijakan dalam mengembangkan alat ukur tersebut. Kisi-kisi memuat materi yang akan diujikan, bentuk soal, tingkat berfikir yang terlibat, bobot soal, dan cara penskoran (Mardapi, 2008). Materi yang ada dalam kisi-kisi harus mewakili keseluruhan materi yang telah diajarkan oleh guru. Hal ini perlu diperhatikan karena sangat tidak mungkin seorang guru mengujikan keseluruhan materi ajar yang telah diajarkan sebelumnya. Karena alasan ini, guru perlu menentukan sampel materi yang harus dinyatakan dalam kisi-kisi dan diujikan kepada peserta didik. Ada beberapa pertimbangan yang harus diperhatikan oleh guru dalam memilih materi atau menentukan sampel materi yang harus diujikan kepada peserta didik. Pertimbangan itu diantaranya adalah (1) urgensi, yaitu kompetensi inti, kompetensi dasar, dan indikator yang secara teoritis mutlak harus dikuasai oleh peserta didik; (2) kontinuitas, yaitu kompetensi inti, kompetensi dasar, dan indikator lanjutan yang merupakan pendalaman materi sebelumnya; (3) relevansi, yaitu materi itu diperlukan untuk mempelajari bidang studi atau mata pelajaran lainnya; dan (4) keterpakaian, yaitu materi itu memiliki nilai terapan yang tinggi dalam kehidupan sehari-hari. Berbagai pertimbangan tersebut penting untuk diperhatikan pada saat mengembangkan alat ukur (tes) sebagai upaya untuk mencapai validitas alat ukur, paling tidak validitas isi. Dengan demikian, alat ukur itu dapat digunakan secara akurat untuk mengukur kompetensi peserta didik yang semestinya akan diukur.

C. Asesmen / Penilaian

Dalam bahasa Indonesia, istilah asesmen sering dipadankan dengan penilaian. Asesmen mengacu pada berbagai macam cara untuk mengumpulkan informasi yang berkaitan dengan kompetensi dan prestasi peserta didik (Dorobat, 2007). Asesmen juga diartikan sebagai upaya untuk menginterpretasikan hasil pengukuran menjadi informasi yang lebih bermakna bagi peserta didik. Misalnya, Amir pada saat mengikuti ulangan

fisika mendapat skor 50. Berdasarkan skor ini berarti Amir belum menguasai materi dengan baik. Di lain pihak, Siti mendapat skor 70 yang berarti Siti telah cukup menguasai materi ajar. Sedangkan Arnold mendapat skor 80 yang berarti Arnold telah menguasai materi ajar dengan baik. Interpretasi skor dalam bentuk angka menjadi informasi kualitatif ini merupakan hakikat dari asesmen.

Selama ini, istilah tes dan asesmen (penilaian) sering digunakan oleh banyak orang secara bergantian untuk mengacu pada hal yang sama, tetapi sebenarnya istilah asesmen merupakan payung yang mencakup alat ukur berupa tes dan juga metode kualitatif yang digunakan untuk memonitor dan mencatat proses belajar peserta didik, seperti observasi, simulasi, dan kerja peserta didik berbasis proyek. Dengan kata lain, asesmen dan tes memiliki hubungan yang erat. Seorang guru sulit melakukan asesmen tanpa didahului oleh tes yang kemudian dari tes itu akan diperoleh skor. Skor yang diperoleh oleh peserta didik ini merupakan dasar pijakan untuk melakukan asesmen terhadap prestasi peserta didik.

Asesmen terhadap prestasi peserta didik harus dilakukan mengacu pada apa yang telah dipelajari oleh setiap peserta didik dan berkaitan dengan konten dan tujuan mata pelajaran tertentu. Ini berarti bahwa asesmen tidak boleh dilakukan terhadap penguasaan dan prestasi peserta didik di luar materi yang belum diajarkan. Berdasarkan kapan asesmen itu harus dilakukan, ada dua jenis asesmen dalam proses pembelajaran, yaitu asesmen formatif dan asesmen sumatif. Asesmen formatif dilakukan oleh guru selama proses pembelajaran berlangsung yang dimaksudkan untuk memperbaiki dan meningkatkan kualitas pembelajaran. Sedangkan asesmen sumatif dilakukan pada akhir program pembelajaran yang dimaksudkan untuk memberikan informasi tentang efektivitas program pembelajaran itu untuk sekolah atau lembaga pendidikan yang lebih tinggi.

Pada dasarnya, ada tiga bentuk asesmen yang perlu dilakukan oleh guru dalam rangka mengetahui perkembangan hasil belajar peserta didik yaitu, asesmen formal, asesmen informal, dan asesmen diri (Harris and McCann, 1994). Asesmen formal adalah bentuk penilaian yang biasa disebut dengan tes. Dalam asesmen formal, guru harus menciptakan kondisi yang sedemikian rupa agar tes sebagai bentuk dari asesmen formal dapat dilaksanakan dengan baik. Penciptaan kondisi ini meliputi menyiapkan soal, melaksanakan tes, megawasi pelaksanaan tes dan sebagainya. Di lain pihak, asesmen informal adalah bentuk penilaian yang dilakukan oleh guru tidak di bawah kondisi tes khusus, tetapi penilaian itu dilaksanakan dalam lingkungan kelas yang normal. Pada saat melaksanakan asesmen atau penilaian informal, guru tidak perlu menyiapkan dan menciptakan kondisi khusus yang memungkinkan tes dapat dilaksanakan secara formal. Teknik

utama dalam asesmen informal adalah melalui observasi kelas yang dilakukan oleh guru itu sendiri. Guru juga tidak perlu secara khusus menyiapkan soal yang harus dikerjakan oleh peserta didik. Selanjutnya, asesmen diri adalah proses penilaian yang dilakukan oleh peserta didik sendiri terhadap kemajuan belajarnya dan juga kendala belajar yang mungkin dihadapi. Bagaimana seharusnya ketiga bentuk asesmen tersebut dilakukan oleh guru adalah sebuah pertanyaan penting yang harus dicari solusinya, karena pemahaman terhadap ketiga bentuk asesmen tersebut oleh sebagian guru masih sering samar-samar.

1. Asesmen formal

Seperti telah dijelaskan bahwa asesmen formal dilakukan oleh guru dalam bentuk tes. Tes tentu saja memuat serangkaian atau seperangkat pertanyaan dan mungkin juga pernyataan yang harus direspon oleh setiap peserta didik. Banyak ahli menyebut bentuk asesmen formal sebagai bentuk asesmen konvensional karena memang bentuk asesmen ini sejak lama telah banyak digunakan di dunia pendidikan. Sebelum asesmen formal dilaksanakan, guru perlu mempersiapkan banyak hal berkaitan dengan soal yang harus dikerjakan oleh peserta didik, ruangan yang harus diatur sedemikian rupa agar peserta tes (*testee*) tidak dapat saling bertindak curang, menyiapkan seperangkat tes, menyiapkan rubrik penskoran dan sebagainya. Asesmen formal tidak mungkin dilaksanakan tanpa mempersiapkan kondisi yang dapat mendukung kelancaran pelaksanaan asesmen tersebut.

Asesmen formal, walaupun bentuk asesmen ini dikategorikan sebagai asesmen konvensional, masih mempunyai peran penting untuk mengetahui tingkat perkembangan belajar peserta didik. Harris and McCann (1994) menjelaskan ada beberapa alasan mengapa tes sebagai bentuk asesmen formal perlu dilakukan, diantaranya:

- a. Guru mungkin ingin mengetahui tingkat pengetahuan awal setiap peserta didik sebelum mengikuti pembelajaran untuk materi tertentu sehingga peserta didik dapat dikelompokkan sesuai tingkat kemampuannya dan kemudian metode dan teknik pembelajaran dapat disesuaikan dengan gaya belajar peserta didik.
- b. Guru ingin mengetahui daya serap peserta didik terhadap materi yang telah diajarkan selama periode tertentu.
- c. Guru ingin membandingkan kinerja setiap peserta didik dengan kinerja peserta didik lain di kelas tertentu.

- d. Guru ingin mengetahui sejauhmana kemajuan belajar setiap peserta didik setelah proses pembelajaran berlangsung dibandingkan dengan sebelum proses pembelajaran berlangsung.

2. Asesmen Informal

Asesmen informal adalah upaya untuk mengumpulkan informasi mengenai penguasaan materi, kinerja (*performance*), sikap, kemampuan bekerja-sama, kemandirian, dan kreativitas peserta didik pada kondisi kelas normal. Asesmen ini dilakukan tanpa harus menciptakan kondisi tertentu seperti yang dilakukan pada saat pelaksanaan asesmen formal pada umumnya. Asesmen informal biasa juga disebut sebagai asesmen berkelanjutan karena dilaksanakan sepanjang periode tertentu, misalnya sepanjang jam pelajaran berlangsung atau dapat juga sepanjang semester terhadap kinerja peserta didik. Ketika guru dan peserta didik secara bersama-sama berada di kelas, guru secara intuitif dapat melakukan penilaian terhadap kinerja peserta didik, misalnya pada saat mereka mendengarkan dan mencatat penjelasan guru, pada saat membaca buku, dan pada saat menyampaikan tanggapan atau pendapat. Melalui proses pengamatan, guru dapat menilai mana peserta didik yang telah memiliki kinerja baik dan mana yang masih memiliki kendala belajar. Melalui asesmen informal, guru pada akhirnya dapat mengetahui bagaimana sikap peserta didik, seberapa besar usaha mereka untuk belajar dan sejauhmana tingkat partisipasi mereka selama kegiatan pembelajaran berlangsung. Dengan demikian dapat dipahami bahwa asesmen informal dapat dilakukan oleh guru melalui proses observasi yang sistematis selama dan setelah pembelajaran berlangsung.

Observasi merupakan metode yang paling krusial bagi guru untuk melaksanakan asesmen informal. Agar asesmen informal dapat berjalan efektif dan efisien, ada beberapa hal yang perlu disiapkan oleh guru sebelum observasi dilaksanakan, diantaranya:

- a. Guru perlu menentukan apa saja yang akan dinilai dari sekian banyak kinerja peserta didik, karena pada dasarnya guru tidak mungkin dapat menilai seluruh kinerja yang dimiliki peserta didik secara bersamaan dalam kurun waktu tertentu.
- b. Guru harus menentukan kriteria yang jelas untuk menilai peserta didik sebelum penilaian dilaksanakan dan penilaian tidak boleh hanya mengandalkan kesan yang kasat mata saja. Kriteria ini juga bermanfaat untuk mengurangi subjektivitas guru pada saat melakukan penilaian.
- c. Guru harus memikirkan bagaimana menghubungkan atau mengaitkan hasil penilaian informal dengan penilaian formal dan

penilaian diri yang dilakukan oleh peserta didik. Upaya ini perlu dilakukan agar hasil penilaian informal dapat mendukung hasil penilaian formal dan penilaian diri. Jadi hasil dari ketiga penilaian tersebut dapat saling melengkapi dan bukan sesuatu yang saling asing (*mutually exclusive*).

Ada beberapa hal yang harus dipahami berkaitan dengan penilaian informal, yaitu:

- a. Asesmen atau penilaian informal bukanlah pengganti dari jenis penilaian lain, seperti penilaian formal dan penilaian diri. Oleh karena itu, penilaian informal semestinya tidak dianggap sebagai satu-satunya cara untuk mendapatkan dan memberikan informasi tentang peserta didik.
- b. Penilaian informal dilakukan bukan sebagai upaya untuk menghindari tes sehingga guru tidak boleh membuat keputusan tentang hasil belajar peserta didik hanya semata-mata berdasarkan hasil pengamatan informal.
- c. Penilaian informal bukanlah bentuk penilaian tanpa kriteria. Ini berarti bahwa semua keputusan berkaitan dengan kemajuan belajar peserta didik harus didasarkan pada kriteria-kriteria yang telah ditentukan sebelum penilaian informal dilaksanakan.
- d. Penilaian informal bukan merupakan sistem tunggal yang dapat digunakan oleh banyak guru dari sekolah-sekolah yang memiliki latar-belakang dan konteks yang berbeda. Hal ini berarti, setiap guru atau kelompok guru harus mencari format penilaian informal yang sesuai dengan kondisi sekolah masing-masing.
- e. Penilaian informal mestinya tidak dianggap sebagai bentuk penilaian yang terpisah dari bentuk penilaian lain, tetapi bentuk penilaian ini digunakan untuk melengkapi dan mendukung hasil dari bentuk penilaian lain, misalnya penilaian formal, agar pengambilan keputusan tentang hasil belajar peserta didik menjadi lebih akurat.

Setelah melaksanakan penilaian informal, guru akan mendapatkan informasi berkaitan dengan tingkat kemajuan belajar peserta didik. Misalnya dalam pembelajaran bahasa Inggris, guru akan mendapatkan informasi berkaitan dengan kemampuan *listening*, *speaking*, *reading*, dan *writing* secara bersamaan. Selain itu, guru juga mendapat informasi berkaitan dengan tingkat kemandirian dan kemampuan peserta didik dalam bekerja sama dengan anggota dalam kelompoknya. Agar informasi ini dapat

bermanfaat dalam konteks pembelajaran, hasil penilaian informal perlu dicatat dan diberitahukan kepada setiap peserta didik.

Karena hasil penilaian informal merupakan bagian dari keseluruhan penilaian yang diperoleh dari penilaian formal dan penilaian diri, maka hasil itu secara periodik perlu dibandingkan dengan hasil dari bentuk penilaian lain. Jika hasil dari tiga bentuk penilaian tersebut memiliki korelasi yang tinggi, yaitu hasil ketiga bentuk penilaian tersebut menghasilkan informasi yang relatif sama, maka reliabilitas kemampuan peserta didik telah terbangun dan akhirnya guru dapat memberi keputusan tentang kemampuan peserta didik berdasarkan landasan yang kuat. Seandainya hasil penilaian salah satu peserta didik tidak berkorelasi secara signifikan, misalnya hasil penilaian informal sangat berbeda dengan hasil penilaian formal, guru harus mencari penyebab mengapa hal itu terjadi. Mungkin saja pada saat mengikuti penilaian formal (tes), peserta didik itu sedang dalam keadaan sakit atau kecapekan yang tentu hasilnya akan berbeda ketika sedang sehat dan segar bugar. Tetapi kalau perbedaan itu terjadi pada hampir seluruh peserta didik, guru perlu mengecek kembali validitas dan reliabilitas tes formalnya. Sebaliknya jika hasil tes lebih konsisten daripada hasil penilaian informal, maka guru perlu mengecek kembali prosedur penilaian informal yang telah dilakukan. Mungkin ada prosedur yang salah pada saat penilaian informal itu dilakukan dan pada tahap berikutnya guru harus memperbaiki prosedur itu.

Secara substansial, asesmen informal dan asesmen formal berbeda berkaitan kapan kedua bentuk asesmen tersebut harus dilaksanakan. Asesmen formal dilakukan pada saat proses pembelajaran untuk materi tertentu telah selesai. Bisa saja materi itu harus dibahas dalam beberapa kali proses pembelajaran dan asesmen formal baru bisa dilaksanakan. Tujuannya untuk mengetahui tingkat daya serap peserta didik terhadap materi yang telah diajarkan. Hasil asesmen formal dapat diberikan untuk memberikan *feedback* baik bagi guru maupun peserta didik. Di lain pihak, asesmen informal dilaksanakan kapan saja saat proses pembelajaran sedang berlangsung. Tujuannya untuk mengetahui cara belajar setiap peserta didik dan mengidentifikasi kesulitan belajar peserta didik.

3. Penilaian Diri

Setelah memahami bagaimana melakukan penilaian terhadap hasil belajar peserta didik, baik secara formal dengan menggunakan tes dan secara informal melalui pengamatan yang sistematis terhadap proses belajar peserta didik di kelas, selanjutnya guru perlu menerapkan pula teknik penilaian yang lain, yaitu penilaian diri (*self-assessment*). Penilaian diri adalah sebuah proses penilaian dimana peserta didik menilai dirinya sendiri akan hasil

belajar yang telah dicapai. Akan tetapi, guru pada umumnya masih mengabaikan peran penilaian diri dalam rangka mengidentifikasi kelemahan dan kendala belajar peserta didik. Penilaian diri mempunyai peran yang sangat besar guna mendorong setiap peserta didik untuk lebih bertanggung-jawab terhadap proses belajar yang mestinya mereka lakukan. Melalui penilaian diri, peserta didik dapat didorong untuk melakukan refleksi terhadap proses dan hasil belajar yang telah dicapai dan kemudian hasil refleksi itu digunakan untuk mendorong dirinya untuk lebih banyak terlibat selama proses pembelajaran berlangsung. Selain itu, peserta didik dapat belajar dari kesalahan proses belajar yang telah dilakukan sebelumnya dan selanjutnya mereka diharapkan mampu memperbaikinya pada saat proses belajar selanjutnya. Dengan demikian, penilaian diri dapat dimanfaatkan oleh guru untuk mendapatkan informasi yang berharga berkaitan dengan harapan dan kebutuhan belajar setiap peserta didik; masalah dan kecemasan peserta didik selama melakukan proses belajar; bagaimana perasaan peserta didik terhadap tingkat kemajuan belajarnya; reaksi setiap peserta didik terhadap materi ajar dan metode mengajar yang digunakan oleh guru; dan juga pandangan peserta didik terhadap mata pelajaran yang diajarkan oleh guru. Singkatnya, sama seperti tes diagnostik, penilaian diri dapat digunakan oleh guru untuk mengungkap materi apa yang sudah dan belum dikuasai oleh setiap peserta didik.

Manfaat yang paling penting dari penilaian diri adalah jenis penilaian ini berdampak terhadap cara belajar peserta didik. Harris & McCann (1994) menyatakan bahwa penilaian diri adalah bagian integral dari proses belajar peserta didik. Untuk mempelajari sesuatu secara lebih efektif, peserta didik perlu menilai terlebih dahulu apa yang telah diketahui dan bagaimana mereka akan meningkatkan pengetahuannya itu. Konkritnya, untuk mempelajari materi ajar tertentu, peserta didik perlu diminta mengidentifikasi apa yang sudah dan belum dipahami berkaitan dengan materi ajar itu dan selanjutnya guru selama proses pembelajaran akan memberikan penekanan yang lebih banyak pada materi ajar yang belum banyak dikuasai oleh peserta didik.

Penilaian diri adalah bagian yang sangat penting dari keseluruhan proses pembelajaran. Bahkan para pakar penilaian sepakat bahwa ketika proses penilaian hanya dilakukan oleh guru secara sepihak tanpa melibatkan peserta didik, maka guru itu baru mampu menjalankan sebagian tugasnya dalam menilai hasil belajar peserta didik. Melalui penilaian diri, guru mata pelajaran dapat mendiagnosis berbagai masalah pembelajaran yang dialami oleh peserta didik dan sekaligus mengukur kemajuan belajar mereka. Informasi yang didapat dari penilaian diri dapat dimanfaatkan oleh guru untuk memperbaiki kualitas pembelajaran yang dilakukan. Selama ini

peserta didik hanya pasif dan sekedar menunggu informasi dari guru berkaitan dengan apakah hasil belajar mereka baik atau buruk atau bahkan gagal. Mereka tidak banyak diberi kesempatan untuk menilai sendiri kemajuan belajar mereka dan menilai mengapa mereka telah gagal dalam belajarnya. Guru semestinya mendorong setiap peserta didik untuk memikirkan apa yang mereka telah capai setelah melakukan proses belajarnya melalui penerapan penilaian diri. Peserta didik akan belajar dari kesalahan mereka apabila mereka diberi kesempatan untuk memikirkan kesalahan itu dan kemudian diminta memperbaikinya. Lebih lanjut Harris & McCann (1994) menjelaskan bahwa penilaian diri dapat membantu peserta didik untuk berpikir tentang kemajuan belajarnya serta masalah yang dihadapinya selama proses belajarnya dan selanjutnya mencari cara untuk mengubah, mengadaptasi dan memperbaiki cara belajar itu. Jika setiap peserta didik didorong untuk menilai kinerja dan hasil belajar mereka sendiri, diminta memikirkan kesalahan yang pernah mereka lakukan, dan kemudian diminta mencoba memperbaikinya, maka mereka akan mempunyai peluang besar untuk mampu memperbaiki kesalahannya dan hal ini lebih efektif dibandingkan jika penilaian dan koreksi hanya semata-mata dilakukan oleh guru.

Ilustrasi yang tampak seperti pada gambar 1.1 dan gambar 1.2 dapat dipahami bahwa ketika penilaian hanya dilakukan oleh guru, maka hasil penilaian itu kemungkinan hanya berdampak pada perbaikan kualitas metode, teknik, strategi dan prosedur mengajar yang dilakukan oleh guru saja. Sedangkan, peserta didik belum tentu melakukan perbaikan cara belajarnya. Kondisi berbeda jika proses penilaian itu mengintegrasikan kedua proses penilaian sekaligus, yaitu penilaian yang dilakukan oleh guru dan sekaligus juga penilaian yang dilakukan oleh peserta didik melalui penilaian diri. Berdasarkan hasil dari penilaian diri, peserta didik dimungkinkan akan memperbaiki proses belajarnya. Selain itu, peserta didik juga akan terbantu untuk memikirkan sejauhmana kemajuan belajar yang telah dicapai dan juga sekaligus belajar dari kesalahan yang telah dilakukan untuk memperbaiki cara belajar berikutnya.



Gambar 1.1
Dampak penilaian yang hanya dilakukan oleh guru



Gambar 1.2
Dampak penilaian oleh guru yang diintegrasikan dengan penilaian diri oleh peserta didik

Ada beberapa teknik penilaian diri yang dapat diimplementasikan oleh guru di kelas. Diantara beberapa teknik penilaian diri berikut ini sebenarnya sama seperti jika penilaian itu dilakukan oleh guru (*teacher*

assessment). Perbedaannya adalah seluruh proses dalam beberapa teknik penilaian ini secara keseluruhan dilakukan oleh peserta didik sendiri (*student self-assessment*). Beberapa teknik penilaian diri itu dijelaskan dalam bagian berikut (Harris & McCann, 1994).

a. Deskripsi diri

Melalui teknik ini, guru meminta setiap peserta didik untuk mendeskripsikan dirinya sendiri berkaitan dengan kemajuan belajar yang telah dicapai secara faktual. Setiap peserta didik diminta menuliskan materi apa yang telah dikuasai dan sekaligus yang belum dikuasai. Ada beberapa cara yang dapat dilakukan oleh guru dalam rangka mengimplementasikan teknik deskripsi diri ini, misalnya:

- 1) Meminta peserta didik untuk menulis profil dirinya, yaitu menulis laporan berkaitan dengan kemajuan belajarnya dan kemudian laporan itu diminta untuk diserahkan kepada guru untuk dianalisis;
- 2) Meminta peserta didik untuk menulis *diary* yang berisi berbagai catatan tentang apa yang telah peserta didik lakukan selama proses pembelajaran di kelas, apa yang telah mereka pelajari di kelas dan masalah-masalah apa yang mereka hadapi selama proses pembelajaran di kelas.
- 3) Melakukan sesi *counseling*, yaitu guru mengadakan sesi pembicaraan / tanya jawab dengan peserta didik secara individual untuk menanyakan sejauhmana tingkat kemajuan belajarnya.

Guru dapat memilih salah satu teknik tersebut sesuai dengan kondisi peserta didik dan juga kesiapan guru dalam menganalisis hasil penilaian diri yang telah dilakukan oleh peserta didik.

b. Memberikan angka (*rate*) berdasarkan standar tertentu.

Teknik ini menghendaki peserta didik untuk memberikan angka-angka menurut skala tertentu. Dalam hal ini peserta didik cukup melingkari atau mencontreng angka-angka yang menunjukkan gradasi kemampuan mereka. Ada beberapa cara yang dapat dilakukan oleh guru dalam rangka mengimplementasikan teknik ini, misalnya:

- 1) Skala angka (*rating scale*)

Peserta didik diminta memberikan penilaian akan kemampuannya sendiri dengan menggunakan skala angka dimana setiap angka itu memiliki deskriptor yang menggambarkan tingkat kemampuan / pengetahuan yang telah mereka capai, misalnya skala itu bersifat kontinum 1 – 5.

- 2) Memberikan peringkat (*general rating*)
Peserta didik diminta memberikan angka yang menggambarkan tingkat keterampilan dan pengetahuan yang dimiliki.
- 3) Menggambar grafik (*graphing*)
Peserta didik diminta menilai sendiri seberapa banyak materi-materi yang telah dipahami.
- 4) Menuliskan sifat diri
Peserta didik diminta menuliskan daftar kata sifat yang untuk mendeskripsikan sifat dan sikapnya sendiri, misalnya 'kerja keras' yang berarti bahwa peserta didik itu memiliki sifat pekerja keras; 'rileks' yang bermakna bahwa peserta didik itu memiliki sifat rileks atau santai dalam belajar; 'malas' yang dimaksudkan untuk menunjukkan bahwa peserta didik itu mempunyai sifat malas belajar.

c. Monitoring

Peserta didik diminta untuk memonitor sendiri terhadap kemajuan belajar yang telah dicapai setelah mengikuti proses pembelajaran selama waktu tertentu. Monitoring dapat dilakukan melalui beberapa aktivitas sebagai berikut.

- 1) Mengedit hasil kerjanya sendiri (*self-editing*)
Dalam mata pelajaran bahasa (bahasa Inggris, bahasa Indonesia, dll), peserta didik diminta mengamati hasil karangannya, kemudian diminta mengidentifikasi dan mengoreksi kesalahan-kesalahan yang mungkin ditemukan dalam karangan itu.
- 2) Memberikan kode hasil koreksi (*correction code*)
Peserta didik diminta untuk menggunakan berbagai kode koreksi ketika mereka diminta untuk mencermati kembali tugas / karya yang telah dibuat. Misalnya, kode rs : rumus salah; sk : salah ketik; pr : preposisi dan sebagainya.
- 3) Merekam (*taping*)
Dalam pembelajaran bahasa, peserta didik diminta untuk berpidato dan merekam pidato yang telah dilakukan itu untuk didengarkan ulang, selanjutnya peserta didik diminta mengidentifikasi kemungkinan kesalahan yang telah dilakukan.
- 4) Memeriksa kesalahan (*grading mistakes*)
Peserta didik diminta memeriksa atau mengidentifikasi kesalahan yang mungkin telah dilakukan selama berpartisipasi dalam proses pembelajaran. Aktivitas ini dapat dilakukan dengan tujuan meningkatkan keseriusan peserta selama mengikuti proses pembelajaran.

5) Menguji diri (*test yourself*)

Peserta didik diminta untuk mengerjakan tes pendek untuk menguji dirinya sendiri (*self-check*) dengan maksud untuk mengetahui sejauhmana mereka telah memahami dan menguasai materi pelajaran yang telah dipelajari.

d. Menggunakan kuesioner

Guru menyiapkan kuesioner untuk mengungkap perilaku atau kebiasaan belajar peserta didik dan kemudian mereka diminta menjawab pertanyaan-pertanyaan yang ada dalam kuesioner itu. Contoh pertanyaan dalam kuesioner:

1. Seberapa sering anda belajar matematika dalam satu minggu?
 - a. Sangat sering
 - b. Sering
 - c. Kadang-kadang
 - d. Tidak pernah
2. Bagaimana pendapat anda tentang materi ‘matriks’ dalam pelajaran matematika?
 - a. Sangat sulit
 - b. Sulit
 - c. Cukup mudah
 - d. Mudah

Hasil penilaian diri oleh peserta didik tersebut dapat dimanfaatkan oleh guru untuk memperbaiki metode, teknik, strategi dan prosedur pembelajaran agar metode, teknik, strategi dan prosedur pembelajaran itu sesuai dengan kondisi peserta didik di kelasnya. Peserta didik pun dapat menggunakan hasil penilaian diri itu sebagai input untuk memperbaiki strategi belajarnya sehingga mereka mampu menguasai materi ajar dengan baik. Dengan demikian, penilaian diri oleh peserta didik memiliki peran ganda, yaitu dimanfaatkan oleh guru untuk memperbaiki metode, teknik, strategi dan prosedur pembelajaran agar metode, teknik, strategi dan prosedur pembelajaran itu sesuai dengan kondisi dan kebutuhan peserta didik. Selain itu, hasil penilaian diri dapat digunakan oleh peserta didik untuk memperbaiki strategi dan motivasi belajarnya.

D. Evaluasi

Berbeda dengan asesmen atau penilaian, evaluasi merupakan aktivitas yang lebih luas dan melibatkan proses yang sistematis untuk menentukan suatu keputusan berkaitan dengan hasil belajar peserta didik berdasarkan kriteria dan bukti-bukti tertentu. Proses evaluasi melibatkan

banyak aktivitas yang harus dilakukan oleh seorang evaluator (guru) dan juga peserta didik, misalnya aktivitas observasi, presentasi, kerja kelompok, mengerjakan tugas, dan banyak hal berkaitan dengan berbagai bentuk tes formal. Dengan demikian, seorang guru yang akan mengevaluasi hasil belajar peserta didik, ada banyak aktivitas yang harus dilakukan oleh guru itu agar hasil evaluasi merupakan informasi yang akurat berkaitan dengan hasil belajar peserta didik itu. Lalu apa yang dimaksud dengan evaluasi pembelajaran? Evaluasi pembelajaran merupakan prosedur sistematis untuk menentukan apakah setiap peserta didik telah mampu memenuhi atau mencapai berbagai kriteria hasil belajar yang telah ditentukan. Dengan demikian, tujuan utama dari proses evaluasi pembelajaran adalah untuk menentukan keputusan apakah peserta didik telah mencapai berbagai kriteria hasil belajar yang telah ditentukan sesuai dengan kurikulum yang menjadi pijakan selama proses pembelajaran berlangsung.

Evaluasi pembelajaran merupakan sebuah proses yang melibatkan serangkaian tahapan, yaitu menentukan tujuan, mengklasifikasi tujuan, membatasi tujuan, menentukan indikator, dan membandingkan data hasil evaluasi dengan tujuan yang telah ditentukan. Akhir dari aktivitas evaluasi adalah sebuah keputusan berkaitan dengan kualitas hasil belajar peserta didik setelah belajar selama periode tertentu, misalnya setelah peserta didik belajar setelah satu semester. Dengan kata lain, evaluasi merupakan sebuah proses untuk mendapatkan, menganalisis, dan menginterpretasi informasi untuk menentukan sejauh mana peserta didik telah mencapai tujuan pembelajaran yang telah ditentukan.

Karena evaluasi pembelajaran dilakukan setelah peserta didik belajar selama periode tertentu, maka ruang lingkup evaluasi pembelajaran mencakup hal-hal yang lebih luas, yaitu program pembelajaran, proses pembelajaran, dan hasil pembelajaran. Ini berarti bahwa evaluasi pembelajaran tidak lagi hanya berkutat pada satu kompetensi dasar yang harus dikuasai oleh peserta didik saja, tetapi evaluasi menyangkut berbagai aspek pembelajaran.

E. Hubungan antara Tes, Pengukuran, Asesmen dan Evaluasi

Kebanyakan orang menggunakan istilah tes, pengukuran, asesmen dan evaluasi secara bergantian untuk mengacu hal yang sama, padahal keempat istilah tersebut sebenarnya mengacu pada aktivitas yang berbeda tetapi saling berkaitan. Tes, pengukuran, asesmen dan evaluasi merupakan aktivitas yang bersifat skuensial. *Testing* (pengujian) adalah aktivitas pertama yang harus dilakukan oleh guru agar dapat melakukan pengukuran terhadap hasil belajar peserta didik. Jadi seorang guru tidak mungkin mampu

melakukan pengukuran terhadap hasil belajar peserta didik tanpa menyiapkan dan memiliki instrumen atau alat ukur berupa tes terlebih dulu. Selanjutnya, data yang berupa angka-angka atau skor hasil pengukuran diinterpretasikan untuk memahami makna dari hasil pengukuran tersebut. Misalnya angka atau skor enam yang diperoleh seorang peserta didik dapat diinterpretasikan atau dimaknai bahwa peserta didik tersebut berkemampuan cukup dalam menguasai materi pelajaran. Peserta didik yang mendapat skor delapan dapat dimaknai bahwa peserta didik tersebut memiliki kemampuan baik dalam menguasai materi pelajaran. Interpretasi skor atau angka-angka hasil pengukuran tersebut sebenarnya merupakan aktivitas penilaian. Jadi, penilaian tidak dapat dilakukan oleh guru sebelum guru mengembangkan alat ukur (tes) yang akan digunakan untuk mengukur hasil belajar peserta didik. Penilaian dilakukan untuk cakupan materi yang terbatas, misalnya satu kompetensi dasar. Kemudian, sesuai dengan definisinya evaluasi merupakan aktivitas yang lebih luas dan melibatkan proses yang sistematis untuk menentukan suatu keputusan berkaitan dengan hasil belajar peserta didik berdasarkan kriteria dan bukti-bukti tertentu. Hasil evaluasi dapat digunakan oleh guru, misalnya untuk memutuskan apakah seorang peserta didik naik kelas atau tidak. Kegiatan evaluasi yang dilakukan oleh guru tidak mungkin dilakukan tanpa melibatkan aktivitas pengukuran dan penilaian. Dengan demikian, dapat dipahami bahwa tes, pengukuran, penilaian, dan evaluasi adalah rangkaian aktivitas yang saling berkaitan untuk menggambarkan tingkat keberhasilan peserta didik dalam menguasai pengetahuan dan keterampilan yang telah dipelajari.

Bab 2

Tenik Pengembangan Tes Prestasi

A. Tes Prestasi

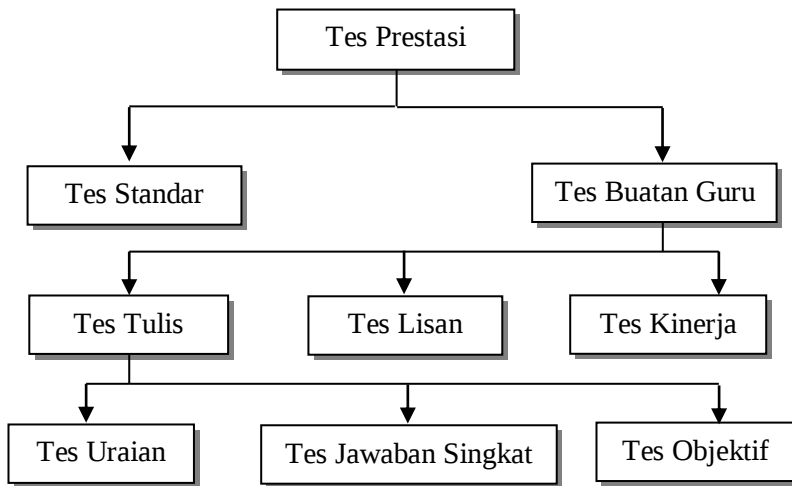
Tes merupakan serangkaian pertanyaan baik lisan maupun tulis yang harus direspon oleh peserta didik dan respon itu dapat dikategorikan menjadi benar dan salah. Selain itu, tes dapat pula dimaknai sebagai instrumen yang digunakan untuk mengukur kompetensi peserta didik dalam menguasai bahan ajar setelah mereka berpartisipasi dalam proses pembelajaran selama periode tertentu. Guru dapat menggunakan tes sebagai salah satu upaya untuk mengidentifikasi kemajuan hasil belajar (prestasi) peserta didiknya. Karena tes dimaksudkan untuk mengukur kemajuan belajar, maka sebuah tes harus mencakup materi pelajaran sesuai dengan tuntutan silabus yang menjadi dasar pijakan guru dan peserta didik dalam melakukan proses pembelajaran. Oleh karena itu, semua materi yang diteskan harus telah diajarkan oleh guru terlebih dahulu sebelum tes dilakukan. Tes yang dimaksudkan untuk mengukur kemajuan belajar atau prestasi peserta didik tidak boleh untuk menguji materi yang belum pernah diajarkan oleh guru.

Tes mempunyai fungsi sentral dalam proses pembelajaran karena berangkat dari tes, keputusan tentang kemajuan belajar peserta didik dapat dibuat. Guru dapat memanfaatkan tes sebagai upaya untuk mengetahui sejauhmana tujuan pembelajaran telah dicapai oleh sebagian besar peserta didik. Dengan demikian, guru memberikan tes kepada peserta didik bukan tanpa alasan. Ada beberapa alasan mengapa guru perlu memberikan tes kepada peserta didik, diantara tujuan tes itu adalah untuk:

1. memberikan bahan catatan dan pertimbangan bagi guru untuk memberikan nilai;
2. memberikan pengalaman belajar bagi peserta didik;

3. memotivasi peserta didik untuk lebih giat belajar;
4. memberikan dasar dan panduan bagi peserta didik untuk studi lanjut;
5. menilai seberapa baik peserta didik telah mencapai tujuan pembelajaran yang telah ditentukan; dan
6. memberikan kesempatan kepada guru untuk memperkuat pencapaian tujuan pembelajaran dan memfokuskan pada materi apa yang penting bagi peserta didik untuk dikuasai.

Tes yang digunakan oleh guru untuk mengukur kemajuan belajar peserta didik di kelas dikategorikan sebagai tes prestasi (*achievement test*). Tes prestasi dimaksudkan untuk menunjukkan seberapa banyak materi yang telah dipelajari dan dikuasai oleh peserta didik sesuai dengan tujuan pembelajaran yang telah ditentukan. Hal ini berbeda dengan tes kemahiran (*proficiency test*) yang dimaksudkan untuk mengukur kompetensi global yang dimiliki oleh peserta didik. Dalam bidang bahasa, misalnya, tes kemahiran dimaksudkan untuk mengukur kemampuan berbahasa peserta didik secara global yang meliputi kemampuan untuk menguasai tata bahasa, kosa kata, serta keterampilan mendengarkan dan membaca. Selain itu, bahan atau materi yang harus diujikan dengan tes jenis ini tidak perlu diajarkan terlebih dahulu. Jadi bahan ajar atau materi yang diujikan tidak terdapat dalam kurikulum maupun silabus pembelajaran. Namun demikian, tes ini biasanya termasuk dalam tes yang sudah terstandar. Contoh tes ini misalnya TOEFL (*test of English as a foreign language*), TOEIC (*test of English as international communication*), IELTS (*international English language testing system*) dan sebagainya. Karena tes yang digunakan oleh guru di kelas adalah tes prestasi, maka materi atau bahan yang harus diteskan kepada setiap peserta didik harus sesuai dengan materi yang tercantum dalam silabus dan telah diajarkan terlebih dahulu kepada peserta didik. Tes prestasi belajar terdiri dari beberapa jenis tes, diantaranya adalah tes standar (*standardized test*), tes buatan guru (*teacher-made test*), tes tulis, tes lisan, tes kinerja, tes uraian, tes jawaban singkat, ojektif tes, dan sebagainya. Secara hirarkis, Arifin (2010) menjelaskan bahwa hubungan dari berbagai jenis tes tersebut digambarkan dalam diagram 2.1.



Gambar 2.1 Berbagai macam tes prestasi

Tes prestasi sebagaimana telah dijelaskan sebelumnya adalah sebuah jenis tes yang dimaksudkan untuk mengukur kemajuan belajar atau prestasi peserta didik. Secara spesifik tes prestasi didesain untuk mengukur tingkat keterampilan, capaian belajar dan pengetahuan peserta didik pada pokok bahasan tertentu dari suatu mata pelajaran. Beberapa contoh jenis tes ini, misalnya tes matematika yang isi atau materi tesnya meliputi materi ajar pada bab tertentu dari buku paket matematika; tes kemampuan berbicara dalam bahasa Inggris; tes untuk mendemonstrasikan kemampuan bela diri dalam mata pelajaran olah raga dan sebagainya. Masing-masing tes ini didesain untuk mengukur sejauhmana peserta didik telah menguasai bahan ajar yang telah diajarkan oleh guru dalam jangka waktu tertentu. Tes prestasi bukan dimaksudkan untuk mengukur kemahiran peserta didik pada bidang tertentu, tetapi tes prestasi didesain untuk mengetahui penguasaan peserta didik terhadap bahan ajar tertentu sesuai tuntutan silabus dan kurikulum.

Tes standar adalah berbagai macam bentuk tes yang menghendaki semua *testee* (peserta didik) untuk menjawab berbagai pertanyaan yang sama dengan cara yang sama pula. Jadi, tes standar adalah suatu tes dimana semua peserta didik menjawab pertanyaan-pertanyaan yang sama dari sebagian besar pertanyaan dikerjakan dengan mengikuti petunjuk yang sama dan dalam batasan waktu yang sama pula. Selain itu, hasil tes itu akan diberi

skor dengan cara yang ‘standar’ yang menggambarkan kemampuan relatif peserta didik, baik secara individu atau secara kelompok. Tes standar biasanya disusun oleh satu tim ahli, atau disusun oleh lembaga yang khusus menyelenggarakan tes secara profesional. Tes ini dapat digunakan dalam waktu yang relatif lama dan dapat diterapkan pada beberapa obyek mencakup wilayah yang luas. Istilah tes standar dapat juga dimaknai sebagai sebuah tes yang telah diketahui berkualitas baik dilihat dari aspek validitas, daya beda, tingkat kesulitan, dan reliabilitas. Sebelum digunakan untuk mengukur tingkat capaian belajar, keterampilan dan pengetahuan peserta didik, tes ini telah terlebih dahulu diujicobakan secara repetitif untuk mengetahui kualitas tes dilihat dari berbagai aspek tersebut. Bentuk tes standar yang paling umum adalah tes pilihan ganda, walaupun tidak menutup kemungkinan dapat juga berupa pertanyaan benar – salah, jawaban singkat, *essay*, atau campuran dari berbagai bentuk tes tersebut.

Tes buatan guru adalah sebuah tes yang didesain oleh guru sendiri berdasarkan silabus dan kurikulum mata pelajaran tertentu. Dengan kata lain, tes buatan guru adalah tes yang dibuat oleh guru untuk mengukur kemajuan belajar peserta didik di kelas dan sekolahnya sendiri. Tes ini dapat digunakan oleh guru untuk ulangan harian, tes formatif, dan tes sumatif. Berbeda dengan tes standar, kualitas tes buatan guru belum menjadi jaminan. Jika tes standar dapat digunakan dalam waktu yang relatif lama, tes buatan guru sering bersifat temporer sehingga pada kesempatan berbeda belum tentu tes itu masih dapat digunakan lagi sebagai akibat dari perbedaan konteks dan karakteristik peserta didik. Selain untuk mengukur tingkat capaian belajar, keterampilan dan pengetahuan peserta didik, tes ini juga dimanfaatkan oleh guru untuk melakukan refleksi terhadap kualitas metode dan strategi pembelajaran yang digunakan oleh guru itu. Ketika metode dan strategi pembelajaran itu diidentifikasi tidak berjalan dengan baik dan belum mampu memfasilitasi proses belajar peserta didik dengan baik, guru kemudian perlu memodifikasi atau bahkan mengganti metode dan strategi itu dengan yang sama sekali baru. Dengan demikian ada perbedaan prinsip antara tes standar dan tes buatan guru seperti dapat diamati pada tabel 2.1.

Tabel 2.1 Perbedaan tes standard dan tes buatan guru

Tes Buatan Guru	Tes Standar
Dikembangkan oleh guru sendiri sebagai ulangan harian, tes formatif dan sumatif	Dikembangkan oleh ahli atau lembaga pengujian tertentu
Saat pelaksanaan tes tidak ada prosedur pelaksanaan yang seragam antara satu sekolah dengan sekolah yang lain	Saat pelaksanaan tes mengikuti prosedur pelaksanaan yang standar
Cakupan materi ditentukan sendiri oleh guru sesuai dengan materi yang telah diajarkan di kelas	Cakupan materi ditentukan oleh otoritas tertentu, misalnya Kemendikbud
Proses penskoran sering dipengaruhi oleh subjektivitas guru	Proses penskoran lebih objektif
Bertujuan untuk mengukur capaian belajar sesuai dengan tujuan pembelajaran yang telah ditentukan dan capaian itu dibandingkan dengan peserta didik dalam satu kelas	Bertujuan untuk mengukur tujuan yang lebih luas dan digunakan untuk membandingkan capaian belajar dalam konteks antarkelas, antar sekolah, dan bahkan tingkat nasional

Tes tulis adalah jenis tes yang menghendaki peserta didik untuk menjawab semua pertanyaan dengan menggunakan pensil atau bolpoin (*paper and pencil test*). Hal ini berbeda dengan tes lisan yang menghendaki peserta didik menjawab pertanyaan secara lisan. Perbedaan kedua tes ini hanya lebih terletak pada cara peserta didik dalam menjawab pertanyaan yang diberikan oleh guru atau pengujinya. Perbedaan lain, tes tulis memberikan lebih luas kesempatan kepada peserta didik untuk berpikir

dalam menjawab pertanyaan. Sedangkan tes lisan sering meghendaki peserta didik untuk berpikir secara spontan dan kondisi ini sering mengakibatkan peserta didik menjadi kurang konsentrasi apabila mereka tidak memiliki pemahaman konsep yang baik terhadap materi yang diujikan, mereka biasanya menjadi gugup dalam menjawab pertanyaan yang pada akhirnya jawaban sering tidak akurat.

Tes kinerja (*performance test*) adalah tes yang dimaksudkan untuk menguji kinerja seseorang (peserta didik) melalui cara-cara yang sistematis. Cara sistematis ini mengandung makna bahwa prosedur tes dilaksanakan sesuai dengan kriteria dan langkah-langkah yang telah ditentukan dan setiap peserta didik harus menunjukkan kinerjanya mengikuti prosedur yang seragam antara satu peserta didik dengan peserta didik yang lain atau antara satu kelompok dengan kelompok lain. Respon yang diberikan oleh peserta didik dapat berupa perilaku, tindakan, dan perbuatan. Dengan demikian dapat dipahami bahwa tes kinerja adalah suatu bentuk tes dimana peserta didik diminta untuk melakukan kegiatan khusus di bawah pengawasan guru atau penguji yang akan mengobservasi penampilannya dan kemudian membuat keputusan tentang kualitas hasil belajar yang didemonstrasikan oleh peserta didik sesuai dengan apa yang diperintahkan dan ditanyakan oleh guru. Ada beberapa alasan mengapa guru perlu memberikan tes kinerja kepada peserta didik, yaitu:

1. untuk mengetahui kinerja peserta didik dalam mempraktikkan pengetahuannya;
2. untuk mengetahui kualitas, kuantitas, dan luaran kinerja peserta didik; dan
3. untuk mengetahui kemampuan bekerja-sama antara peserta didik.

Contoh tes kinerja ini misalnya, tes berbicara dalam bahasa Inggris, tes melakukan lempar lembing; tes memasang karburator sepeda motor; tes membaca puisi; tes membuat bahan eksperimen di laboratorium fisika, kimia dan biologi; tes membuat pola dan menjahit busana; dan sebagainya.

Tes uraian adalah jenis tes yang meghendaki peserta didik untuk menyusun jawaban dari berbagai pertanyaan yang diberikan dan biasanya panjang jawaban itu dapat mencapai beberapa paragraf. Pertanyaan-pertanyaan dalam tes uraian biasanya digunakan untuk mengukur proses dan kemampuan berpikir tingkat tinggi (*higher-order thinking skills*) peserta didik pada domain analisis, sintesis, evaluasi, dan kreativitas. Tes ini

menghendaki peserta didik untuk menjawab berbagai butir tes secara bebas sesuai tingkat kemampuan berpikir mereka. Peserta didik bebas memilih, menghubungkan, dan mempresentasikan idea atau gagasannya dalam rangkaian kata yang luas dan mendalam. Tes ini pula dapat dimanfaatkan oleh guru untuk mengukur kemampuan peserta didik untuk mengingat bahan ajar yang telah mereka pelajari, kemudian mengorganisir dan mengintegrasikan idea mereka menjadi suatu pemahaman dan pengetahuan yang lebih bermakna.

Bentuk tes uraian dapat berupa tes uraian terbatas (*restricted response test*) dan tes uraian bebas (*extended response test*). Tes uraian terbatas menghendaki peserta didik untuk membatasi jawaban yang diberikan sesuai dengan topik atau pokok bahasan yang telah dipelajarinya. Tes uraian terbatas sangat berguna untuk mengukur hasil belajar peserta didik yang menghendaki mereka untuk menginterpretasikan dan mengaplikasikan berbagai pengalaman dan pengetahuan mereka pada konteks yang lebih spesifik. Tes uraian terbatas memiliki beberapa kelebihan, yaitu:

1. jawaban yang diberikan oleh peserta didik lebih terstruktur;
2. tes ini hanya mengukur kemajuan belajar peserta didik pada konteks yang terbatas;
3. guru lebih mudah untuk mengoreksi jawaban yang diberikan oleh peserta didik; dan
4. proses penskoran lebih objektif.

Sedangkan keterbatasan dari tes uraian terbatas adalah tes ini hanya menghendaki jawaban yang terbatas dan berkaitan dengan lingkup pokok bahasan yang pernah dipelajari oleh peserta didik sehingga peserta didik tidak leluasa untuk mendemostrasikan kemampuan berpikir logis secara lebih luas dan mendalam.

Di lain pihak, tes uraian bebas menghendaki peserta didik untuk memberikan jawaban yang luas dan mengarahkan mereka untuk memilih berbagai informasi yang relevan guna mendukung dan memperkuat jawaban yang diberikan, mengorganisasikan jawaban itu sesuai dengan kemampuan berpikir logis yang mereka miliki, dan kemudian mengintegrasikan dan mengevaluasi gagasan atau jawaban yang mereka miliki menjadi informasi yang lebih bermakna. Peserta didik dapat pula mengutip suatu teori yang relevan untuk memperkuat gagasan yang mereka berikan. Kelebihan dari tes

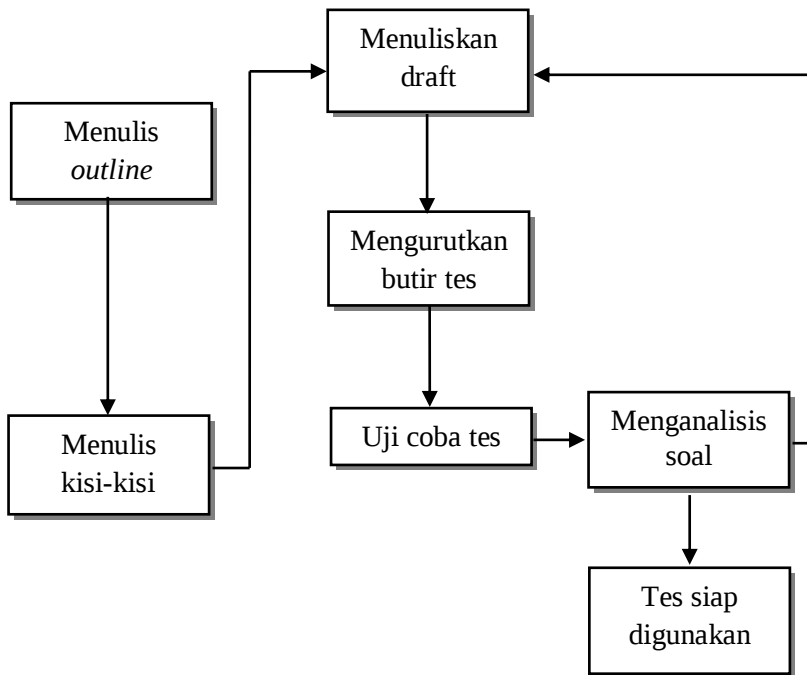
ini adalah tes ini sangat bermanfaat untuk mengukur hasil belajar peserta didik pada level berpikir tinggi. Tes ini juga mampu membedakan secara signifikan kemampuan setiap peserta didik secara individu dalam bersikap, berperilaku dan berkreativitas. Sedangkan kelemahan dari tes ini adalah tes ini tidak begitu efektif untuk mengukur pengetahuan faktual karena tes ini biasanya menghendaki peserta didik untuk memberikan respon secara detail. Selain itu, proses penskorannya juga relatif sulit dan sering pula tidak menghasilkan skor yang reliabel.

Tes objektif adalah sebuah tes yang menghendaki respon atau jawaban dari peserta didik dan respon itu dapat dinilai secara objektif. Tes objektif dirancang untuk menghindari unsur subjektif dari para penilai dengan menghadirkan sejumlah pertanyaan faktual yang harus dijawab oleh peserta didik dalam beberapa kata saja atau dengan memberikan tanda centang. Tidak ada penilaian subjektif yang dilakukan oleh pemeriksa jawaban tes ketika menilai jawaban dalam tes objektif. Berbeda dengan tes subjektif yang penilaiannya sangat dipengaruhi oleh pendapat atau subjektivitas penilainya. Bahkan dalam tes objektif, pemeriksaan hasil tes dapat dilakukan oleh siapapun tanpa pengetahuan yang memadai tentang materi yang diujikan. Di bidang pendidikan, tes objektif sangat populer karena tes ini dapat mencakup materi yang luas, mudah dalam pengoreksian, dan dapat memberikan hasil yang terukur dan konkrit. Ada beberapa variasi tes objektif yang biasa digunakan oleh guru untuk mengukur hasil belajar peserta didik, misalnya tes bentuk soal pilihan ganda, soal menjodohkan, soal jawaban singkat, dan soal benar salah.

B. Tahapan Pengembangan Tes Prestasi

Tes yang digunakan sebagai instrumen untuk mengukur hasil belajar peserta didik harus memiliki kualitas yang baik sehingga informasi tentang kemajuan belajar peserta didik yang diperoleh melalui penggunaan tes itu merupakan informasi yang akurat dan tidak bias. Apabila tes yang digunakan untuk mengukur hasil belajar peserta didik bukan merupakan tes yang baik, maka informasi atau data yang diperoleh pun tidak akan akurat. Agar diperoleh tes yang baik, ada beberapa tahapan yang harus diperhatikan dan diikuti oleh setiap penulis tes agar tes yang dikembangkan memenuhi kriteria baik. Secara rinci, tahapan-tahapan dalam mengembangkan tes

sebagai instrumen untuk mengukur hasil belajar peserta didik dapat diamati pada gambar 2.2 seperti yang dilukiskan berikut ini.



Gambar 2.2 Tahapan-tahapan pengembangan tes

Seperti tampak pada gambar 2.2, proses pengembangan tes sebagai instrumen untuk mengukur hasil belajar peserta didik perlu dilakukan melalui beberapa tahapan yang bersifat hirarkis dan sistematis. Tahapan pertama sangat mempengaruhi pada tahapan kedua, tahapan kedua juga sangat mempengaruhi tahapan ketiga dan seterusnya. Dengan demikian, untuk mendapatkan tes yang berkualitas, pengembang tes tidak boleh hanya sekedar menulis butir-butir soal tanpa mengikuti tahapan-tahapan hirarkis dan sistematis seperti yang disarankan dalam gambar tersebut. Lalu, apa yang harus dilakukan oleh pengembang (penulis) soal pada setiap tahapan.

Berikut diuraikan satu-persatu berkaitan dengan apa yang harus dilakukan oleh penulis soal pada setiap tahapan tersebut.

1. Menulis *outline*

Membuat *outline* merupakan salah satu tahapan penting yang harus dilalui oleh penulis soal. Semua butir soal harus selalu berkaitan dengan *outline* yang telah terlebih dahulu ditentukan oleh penulis soal. Tentunya *outline* ini juga tidak dapat dipisahkan dari proses pembelajaran yang telah dilakukan sebelumnya. Pada tahapan ini, ada paling tidak dua sub-tahapan yang harus dilakukan oleh penulis soal, yaitu menganalisis tujuan pembelajaran dan memilih materi yang perlu dicakup dalam tes. Menganalisis tujuan pembelajaran merupakan hal yang sangat penting untuk dilakukan oleh penulis soal karena pada dasarnya sebuah tes dimaksudkan untuk mengukur sejauhmana tujuan pembelajaran itu telah dicapai oleh peserta didik. Sedangkan pemilihan materi yang perlu dicakup dalam tes akan mengarahkan penulis soal dapat mengidentifikasi dan menentukan bagian-bagian materi mana yang perlu diteskan karena ada kemungkinan sebuah tes tidak dapat mencakup semua materi yang telah diajarkan sebelumnya.

a. Menganalisis tujuan pembelajaran

Tujuan pembelajaran merupakan pernyataan yang menjelaskan tentang apa yang harus dicapai oleh peserta didik setelah proses pembelajaran berakhir. Penjelasan senada disampaikan oleh Arreola (1998) yang menyatakan bahwa tujuan pembelajaran merupakan sebuah pernyataan tentang apa yang akan mampu dilakukan oleh peserta didik ketika mereka telah mengikuti proses pembelajaran selama periode waktu tertentu. Selain itu, tujuan pembelajaran memuat penjelasan rinci tentang apa yang harus dikuasai dan dilakukan oleh setiap peserta didik setelah proses pembelajaran berakhir. Tujuan pembelajaran merupakan strategi sentral dalam proses pembelajaran berbasis kompetensi dan mengacu pada sistem pembelajaran itu sendiri dan sistem penilaian. Arreola (1998) lebih lanjut menyatakan bahwa tujuan pembelajaran paling tidak memuat tiga komponen utama, yaitu:

- 1) Deskripsi tentang apa yang akan dapat dilakukan oleh peserta didik setelah berpartisipasi dalam proses pembelajaran.

- 2) Kondisi dimana peserta didik akan menerapkan kemampuannya setelah berpartisipasi dalam proses pembelajaran.
- 3) Kriteria untuk mengevaluasi performans peserta didik setelah berpartisipasi dalam proses pembelajaran.

Analisis tentang tujuan pembelajaran menjadi aktivitas yang sangat penting untuk dilakukan oleh penulis atau pengembang soal. Ada beberapa alasan mengapa analisis tujuan pembelajaran ini sangat perlu untuk dilakukan sebelum mengembangkan atau menulis butir-butir soal, diantaranya adalah:

- 1) Hasil analisis tujuan pembelajaran dapat digunakan sebagai dasar untuk memilih materi esensial yang perlu dicakup dalam soal.
- 2) Sebagai dasar untuk mengembangkan strategi pembelajaran yang sesuai dengan kondisi dan kebutuhan peserta didik.
- 3) Sebagai dasar untuk mengembangkan dan memilih bahan pelajaran.
- 4) Sebagai dasar untuk menyusun butir-butir soal yang akan digunakan untuk mengukur dan mengevaluasi hasil belajar peserta didik.

Mengingat pentingnya tujuan pembelajaran sebagai dasar untuk mengembangkan butir-butir soal, maka guru harus menentukan dan menyusun tujuan pembelajaran secara seksama dan terukur. Ada beberapa hal yang harus diperhatikan oleh guru ketika akan menentukan tujuan pembelajaran, yaitu:

- 1) Tujuan pembelajaran harus fokus pada performans peserta didik, bukan pada performans guru.
- 2) Tujuan pembelajaran harus fokus pada produk dari hasil pembelajaran, bukan pada proses pembelajaran itu sendiri.
- 3) Tujuan pembelajaran harus fokus pada perilaku akhir yang harus dilakukan peserta didik setelah proses pembelajaran berakhir, bukan fokus pada isi pelajaran.

Berikut adalah beberapa contoh rumusan tujuan pembelajaran untuk beberapa mata pelajaran:

- 1) Setelah berakhirnya kegiatan pembelajaran ini, peserta didik kelas X mampu menjawab soal-soal matematika berkaitan dengan persamaan linier secara tepat.
- 2) Peserta didik kelas XI dapat menjelaskan hukum termodinamika dan aplikasinya secara benar.
- 3) Setelah berakhirnya pembelajaran, peserta didik dapat menjelaskan proses respirasi pada manusia secara sistematis dan tepat.

Tujuan pembelajaran yang baik harus dapat diamati (*observable*) dan terukur (*measurable*) dengan baik. Dapat diamati mengandung makna bahwa tujuan pembelajaran merupakan aktivitas atau perilaku konkrit, nyata, bersifat empiris dan tidak bersifat abstrak. Sedangkan, dapat diukur berarti tujuan pembelajaran itu harus mengarah perilaku yang dapat diukur untuk menunjukkan ketercapaian kompetensi dasar tertentu yang menjadi acuan dalam melakukan penilaian pada mata pelajaran itu. Agar sebuah tujuan pembelajaran dapat memenuhi kriteria terukur dan teramati, maka sebuah tujuan pembelajaran harus dirumuskan dengan mengintegrasikan empat aspek, yaitu *audience* (A), *behavior* (B), *condition* (C), dan *degree* (D). Keempat aspek tersebut memiliki karakteristik sendiri-sendiri yang harus diperhatikan oleh guru ketika mereka akan mengembangkan sebuah tes prestasi.

1) Audience

Audience mengacu pada peserta didik, mahasiswa, *testee*, petatar dan sasaran didik lainnya yang menjadi pelaku atau sasaran dalam keseluruhan proses pembelajaran. Penulis tes harus menjelaskan secara eksplisit siapa yang menjadi objek tes dan akan diukur tingkat kompetensinya. Siapa yang menjadi objek tes harus dijelaskan secara spesifik, misalnya peserta didik kelas berapa, semester berapa, kelas berapa dan sebagainya. Batasan ini sangat penting agar setiap *testee* menyadari bahwa bahan ajar yang diujikan itu sesuai atau tidak sesuai dengan tingkat kompetensinya.

2) Behavior

Behavior mengacu pada perilaku dan kompetensi yang harus dimiliki oleh *testee* setelah mengikuti proses pembelajaran selama periode waktu tertentu. Perilaku ini paling tidak mengandung dua unsur penting, yaitu

kata kerja operasional dan objek yang harus dikuasai oleh *testee*. Kata kerja mengacu pada bagaimana *testee* mampu mendemonstrasikan suatu pengetahuan, sikap dan keterampilan tertentu, misalnya mendeskripsikan, mengidentifikasi, membandingkan, mengamati, mempresentasikan, menjelaskan, mendiskusikan, mengidentifikasi fakta dan sebagainya. Sedangkan objek mengacu pada apa yang mampu didemonstrasikan oleh *testee*, misalnya persamaan linier (mata pelajaran matematika), hukum termodinamika (mata pelajaran fisika), *conditional sentences*, *recount text* (mata pelajaran bahasa Inggris), sistem respirasi (mata pelajaran biologi) dan sebagainya.

3) Condition

Condition adalah suatu kondisi yang menjadi prasyarat pada saat *testee* diuji untuk mengetahui sejauhmana mereka telah menguasai materi ajar yang telah dipelajari. Suatu kondisi tes yang dirumuskan secara jelas akan menjadi petunjuk bagaimana *testee* telah mencapai perilaku dan kompetensi tertentu yang diharapkan muncul setelah proses pembelajaran. Oleh karena itu, seorang pengembang tes, walaupun tidak mutlak, perlu merumuskan kondisi bagaimana seorang *testee* telah mampu mencapai suatu perilaku dan kompetensi tertentu agar mudah diamati dan diukur.

4) Degree

Degree mengandung makna derajat atau tingkatan keberhasilan *testee* untuk mencapai suatu perilaku dan kompetensi tertentu yang telah ditargetkan oleh penguji untuk dikuasai oleh *testee* sebagai hasil dari proses belajar mereka. Target perilaku dan kompetensi yang telah ditargetkan oleh penguji itu, misalnya melakukan tanpa salah, menunjukkan dengan benar, menyelesaikan dengan tepat waktu, berenang sejauh 100 meter dalam waktu tertentu, atau ukuran tingkat keberhasilan lainnya.

Keempat aspek tersebut merupakan hal yang penting untuk diperhatikan ketika penulis soal ingin merumuskan tujuan tes agar tujuan tersebut dapat diukur dan diamati secara akurat dan tepat. Sebagai gambaran dari implementasi keempat aspek tersebut dalam sebuah tujuan tes, dapat diamati contoh berikut.

Setelah berakhirnya kegiatan pembelajaran ini, peserta didik kelas X mampu menjawab soal-soal matematika berkaitan dengan persamaan linier secara tepat.

Audience : Peserta didik kelas X

Behavior : mampu menjawab soal-soal matematika
berkaitan dengan persamaan linier

Condition : Setelah berakhirnya kegiatan pembelajaran ini

Degree : secara tepat.

Rumusan tujuan pembelajaran tersebut tentu akan memudahkan para penulis soal untuk melakukan pengamatan dan pengukuran terhadap perilaku dan kompetensi yang diharapkan muncul dan kuasai oleh testee.

b. Memilih Materi

Guru mungkin saja telah mengajarkan materi atau bahan ajar yang cakupannya sangat luas dalam suatu periode proses pembelajaran tertentu. Misalnya dalam satu semester, guru telah mengajarkan berbagai materi yang dijabarkan dalam beberapa kompetensi dasar yang harus dipahami dan dikuasai oleh peserta didik. Karena luasnya cakupan materi ajar itu, guru seringkali tidak memungkinkan untuk mengujikan semua materi ajar itu dalam satu kali kesempatan tes. Oleh karena itu, dalam suatu kesempatan tes, guru harus memilih atau melakukan sampel materi mana yang harus diujikan sebagai akibat dari luasnya cakupan materi itu. Ada beberapa pertimbangan yang harus diperhatikan oleh guru sebagai penulis soal dalam menentukan materi mana yang harus dipilih untuk bahan ujian. Beberapa pertimbangan itu menyangkut empat hal, yaitu urgensi, kontinuitas, relevansi, dan keterpakaian.

1) Urgensi

Urgensi berarti kompetensi dasar dan materi itu secara teoritis mutlak harus dikuasai oleh peserta didik. Dengan kata lain, urgensi mengacu pada materi ajar yang dikategorikan materi esensial, inti, penting atau materi yang mendasar / hakiki dan perlu sekali untuk dikuasai peserta didik. Permasalahan yang kemudian sering muncul adalah guru yang bertugas mengembangkan soal ujian kesulitan dalam menentukan materi esensial yang harus dipilih menjadi bahan ujian. Bahkan ada yang

berasumsi bahwa semua materi yang telah diajarkan adalah penting dan tidak ada materi yang tidak penting. Oleh karena itu, pengembang soal harus banyak memanfaatkan pengalamannya agar mampu memilih materi mana yang harus dipilih sebagai bahan ujian.

2) Kontinuitas

Kontinuitas mengacu pada kompetensi dasar dan materi lanjutan yang merupakan pendalaman materi sebelumnya. Adakalanya suatu bahan ajar menjadi prasyarat untuk dapat memahami bahan ajar yang lain. Misalnya, dalam mata pelajaran fisika, agar peserta didik mampu menguasai hukum newton, maka peserta didik itu harus terlebih dulu memahami vektor. Oleh karena itu, vektor adalah materi yang harus diujikan terlebih dahulu sebelum peserta didik belajar tentang hukum newton. Dalam mata pelajaran bahasa, fonologi merupakan materi yang harus dipilih dan diujikan terlebih dulu sebelum peserta didik mempelajari morfologi karena fonologi adalah prasyarat yang harus dikuasai oleh peserta didik sebelum mereka belajar morfologi dan demikian juga untuk mata pelajaran yang lain.

3) Relevansi

Relevansi berarti bahwa materi itu sangat diperlukan untuk mempelajari bidang studi atau mata pelajaran lain. Sebagai contoh materi dalam mata pelajaran fisika tentang 'kinematika gerak' sangat dipengaruhi oleh kemampuan peserta didik dalam menguasai 'persamaan integral dan diferensial' dalam mata pelajaran matematika. Oleh karena itu, kedua materi tersebut mempunyai relevansi dalam dua mata pelajaran sekaligus, yaitu matematika dan fisika dan kedua materi tersebut harus dipilih menjadi bahan ujian.

4) Keterpakaian

Keterpakaian (*practicality*) berarti materi itu memiliki nilai terapan yang tinggi dalam kehidupan sehari-hari. Materi ajar yang memiliki tingkat terapan yang tinggi dalam kehidupan nyata peserta didik harus dipilih menjadi bahan ujian. Hal ini penting karena penguasaan materi yang memiliki tingkat praktikalitas yang tinggi akan dapat digunakan oleh

peserta didik untuk mengatasi berbagai permasalahan yang dihadapi oleh peserta didik itu dalam kehidupan sehari-hari.

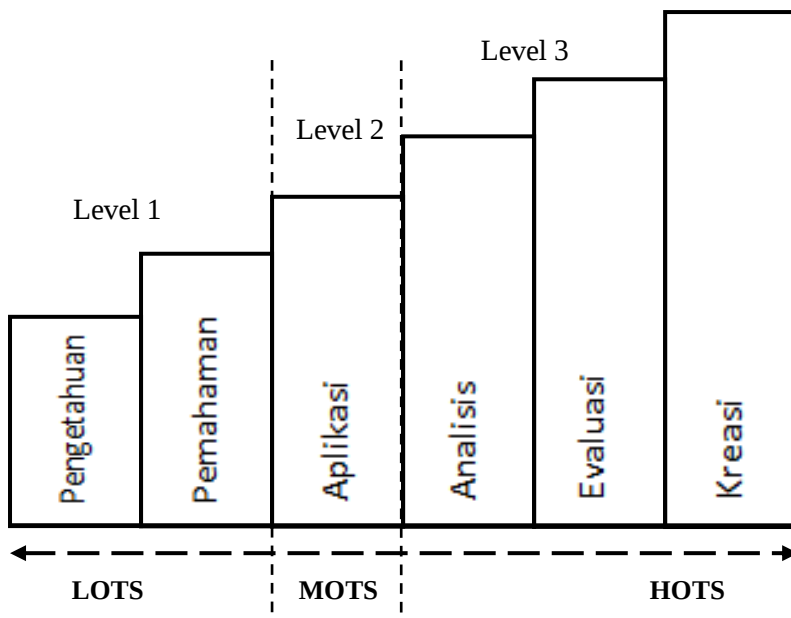
Keempat hal tersebut harus menjadi pertimbangan utama bagi guru sebagai penulis soal ketika guru harus memilih materi yang harus diujikan. Penulis soal tidak boleh sembarangan dalam memilih materi ajar yang harus diujikan agar soal-soal yang diujikan itu selanjutnya memiliki tingkat validitas, terutama validitas isi, yang tinggi. Validitas adalah komponen yang sangat penting dan merepresentasikan kualitas soal yang telah dikembangkan.

2. Menulis Kisi-kisi Soal

Kisi-kisi adalah sebuah format yang berisi kriteria butir-butir soal yang harus disusun oleh pengembang soal. Kisi-kisi juga dapat diartikan sebagai bagan dua arah yang mengkaitkan hasil pembelajaran yang diharapkan dan materi pembelajaran. Sebuah kisi-kisi memuat kompetensi dasar, cakupan materi, indikator pencapaian hasil belajar, jenis soal, nomor soal dan tingkat kesulitan butir soal. Kompetensi dasar adalah sebuah pernyataan yang memuat pengetahuan, keterampilan dan sikap minimal yang harus dikuasai oleh peserta didik setelah mengikuti proses pembelajaran. Kompetensi dasar merupakan penjabaran dari standar kompetensi pada suatu mata pelajaran tertentu. Sedangkan indikator pencapaian hasil belajar merupakan pernyataan yang memuat petunjuk pencapaian kompetensi dasar yang ditandai oleh perubahan perilaku yang dapat diamati dan diukur dan mencakup sikap, pengetahuan, dan keterampilan yang harus dikuasai peserta didik.

Komponen lain yang harus ada dalam kisi-kisi adalah level kognitif. Level kognitif merupakan tingkat kemampuan yang harus dimiliki oleh setiap peserta didik yang dapat dijabarkan menjadi tiga level yang berbeda, yaitu level satu (L1), level dua (L2) dan level tiga (L3). Level satu mengacu pada kemampuan berpikir tingkat rendah yang meliputi pengetahuan dan pemahaman. Level dua merupakan tingkat kemampuan medium yang meliputi kemampuan untuk menerapkan sesuatu. Sedangkan kemampuan level tiga adalah tingkat kemampuan tingkat tinggi, yaitu kemampuan peserta didik untuk melakukan penalaran. Kemampuan ini meliputi kemampuan untuk melakukan kognitif analisis, sintesis, evaluasi dan kreasi.

Penentuan ketiga level tersebut dibuat mengacu pada taksonomi Bloom. Penulis soal perlu memperhatikan kata kerja operasional yang menjadi karakteristik unik untuk setiap domain (kognitif, afektif, dan psikomotorik). Menurut Anderson & Krathwohl (2001), ketiga level tersebut sebenarnya merepresentasikan kemampuan berpikir peserta didik dari kemampuan berpikir tingkat rendah (*lower order thinking skills* atau *LOTS*) menuju kemampuan berpikir tingkat tinggi (*higher order thinking skills* atau *HOTS*). Secara visual, ketiga level tersebut dapat digambarkan sebagai berikut.



Gambar 2.3 Level kognitif menurut Anderson & Krathwohl

Penyusunan kisi-kisi disusun tidak boleh lepas dari tujuan penggunaan tes. Penyusunan kisi-kisi merupakan tahapan yang sangat penting dari keseluruhan langkah penyusunan tes. Dengan berpatokan pada sebuah kisi-kisi, dua atau lebih penulis soal akan mampu menyusun soal yang relatif sama (*parallel*) dilihat dari kedalaman dan cakupan materi yang ditanyakan dan tingkat kesulitan butir-butir soalnya. Sebuah kisi-kisi akan memungkinkan seorang guru sebagai penulis soal untuk dapat menyiapkan atau menyusun soal yang memuat sampel materi yang representatif. Oleh

karena itu, sebelum menyusun tes, penulis soal perlu menulis terlebih dulu kisi-kisi yang baik. Karakteristik kisi-kisi yang baik meliputi 1) kisi-kisi harus mewakili isi kurikulum yang akan diujikan; 2) berbagai komponen (kompetensi dasar, materi, indikator dan lain-lain) yang ada dalam kisi-kisi ditulis secara rinci, jelas, dan mudah dipahami; dan 3) indikator soal dinyatakan secara jelas, tidak menimbulkan misinterpretasi bagi penulis soal, dan menggunakan kata kerja operasional sesuai perilaku peserta didik yang akan diukur.

Pemilihan kata kerja operasional secara tepat dalam indikator pencapaian hasil belajar menjadi urusan yang sangat penting ketika menulis kisi-kisi soal. Menjadi urusan penting karena kata kerja operasional yang dipilih dan digunakan dalam sebuah indikator akan mencerminkan tingkat kesulitan soal dan level kognitif yang harus dikuasai oleh peserta didik. Berikut adalah kata kerja operasional yang dapat dipilih oleh penulis soal serta aktivitas pembelajaran yang harus dilakukan oleh peserta didik menurut taksonomi Bloom.

Tabel 2.2 Taksonomi Bloom, aktivitas belajar dan kata kerja operasional yang menyertainya

Domain Kognitif Menurut Bloom	Aktivitas Belajar Peserta Didik	Kata Kerja Operasional Dalam Indikator Soal
Pengetahuan	Mengingat fakta, istilah, konsep, denisi, dan prinsip	Mendefinisikan, menyebutkan, menyatakan, menyebut nama, tempat, waktu dan orang
Pemahaman	Menjelaskan atau menginterpretasikan sesuatu	Menjelaskan, menginterpretasikan, memprediksikan, menyimpulkan, mengkonversikan, menerjemahkan, member contoh, membuat parafrase
Aplikasi	Menggunakan suatu konsep atau prinsip	Menerapkan, memecahkan, menunjukkan,

Domain Kognitif Menurut Bloom	Aktivitas Belajar Peserta Didik	Kata Kerja Operasional Dalam Indikator Soal
	untuk memecahkan suatu masalah tertentu	memanfaatkan, memodifikasi, mendemonstrasikan, menghitung
Analisis	Menguraikan sesuatu menjadi komponen-komponen yang lebih kecil guna melihat tata hubungan atau hirarki dari suatu gagasan	Membedakan, membedakan/ mengkontraskan, mebedakan antara dan ..., bagaimana menghubungkan antara ... dengan ..., mengapa dapat berjalan
Evaluasi	Membuat penilaian berdasarkan criteria yang sudah ditentukan sebelumnya	Menilai, mengevaluasi, menjustifikasi, mengkritik, menentukan mana yang lebih baik/buruk
Kreasi	Menghasilkan sesuatu yang baru dan orisinil dari suatu komponen-komponen tertentu	Mendesain, mengkonstruksi, mengembangkan, menformulasikan, mengimajinasikan, mengkreasikan, merubah, menulis puisi atau cerita pendek, membuat tekonoigi tepat guna.

Ada beberapa hal yang harus diperhatikan oleh penulis soal ketika akan menulis kisi-kisi agar dapat digunakan sebagai panduan praktis dalam mengembangkan soal. Penulis soal semestinya memperhatikan dengan seksama beberapa teknik menulis kisi-kisi agar nantinya kisi-kisi itu dapat digunakan sebagai panduan dalam menulis butir-butir soal secara konsisten

dan tidak menimbulkan misinterpretasi bagi penulis soal atau tes prestasi. Berikut adalah beberapa teknik menulis kisi-kisi yang baik.

- 1) Kisi-kisi sebaiknya dibuat terbuka atau tidak dibuat terlalu detail. Kisi-kisi yang dibuat terlalu detail akan mengakibatkan penulis soal tidak mampu mengembangkan soal yang variatif. Bahkan penulis soal akan mengalami stagnasi dalam memilih bahan ajar yang akan diujikan.
- 2) Kisi-kisi lebih baik dibuat untuk mengidentifikasi gagasan dan keterampilan yang bersifat umum daripada gagasan dan keterampilan yang bersifat detail dan spesifik.
- 3) Kisi-kisi ditulis dengan menggunakan taksonomi kognitif yang paling sesuai dengan mata pelajaran tertentu. Kemampuan kognitif peserta didik pada jenjang sekolah yang berbeda, misalnya peserta didik jenjang SMP dan SMA, tentu akan berbeda pula. Karena itu, pemilihan taksonomi kognitifnya pun harus dibedakan.
- 4) Kisi-kisi ditulis dengan mempertimbangkan tingkat kemampuan berpikir peserta didik, tujuan tes, dan alokasi waktu yang tersedia.
- 5) Tingkat kesulitan pertanyaan disesuaikan dengan tingkat keterampilan berpikir peserta didik.

Ada beberapa format kisi-kisi yang dapat diikuti dan digunakan oleh pengembang soal. Tidak ada format baku yang harus menjadi patokan bagi para penulis soal dalam menuliskan kisi-kisi. Tetapi, bagaimanapun format kisi-kisi yang akan ditulis sebagai panduan dalam penyusunan tes, sebuah kisi-kisi harus memuat komponen-komponen pokok, yaitu kompetensi dasar, materi ajar, dan indikator pencapaian hasil belajar. Oleh karena itu, ketika mengembangkan sebuah kisi-kisi soal, penulis kisi-kisi tidak boleh mengabaikan ketiga komponen tersebut agar validitas isi (*content validity*) tes atau soal selanjutnya dapat dicapai. Hal ini penting karena kisi-kisi dapat dijadikan panduan utama untuk mengembangkan seperangkat tes yang memiliki validitas isi yang baik. Salah satu model kisi-kisi sebagai panduan pengembangan soal dapat diamati contoh berikut.

Contoh Kisi-kisi Soal
Kisi-Kisi Penilaian Akhir Semester
Tahun Pelajaran 2017/2018

Sekolah : SMA Negeri 1 Mojokulon
Peminatan : MIPA / IPS / BB
Mata Pelajaran : Bahasa Inggris

Kelas/Smt : XI / Gasal
Jumlah Soal : 50 Butir Soal
Waktu : 90 menit

No.	Komp. Dasar	Materi	Indikator Butir Soal	Level Kognitif	Nomor Soal	Bentuk Soal
1.	Memahami teks fungsional pendek dan teks esai berbentuk <i>announcement</i> dalam konteks kehidupan sehari-hari dan untuk mengakses ilmu pengetahuan.	Gambaran umum	Diberikan teks berbentuk <i>announcement</i> , peserta didik mampu menentukan jawaban atas pertanyaan tentang gambaran umum isi bacaan.	L1/C2	1	PG
2.		Informasi tersurat	Diberikan teks berbentuk <i>announcement</i> , peserta didik dapat menjawab pertanyaan tentang informasi tersurat dari bacaan.	L2/C1	2	PG
3.	Peserta didik mampu menyebutkan informasi tertentu dari teks <i>exposition</i>	Struktur teks <i>exposition</i>	Disediakan sebuah teks <i>exposition</i> , peserta didik dapat menentukan informasi tersirat	L3/C3	3	PG
4.		Struktur teks <i>exposition</i>	Disediakan sebuah teks <i>exposition</i> , peserta didik dapat alasan	L2/C3	4	PG

No.	Komp. Dasar	Materi	Indikator Butir Soal	Level Kognitif	Nomor Soal	Bentuk Soal
dst. —	-----	-----	mengapa suatu kegiatan perlu dilakukan	-----	-----	-----

Gambar 2.4 Model kisi-kisi penulisan soal

3. Menuliskan Draft Tes

Setelah menulis kisi-kisi tes, tahap selanjutnya adalah menulis draft tes. Tes atau soal adalah sekumpulan butir-butir tes yang dikembangkan untuk mengukur hasil belajar peserta didik. Butir tes mengacu pada satuan pertanyaan yang harus direspon oleh peserta didik dan respon atau jawaban yang diberikan oleh peserta didik dapat dikategorikan menjadi jawaban benar dan salah. Penulisan draft tes menjadi tahapan yang penting karena draft tes yang telah dibuat dapat digunakan sebagai dasar pijakan untuk menghasilkan tes yang berkualitas. Draft tes adalah rencana atau rancangan tes yang masih bersifat sementara. Penulisan draft tes harus mengacu pada kisi-kisi yang sudah dibuat sebelumnya oleh pengembang soal. Draft tes berisi gagasan tentang materi yang akan diujikan, rumusan stem (pokok soal), instruksi untuk mengerjakan soal, pilihan atau alternatif jawaban (untuk tes pilihan ganda), rubrik penskoran dan lain sebagainya.

Penulisan draft tes merupakan tahapan yang urgen dan berkontribusi terhadap kualitas tes yang akan dibuat oleh pengembang soal. Penulisan draft tes ini selanjutnya dimaksudkan untuk memastikan apakah butir-butir tes yang dikembangkan merepresentasikan berbagai kriteria yang diinginkan seperti yang telah dijabarkan dalam kisi-kisi soal. Untuk memastikan apakah draft tes telah sesuai dengan kisi-kisi, selanjutnya draft itu perlu dianalisis. Proses analisis draft tes dilakukan dengan menjawab beberapa pertanyaan sebagai erikut:

- Apakah perintah untuk mengerjakan setiap bagian tes telah ditulis secara jelas?
- Apakah setiap butir tes telah mengukur aspek pengetahuan tertentu secara jelas?

- c. Apakah setiap butir tes telah dinyatakan secara ringkas dan jelas?
- d. Apakah setiap butir tes pilihan ganda telah memiliki distraktor (pengecoh) yang tepat, yaitu pengecoh yang merepresentasikan jawaban salah dan akhirnya peserta didik yang tidak mempunyai pemahaman konsep yang baik terhadap materi yang diujikan 'tertarik' untuk memilih pengecoh tersebut?
- e. Apakah tingkat kesulitan dari setiap butir tes sesuai dengan usia dan kemampuan berpikir peserta didik yang akan diuji?
- f. Apakah bahasa yang digunakan untuk menuliskan setiap butir tes jelas dan bahasa itu tidak berlaku lokal untuk daerah tertentu saja?
- g. Apakah butir-butir tes telah merepresentasikan tujuan pembelajaran yang telah ditentukan?

Dengan menjawab berbagai pertanyaan tersebut, diharapkan tes atau soal yang dikembangkan akan memiliki kualitas yang baik dilihat dari validitas isi dan tingkat kesulitan tes yang proporsional.

4. Mengurutkan Butir-butir Tes

Setelah butir-butir tes dibuat berdasarkan kisi-kisi yang telah ditulis sebelumnya dan dituangkan dalam bentuk draft, urutan butir-butir tes itu belum tentu sesuai dengan urutan materi yang telah diajarkan atau bisa juga urutan tingkat kesulitan butir tes belum proporsional dan logis. Oleh karena itu, tahap selanjutnya adalah mengurutkan butir-butir tes sesuai dengan kriteria tertentu, misalnya diurutkan sesuai urutan materi yang diajarkan atau sesuai dengan tingkat kesulitan. Urutan butir-butir tes ini sangat berpengaruh pada motivasi dan cara peserta didik dalam menjawab tes secara keseluruhan. Jika butir-butir tes diurutkan dari butir-butir yang sulit terlebih dulu dan butir-butir mudah dibagian akhir dari keseluruhan tes, maka peserta didik kemungkinan akan merasa berkecil hati dan tidak termotivasi untuk mengerjakan tes dengan baik.

Pengembang tes perlu mengikuti strategi tertentu dalam mengurutkan butir-butir tes. Berikut adalah strategi atau panduan dalam mengurutkan butir-butir tes.

- a. Tempatkan butir-butir tes yang mudah di bagian awal tes dilanjutkan ke butir-butir yang lebih sulit.
- b. Kelompokkan jenis butir soal tertentu secara bersama-sama; misalnya jenis pilihan ganda, menjodohkan, benar-salah, uraian dan sebagainya, dibarengi dengan instruksi pengerjaan soal yang jelas dan mudah dipahami oleh peserta didik.
- c. Jika diperlukan, urutkan butir-butir soal secara logis berdasarkan urutan materi yang telah diajarkan dan dipelajari oleh peserta tes.

5. Uji-coba Tes

Setelah seperangkat draft test telah dibuat dan diurutkan secara logis, butir-butir tes itu kemudian perlu diujicobakan kepada peserta didik (*testee*). Uji coba tes ini tidak dimaksudkan untuk mendapatkan skor atau nilai dan tidak juga dimaksudkan untuk mengukur kompetensi dari peserta didik, tetapi semata-mata dilakukan untuk mengevaluasi kualitas tes yang telah dikembangkan. Secara spesifik, uji-coba tes lebih dimaksudkan untuk mengumpulkan informasi tentang seberapa baik tes yang telah dikembangkan agar nanti dapat digunakan untuk mengukur hasil belajar peserta didik secara akurat. Agar uji-coba tes itu mampu menghasilkan informasi yang akurat tentang kualitas tes, maka uji-coba tes itu harus melibatkan *testee* yang tepat. *Testee* yang dipilih untuk menguji-cobakan tes harus representatif atau mewakili keseluruhan populasi peserta didik. Dengan kata lain, *testee* yang dipilih untuk mengujicobakan butir-butir tes harus memiliki karakteristik yang sama dengan *testee* yang sebenarnya akan diukur hasil belajarnya. *Testee* yang diminta untuk uji-coba tes perlu juga dimotivasi agar mengerjakan tes sebaik mungkin agar dapat dihasilkan informasi akurat tentang kualitas tes yang diuji-cobakan. Informasi yang diperoleh dari uji-coba tes kemudian digunakan untuk melakukan analisis dan review butir soal sebelum soal itu secara operasional dapat digunakan untuk mengukur hasil belajar peserta didik.

Ada beberapa strategi yang dapat dilakukan oleh pengembang atau penulis soal untuk menguji-cobakan tes yang telah dikembangkan. Beberapa strategi itu meliputi uji awal (*pilot testing*), dan uji lapangan (*field testing*). Berbagai strategi itu memiliki karakteristik yang berbeda satu sama lain.

a. Uji-coba Awal (*pilot testing*)

Uji awal tes dilakukan pada sampel peserta didik (*testee*) yang jumlahnya kecil dan terbatas. *Pilot testing* sangat berguna untuk mengidentifikasi berbagai kelemahan dan kekurangan dari tes yang telah dikembangkan secara cepat dan ekonomis. Berbagai kelemahan dan kekurangan yang mungkin ditemukan dapat segera dikoreksi sebelum uji lapangan dilakukan. Oleh karena itu, tujuan utama dari uji-coba awal ini dimaksudkan untuk mendapatkan data kualitatif dari tes yang telah dikembangkan. Uji coba awal ini cukup melibatkan *testee* sejumlah 15 – 30 orang saja. *Testee* yang dipilih untuk subjek uji-coba harus memiliki karakteristik dan jangkauan penguasaan materi yang sama dengan populasi *testee* yang sebenarnya akan diukur hasil belajarnya. Uji coba awal dapat digunakan sebagai sarana untuk mengevaluasi instruksi pengerjaan tes, apakah instruksi itu sudah jelas dan mudah dipahami oleh peserta didik atau belum. Sebelum melaksanakan uji-coba tes ini, peserta didik perlu diinformasikan bahwa yang akan dievaluasi adalah kualitas tes yang dikerjakan, bukan kemampuan peserta dalam menguasai materi ajar. Selama uji-coba dilaksanakan, batasan waktu dan prosedur standar pelaksanaan tes harus diikuti secara konsisten. Ketika mengerjakan tes uji-coba, peserta didik harus diamati untuk mengidentifikasi apakah terjadi kebingungan dalam memahami butir-butir tes dan apakah ada butir-butir soal tertentu yang bermasalah atau tidak. Setelah uji-coba selesai segera dilakukan analisis terhadap kualitas tes tersebut. Peserta didik perlu juga dilakukan interview atau tanya-jawab untuk mengetahui pendapat dan reaksi mereka terhadap tes yang telah mereka kerjakan.

b. Uji lapangan (*field testing*)

Uji lapangan dilakukan untuk menguji-cobakan tes pada situasi nyata dan peserta didik yang sebenarnya. Uji lapangan dilakukan pada subjek uji-coba (peserta didik) yang lebih banyak dibandingkan saat uji-coba awal. Jumlah peserta didik pada uji lapangan ini antara 100 – 200 peserta didik sehingga ukuran sampel yang mencukupi itu pada akhirnya akan diperoleh data statistik yang stabil terhadap kualitas tes yang telah dikembangkan. Uji lapangan ini lebih dimaksudkan untuk mendapatkan data kuantitatif (tingkat kesulitan, daya beda, validitas, reliabilitas tes) dari tes yang telah dikembangkan. Pada tahap ini, perlu ditegaskan pula kepada peserta

didik bahwa tujuan pelaksanaan tes ini bukan untuk mengukur prestasi belajar, tetapi semata-mata untuk mengetahui kualitas butir-butir tes yang mereka kerjakan. Selain itu, partisipasi mereka dalam mengerjakan tes ini bersifat sukarela, bukan suatu kewajiban. Karena bersifat sukarela, kadang-kadang motivasi mereka untuk berpartisipasi dalam uji-coba tes ini kemungkinan juga akan rendah. Oleh karena itu, pengembang tes perlu memberikan insentif kepada peserta didik yang telah meluangkan waktunya untuk berpartisipasi dalam program uji-coba tes ini. Pemberian insentif ini dipercaya akan dapat memotivasi mereka untuk berpartisipasi secara serius dalam program uji-coba tes ini. Keseriusan peserta didik kemungkinan akan meningkat lebih baik lagi jika mereka yang memperoleh skor tertinggi diapresiasi dengan kuantitas dan kualitas insentif yang berbeda dengan peserta didik lainnya.

6. Menganalisis Soal

Setelah tes diuji-cobakan, hasil uji coba kemudian digunakan sebagai dasar untuk melakukan analisis terhadap kualitas tes yang telah dikembangkan. Analisis dilakukan untuk mendapatkan informasi tentang kondisi tes, baik secara kualitatif maupun kuantitatif. Secara kualitatif, aktivitas analisis ini dimaksudkan untuk mendapatkan informasi berkaitan dengan apakah pertanyaan-pertanyaan dalam tes itu telah disusun secara baik atau belum; apakah instruksi untuk mengerjakan tes itu mudah dipahami oleh peserta didik atau tidak; apakah cakupan materi yang diujikan telah representatif; dan apakah alternatif atau opsi dalam soal pilihan ganda sudah homogen, sama panjang atau belum; dan lain sebagainya.

Analisis kuantitatif dimaksudkan untuk mengidentifikasi berbagai aspek berkaitan koefisien reliabilitas tes, koefisien validitas, indeks daya beda, tingkat kesulitan butir soal dan sebagainya. Guna menganalisis berbagai aspek tersebut, dengan menggunakan formula matematis tertentu, pengembang tes akan dapat lebih mudah mengidentifikasi koefisien reliabilitas tes, koefisien validitas, indeks daya beda, dan tingkat kesulitan butir soal. Dengan melakukan analisis butir soal ini, pengembang tes akan mampu menentukan butir soal mana yang harus dipertahankan; atau harus dimodifikasi; atau bahkan harus dibuang. Dari hasil analisis ini pula, pengembang tes dapat memfinalisasi apakah keseluruhan butir soal itu sudah layak atau belum untuk mengukur hasil belajar peserta didik. Dengan

demikian, analisis keseluruhan butir soal menjadi tahapan yang sangat penting karena hasil analisis ini akan menentukan kualitas soal secara keseluruhan. Apabila hasil analisis tes dapat diketahui bahwa tes secara keseluruhan sudah berkualitas baik, maka tes itu kemudian dapat langsung digunakan sebagai instrumen untuk mengukur hasil belajar peserta didik. Tetapi apabila hasil analisis menunjukkan bahwa tes yang dikembangkan belum baik, maka pengembang tes harus kembali ke tahapan sebelumnya hingga dapat dihasilkan tes yang berkualitas.

Bab 3

Variasi Bentuk Tes Prestasi

Secara garis besar, bentuk tes dapat dibedakan menjadi dua, yaitu tes objektif dan tes subjektif. Tes objektif dapat dipahami sebagai sebuah tes dimana proses penilaiannya tanpa dipengaruhi oleh pertimbangan pribadi dari penguji. Dengan kata lain, proses penilaian dilakukan secara objektif tanpa dipengaruhi oleh faktor-faktor subjektif dari penguji. Tingkat akurasi dari jawaban peserta didik yang diuji (*testee*) semata-mata didasarkan pada kriteria yang telah ditentukan sebelumnya. Jawaban peserta didik kemudian dikategorikan sebagai jawaban benar atau salah setelah dibandingkan dengan kunci jawaban yang telah disiapkan oleh guru sebagai penguji.

Sebaliknya tes subjektif adalah sebuah bentuk tes dimana proses penilaiannya sering dipengaruhi oleh pertimbangan pribadi dari penguji. Pada bentuk tes ini, subjektivitas penguji sulit sekali untuk dapat dihilangkan, bahkan tidak mungkin untuk dihilangkan. Hal yang dapat dilakukan adalah bagaimana meminimalisir subjektivitas penguji agar hasil penilaiannya mendekati objektif. Guna meminimalisir subjektivitas itu, ada beberapa hal yang harus dilakukan oleh guru atau penguji. Pertama, guru sebelum melakukan proses penilaian harus telah menyiapkan rubrik penilaian yang baik sebagai pedoman untuk memberikan nilai terhadap hasil belajar peserta didik. Kedua, proses penilaian dilakukan dengan melibatkan lebih dari satu penguji (*inter-rater*) sehingga dapat diketahui sejauhmana konsistensi hasil penilaian itu. Semakin konsisten hasil penilaian yang dihasilkan oleh dua atau lebih penguji itu, berarti semakin objektif hasil penilaian itu. Ketiga, dilakukan teknik pengujian ulang (*mark-remark technique of scoring*), yaitu penguji yang sama melakukan dua kali proses penilaian terhadap tes yang sama dan proses penilaian kedua dilakukan beberapa saat setelah proses penilaian pertama selesai dilakukan (biasanya

tidak lebih dari satu atau dua minggu dari proses penilaian pertama). Rentang waktu ini dimaksudkan untuk menghindari *carry over effect*, yaitu efek atau kondisi dimana pada saat guru memeriksa hasil tes peserta didik pada kesempatan yang berbeda masih dipengaruhi oleh efek atau kondisi pada saat guru memeriksa hasil tes peserta didik pada kesempatan pertama. Pada bagian berikut selanjutnya diuraikan variasi dari tes objektif dan tes subjektif secara lengkap beserta kelebihan dan kekurangan dari tes tersebut.

A. Tes Objektif

Sesuai dengan namanya, tes objektif merupakan sebuah bentuk tes yang menghendaki objektivitas yang tinggi selama proses penilaian dilakukan. Tes bentuk ini paling banyak digunakan oleh guru-guru di dsekolah. Ada paling tidak dua alasan mengapa tes bentuk ini banyak digunakan di sekolah. Pertama, tes bentuk ini sangat praktis karena proses penilaian dapat dilakukan secara cepat dan mudah. Bahkan proses penilaian dapat dilakukan oleh pihak-pihak yang tidak memiliki pengetahuan dan pemahaman yang baik terhadap materi yang diujikan. Kedua, reliabilitas atau konsistensi hasil penilaian dapat dicapai lebih mudah. Konsistensi hasil penilaian relatif mudah dicapai karena ketepatan atau akurasi jawaban peserta didik semata-mata didasarkan pada kunci jawaban yang telah ditentukan sebelum ujian dilakukan. Jawaban peserta didik dikategorikan benar jika sesuai dengan kunci jawaban dan dikategorikan salah jika tidak sesuai dengan kunci jawaban.

Pada umumnya, guru menggunakan tes objektif untuk mengukur kemampuan berpikir peserta didik pada aspek pengetahuan dengan cakupan materi yang luas. Sebenarnya, tes bentuk ini dapat dikembangkan dan digunakan pula untuk mengukur proses berpikir peserta didik tingkat tinggi yang dimulai dari kemampuan peserta didik untuk memahami, menerapkan, menganalisis dan mengevaluasi terhadap suatu objek pengetahuan. Agar tes objektif tidak sekedar untuk mengukur kemampuan berpikir tingkat rendah, pengembang tes perlu memiliki pengetahuan dan pengalaman yang memadai dalam menuliskan tes tersebut. Dorobat (2007) menegaskan ada beberapa tahapan sebagai panduan bagi guru dalam mengembangkan tes objektif yang baik. Panduan dalam mengembangkan tes ojektif tersebut diuraikan pada bagian berikut.

1. Guru cukup memilih materi yang berkaitan dengan fakta dan pengetahuan penting dari keseluruhan materi yang telah diajarkan untuk diujikan kepada peserta didik.
2. Sesuaikan pertanyaan-pertanyaan dengan usia dan tingkat kemampuan peserta didik serta sesuaikan pula dengan tujuan dari tes itu.
3. Guru harus menulis butir-butir pertanyaan secara jelas dan mudah dipahami oleh setiap peserta didik. Guru perlu menghindari pernyataan-pernyataan yang ambigu atau bermakna ganda agar peserta didik tidak memiliki pemahaman yang berbeda-beda terhadap butir-butir pertanyaan tersebut.
4. Untuk menghindari ambiguitas pertanyaan, guru harus menuliskan butir-butir pertanyaan tersebut dengan menggunakan tata-bahasa yang baik dan benar.
5. Guru harus menghindari menulis butir-butir pertanyaan yang saling berkaitan dengan butir-butir pertanyaan pada nomor-nomor sebelumnya.
6. Jika tes dalam bentuk pilihan ganda, butir pertanyaan harus memuat hanya satu pilihan jawaban yang benar.

Panduan tersebut harus diikuti secara konsisten oleh guru sebagai pengembang tes objektif agar tes yang dikembangkan benar-benar dapat mengukur kemampuan peserta didik yang semestinya untuk diukur. Dengan kata lain, panduan tersebut perlu diikuti agar tes yang dikembangkan memiliki validitas isi yang baik.

1. Keunggulan dan kelemahan tes objektif

Sebagai sebuah bentuk tes, tes objektif tentu memiliki kelebihan dan kelemahan. Dorabot (2007) secara detail menguraikan kekuatan dan kelemahan dari tes objektif ini. Diantara beberapa kelebihan atau keunggulan tes objektif tersebut diuraikan pada bagian berikut.

- a. Tes bentuk ini hanya memiliki satu alternatif jawaban yang benar sehingga objektivitas penilaian dapat dengan mudah dicapai.
- b. Tes objektif dapat dinilai secara mekanis atau menggunakan mesin. Pihak-pihak yang memeriksa tes objektif tidak harus melakukan pelatihan secara khusus dan tidak harus memiliki

pengetahuan yang memadai berkaitan dengan materi yang diujikan kepada peserta didik.

- c. Tes ini menghendaki setiap peserta didik untuk menjawab setiap butir pertanyaan dalam situasi yang terstruktur atau terkontrol ketat dan menjawab butir-butir soal dengan cakupan materi yang luas.
- d. Peserta didik biasanya mampu menyelesaikan untuk menjawab butir-butir pertanyaan dalam periode waktu yang disediakan.
- e. Sampel materi yang representatif dari keseluruhan materi yang telah diajarkan mudah dicapai karena tes ini mencakup materi yang lebih luas dibandingkan dengan tes subjektif.
- f. Butir-butir pertanyaan tes objektif mampu mendorong peserta didik untuk mengembangkan basis pengetahuan, keterampilan, dan kemampuan yang luas.
- g. Tes ini memungkinkan untuk dikoreksi dengan bantuan komputer.
- h. Dalam proses pengembangannya, tes ini memerlukan persiapan yang lebih matang dan teliti dibandingkan dengan proses penyiapan tes subjektif.
- i. Tes ini dapat dikembangkan pada tingkat kesulitan yang serendah atau setinggi mungkin sesuai kehendak pembuat atau pengembang soal.
- j. Tes objektif memungkinkan untuk diuji-cobakan terlebih dulu guna mengetahui kualitas butir-butir pertanyaannya sebelum digunakan dalam konteks yang lebih luas.

Selain memiliki beberapa kelebihan seperti diuraikan di atas, tes objektif juga memiliki beberapa kelemahan. Diantara kelemahan-kelemahan tes objektif itu seperti dijelaskan oleh Dorobat (2007) adalah sebagai berikut.

- a. Tes bentuk objektif biasanya hanya mengukur kemampuan berpikir tingkat rendah, bahkan hanya mengukur kemampuan peserta didik pada tingkat hafalan atau ingatan saja.
- b. Tes bentuk ini banyak memberi peluang kepada peserta didik untuk menebak jawaban yang dianggap benar jika mereka tidak memiliki pemahaman yang memadai terkait materi yang diujikan.

- c. Tes bentuk objektif banyak mengabaikan kemampuan peserta didik untuk mengungkapkan gagasannya, baik secara tertulis maupun secara lisan.

Mengingat berbagai kelemahan tes bentuk objektif seperti diuraikan diatas, guru sebagai pengembang tes perlu terus mengembangkan pengalamannya dalam membuat butir-butir pertanyaan sehingga butir-butir pertanyaan yang dikembangkan tidak sekedar mengukur kemampuan peserta didik pada tingkat hafalan, tetapi tes diharapkan juga digunakan untuk mengukur proses berpikir tingkat tinggi, seperti kemampuan untuk memahami, mengaplikasikan, menganalisis, dan mengevaluasi.

2. Jenis-jenis tes objektif

Ada beberapa jenis tes objektif yang biasa digunakan oleh guru untuk mengukur hasil belajar peserta didik, misalnya bentuk pilihan ganda (*multiple-choice test*), menjodohkan (*matching type test*), benar-salah (*true-flase*), dan jawaban singkat (*short-answer test*). Berbagai jenis tes tersebut memiliki karakteristik masing-masing yang harus diperhatikan oleh pengembang tes.

a. Tes bentuk pilihan ganda

Bentuk tes pilihan ganda adalah suatu bentuk tes dimana butir-butir soalnya memiliki beberapa alternatif jawaban yang harus dipilih oleh peserta didik (*testee*). Biasanya alternatif jawaban yang ada dalam sebuah butir soal berjumlah tiga hingga lima alternatif. Tes pilihan ganda paling tidak memuat dua bagian utama, yaitu pokok soal (*stem*) dan alternatif jawaban (*option*). Soal pilihan ganda dapat digunakan untuk mengukur kemampuan peserta didik pada aspek pengetahuan, pemahaman, aplikasi, analisis, dan evaluasi. Aspek-aspek tersebut tentunya memiliki karakteristik yang berbeda satu sama lain, sehingga guru sebagai pengembang tes perlu memperhatikan berbagai karakteristik tersebut agar tes yang dikembangkan benar-benar mampu mengukur tingkat berpikir peserta didik yang diharapkan.

Tes untuk mengukur pengetahuan adalah tes yang dimaksudkan untuk mengidentifikasi kemampuan berpikir peserta didik pada tingkat rendah. Tes ini diarahkan untuk mengukur sejauhmana peserta didik mampu mengingat kembali berbagai hal yang berkaitan dengan istilah, fakta, konsensus,

definisi, pengertian, klasifikasi, kategori, kriteria dan sebagainya. Tes untuk mengukur kemampuan memahami adalah tes yang dimaksudkan untuk mengukur kemampuan peserta didik dalam memahami materi yang telah dipelajari. Kemampuan memahami ini dimanifestasikan dalam bentuk kemampuan peserta didik untuk meringkas, memberikan contoh, memparafrase, melakukan translasi (mentransfer suatu informasi); menginterpretasi (menjelaskan suatu informasi); dan mengestrapolasi (memprediksi dan memaknai suatu informasi). Selanjutnya, tes untuk mengukur kemampuan aplikasi adalah butir-butir tes yang dimaksudkan untuk mengidentifikasi kemampuan peserta didik dalam menerapkan suatu informasi atau pengetahuan untuk mengatasi berbagai permasalahan dalam situasi nyata; mencari solusi dari suatu masalah; menunjukkan contoh; memodifikasi, serta menghitung dan memanfaatkan suatu objek pengetahuan. Selanjutnya, tes analisis adalah suatu tes yang dimaksudkan untuk mengukur kemampuan peserta didik dalam menguraikan suatu materi yang telah dipelajari menjadi bagian-bagian yang lebih rinci; membedakan dan menkontraskan suatu informasi; dan menghubungkan satu informasi dengan informasi lainnya. Akhirnya, tes untuk mengukur kemampuan evaluasi adalah butir-butir tes yang dimaksudkan untuk mengukur kemampuan peserta didik dalam menilai dan menjustifikasi suatu manfaat dari suatu pengetahuan yang dimiliki untuk tujuan tertentu berdasarkan kriteria yang telah ditentukan.

Seperti bentuk soal yang lain, tes pilihan ganda memiliki kelebihan dan kekurangan. Diantara beberapa kelebihan soal bentuk pilihan ganda tersebut diuraikan dalam bagian berikut:

1. Cakupan materi yang diujikan lebih banyak dibandingkan dengan tes bentuk subjektif.
2. Proses pelaksanaan dan analisis tes lebih mudah.
3. Proses penilaian dapat dilakukan lebih objektif.
4. Tes pilihan ganda dapat digunakan untuk mengukur berbagai level tujuan pembelajaran, mulai dari yang paling sederhana hingga yang paling kompleks.
5. Tes pilihan ganda dapat memberikan informasi lebih lengkap tentang latar-belakang pengetahuan dan pengalaman peserta didik.
6. Kualitas butir soal mudah dianalisis guna mengetahui tingkat kesukaran, validitas, reliabilitas, dan daya beda dari tes tersebut.

Adapun kelemahan tes bentuk pilihan ganda adalah sebagai berikut:

1. Tes pilihan ganda lebih banyak menguji aspek pengetahuan dan kurang tepat jika digunakan untuk mengukur kemampuan peserta didik untuk mendemonstrasikan keterampilan dan pengalamannya.
2. Apabila peserta didik tidak memiliki pengetahuan dan pemahaman yang memadai terhadap materi yang diujikan, mereka memiliki peluang untuk sekedar menebak atau asal pilih alternatif jawaban yang dianggap benar.
3. Peserta didik memiliki peluang lebih mudah untuk berbuat curang (*cheating*) selama mengikuti tes.
4. Materi tes hanya terbatas pada materi-materi yang memungkinkan untuk diujikan secara objektif, sehingga hasilnya sering tidak mencerminkan kemampuan peserta didik yang sebenarnya.
5. Proses penyusunan butir-butir soal relatif lebih rumit dibandingkan dengan penyusunan soal bentuk subjektif.

Ada beberapa jenis soal bentuk pilihan ganda. Diantaranya adalah pilihan ganda analisis hubungan antar-hal, pilihan ganda analisis kasus, pilihan ganda asosiasi, dan pilihan ganda dengan grafik, tabel dan diagram.

1) Pilihan ganda analisis hubungan antar-hal

Pilihan ganda analisis hubungan antar-hal biasa juga disebut dengan pilihan ganda sebab-akibat. Pada pilihan ganda jenis ini, peserta didik diharapkan mampu mengidentifikasi hubungan sebab-akibat yang dinyatakan dalam dua pernyataan. Pernyataan pertama berperan sebagai 'sebab' dan pernyataan kedua sebagai 'alasan atau akibat'. Ada beberapa kemungkinan dari hubungan kedua pernyataan tersebut. Pertama, pernyataan pertama secara konseptual benar dan pernyataan kedua juga benar dan keduanya memiliki hubungan sebab-akibat. Kedua, pernyataan pertama benar dan pernyataan kedua juga benar, tetapi keduanya tidak memiliki hubungan sebab-akibat. Ketiga, pernyataan pertama benar, tetapi pernyataan kedua salah. Keempat, pernyataan pertama salah, tetapi pernyataan kedua benar. Kelima, pernyataan pertama dan kedua sama-sama salah. Untuk dapat memahami lebih jauh dari pilihan ganda analisis antar-hal ini, dapat diperhatikan contoh soal berikut.

Petunjuk soal:

Untuk soal-soal berikut, pilihlah:

- A. Jika pernyataan benar, alasan benar, dan alasan merupakan sebab dari pernyataan
- B. Jika pernyataan benar, alasan benar, tetapi alasan bukan merupakan sebab dari pernyataan
- C. Jika pernyataan benar tetapi alasan salah
- D. Jika pernyataan salah tetapi alasan benar
- E. Jika pernyataan salah dan alasan salah

Soal:

Harga kendaraan bekas pakai, seperti mobil dan sepeda motor di masa pandemi Covid-19 turun drastis dan lebih murah

SEBAB

Permintaan terhadap mobil dan sepeda motor berkurang drastis sebagai dampak dari menurunnya penghasilan konsumen akibat Covid-19.

Jawaban dari butir soal tersebut adalah (A) karena pernyataan pertama dan kedua sama-sama benar dan keduanya memiliki hubungan sebab-akibat.

2) Pilihan ganda analisis kasus

Pada jenis pilihan ganda analisis khusus ini, peserta didik akan disajikan sebuah kasus faktual berupa peristiwa tertentu yang diambil dari surat kabar, majalah, jurnal, artikel penelitian dan sebagainya. Berdasarkan peristiwa yang disajikan itu, peserta didik kemudian diberikan satu atau beberapa pertanyaan. Pertanyaan-pertanyaan yang diajukan untuk dijawab oleh peserta didik diarahkan untuk mendorong peserta didik itu untuk melakukan analisis dari peristiwa yang disajikan itu. Jadi pilihan ganda jenis ini dibuat bukan sekedar untuk mengarahkan peserta didik untuk mengingat dan menghafal berbagai objek dalam peristiwa yang disajikan, tetapi peserta didik lebih diarahkan untuk membedakan, membandingkan, mengkontraskan, dan menghubungkan peristiwa satu dengan peristiwa atau kasus yang lain. Dapat juga butir-butir pertanyaan itu diarahkan agar peserta didik mampu mencari alasan mengapa peristiwa itu bisa terjadi.

Contoh soal mata pelajaran bahasa Indonesia:

Instruksi: Jawablah soal-soal berikut berdasarkan kasus yang diberikan.
Pahamilah setiap kasus secara cermat sebelum anda menjawab pertanyaan-pertanyaan yang disediakan.

JAKARTA, KOMPAS.com - Komisi Perlindungan Anak Indonesia (KPAI) menerima sejumlah aduan terkait keluhan sistem Penerimaan Peserta Didik Baru (PPDB) DKI Jakarta. "Dari DKI Jakarta soal usia ada delapan pengaduan per Rabu lalu, dan hari ini masuk puluhan pengaduan," ujar Komisioner Bidang Pendidikan KPAI Retno Listyarti ketika dihubungi Kompas.com, Jumat (26/6/2020). Retno belum dapat merinci berapa jumlah laporan terkait PPDB DKI yang diterima KPAI sampai Jumat ini. Namun, dia menyebut bahwa sebagian besar laporan tersebut mengenai sistem seleksi PPDB berdasarkan usia yang dikeluhkan para orangtua. "Dari tadi malam masuk banyak pengaduan sampai tadi pagi soal usia (PPDB) dari DKI Jakarta. Saya baru bisa update data semalam dan hari ini jam 17.00 nanti," ujar dia.

.....
(Kompas, 26/6/2020).

Pertanyaan:

Apa yang menjadi permasalahan utama dalam proses Penerimaan Peserta Didik Baru (PPDB) di DKI Jakarta?

- A. Banyaknya calon peserta didik yang mengeluh karena tidak tertampung di sekolah negeri.
- B. Banyaknya calon peserta didik sulit mendapatkan informasi tentang sistem PPDB.
- C. Adanya kemungkinan calon peserta didik tidak diterima di sekolah baru karena faktor usia.
- D. Orang tua calon peserta didik protes ke pemerintah provinsi DKI karena sistem PPDB tidak transparan.
- E. KPAI tidak mampu menampung semua aduan dari orang tua calon peserta didik.

Jawaban dari butir soal tersebut adalah (C). Untuk dapat menentukan jawaban secara tepat terhadap butir soal tersebut, peserta didik harus mampu melakukan analisis terhadap peristiwa yang disajikan.

3) Pilihan ganda asosiasif

Jenis tes ini sama dengan pilihan ganda jenis yang lain, yaitu tes ini juga menyediakan alternatif jawaban dan peserta didik memilih salah satu jawaban yang paling tepat. Bedanya adalah jika jenis pilihan ganda yang lain hanya menyediakan satu pilihan jawaban yang benar, jenis tes pilihan ganda asosiasif menyediakan beberapa alternatif jawaban yang benar. Jadi, untuk jenis ini akan ada lebih dari satu jawaban yang benar, misalnya dalam satu butir soal ada dua alternatif jawaban yang benar, tiga alternatif jawaban benar, atau bahkan ada empat alternatif jawaban yang benar. Dengan demikian, kejelasan petunjuk atau instruksi untuk mengerjakan soal menjadi hal yang sangat penting karena jawaban yang tepat dan harus dipilih oleh peserta didik sangat dipengaruhi oleh rumusan petunjuk pengerjaan soal itu.

Contoh soal mata pelajaran biologi:

Petunjuk: Untuk soal-soal berikut ini, pilihlah:

- A. Jika (1), (2), dan (3) benar
- B. Jika (1) dan (3) benar
- C. Jika (2) dan (4) benar
- D. Jika hanya (4) yang benar
- E. Jika semuanya benar

Soal:

Berikut adalah dampak dari terjadinya hujan asam bagi lingkungan:

- (1) Dapat melarutkan nutrisi penting dalam tanah yang dibutuhkan oleh tanaman.
- (2) Dapat mempengaruhi habitat perairan karena banyak biota air mati akibat kadar air yang terlalu asam.
- (3) Dapat mengakibatkan kerusakan bangunan karena hujan asam bersifat korosif.
- (4) Hujan asam mengakibatkan udara banyak mengandung H_2O sehingga berbahaya untuk pernafasan.

Jawaban dari soal tersebut adalah (A) karena pilihan (1), (2), dan (3) adalah mengacu pada dampak negatif dari hujan asam terhadap lingkungan.

4) Pilihan ganda dengan grafik, tabel, dan diagram

Jenis pilihan ganda ini sebenarnya tidak jauh berbeda dengan pilihan ganda analisis kasus. Bedanya adalah tes ini tidak menyajikan kasus dalam bentuk peristiwa atau cerita, tetapi kasus diwujudkan dalam bentuk tabel, diagram, atau grafik. Kemampuan peserta didik untuk menganalisis tabel, grafik, dan diagram menjadi hal yang sangat penting karena kemampuan itu akan berpengaruh terhadap ketepatan dalam memilih alternatif jawaban yang tepat.

Contoh soal mata pelajaran sejarah:

Perhatikan tabel berikut!

No	X	Y
1)	Soekarno	Hal-hal yang mengenai pemindahan kekuasaan dan lain-lain diselenggarakan dengan cara seksama dan dalam tempo yang sesingkat-singkatnya.
2)	Moh. Hatta	Kami bangsa Indonesia dengan ini menyatakan kemerdekaan Indonesia
3)	Achmad Soebarjo	Djakarta, hari 17 boelan 08 tahun 05 Atas nama bangsa Indonesia, Soekarno-Hatta

Jawaban yang tepat untuk melengkapi peran tokoh dalam perumusan naskah proklamasi dan gagasannya yang tepat ditunjukkan oleh

- A. X1-Y1 dan X2-Y2
- B. X1-Y2 dan X2-Y3
- C. X1-Y3 dan X2-Y1
- D. X2-Y1 dan X3-Y1
- E. X2-Y2 dan X3-Y2

Jawaban dari pertanyaan tersebut adalah C. Agar mampu menjawab pertanyaan tersebut dengan benar, peserta didik harus memiliki kemampuan yang baik untuk menganalisis informasi yang disajikan dalam tabel.

Mengembangkan atau membuat soal bentuk pilihan ganda bukan suatu pekerjaan yang mudah. Perlu ketelitian agar tes pilihan ganda yang

dibuat oleh guru yang berperan sebagai pengembang soal memiliki kualitas yang baik. Ketika guru akan mengembangkan soal pilihan ganda, ada beberapa teknik yang harus diperhatikan. Berikut diuraikan beberapa teknik dalam mengembangkan tes bentuk pilihan ganda.

- 1) Pokok soal (*stem*) harus dinyatakan dengan jelas, mudah dipahami, dan disusun dengan menggunakan tata-bahasa yang benar.

Contoh mata pelajaran geografi:

Diketahui daerah A memiliki jumlah penduduk seluruhnya 5,2 juta jiwa, terdiri dari penduduk laki-laki berjumlah 48%. Berapa sex ratio penduduk daerah A tersebut?

- A. 82
- B. 92
- C. 92,4
- D. 108
- E. 108,3

- 2) Apabila pokok soal mengandung makna negatif, maka kata yang bermakna negatif itu dicetak tebal atau miring.

Contoh soal mata pelajaran biologi:

- ✓ Manakah pernyataan berikut yang **TIDAK BENAR** berkaitan dengan dampak negatif dari hujan asam bagi lingkungan?
- ✓ Berikut ini adalah dampak negatif dari hujan asam bagi lingkungan hidup, **KECUALI**

- 3) Hindari menggunakan alternatif jawaban ‘Semua jawaban di atas benar’ atau ‘Semua jawaban di atas salah’. Jika alternatif jawaban itu terpaksa harus dipakai, alternatif jawaban itu sebaiknya bukan kunci jawaban.

Contoh alternatif jawaban yang harus dihindari:

Mitigasi bencana :

- (1) Turun lewat tangga darurat;
- (2) Melindungi kepala dengan tangan/tas/buku;
- (3) Lari menuju lapangan terbuka;
- (4) Menjauhi bangunan tinggi

Pada saat terjadi gempa bumi dan kita berada di lantai dasar suatu bangunan, maka tindakan yang harus dilakukan untuk mengurangi korban antara lain

- A. (1) dan (3)
- B. (1) dan (4)
- C. (2) dan (5)
- D. (2) dan (5)
- E. Semua tindakan di atas benar

- 4) Semua alternatif jawaban harus homogen dan panjang / pendeknya jawaban harus relatif sama.

Contoh soal mata pelajaran sejarah:

Para pedagang Eropa yang berasal dari Inggris, Portugis, Spanyol dan Belanda mencari kepulauan Indonesia sekitar abad ke-15 dan ke-16 karena adanya komoditas

- A. minyak bumi dan tembakau
- B. rempah-rempah dan kayu cendana
- C. batu permata dan cengkeh
- D. mutiara dan keramik
- E. kain tenun dan minyak kelapa

- 5) Sebuah pokok soal hanya mengandung satu gagasan utama.

Contoh soal sosiologi:

Di sebuah kota, telah terjadi peningkatan kasus kriminalitas yang disebabkan oleh faktor ekonomi. Untuk mendapatkan data tentang berbagai macam penyimpangan sosial yang terkait dengan aksi kriminalitas di sebuah kota tersebut, seorang sosiolog melakukan penelitian yang akan digunakan sebagai bahan kajian untuk menentukan kebijakan pemerintah. Peran penting seorang sosiolog yang terlihat dalam ilustrasi tersebut adalah

- A. sebagai ahli dalam penelitian sosial
- B. sebagai konsultan kebijakan pemerintah
- C. memberi masukan kepada pemerintah
- D. mengajar dan mendidik masyarakat
- E. membantu membuat kebijakan publik

- 6) Jika pokok soal berupa pernyataan tidak lengkap dan peserta didik harus melengkapi pernyataan itu, tempatkan titik-titik di bagian akhir pernyataan, bukan di awal pernyataan.

Contoh soal bahasa Inggris:

The least severe form of behavior disorder is

- A. Psychosis
- B. Panic disorder
- C. Neurasthenia
- D. Neurosis

- 7) Hindari pokok soal yang menggunakan kata-kata yang tidak perlu (*wordiness*), karena hal itu akan membingungkan peserta didik.

Contoh mata pelajaran bahasa Inggris:

Sheldon developed a highly controversial theory of personality based on body type and temperament of the individual. Which of the following is a criticism of Sheldon's work?

- A. He was influenced too much by the Freudian psychoanalysis.
- B. His rating of physique and temperament were not independent.
- C. He failed to use empirical approach.
- D. His research sample was improperly selected.

Pokok soal tersebut menggunakan banyak kosa kota yang tidak perlu dan informasi yang kurang relevan. Pokok soal tersebut mestinya cukup ditulis:

Which of the following is a criticism of Sheldon's theory of personality?

- 8) Upayakan alternatif jawaban yang berperan sebagai pengecoh (*distractor*) dapat berfungsi dengan baik.

Contoh soal mata pelajaran seni musik:

Instrumen musik yang menggunakan selaput tipis sebagai sumber bunyi disebut

- A. chordophone
- B. membranophone
- C. idiophone
- D. aerophone
- E. electrophone

Distraktor dalam soal tersebut berpeluang dapat berfungsi dengan baik dan dapat mengecoh peserta didik yang tidak memiliki pengetahuan yang baik tentang jenis-jenis alat musik, karena setiap alternatif jawaban memiliki kemiripan.

- 9) Apabila alternatif jawaban berupa angka-angka, urutkan alternatif jawaban dari angka kecil ke angka yang besar.

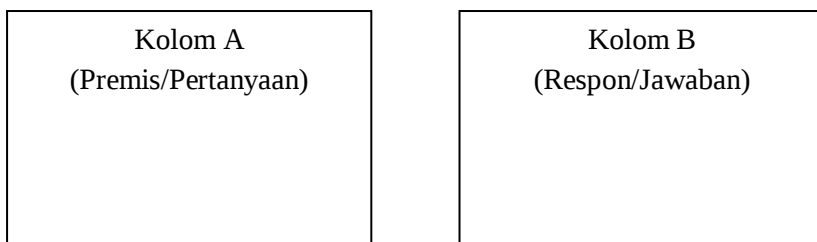
Contoh soal mata pelajaran biologi:

Jumlah fotoreseptor pada retina manusia adalah berjumlah sekitar

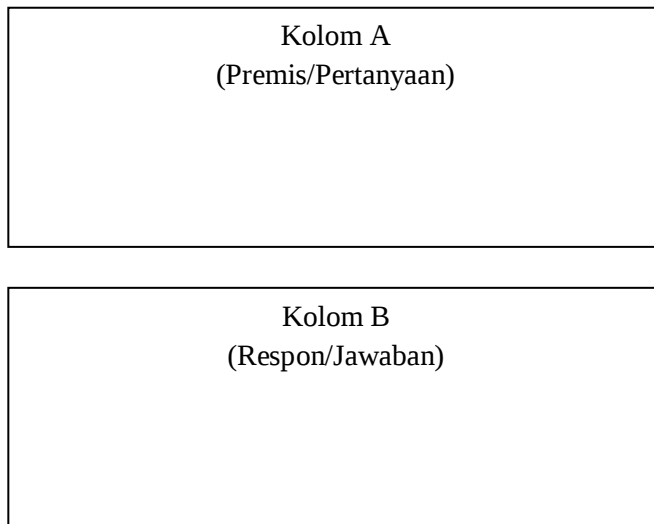
- A. 5 juta
- B. 35 juta
- C. 65 juta
- D. 115 juta

b. Tes bentuk menjodohkan

Tes bentuk menjodohkan adalah suatu bentuk tes yang terdiri dari dua kolom. Kolom pertama biasa disebut premis dan kolom kedua disebut respon. Kolom pertama biasa juga disebut dengan pokok soal atau pertanyaan dan kolom kedua sebagai kolom jawaban. Sebagai salah satu bentuk tes objektif, tes bentuk ini menyediakan beberapa pertanyaan atau pernyataan yang ditulis dalam kolom premis dan setiap pertanyaan atau pernyataan itu disediakan satu jawaban yang ditulis dalam kolom respon. Setiap peserta didik harus menjodohkan pertanyaan atau pernyataan itu dengan satu jawaban yang tepat. Seri jawaban yang disediakan dalam kolom jawaban harus memiliki jumlah yang lebih banyak dibandingkan dengan jumlah pertanyaan. Tetapi, jika satu jawaban yang tersedia dalam kolom jawaban dapat dipilih lebih dari sekali sebagai jawaban yang benar, maka jumlah jawaban yang tersedia boleh sama dengan jumlah pertanyaan. Ada dua model kolom yang biasa digunakan untuk bentuk tes menjodohkan ini, yaitu model horizontal dan model vertikal.



Gambar 3.1
Tes bentuk menjodohkan model horizontal



. Gambar 3.2
Tes bentuk menjodohkan model vertical

Kedua model tersebut dapat digunakan oleh guru ketika ingin mengembangkan tes bentuk menjodohkan. Tetapi model tersebut sebenarnya hanya sekedar contoh agar model tes bentuk menjodohkan tidak terkesan monoton dan kurang variatif.

1) Teknik mengembangkan tes bentuk menjodohkan

Ada beberapa teknik yang harus diperhatikan ketika guru ingin mengembangkan tes bentuk menjodohkan. Diantara beberapa teknik dalam mengembangkan tes bentuk menjodohkan tersebut diuraikan dalam bagian berikut.

- Daftar jawaban yang disediakan dalam kolom jawaban sebaiknya dinyatakan dalam kalimat atau frasa yang relatif pendek.
- Pilihan jawaban yang disediakan sebaiknya disusun secara alfabetis atau numerik.
- Perintah atau instruksi untuk mengerjakan soal harus mengindikasikan secara jelas bahwa tes tersebut adalah tes bentuk menjodohkan.

Hal-hal yang perlu diperhatikan dalam menuliskan perintah untuk mengerjakan tes menjodohkan adalah:

- ✓ apakah jawaban-jawaban yang disediakan dalam kolom jawaban dapat digunakan sekali saja atau lebih dari satu kali;
 - ✓ jelaskan dimana peserta didik harus meletakkan jawaban yang mereka berikan, di samping kiri pertanyaan, di samping kanan pertanyaan, atau di tempat lain;
 - ✓ apakah peserta didik harus menjawab setiap pertanyaan berdasarkan hubungan antara pertanyaan dan jawaban, atau berdasarkan logika nyata.
- d) Posisi penjodohan antara pertanyaan dan jawaban harus bervariasi. Hindari posisi pertanyaan dan jawaban dalam bentuk pola yang teratur untuk menghindari peluang peserta didik menebak posisi jawaban dari pertanyaan yang sebenarnya mereka tidak kuasai.
 - e) Posisi pertanyaan dan jawaban sebaiknya ada dalam satu halaman yang sama pada lembar soal.
 - f) Jumlah daftar jawaban harus lebih banyak daripada daftar pertanyaan, jika satu jawaban yang tersedia tidak dapat dipilih lebih dari sekali sebagai jawaban benar.
 - g) Hindari menggunakan kalimat atau pernyataan berupa kutipan langsung dari buku, artikel atau dokumen lainnya sebagai premis.

Pengembang soal perlu secara konsisten memerhatikan teknik pengembangan tes menjodohkan tersebut sebagai upaya untuk mendapatkan bentuk tes yang berkualitas. Pada bagian berikut dapat diamati dua contoh tes bentuk menjodohkan yang dikategorikan sebagai tes yang baik dan kurang baik. Dengan mengamati kedua contoh ini, pengembang soal diharapkan memiliki pemahaman yang baik berkaitan dengan bagaimana seharusnya tes menjodohkan dikembangkan.

Contoh 1:

Jodohkan berikut ini.

- | | |
|--------------------|-------------------------|
| 1. Makanan | A. Penguat primer |
| 2. Psikoanalisis | B. Sigmund Freud |
| 3. B.F. Skinner | C. Operant Conditioning |
| 4. Standar deviasi | D. Ukuran variabilitas |
| 5. Skizofrenia | E. Halusinasi |

Contoh tes bentuk menjodohkan tersebut dikategorikan sebagai tes yang kurang baik karena beberapa alasan. Pertama, perintah untuk mengerjakan tes tidak jelas. Apakah daftar jawaban yang tersedia dalam kolom jawaban harus dipilih sekali saja atau dapat dipilih lebih dari sekali. Kedua, materi yang diujikan dalam soal tersebut tidak homogen karena ada materi bidang psikologi dan sekaligus bidang statistik. Ketiga, jika jawaban-jawaban yang tersedia dalam kolom jawaban tidak dapat dipilih lebih dari sekali, mestinya jumlah daftar jawaban lebih banyak daripada jumlah pertanyaan. Keempat, peserta didik tidak mendapatkan gambaran dimana mereka harus meletakkan jawaban yang ingin mereka berikan.

Contoh 2

Jodohkan teori-teori di kolom A dengan penggagasnya di kolom B dengan cara menuliskan huruf jawaban yang benar di sisi kiri teori!

- | | |
|-----------------------------------|--------------------|
| ____ 1. Teori psikodinamika | A. Albert Bandura |
| ____ 2. Teori sifat | B. B.F. Skinner |
| ____ 3. Teori behaviorisme | C. Carl Rogers |
| ____ 4. Teori humanisme | D. Gordon Allport |
| ____ 5. Teori pembelajaran sosial | E. Karn Honey |
| | F. Raymond Cattell |
| | G. Sigmund Freud |

Contoh 2 merupakan tes bentuk menjodohkan yang lebih baik karena perintah untuk mengerjakan tes lebih jelas dan mudah dipahami. Selain itu, materi yang diujikan juga homogen. Kemudian jawaban yang tersedia di kolom B juga lebih banyak daripada premis di kolom A. peserta didik juga mendapatkan gambaran yang jelas dimana mereka harus meletakkan jawaban yang dipilih.

2) Kelebihan dan kekurangan tes bentuk menjodohkan

Seperti bentuk tes objektif lainnya, tes bentuk menjodohkan juga memiliki kelebihan dan kekurangan. Beberapa kelebihan bentuk tes ini diantaranya adalah:

- a) Tes ini mudah dalam penyusunannya sehingga guru tidak memerlukan waktu yang lama mulai persiapan hingga tes siap digunakan.
- b) Mudah bagi guru untuk melakukan penilaian.

- c) Tes ini sangat cocok untuk menguji pengetahuan peserta didik yang berkaitan dengan definisi, konvensi, istilah dan sejenisnya.
- d) Tes ini dapat digunakan untuk mengukur kemampuan peserta didik dalam menghubungkan dua hal, baik yang berhubungan langsung maupun tidak langsung.
- e) Cocok untuk semua mata pelajaran.

Sedangkan kekurangan tes bentuk menjodohkan adalah sebagai berikut:

- a) Tes ini lebih cocok untuk sekedar mengukur kemampuan peserta didik pada aspek ingatan.
- b) Tidak cocok untuk mengukur kemampuan peserta didik pada level berpikir kritis.
- c) Tes ini sering memberi peluang bagi peserta didik untuk menebak jawaban yang benar dengan hanya mengenali kata kunci yang ada pada premis.

Walaupun tes bentuk ini lebih cocok untuk sekedar mengukur kemampuan peserta didik pada aspek ingatan, tes ini mampu memberikan variasi bentuk soal, terutama bagi peserta didik pada tingkat sekolah dasar.

c. Tes bentuk benar-salah

Tes benar-salah adalah sebuah tes yang menghendaki peserta didik untuk menentukan apakah sebuah pernyataan itu benar atau salah menurut kriteria tertentu. Oleh karena itu, tes ini pada umumnya berupa pernyataan-pernyataan yang mengandung aspek kebenaran atau kesalahan. Tes benar-salah secara umum terdiri dari dua jenis, yaitu tes benar-salah tanpa koreksi dan benar-salah dengan koreksi. Tes benar-salah tanpa koreksi adalah jenis tes dimana ketika peserta didik menentukan jawaban bahwa pernyataan itu salah, maka peserta didik tidak perlu melakukan pembetulan. Sebaliknya tes benar-salah dengan koreksi adalah tes yang menghendaki adanya pembetulan dari peserta didik jika mereka menentukan bahwa pernyataan itu salah.

Ada beberapa teknik yang harus diperhatikan oleh penulis atau pengembang tes benar-salah ketika ia memulai tugasnya untuk mengembangkan tes bentuk ini. Berikut adalah teknik dalam mengembangkan tes benar-salah:

- 1) Setiap pernyataan harus secara tegas menunjukkan kebenaran atau kesalahan.
- 2) Pernyataan harus ditulis secara tepat tanpa mengandung makna ganda.
- 3) Pernyataan dalam tes benar-salah bukan merupakan kutipan langsung dari sebuah teks tertentu.
- 4) Hindari menggunakan kata-kata tertentu yang memberi petunjuk bahwa pernyataan itu adalah benar atau salah. Kata-kata seperti '*semua, selalu, tidak pernah, tidak satupun, hanya*' biasanya memberi petunjuk bahwa pernyataan itu adalah salah. Sedangkan kata-kata seperti '*umumnya, kadan-kadang, biasanya, mungkin, sering*' biasanya memberi petunjuk bahwa pernyataan itu adalah benar.
- 5) Hindari menggunakan pernyataan yang mengandung makna negatif, karena hal itu memungkinkan akan mengecilkan hati peserta didik.
- 6) Apabila terpaksa menggunakan pernyataan kontroversial yang diungkapkan oleh otoritas atau pihak tertentu, maka pihak-pihak itu perlu disebut dalam pernyataan.

d. Tes bentuk jawaban singkat

Tes isian atau jawaban singkat adalah bentuk tes yang menghendaki peserta didik untuk menjawab pertanyaan tertentu hanya dalam bentuk kata, angka, frasa, kalimat, atau simbol tertentu dan jawaban itu dapat dikategorikan menjadi jawaban benar atau jawaban salah secara mutlak. Tes ini biasanya hanya mengukur kemampuan peserta didik untuk menyebutkan istilah, fakta, dan definisi; menginterpretasikan data sederhana, memecahkan soal-soal hitungan yang menghasilkan angka tertentu; memanipulasi simbol-simbol matematis, dan melengkapi persamaan matematis tertentu. Berikut adalah teknik penulisan soal jawaban singkat.

- 1) Butir tes harus sesuai dengan kompetensi yang ingin diukur.
- 2) Butir tes harus dirumuskan dengan menggunakan kalimat yang singkat dan mudah dimengerti oleh setiap peserta didik.
- 3) Rumusan butir tes hanya diarahkan untuk mengukur kemampuan peserta didik untuk mengingat fakta, angka, simbol, kata, frase, tahun, tempat, nama dan sebagainya.

- 4) Hindari butir tes berupa kutipan langsung dari suatu teks tertentu.
- 5) Jika butir tes berupa kalimat tidak lengkap karena ada kata tertentu yang dihilangkan, maka kata yang boleh dihilangkan hanyalah kata-kata penting saja (*significant word only*).
- 6) Panjang jawaban yang harus diberikan oleh peserta didik hanya terbatas pada satu kata atau frase.

B. Tes Subjektif

Menurut sejarahnya, tes subjektif sudah lebih dulu ada dan banyak digunakan oleh guru untuk mengukur hasil belajar peserta didik sebelum tes objektif muncul. Tes subjektif adalah sebuah bentuk tes dimana proses penilaiannya sering dipengaruhi oleh kesan atau pertimbangan pribadi oleh pihak-pihak yang melakukan proses penilaian. Tes ini menghendaki peserta didik untuk tidak sekedar menghafal dan mengingat kembali materi yang sudah dipelajari, tetapi lebih jauh tes ini juga menghendaki peserta didik untuk menjelaskan, memprediksikan, membedakan, membandingkan, mengkontraskan, mengkonversikan dan menghubungkan suatu informasi berdasarkan kemampuan analisis dan berpikir kritis yang dimiliki oleh peserta didik. Tes subjektif dapat digunakan untuk mengukur kemampuan peserta didik yang tidak dapat dilakukan dengan menggunakan tes objektif karena tes ini dapat diarahkan agar peserta didik melakukan eksplorasi dan diskusi secara kritis mengenai suatu permasalahan tertentu dengan menggunakan bahasa mereka sendiri. Jumlah butir pertanyaan dalam tes ini umumnya juga lebih sedikit dibandingkan dengan tes objektif dan biasanya hanya berjumlah 5 – 10 butir pertanyaan saja. Dengan demikian, respon yang diberikan oleh peserta didik terhadap pertanyaan-pertanyaan dalam tes subjektif dapat berupa uraian yang panjang, kritis, analitis, dan evaluatif.

Tes subjektif berpeluang lebih besar digunakan untuk mengukur kemampuan berpikir tingkat tinggi yang dimiliki oleh peserta didik. Brookhart (2010) lebih jauh menjelaskan bahwa kemampuan berpikir tingkat tinggi adalah kemampuan berpikir yang melibatkan paling tidak tiga kategori berpikir, yaitu kemampuan untuk melakukan transfer informasi, berpikir kritis, dan pemecahan masalah. Kemampuan melakukan transfer adalah sebuah kemampuan peserta didik yang tidak hanya sekedar mengingat materi yang telah dipelajari, tetapi juga bagaimana peserta didik mampu memaknai dan memanfaatkan materi yang telah dipelajari tersebut

dalam kehidupan nyata. Selanjutnya, kemampuan berpikir kritis dimaknai sebagai kemampuan untuk bernalar dan berpikir reflektif guna memutuskan sesuatu yang dipercayai untuk dilakukan. Sebagaimana diuraikan oleh Barahal (2008), berpikir kritis dapat juga dipahami sebagai proses berpikir yang melibatkan kemampuan untuk bernalar, bertanya, menyelidiki, mengamati dan mendeskripsikan, membandingkan dan menghubungkan, menemukan kompleksitas, dan mengeksplorasi suatu sudut pandang. Sedangkan, kemampuan untuk memecahkan masalah adalah kemampuan peserta didik untuk memahami suatu masalah pembelajaran yang dihadapi dan kemudian menemukan solusi yang tepat terhadap permasalahan tersebut dalam rangka mencapai tujuan pembelajaran yang diharapkan. Masalah pembelajaran ini dapat menghambat peserta didik untuk mencapai tujuan pembelajaran yang telah ditentukan, maka kemampuan memecahkan masalah menjadi hal yang sangat penting.

Ada beberapa faktor yang mempengaruhi subjektivitas guru dalam melakukan proses penilaian untuk tes subjektif, yaitu *halo effect*, *carry-over effect*, *generosity effect*, dan *recency effect*. *Halo effect* mengacu pada bias hasil penilaian yang diakibatkan adanya kesan pertama (positif atau negatif) yang dimiliki oleh guru terhadap peserta didik. *Halo effect* menggambarkan adanya kecenderungan bagi guru untuk memberi nilai tinggi atau rendah berdasarkan kesan (menarik, tidak sukai, rajin, tertib, dan sebagainya) terhadap seorang peserta didik. Selanjutnya *carry-over effect* adalah efek yang ‘dibawa’ dari suatu kondisi ke kondisi yang lain. Sebagai gambaran, *carry-over effect* terjadi jika guru telah memberikan nilai tinggi pada butir soal no. 1, maka ia juga akan cenderung memberi nilai tinggi pada nomor-nomor selanjutnya karena guru menganggap bahwa kualitas jawaban nomor-nomor selanjutnya juga sama dengan kualitas jawaban pada nomor sebelumnya yang telah diperiksa. Kemudian, *generosity effect* mengacu pada sikap guru yang ingin mendapatkan predikat guru yang baik dengan cara memberikan nilai yang tinggi kepada mayoritas peserta didik. *Recency effect* adalah bias penilaian yang diakibatkan oleh kecenderungan guru untuk selalu mengingat informasi terbaru yang disajikan oleh peserta didik.

Subjektivitas guru dalam melakukan penilaian tentu akan mengakibatkan informasi berkaitan dengan hasil belajar peserta didik menjadi bias dan tidak menggambarkan kompetensi yang sebenarnya. Lalu bagaimana cara guru mengurangi subjektivitasnya selama melakukan proses

penilaian? Ada beberapa cara yang dapat dilakukan untuk mengurangi subjektivitas guru dalam melakukan penilaian untuk tes subjektif, yaitu

1. Guru harus menyadari bahwa ada subjektivitas mungkin saja terjadi dan akan berpengaruh terhadap hasil penilaian. Kesadaran terhadap eksistensi subjektivitas ini penting karena tanpa adanya kesadaran ini bias penilaian akan terus terjadi.
2. Guru harus menyiapkan kriteria penilaian sebelum proses penilaian dilakukan. Kriteria ini sangat penting sebagai panduan bagi guru untuk meningkatkan objektivitas hasil penilaian.
3. Melibatkan lebih dari satu penilai (*inter-rater*). Hasil penilaian kemudian dibandingkan untuk mengetahui konsistensi hasil penilaian dari dua atau lebih penilai tersebut.
4. Dilakukan penilaian dua kali (*mark-remark*), yaitu guru yang sama melakukan penilaian terhadap tes yang sama dalam waktu yang berbeda, kemudian hasilnya disandingkan untuk mengetahui konsistensi hasil penilaian.

Beberapa cara untuk mengurangi subjektivitas tersebut dapat dilakukan, baik secara simultan atau hanya melibatkan satu atau dua poin saja mengingat keterbatasan waktu yang tersedia. Misalnya guru hanya berpedoman pada kriteria penilaian poin no. 2 saja tanpa harus menggunakan teknik *inter-rater* untuk mengurangi subjektivitasnya.

1. Jenis-jenis tes subjektif

Ada beberapa jenis tes subjektif yang biasa digunakan oleh guru untuk mengukur hasil belajar peserta didik. Dilihat dari cara menjawab pertanyaan dan cakupan materi yang harus diuraikan dalam jawaban peserta didik, tes subjektif dibedakan mejadi dua, yaitu tes uraian terbatas (*restricted-response test*) dan tes uraian bebas (*extended-response test*).

a. Tes uraian terbatas

Tes uraian terbatas adalah tes dimana peserta didik dalam menjawab pertanyaan-pertanyaan dibatasi oleh beberapa hal, seperti panjang-pendeknya jawaban, cakupan materi yang harus ada dalam jawaban, dan pengorganisasian jawaban. Dalam tes jenis ini, peserta didik cukup mengemukakan jawaban secara singkat dan jelas sesuai dengan kriteria yang

telah ditentukan oleh guru. Setiap peserta didik kemungkinan akan mengemukakan dan mengorganisasikan jawabannya sesuai dengan gaya dan bahasa masing-masing, tetapi setiap jawaban itu harus memuat pokok-pokok gagasan yang terdapat dalam sistematika kunci jawaban dan sesuai dengan batasan-batasan yang telah ditentukan dan dikehendaki oleh soal yang sedang mereka kerjakan.

Contoh soal uraian terbatas:

Berikan dan jelaskan tiga alasan mengapa tata kebiasaan baru (*new normal*) seharusnya atau tidak seharusnya diterapkan oleh pemerintah di era pandemi covid-19 ini!

b. Tes uraian bebas

Tes uraian bebas adalah tes yang dapat dijawab oleh peserta didik dengan cara apapun yang mereka inginkan. Peserta didik dapat menjawab pertanyaan seluas-luasnya sesuai dengan sistematika dan bahasa mereka sendiri. Latar belakang pengetahuan dan kemampuan peserta didik sangat mempengaruhi kedalaman jawaban yang mereka berikan. Dalam tes ini, peserta didik pun kemungkinan akan menjawab pertanyaan secara bebas dan memuat cakupan materi yang luas. Walaupun demikian, guru tetap harus menentukan patokan dan kriteria yang jelas sebagai acuan ketika mengoreksi setiap jawaban yang diberikan oleh peserta didik. Patokan dan kriteria ini juga dimaksudkan agar subjektivitas guru sebagai korektor soal tidak terlalu banyak bermain dalam menentukan skor atau nilai akhir.

Contoh soal uraian bebas:

Menurut anda, apakah pemerintah perlu memberikan bantuan sosial tunai (BST) kepada masyarakat yang terdampak pandemik Covid-19? Jelaskan alasan anda secara detail dan lengkap!

2. Kelebihan dan kekurangan tes subjektif

Seperti halnya bentuk tes lainnya, tes subjektif juga memiliki kelebihan dan kekurangan sebagai sebuah tes untuk mengukur hasil belajar peserta didik. Disha (2020) secara lengkap menguraikan kelebihan dan kekurangan dari tes subjektif ini. Diantara kelebihan tes subjektif adalah:

- a. Tes ini relatif mudah dalam proses pengembangan dan pelaksanaannya.
- b. Dapat digunakan untuk mengukur kemampuan peserta didik dalam mengorganisir dan menyajikan gagasannya secara logis dan koheren.
- c. Secara praktis, dapat digunakan untuk mengukur kemampuan peserta didik untuk semua mata pelajaran.
- d. Dapat digunakan untuk mengukur kemampuan peserta didik dalam mengorganisir gagasannya secara efektif, mengkritisi dan menginterpretasi suatu informasi dengan baik.
- e. Dapat digunakan untuk mengukur kemampuan peserta didik untuk berpikir logis dan bernalar kritis.
- f. Jawaban yang diberikan oleh peserta didik tidak harus dikategorikan sebagai jawaban yang benar-benar salah atau juga benar mutlak.
- g. Peserta didik tidak dimungkinkan menjawab soal dengan sekedar menebak.

Sedangkan kekurangan dari tes bentuk subjektif adalah sebagai berikut:

- a. Salah satu masalah serius dari tes subjektif adalah tes ini tidak mampu mencakup sampel materi yang luas. Jadi cakupan materi yang dapat diujikan melalui tes subjektif sangat terbatas.
- b. Untuk tes bidang bahasa, tes ini menghendaki pemilihan teks atau bacaan yang sesuai dengan level kognitif peserta didik.
- c. Seringkali penilai dalam memberikan skor dipengaruhi oleh tingkat keterbacaan tulisan tangan, ejaan, kerapian tulisan, tata-bahasa, panjang atau pendeknya jawaban dan sebagainya.
- d. Jawaban panjang yang diberikan oleh peserta didik sering mengakibatkan penskoran kurang valid dan kurang reliabel sehingga memiliki nilai prediksi yang kecil.
- e. Peserta didik memerlukan waktu yang cukup lama untuk menuliskan jawabannya dan proses pemeriksaan jawaban oleh guru juga memakan waktu dan melelahkan.
- f. Proses pemeriksaan jawaban hanya dapat dilakukan oleh pihak-pihak yang memiliki kompetensi yang memadai terhadap materi yang diujikan.

- g. Proses penilaian sering dipengaruhi *halo-effect* karena adanya kesan positif atau negatif terhadap peserta didik.
- h. Skor yang dihasilkan sering pula dipengaruhi oleh keberpihakan, bias pribadi, pandangan, dan cara memahami butir-butir pertanyaan dari peserta didik.

Uraian tentang kekurangan dari tes subjektif tersebut diharapkan memberikan pemahaman yang lebih baik kepada pihak-pihak pengembang tes. Tetapi secara ringkas dapat dipahami bahwa kekurangan atau kelemahan yang potensial dari tes bentuk subjektif adalah tes ini memiliki validitas prediktif yang lemah, sampel materi yang diujikan sangat terbatas, hasil penilaian sering tidak reliabel, dan adanya beberapa kendala penilaian.

3. Teknik meningkatkan kualitas tes subjektif

Guru sebagai pengembang tes subjektif seringkali menuliskan butir-butir pertanyaan tanpa memperhatikan bagaimana teknik yang benar dalam mengembangkan butir-butir pertanyaan tersebut. Disha (2020) memberikan arahan secara rinci bagaimana meningkatkan kualitas butir-butir pertanyaan tes subjektif agar tes itu benar-benar dapat dimanfaatkan oleh guru untuk menggali informasi berkaitan dengan kemampuan peserta didik, kesulitan dan cara berpikir peserta didik, dan kemudian informasi tersebut dapat digunakan sebagai dasar untuk memberikan bimbingan yang tepat terhadap proses belajar mereka. Selanjutnya, Disha (2020) menguraikan berbagai upaya untuk meningkatkan kualitas tes subjektif itu yang dimulai dari saat guru mengembangkan butir-butir pertanyaan hingga saat guru melakukan penskoran. Masing-masing dari upaya itu diuraikan secara lengkap pada bagian berikut.

a. Upaya perbaikan kualitas tes pada saat perumusan butir-butir pertanyaan

Ada beberapa upaya yang perlu dilakukan oleh pengembang tes pada saat menyusun butir-butir pertanyaan. Diantara upaya itu diuraikan sebagai berikut.

1. Sediakan waktu yang cukup untuk mempersiapkan dan menyusun butir-butir pertanyaan, sehingga guru dapat meneliti kembali, merevisi, dan mengedit butir-butir pertanyaan yang telah dikembangkan sebelum digunakan untuk mengukur hasil belajar

peserta didik. Hal ini juga sebagai upaya untuk meningkatkan validitas butir tes, terutama validitas isi.

2. Butir-butir pertanyaan ditulis dan diarahkan untuk memunculkan jenis perilaku tertentu dari peserta didik yang diinginkan oleh guru untuk diukur. Misalnya, jika guru ingin mengukur aspek pengetahuan, guru tidak boleh menulis butir pertanyaan untuk mengukur aspek analisis. Contoh:

“Bandingkan kondisi negara pada saat sebelum dan sesudah pandemi Covid-19 dilihat dari keadaan ekonomi makro.”

Butir soal tersebut tampaknya bukan untuk mengukur aspek pengetahuan, tetapi lebih dimaksudkan untuk mengukur aspek analisis.

3. Dalam merumuskan butir pertanyaan, gunakan kata-kata yang dimaksudkan untuk mengarahkan pencapaian perilaku tertentu, misalnya *definiskan, ilustrasikan, uraikan, klasifikasikan, pilihlah, ringkaslah* dan sebagainya. Hindari menggunakan kata-kata seperti *diskusikan, komentarkan, terangkan*, dan sebagainya.
4. Berikan perintah atau instruksi pengerjaan tes yang spesifik kepada peserta didik agar mereka mampu memberikan respon yang diinginkan.
5. Informasikan kepada peserta didik bobot setiap butir pertanyaan dan berapa lama waktu untuk menyelesaikan tes secara keseluruhan.
6. Jangan memberikan butir-butir pertanyaan opsional yang harus dikerjakan oleh peserta didik. Contoh:

“Kerjakan empat butir soal dari enam butir soal yang tersedia yang anda anggap lebih mudah!”

Instruksi tersebut harus dihindari karena beberapa alasan. Pertama, tidak semua butir pertanyaan memiliki tingkat kesulitan yang sama, sehingga jika ada peserta didik yang mengerjakan butir pertanyaan yang memiliki tingkat kesulitan tinggi, skor yang diperoleh mestinya berbeda dengan butir pertanyaan dengan kesulitan rendah. Kedua, peserta didik akan mengalami kesulitan untuk memilih butir-butir pertanyaan yang memungkinkan mereka

mampu memberikan jawaban terbaik. Ketiga, peserta didik yang pandai mungkin merasa dirugikan karena mereka biasanya tertantang untuk mengerjakan butir-butir pertanyaan yang memiliki tingkat kesulitan tinggi.

7. Sesuaikan kompleksitas dan panjang / pendeknya jawaban yang dikehendaki dengan tingkat kematangan berpikir peserta didik.
8. Pilihan kata yang digunakan dalam butir-butir pertanyaan harus jelas dan tidak ambigu.
9. Butir-butir pertanyaan harus diarahkan sebagai *power test* bukan sekedar *speed test*.
10. Butir-butir pertanyaan harus disusun dari yang memiliki tingkat kesulitan rendah menuju butir-butir pertanyaan dengan kesulitan tinggi.

b. Upaya perbaikan kualitas tes pada saat penskoran

Ada beberapa upaya yang dapat dilakukan oleh guru untuk meningkatkan kualitas dan efektivitas tes pada saat proses penskoran dilakukan.

1. Guru perlu mempersiapkan terlebih dulu kriteria penilaian sebagai panduan selama melakukan proses penskoran. Pada tahapan ini perlu ditetapkan pula bobot skor untuk setiap butir pertanyaan dan kata-kata kunci yang harus ada dalam jawaban yang diberikan oleh peserta didik.
2. Guru harus mentukan metode penskoran yang tepat untuk mengurangi bias penilaian; memperhatikan apakah aspek-aspek penting dan relevan muncul dalam jawaban peserta didik; berhati-hati terhadap subjektivitas personal yang akan mempengaruhi hasil penilaian; dan menerapkan standar yang sama untuk menilai jawaban dari setiap peserta didik.
3. Sedapat mungkin sembunyikan identitas peserta didik untuk menghindari *halo-effect* dan bias yang dapat mempengaruhi hasil penilaian.
4. Ketika proses penilaian telah dilakukan, hindari untuk mengubah kriteria penilaian yang telah ditetapkan agar setiap peserta didik mendapatkan hasil penilaian dengan kriteria yang sama.

5. Penilaian semata-mata difokuskan pada substansi materi yang ada pada jawaban yang diberikan peserta didik, bukan pada tata-bahasa, ejaan, kerapian tulisan, dan sejenisnya.
6. Jika memungkinkan, guru perlu melakukan teknik *mark-remark* untuk memastikan konsistensi dan *fairness* hasil penilaian.
7. Jika memungkinkan, guru perlu melakukan penilaian terhadap satu nomor yang sama dari keseluruhan lembar jawaban dalam satu waktu untuk menghindari terjadinya *halo-effect* karena saat itu guru dapat fokus hanya pada satu kriteria penilaian saja.

Bab 4

Karakteristik Tes yang Baik

Tes sebagai sebuah alat ukur hasil belajar peserta didik harus mampu memberikan informasi yang akurat, konsisten, jujur, dan dapat dipercaya. Tes yang baik akan mampu memberikan hasil pengukuran tentang capaian belajar peserta didik secara tepat, sehingga interpretasi terhadap hasil pengukuran tersebut tidak bias atau mencerminkan capaian kompetensi yang faktual. Sebaliknya, tes yang kurang baik dipastikan akan memberikan hasil pengukuran yang tidak akurat dan akibatnya interpretasi terhadap hasil pengukuran tersebut akan bias atau tidak mencerminkan kompetensi faktual yang dimiliki oleh peserta didik. Sebagai gambaran, seorang tukang kayu akan mengukur panjang dari beberapa balok kayu yang akan dipakai untuk membuat kaki sebuah meja. Ia memiliki empat buah balok yang akan dipotong dengan ukuran panjang yang sama. Saat mengukur keempat balok tersebut, ia tidak menggunakan alat ukur (meteran) standar. Tetapi ia hanya menggunakan rentang jarinya untuk mengukur panjang balok tersebut. Walaupun pengukuran dilakukan berkali-kali hasilnya sama, hasil pengukuran terhadap panjang balok tersebut dapat dipastikan tidak akurat. Analogi ini memberikan gambaran bahwa untuk mengukur hasil belajar peserta didik pun harus menggunakan alat ukur (tes) yang berkualitas agar hasil pengukuran tepat, sehingga informasi berkaitan dengan capaian hasil belajar peserta didik pun akurat.

Pertanyaan kemudian muncul, bagaimana karakteristik sebuah tes yang baik? Brown (2004) menguraikan bahwa sebuah tes yang baik paling tidak harus memenuhi beberapa kriteria, yaitu valid, reliabel, praktis, otentik, dan memiliki dampak terhadap proses belajar dan mengajar (*washback*). Selain itu, tes yang baik harus memiliki daya beda yang baik dan tingkat kesulitan yang tidak terlalu tinggi dan tidak terlalu mudah bagi

peserta didik. Beberapa karakteristik tes yang baik tersebut akan diuraikan lebih detail dalam bagian berikut.

A. Validitas Tes

Validitas tes mengacu pada sejauhmana sebuah tes mampu mengukur secara akurat apa yang dimaksudkan untuk diukur. Sebuah tes dikategorikan memiliki validitas yang tinggi apabila tes tersebut mampu menjalankan fungsinya dengan baik, yaitu mampu mengukur apa yang semestinya diukur atau hasil pengukuran dengan tes tersebut adalah akurat, dan memiliki kesalahan pengukuran yang kecil. Tes yang valid akan mampu memberikan informasi tentang kompetensi sebenarnya yang dimiliki oleh peserta didik. Ada hal yang perlu diperhatikan bahwa validitas suatu tes juga berkaitan erat dengan tujuan dari tes tersebut untuk mengukur apa. Suatu tes yang dikategorikan valid untuk mengukur kompetensi berbahasa seorang peserta didik, belum tentu tes itu masih dikategorikan valid untuk mengukur kompetensi lainnya. Sebagai analogi, sebuah timbangan yang biasa dipakai untuk menimbang berat badan seseorang akan mampu memberikan informasi yang valid berkaitan dengan berat badan orang tersebut. Tetapi timbangan itu tentu tidak akan mampu memberikan informasi yang valid jika dipakai untuk menimbang berat sebuah cincin emas. Oleh karena itu, tidak ada validitas yang secara universal berlaku untuk semua tujuan tes. Dengan demikian, dapat dipahami bahwa suatu alat ukur (termasuk tes) biasanya akan memberikan hasil pengukuran yang valid untuk satu tujuan pengukuran saja.

Adakalanya guru sebagai pengembang tes kurang memperhatikan langkah-langkah pengembangan tes dan prosedur penskoran yang baik, sehingga faktor-faktor tersebut telah mengakibatkan validitas tes akhirnya tidak dapat dicapai dengan baik pula. Lebih lanjut, Gronlund (1985) menyebutkan bahwa ada beberapa faktor yang dapat mempengaruhi validitas sebuah tes. Pertama, faktor instrumen. Jika instrumen (tes) yang digunakan untuk mengukur hasil belajar peserta didik bukan merupakan tes yang berkualitas baik, maka akurasi pengukuran hasil belajar peserta didik akan kurang baik pula. Kedua, faktor pelaksanaan tes. Kesalahan yang terjadi pada saat pelaksanaan tes, seperti alokasi waktu yang tersedia untuk mengerjakan tes; terjadinya kecurangan yang dilakukan oleh peserta tes; dan kondisi psikologis dan fisik peserta tes saat mengerjakan soal, akan

berdampak serius terhadap validitas hasil pengukuran. Ketiga, faktor penskoran. Kesalahan yang terjadi saat proses penskoran, seperti adanya *halo-effect*, *generosity-effect*, kelelahan dan *bad mood* yang terjadi pada penilai juga akan berpengaruh besar terhadap validitas tes. Keempat, faktor cara peserta tes dalam menjawab butir-butir pertanyaan. Faktor ini terjadi jika peserta didik tergesa-gesa dalam menjawab butir-butir pertanyaan yang tersedia; menjawab butir-butir pertanyaan dengan *trial and error*; dan sebagainya.

1. Jenis-jenis validitas tes

Secara umum, validitas tes dibedakan menjadi dua, yaitu validitas internal dan validitas eksternal. Validitas internal adalah sebuah tipe validitas dimana kriteria validitas instrumen ada dalam instrumen itu sendiri. Validitas ini meliputi beberapa jenis sub-validitas, yaitu validitas isi (*content validity*), validitas wajah (*face validity*), dan validitas konstruk (*construct validity*). Sedangkan validitas eksternal adalah sebuah tipe validitas dimana kriteria validitas instrumen didasarkan pada kriteria eksternal atau kriteria di luar instrumen itu sendiri. Pada umumnya kriteria ini didasarkan pada fakta-fakta empiris yang dapat dibuktikan secara matematis. Validitas eksternal meliputi validitas kurikuler (*curricular validity*), validitas terkait kriteria (*criterion-related validity*), validitas kesejajaran yang ada sekarang (*concurrent validity*), dan validitas prediktif (*predictive validity*). Bagian berikut akan menjelaskan secara lengkap berbagai macam validitas tersebut.

a. Validitas isi

Validitas isi mengacu pada sejauhmana butir-butir soal secara representatif mampu mewakili materi ajar dan perilaku yang akan diukur. Validitas isi merupakan jenis validitas yang sangat penting dan harus lebih diperhatikan oleh pengembang tes, karena validitas ini akan berpengaruh terhadap tercapainya jenis validitas yang lain. Jika elemen-elemen yang ada dalam seperangkat tes tidak relevan dengan konstruk utama yang akan diujikan, maka tes itu bisa dipastikan mengukur sesuatu yang lain sehingga sangat memungkinkan terjadi bias pengukuran. Dengan kata lain, jika seperangkat tes itu menguji materi di luar yang telah dipelajari oleh peserta didik, maka tes itu tidak memiliki validitas isi yang baik.

Validitas isi diperoleh dengan mengadakan sampling (uji petik) yang baik terhadap materi yang akan diujikan. Proses sampling dilakukan dengan memilih butir-butir pertanyaan yang dapat mewakili keseluruhan materi yang telah diajarkan kepada peserta didik. Kesalahan yang biasa terjadi dari upaya untuk mendapatkan validitas isi adalah guru dalam menentukan butir-butir pertanyaan tidak dilakukan secara sistematis, tetapi hanya didasarkan pada logika dan subjektivitas personal. Padahal kesesuaian antara materi dan butir-butir pertanyaan menjadi faktor utama untuk mencapai validitas isi. Seperti telah dijelaskan di bab sebelumnya bahwa pemilihan materi mana yang harus dipilih untuk diujikan kepada peserta didik didasarkan pada empat hal, yaitu materi itu adalah urgen (penting) untuk dikuasai oleh peserta didik, memiliki nilai mafaat yang tinggi (praktis), materi itu memiliki keterkaitan dengan mata pelajaran lain (relevan), dan materi itu merupakan prasyarat untuk dapat menguasai materi selanjutnya (kontinuitas). Untuk mencapai validitas isi, hal-hal yang harus pengembangan tes lakukan adalah menuliskan terlebih dulu kisi-kisi yang memuat kriteria butir-butir soal, deskripsi materi, dan kompetensi yang diinginkan. Selanjutnya, teknik untuk mengidentifikasi sejauhmana seperangkat tes memiliki validitas isi yang baik, pengembang soal cukup membandingkan setiap butir soal dengan indikator yang ada dalam kisi-kisi. Sekiranya setiap butir soal telah sesuai dengan indikator yang ada dalam kisi-kisi, berarti butir-butir soal itu telah memiliki validitas isi yang baik.

b. Validitas wajah

Validitas wajah atau validitas tampak merupakan jenis validitas yang paling rendah. Sebuah tes memiliki validitas wajah yang baik ketika kelihatan benar bagi orang lain (penguji, guru, atau peserta didik). Untuk menguji validitas ini, seseorang hanya mengandalkan kesan subjektif yang tampak dari tes yang telah dikembangkan. Sebagai gambaran, apabila pengembang tes ingin mengembangkan tes bentuk pilihan ganda dan tes yang dikembangkan tampak telah memiliki karakteristik tes bentuk pilihan ganda (ada stem dan beberapa alternatif jawaban) berarti tes tersebut telah dikategorikan memiliki validitas wajah yang baik. Contoh lain, jika tes yang dikembangkan telah terlihat sesuai dengan apa yang ingin diukur, maka tes tersebut sudah memiliki validitas wajah. Pada dasarnya validitas wajah mirip dengan validitas isi. Bedanya adalah validitas wajah melibatkan

penilaian terhadap kelayakan suatu validitas yang dilakukan oleh pengguna tes, sedangkan validitas isi menyangkut penentuan atau penilaian kelayakan suatu validitas yang dilakukan oleh para ahli. Validitas wajah akan meningkatkan motivasi peserta didik (*testee*) untuk berusaha keras dalam mengerjakan tes yang diberikan oleh guru, jika tes yang dikerjakan itu terlihat baik. Ada beberapa teknik untuk meningkatkan validitas wajah, yaitu:

- 1) Gunakan format tes yang disusun secara cermat;
- 2) Tuliskan butir-butir pertanyaan secara jelas dan mudah dipahami;
- 3) Cantumkan instruksi pengerjaan soal yang ringkas, tetapi jelas;
- 4) Pastikan bahwa peserta didik familier dengan butir-butir pertanyaan; dan
- 5) Pastikan bahwa butir-butir soal berkaitan dengan materi pelajaran yang telah dipelajari oleh peserta didik.

c. Validitas konstruk

Validitas konstruk adalah jenis validitas yang mengacu pada sejauhmana sebauah tes mampu benar-benar mengukur apa yang hendak diukur berkaitan dengan konsep atau definisi konseptual yang telah ditetapkan oleh pengembang tes. Tujuan dari validitas konstruk adalah memberikan bukti-bukti bahwa konstruk teori yang mendasari butir-butir soal itu dikembangkan dapat diukur secara valid. Sebagai contoh, seorang guru ingin mengukur tingkat kedisiplinan peserta didik dalam belajar dengan menggunakan kuesioner. Jika butir-butir kuesioner tersebut telah sesuai dengan konsep dan definisi tentang kedisiplinan, maka butir-butir kusioner tersebut telah memiliki validitas konstruk yang baik. Untuk mencapai validitas konstruk, guru harus mampu menyakinkan bahwa setiap indikator dan butir kuesioner benar-benar dikembangkan secara hati-hati berdasarkan pengetahuan yang relevan. Kuesioner yang dikembangkan harus mencakup hanya pertanyaan-pertanyaan yang relevan dengan berbagai indikator tentang kedisiplinan.

d. Validitas kurikuler

Validitas kurikuler mengacu pada sejauhmana butir-butir pertanyaan yang ada dalam seperangkat tes memuat berbagai materi ajar yang tercantum dalam sebuah kurikulum. Tes ini mengacu pada kurikulum yang spesifik, terdefinisi dengan baik, dan telah ditetapkan oleh pemerintah.

Dengan kata lain, tes itu mengukur apa saja yang ada di kurikulum yang digunakan. Pencapaian validitas kurikuler sebagai langkah penting jika seorang guru atau pihak-pihak tertentu ingin membuat kesimpulan tentang efektivitas sebuah program pembelajaran. Validitas kurikuler dianggap sebagai hal yang sangat penting karena dapat digunakan untuk memberikan gambaran terhadap tercapainya kompetensi minimal oleh setiap peserta didik dan kemudian akan digunakan sebagai bahan untuk membuat keputusan apakah seorang peserta didik itu lulus atau tidak setelah mengikuti keseluruhan program pembelajaran.

e. Validitas terkait kriteria

Validitas kriteria mengacu pada validitas suatu tes yang dicapai dengan cara membandingkan tes yang ada (tes yang dikembangkan oleh guru) dengan tes lain yang sudah terstandarisasi dan sudah diketahui kualitasnya baik (valid dan reliabel). Kedua tes tersebut kemudian dikorelasikan dan apabila korelasi keduanya dikategorikan signifikan, maka tes yang dikembangkan tersebut memiliki validitas kriteria yang baik. Masalah utama yang sering terjadi ketika ingin menguji validitas kriteria adalah tidak banyak tersedianya tes standar yang akan digunakan sebagai pembanding. Selain itu, beberapa tes terstandar yang tersedia dan digunakan sebagai pembanding tidak serta merta memberikan estimasi yang benar-benar akurat tentang nilai yang sebenarnya dari suatu fenomena. Validitas kriteria dapat dibedakan menjadi dua jenis validitas, validitas kesejajaran (*concurrent validity*) dan validitas prediktif (*predictive validity*).

f. Validitas kesejajaran

Validitas kesejajaran mengacu pada kemampuan sebuah tes untuk mengukur gejala tertentu pada saat sekarang dan kemudian tes itu dibandingkan dengan tes lain untuk konstruk yang sama. Validitas kesejajaran adalah salah satu bentuk validitas empiris, yaitu sebuah validitas yang dicapai karena sesuai dengan pengalaman. Sebagai gambaran untuk validitas ini, seorang peserta didik dari hasil pengukuran dikategorikan sebagai individu yang disiplin. Selanjutnya, berdasarkan pengalaman yang diamati dari kehidupan kesehariannya, memang peserta didik itu adalah individu yang disiplin. Hal ini berarti tes telah memiliki validitas kesejajaran yang baik. Contoh lain, seorang peserta didik dikatakan sebagai individu

yang inovatif jika berdasarkan pengalaman dapat dibuktikan bahwa peserta didik itu merupakan individu yang selalu menemukan dan mengemukakan gagasan-gagasan baru dan segar guna mengatasi kendala pembelajaran yang sedang dihadapinya. Dari uraian dan contoh-contoh tersebut dapat dipahami bahwa secara empiris validitas tidak dapat dicapai hanya sekedar menyusun tes berdasarkan ketentuan seperti halnya validitas lainnya, tetapi harus dibuktikan melalui pengalaman.

g. Validitas prediktif

Validitas prediktif adalah sebuah validitas suatu tes yang kondisinya sesuai dengan kriteria yang akan terjadi di masa yang akan datang. Sebuah tes dikatakan memiliki validitas prediktif jika tes tersebut memiliki kemampuan untuk memprediksikan apa yang kemungkinan terjadi di masa yang akan datang. Sebagai contoh, tes masuk perguruan tinggi. Apakah tes masuk perguruan tinggi itu mempunyai validitas prediktif atau tidak dapat ditentukan dengan mengkorelasikan hasil tes tersebut dengan prestasi belajar setelah menjadi mahasiswa. Biasanya hasil tes tersebut dikorelasikan dengan prestasi belajar yang diperoleh di semester satu atau dua. Jika tes tersebut memiliki korelasi yang signifikan, berarti tes tersebut memiliki validitas prediktif yang baik. Selanjutnya, tes yang memiliki validitas prediktif yang baik dapat digunakan untuk memprediksikan berapa lama masa studi mahasiswa tersebut untuk dapat menyelesaikan studinya dan seberapa tinggi indeks prestasi yang akan diperoleh selama studi di perguruan tersebut.

2. Uji validitas

Uji validitas tes adalah sebuah uji yang dilakukan untuk mengetahui tingkat ketepatan atau ketelitian sebuah tes yang digunakan untuk mengukur apa yang dimaksudkan untuk diukur. Secara spesifik, uji validitas dimaksudkan untuk menilai dan memutuskan apakah sebuah tes sebagai instrumen untuk mengukur hasil belajar telah tepat mengukur apa yang hendak diukur. Uji validitas butir-butir soal dilakukan dengan mengkorelasikan antara skor tiap butir soal dengan skor semua butir soal yang ada dalam seperangkat soal ujian. Dengan demikian, dapat dipahami bahwa uji validitas butir adalah sebuah uji yang dilakukan untuk menilai dan memutuskan apakah setiap butir soal yang ada dalam seperangkat soal ujian

dapat mendukung soal ujian itu sebagai satu kesatuan yang utuh. Tetapi, uji validitas dapat pula dilakukan untuk menilai apakah seperangkat soal ujian yang terdiri dari beberapa butir soal itu secara keseluruhan valid atau tidak. Jadi, uji validitas dapat dilakukan untuk menguji validitas setiap butir soal dan juga menguji validitas seperangkat tes secara keseluruhan.

Uji validitas dapat dilakukan salah satunya dengan menggunakan rumus korelasi *product moment*. Sedangkan untuk mencari koefisien korelasi *product moment*, ada dua cara yang dapat dilakukan, yaitu korelasi *product moment* dengan angka simpangan dan korelasi *product moment* dengan angka kasar. Rumus yang digunakan untuk mencari korelasi *product moment* dengan angka simpangan adalah sebagai berikut:

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

dimana :

- r_{xy} : koefisien validitas (korelasi *product moment*)
- x : nilai variabel x (nilai hasil tes yang akan diuji validitasnya)
- y : nilai variabel y (nilai hasil tes standar)

Sedangkan rumus yang digunakan untuk mencari korelasi *product moment* dengan angka kasar adalah sebagai berikut:

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

- r_{xy} : koefisien validitas (korelasi *product moment*)
- N : banyaknya peserta tes

Contoh:

Seorang guru bahasa Inggris telah mengembangkan seperangkat tes untuk mengukur kemampuan membaca teks berbentuk naratif. Tes dibuat dalam bentuk pilihan ganda dan dikerjakan oleh 10 peserta didik. Selanjutnya ia

ingin mengetahui validitas tes yang telah dikembangkan itu. Sebagai pembandingan, ia menggunakan hasil tes bahasa Indonesia yang dikerjakan oleh sejumlah peserta didik yang sama. Hasil tes bahasa Inggris (X) dan hasil tes bahasa Indonesia (Y) adalah sebagai berikut.

No.	Nama Peserta didik	X	Y
1.	Rony	7	6
2.	Agus	8	7
3.	Fatimah	6	8
4.	Slamet	7	7
5.	Rendra	8	7
6.	Edy	6	7
7.	Untung	7	8
8.	Titut	8	7
9.	Nina	7	7
10.	Tiwuk	7	8

Cara menghitung koefisien validitas menggunakan rumus korelasi *product moment* dengan angka simpangan adalah sebagai berikut:

No.	X	Y	x	Y	x ²	y ²	Xy
1	8	8	0,7	0,6	0,49	0,36	0,42
2	7	7	-0,3	-0,4	0,09	0,16	0,12
3	8	8	0,7	0,6	0,49	0,36	0,42
4	7	7	-0,3	-0,4	0,09	0,16	0,12
5	8	8	0,7	0,6	0,49	0,36	0,42
6	6	7	-1,3	-0,4	1,69	0,16	0,52
7	7	8	-0,3	0,6	0,09	0,36	-0,18
8	8	7	0,7	-0,4	0,49	0,16	-0,28
9	7	7	-0,3	-0,4	0,09	0,16	0,12
10	7	7	-0,3	-0,4	0,09	0,16	0,12
Σ	73	74			4,1	2,4	1,8
Rerata	7,3	7,4					

$$\text{Rata - rata } X = \frac{\sum X}{N} = \frac{73}{10} = 7,3$$

$$\text{Rata - rata } Y = \frac{\sum Y}{N} = \frac{74}{10} = 7,4$$

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{1,8}{\sqrt{4,1 \times 2,4}} = \frac{1,8}{3,14} = 0,573$$

Selanjutnya, cara menghitung koefisien validitas menggunakan rumus korelasi *product moment* dengan angka kasar adalah sebagai berikut:

No	X	Y	X ²	Y ²	XY
1	8	8	64	64	64
2	7	7	49	49	49
3	8	8	64	64	64
4	7	7	49	49	49
5	8	8	64	64	64
6	6	7	36	49	42
7	7	8	49	64	56
8	8	7	64	49	56
9	7	7	49	49	49
10	7	7	49	49	49
Σ	73	74	537	550	542

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

$$= \frac{10 \cdot 542 - 73 \cdot 74}{\sqrt{\{(10 \cdot 537 - 73^2)(10 \cdot 550 - 74^2)\}}} = \frac{18}{31,21} = 0,577$$

Dapat diamati dari hasil perhitungan koefisien validitas di atas, baik menggunakan *product moment* dengan angka simpangan maupun menggunakan *product moment* dengan angka kasar, keduanya menghasilkan koefisien yang hampir sama. Perbedaannya sangat kecil dan perbedaan ini sangat wajar karena adanya pembulatan angka di belakang koma.

Ada dua cara untuk menafsirkan koefisien validitas dari hasil perhitungan dengan menggunakan rumus *product moment* tersebut. Pertama,

dengan menggunakan angka kesetaraan (*equivalents*). Berikut adalah ekuivalen yang biasa digunakan untuk menafsirkan koefisien validitas. Koefisien validitas antara:

0,80 - 1,00	: validitas sangat tinggi
0,60 - 0,80	: validitas tinggi
0,40 - 0,60	: validitas cukup
0,20 - 0,40	: validitas rendah
0,00 - 0,20	: validitas sangat rendah

Sebagai contoh, hasil uji validitas tes bahasa Inggris diatas diperoleh koefisien validitas sebesar 0,577 dan dapat disimpulkan bahwa tes tersebut memiliki tingkat validitas cukup baik.

Kedua, dengan cara membandingkan hasil perhitungan r_{xy} dengan harga kritik yang ada dalam tabel *r product moment*. Jika harga hitung r_{xy} lebih kecil daripada harga kritik dalam tabel *r product moment*, maka dapat disimpulkan bahwa korelasi dari kedua tes tersebut adalah tidak signifikan, sehingga tes tersebut dikategorikan tidak valid. Sebaliknya, Jika harga hitung r_{xy} lebih besar atau sama dengan harga kritik dalam tabel *r product moment*, maka dapat disimpulkan bahwa korelasi dari kedua tes tersebut adalah signifikan, sehingga tes tersebut dikategorikan valid. Sebagai contoh, harga hitung r_{xy} tes bahasa Inggris diatas adalah 0,577 dan harga kritik dalam tabel dengan $N = 10$ pada taraf signifikansi 5% adalah 0,576. Dengan demikian, $r_{xy} (0,577) > \text{harga } r \text{ tabel } (0,576)$ dan dapat disimpulkan bahwa ada korelasi yang signifikan dari kedua tes tersebut dan tes dikategorikan valid.

Hal yang harus digarisbawahi adalah perhitungan diatas digunakan untuk uji validitas seperangkat tes secara keseluruhan (seperangkat tes yang terdiri dari banyak butir sebagai satu kesatuan yang utuh), bukan untuk untuk menguji validitas butir per butir. Pada prinsipnya, langkah-langkah uji validitas tes secara keseluruhan dengan uji validitas butir adalah sama. Keduanya sama-sama dapat menggunakan rumus korelasi *product moment*. Bisa juga menggunakan rumus yang lain, misalnya korelasi biserial. Uji validitas butir dilakukan dengan mengkorelasikan tiap butir soal dengan skor total yang diperoleh dari hasil sebuah ujian. Sebagai contoh, seorang guru

berusaha mengidentifikasi apakah setiap butir soal pilihan ganda yang telah dikembangkan valid atau tidak. Setiap butir yang dijawab benar oleh peserta didik diberi simbol angka 1 dan butir yang dijawab salah diberi simbol angka 0. Data dari hasil ujian disajikan sebagai berikut.

No	Nama	Nomor Butir Soal										Skor Total
		1	2	3	4	5	6	7	8	9	10	
1	Rony	1	1	0	0	1	1	0	1	1	0	6
2	Agus	0	1	1	1	1	1	1	0	0	0	6
3	Fatimah	0	1	0	1	1	0	0	1	1	0	5
4	Slamet	1	0	0	0	1	1	1	0	1	1	6
5	Rendra	1	0	0	1	1	1	0	0	0	1	5
6	Edy	0	1	1	1	1	1	0	0	0	1	6
7	Untung	0	0	0	0	0	0	1	0	0	1	2
8	Titut	1	1	0	0	0	0	0	1	1	1	5
9	Nina	0	1	1	1	1	0	0	1	1	0	6
10	Tiwuk	0	0	0	0	1	1	1	0	1	1	5
Σ		4	6	3	5	8	6	4	4	6	6	52

Validitas butir no. 1 dapat dihitung dengan cara sebagai berikut.

No.	X	Y	X^2	Y^2	XY
1	1	6	1	36	6
2	0	6	0	36	0
3	0	5	0	25	0
4	1	6	1	36	6
5	1	5	1	25	5
6	0	6	0	36	0
7	0	2	0	4	0
8	1	5	1	25	5
9	0	6	0	36	0
10	0	5	0	25	0
Σ	4	52	4	284	22

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

$$= \frac{10 \cdot 22 - 4 \cdot 52}{\sqrt{\{(10 \cdot 4 - 4^2)(10 \cdot 284 - 52^2)\}}} = \frac{12}{57,131} = 0,210$$

Dengan demikian, dapat disimpulkan bahwa butir soal no. 1 yang memiliki koefisien validitas 0,210 dikategorikan memiliki tingkat validitas rendah. Jika diinterpretasikan dengan membandingkan antara r hitung dengan r tabel, karena r_{xy} hitung (0,210) lebih kecil daripada harga kritik dalam tabel *product moment* (0,576) untuk $N=10$ dan taraf signifikansi 5%. Dengan demikian, r_{xy} (0,210) < r tabel (0,576) dan dapat disimpulkan bahwa tidak ada korelasi yang signifikan dari skor butir soal no.1 dengan skor total sehingga butir tersebut dikategorikan tidak valid. Untuk validitas butir-butir yang lain (no. 2 – 10) dapat dihitung dengan cara yang sama seperti butir no. 1 tersebut.

Langkah-langkah dan rumus yang digunakan di atas adalah langkah-langkah dan rumus untuk menghitung koefisien validitas untuk soal pilihan ganda. Untuk soal uraian digunakan tahapan dan rumus berikut.

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

Hasil hitung dengan menggunakan rumus tersebut kemudian dimasukkan ke dalam rumus uji-t (*t-test*) sebagai berikut.

$$t = r_{xy} \sqrt{\frac{N - 2}{1 - r_{xy}^2}}$$

Contoh:

Berapakah koefisien validitas butir soal no. 1 mata pelajaran bahasa Inggris yang diujikan kepada 10 peserta didik? Data hasil ujian disajikan dalam tabel sebagai berikut.

No	Nama	Nomor Butir Soal					Skor Total
		1	2	3	4	5	
1	Rony	4	5	5	4	6	24
2	Agus	6	7	6	6	7	32
3	Fatimah	7	8	7	7	8	37
4	Slamet	3	9	9	9	6	36
5	Rendra	5	8	7	5	9	34
6	Edy	6	7	9	7	8	37
7	Untung	8	10	5	8	7	38
8	Titut	10	9	7	9	8	43
9	Nina	9	7	8	6	9	39
10	Tiwuk	5	8	9	8	8	38

Tahap pertama untuk menghitung validitas butir no. 1 terlebih dulu dibuat tabel persiapan seperti berikut.

No	Nama	X	Y	X ²	Y ²	XY
1	Rony	4	24	16	576	96
2	Agus	6	32	36	1024	192
3	Fatimah	7	37	49	1369	259
4	Slamet	3	36	9	1296	108
5	Rendra	5	34	25	1156	170
6	Edy	6	37	36	1369	222
7	Untung	8	38	64	1444	304
8	Titut	10	43	100	1849	430
9	Nina	9	39	81	1521	351
10	Tiwuk	5	38	25	1444	190
Σ		63	358	441	13048	2322

Selanjutnya angka-angka dalam tabel tersebut dimasukkan ke rumus berikut:

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

$$= \frac{(10 \times 2322) - (63 \times 358)}{\sqrt{(10 \times 441 - 63^2)(10 \times 13048 - 358^2)}} = \frac{23220 - 22554}{\sqrt{2316}}$$

$$= \frac{666}{1010,62} = 0,659$$

Selanjutnya hasil hitung tersebut dimasukkan ke rumus *t-test* sebagai berikut.

$$t = r_{xy} \sqrt{\frac{N - 2}{1 - r_{xy}^2}} = 0,659 \sqrt{\frac{10 - 2}{1 - 0,659^2}}$$

$$= 0,659 \times \sqrt{\frac{8}{0,341}} = \sqrt{23,460} = 4,843$$

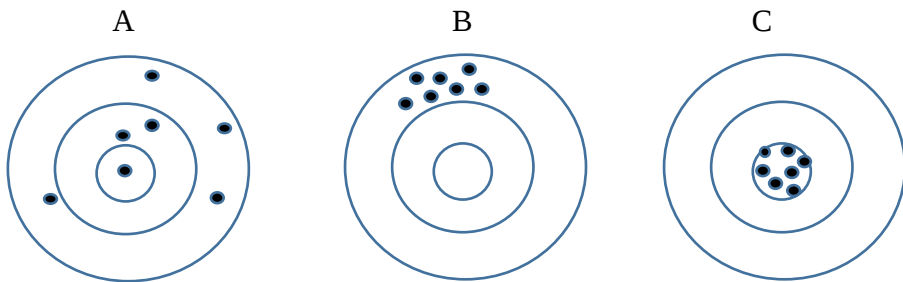
Hasil hitung *t-test* tersebut kemudian dibandingkan dengan nilai *t* tabel. Sebuah butir tes uraian dikategorikan valid apabila nilai *t* hitung lebih besar dari *t* tabel ($t_{hitung} > t_{tabel}$). Berdasarkan data diatas, tabel distribusi *t* pada derajat kebebasan $n - 2$ dan taraf signifikansi 5% diperoleh angka sebesar 1,85955. Hal ini berarti *t* hitung lebih besar dari *t* tabel ($4,843 > 1,85955$). Dengan demikian, dapat disimpulkan bahwa butir soal uraian no. 1 tersebut adalah valid dan butir tersebut mampu mengukur apa yang dimaksudkan untuk diukur.

B. Relibilitas Tes

Reliabilitas mengacu pada konsistensi skor atau jawaban yang diperoleh atau diberikan oleh peserta didik (*testee*) dari satu waktu pelaksanaan tes dan waktu pelaksanaan lainnya. Jika seorang guru memberikan tes yang sama kepada peserta didik yang sama dalam dua waktu pelaksanaan tes yang berbeda dan hasil dari dua pelaksanaan tes tersebut menghasilkan skor yang relatif sama, maka tes itu dikategorikan memiliki reliabilitas yang baik. Brown (2004) lebih lanjut menyatakan bahwa reliabilitas mengacu pada konsistensi atau kehandalan suatu tes untuk mengukur hasil belajar peserta didik. Selanjutnya, reliabilitas dapat juga dipahami bahwa tes yang digunakan untuk mengukur hasil belajar peserta

didik merupakan alat ukur yang dapat dipercaya dan dapat memberikan informasi yang sebenarnya tentang hasil belajar peserta didik.

Reliabilitas dan validitas merupakan dua hal yang berbeda, tetapi keduanya saling berkaitan. Kata kuncinya adalah reliabilitas menghendaki adanya konsistensi hasil penskoran, sedangkan validitas mengacu pada akurasi dari suatu tes untuk mengukur apa yang hendak diukur. Reliabilitas merupakan prasyarat agar sebuah tes menjadi valid. Tetapi sebuah tes yang reliabel tidak menjamin sebuah tes itu juga valid. Keduanya menjadi tolok ukur yang sangat penting agar sebuah tes dapat dikategorikan sebagai sebuah tes yang baik sebagai alat ukur hasil belajar peserta didik. Untuk memperjelas pemahaman tentang perbedaan reliabilitas dan validitas dapat dicermati ilustrasi berikut ini.



Gambar 4.1

Ilustrasi hubungan antara validitas dan reliabilitas

Gambar A pada ilustrasi diatas dapat dipahami bahwa tes yang digunakan untuk mengukur hasil belajar peserta didik dikategorikan sebuah tes yang tidak valid dan tidak reliabel. Ibarat seseorang yang membidikkan anak panah ke bidang sasaran, gambar A pada ilustrasi diatas menunjukkan bahwa anak panah tidak konsisten karena tidak mengenai satu titik pada bidang sasaran, tetapi menyebar di semua bagian bidang sasaran (tidak reliabel), sehingga anak panah tersebut tidak akurat mengenai sasaran yang diharapkan (tidak valid). Berikutnya, gambar B merupakan ilustrasi tes yang reliabel, tetapi tidak valid. Hal itu ditunjukkan bahwa semua anak panah secara konsisten mengenai bidang sasaran dalam satu titik. Gambar C mengilustrasikan sebuah tes yang reliabel dan valid. Hal itu dibuktikan dengan konsistensi anak panah yang konsisten mengenai satu titik pada

bidang sasaran dan secara akurat mengenai area yang memiliki nilai tinggi pada permainan panahan.

1. Faktor-faktor yang mempengaruhi reliabilitas tes

Tinggi atau rendahnya reliabilitas suatu tes dapat dipengaruhi oleh beberapa faktor. Gronlund (1985) mengidentifikasi ada empat faktor yang mempengaruhi reliabilitas tes. Pertama, faktor banyaknya butir soal (*test length*). Ada kecenderungan, semakin banyak butir soal, semakin reliabel tes tersebut. Ini terjadi karena semakin banyak butir soal, semakin banyak cakupan materi yang diujikan dan proporsi jawaban benar yang diberikan oleh peserta didik kemungkinan akan meningkat sehingga kesempatan faktor tebakan akan menurun. Kedua, faktor distribusi skor. Jumlah distribusi skor akan berpengaruh terhadap tingkat reliabilitas suatu tes, karena koefisien reliabilitas yang tinggi diperoleh ketika peserta didik tetap berada pada posisi yang relatif sama dalam kelompoknya untuk satu pengujian dan pengujian berikutnya. Ketiga, faktor tingkat kesulitan butir soal. Butir-butir soal yang terlalu sulit atau terlalu mudah kemungkinan akan menghasilkan reliabilitas yang rendah pula. Butir-butir soal seperti itu akan menghasilkan daya beda kecil pula. Tingkat kesulitan ideal akan diperoleh jika butir-butir soal menghasilkan distribusi skor berbentuk kurva normal. Keempat, faktor objektivitas. Objektivitas mengacu pada peluang bagi setiap peserta didik untuk mendapat skor yang sama jika ia memiliki kemampuan yang sama dan mengerjakan tes yang sama.

Brown (2004) lebih lanjut menguraikan ada beberapa faktor yang mempengaruhi sebuah tes tidak reliabel. Pertama, faktor terkait peserta didik. Peserta didik yang pada saat mengikuti tes berada dalam kondisi sakit, kelelahan, cemas, dan memiliki gangguan fisik atau psikologis tertentu akan menghasilkan skor yang tidak reliabel. Kedua, faktor terkait penilai. Subjektfitas guru dalam memberikan skor kepada setiap peserta didik akan berengaruh pada reliabilitas hasil tes. Ketiga, faktor terkait pelaksanaan tes. Kondisi selama pelaksanaan tes (ramai, berisik, gaduh) juga akan mempengaruhi reliabilitas tes. Keempat, faktor terkait tes itu sendiri. Tes yang memiliki terlalu banyak butir soal akan menyebabkan peserta didik kelelahan dan akhirnya akan buru-buru dalam menjawab setia butir soal. Hal ini tentu akan berpengaruh terhadap reliabilitas tes.

2. Metode pengujian reliabilitas

Ada tiga cara untuk menguji reliabilitas sebuah tes, yaitu dengan metode pengujian berulang (*test-retest method*), metode tes paralel (*parallel forms*), dan metode belah dua (*split-half method*). Ketiga metode pengujian reliabilitas tersebut memiliki karakteristik yang berbeda dan prosedur pengujiannya pun berbeda.

a. Metode pengujian berulang (*test-retest method*)

Metode ini dilakukan ketika seorang guru ingin mengetahui konsistensi hasil pengukuran atau pengujian dengan menggunakan tes yang telah dikembangkan sebelumnya. Proses uji reliabilitas dilakukan dengan melakukan pengujian dua kali dengan menggunakan tes yang sama dan kelompok peserta didik yang sama pula. Pengujian dilakukan dalam lintas waktu yang berbeda. Jarak pengujian pertama dan kedua sebaiknya tidak terlalu lama, biasanya dalam rentang waktu tidak lebih dari dua minggu. Hasil pengujian pertama dan kedua kemudian dibandingkan. Jika seorang peserta didik mendapatkan skor tinggi pada ujian pertama, dan hasil yang sama tinggi pula pada ujian kedua, dan juga sebaliknya, berarti tes tersebut memiliki reliabilitas yang baik.

b. Metode tes paralel (*parallel forms to the same group*)

Metode ini dilakukan dengan mengembangkan dua perangkat tes yang paralel atau sama. Jadi, untuk uji reliabilitas dengan metode ini, seorang guru harus mengembangkan terlebih dulu dua set (perangkat atau paket) tes yang memiliki cakupan materi, tingkat kesulitan, jumlah butir soal, dan rubrik penilaian yang sama. Kedua set soal tersebut kemudian diujikan kepada kelompok peserta didik yang sama pula, tetapi pengujian dilakukan dalam rentang waktu yang berbeda. Hasil pengujian dari kedua perangkat tes tersebut kemudian dibandingkan untuk mengetahui konsistensinya. Kedua tes tersebut dianggap handal apabila keduanya mampu memberikan informasi dalam bentuk skor yang konsisten untuk kelompok peserta didik yang sama. Sebaliknya, kedua tes tersebut dianggap tidak dapat diandalkan apabila keduanya tidak mampu memberikan informasi dalam bentuk skor yang konsisten untuk kelompok peserta didik yang sama.

c. Metode belah dua (*split-half method*)

Metode ini dilakukan dengan membelah satu set atau perangkat tes menjadi dua bagian. Metode ini juga biasa disebut dengan metode untuk mengetahui konsistensi internal dari sebuah perangkat tes karena uji reliabilitas dilakukan hanya pada satu set tes untuk sekali pengujian. Metode belah dua merupakan metode uji reliabilitas yang paling banyak diikuti oleh guru dan pihak-pihak yang bergerak pada bidang penilaian dan pengukuran. Allen & Yan (1979) menjelaskan bahwa ada tiga metode pembelahan tes yang dapat dilakukan untuk membelah tes menjadi dua bagian. Pertama, tes dibelah berdasarkan nomor butir ganjil dan nomor butir genap (*odd and even method*). Belahan pertama adalah butir-butir soal nomor ganjil dan belahan kedua adalah butir-butir soal dengan nomor genap. Kedua, metode pembelahan berdasarkan nomor urut butir-butir soal (*order method*). Misalnya, guru telah mengembangkan tes pilihan ganda yang terdiri dari 50 butir soal. Proses pembelahan dilakukan dengan memilih nomor 1 – 25 sebagai belahan pertama dan nomor 26 – 50 sebagai belahan kedua. Metode ini memiliki kelemahan dibandingkan dengan metode ganjil-genap. Ada kemungkinan peserta didik tidak dapat menyelesaikan nomor-nomor terakhir karena kehabisan waktu, sehingga belahan pertama dan kedua menjadi tidak berimbang. Kelemahan kedua terjadi karena tingkat kesulitan butir-butir soal pada belahan pertama dan kedua juga tidak berimbang karena biasanya butir-butir soal pada nomor-nomor awal adalah butir-butir soal yang lebih mudah dibandingkan nomor-nomor akhir. Ketiga, metode pembelahan acak berpasangan (*matched-random subtest*). Metode ini lebih rumit dibandingkan dua metode sebelumnya dan sangat jarang digunakan sebagai dasar pembelahan butir-butir soal menjadi dua bagian. Rumit karena proses pembelahan didasarkan pada hasil perhitungan statistik setiap butir soal. Skor pada belahan pertama dan kedua kemudian dikorelasikan untuk mengetahui sejauhmana koefisien reliabilitas dari tes tersebut.

3. Uji reliabilitas tes

Tinggi atau rendahnya reliabilitas sebuah tes, secara empirik ditunjukkan dalam bentuk angka yang biasa disebut dengan koefisien reliabilitas. Reliabilitas yang tinggi ditunjukkan dengan adanya korelasi yang signifikan antara perangkat tes yang satu dengan lainnya (*parallel*

forms), antara belahan pertama dan belahan kedua (*split-half*), atau antara tes hasil pengujian pertama dan kedua (*test-retest*) dan ditunjukkan dengan koefisien reliabilitas yang mendekati angka 1. Koefisien reliabilitas yang semakin mendekati angka 0, berarti semakin tidak reliabel tes tersebut. Secara umum, sebuah tes dikatakan reliabel jika tes tersebut memiliki koefisien reliabilitas paling tidak sebesar 0,7 (Kaplan & Saccuzzo, 1982). Ada beberapa cara yang dapat dilakukan untuk melakukan uji reliabilitas tes, misalnya dengan menggunakan rumus *Kuder-Richardson 20* (KR-20), *Kuder-Richardson 21* (KR-21), rumus Flanagan, rumus rulon, rumus Hoyt, dan Spearman-Brown. Dalam buku ini, contoh uji reliabilitas secara khusus diberikan dengan menggunakan rumus spearman-Brown. Langkah-langkah untuk mencari koefisien reliabilitas dengan menggunakan rumus Spearman-Brown diuraikan sebagai berikut.

Tahap pertama yang harus dilakukan adalah membelah tes menjadi dua bagian, misalnya belahan pertama adalah butir-butir soal dengan nomor ganjil dan belahan kedua adalah butir-butir soal dengan nomor genap. Selanjutnya menghitung korelasi skor dari kedua belahan tes tersebut dengan rumus sebagai berikut:

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

Rumus diatas adalah dasarnya adalah rumus *product moment* yang digunakan untuk mengetahui korelasi dari dua belahan tes, belum melihat reliabilitas tes secara keseluruhan. Selanjutnya, untuk mengetahui reliabilitas tes secara keseluruhan, hasil perhitungan dengan menggunakan rumus diatas kemudian dimasukkan ke rumus Spearman-Brown sebagai berikut:

$$r_{nnr} = \frac{2r_{xy}}{1 + (n - 1)r_{xy}}$$

n = 2, karena pada dasarnya jumlah belahan selalu dua (belahan pertama dan belahan kedua)

Secara umum, para ahli mengkategorikan koefisien reliabilitas menjadi beberapa kategori, yaitu reliabilitas sangat tinggi, reliabilitas tinggi, reliabilitas sedang, reliabilitas rendah, dan reliabilitas sangat rendah (tidak

reliabel). Kategori tersebut didasarkan pada koefisien reliabilitas yang diperoleh dari hasil perhitungan dengan menggunakan rumus-rumus reliabilitas yang telah dikembangkan oleh para ahli pula (rumus *Kuder-Richardson 20* (KR-20), *Kuder-Richardson 21* (KR-21), rumus Flanagan, rumus rulon, rumus Hoyt, dan Spearman-Brown dan sebagainya). Sebuah referensi klasik oleh Guilford mengkategorikan kriteria reliabilitas sebagai berikut. Koefisien reliabilitas antara:

0,80 - 1,00	: reliabilitas sangat tinggi
0,60 - 0,79	: reliabilitas tinggi
0,40 - 0,59	: reliabilitas sedang
0,20 - 0,39	: reliabilitas rendah
0,00 - 0,19	: reliabilitas sangat rendah (tidak reliabel)

Contoh uji reliabilitas untuk soal pilihan ganda yang terdiri dari 10 butir soal dan dikerjakan oleh 10 peserta didik. Setiap butir yang dijawab benar diberi skor 1 dan butir yang dijawab salah diberi skor 0.

No	Nama	Nomor Butir Soal									
		1	2	3	4	5	6	7	8	9	10
1	Rony	1	1	0	0	1	1	0	1	1	0
2	Agus	0	1	1	1	1	1	1	0	0	0
3	Fatimah	0	1	0	1	1	0	0	1	1	0
4	Slamet	1	0	0	0	1	1	1	0	1	1
5	Rendra	1	0	0	1	1	1	0	0	0	1
6	Edy	0	1	1	1	1	1	0	0	0	1
7	Untung	0	0	0	0	0	0	1	0	0	1
8	Titut	1	1	0	0	0	0	0	1	1	1
9	Nina	0	1	1	1	1	0	0	1	1	0
10	Tiwuk	0	0	0	0	1	1	1	0	1	1

Data berupa skor tersebut kemudian dibelah menjadi dua bagian, belahan pertama adalah total skor butir-butir soal nomor ganjil (X) dan belahan kedua adalah total skor butir-butir soal nomor genap (Y).

No.	Nama	X	Y	x	y	x ²	y ²	xy
1	Rony	3	3	0,5	0,3	0,25	0,09	0,15
2	Agus	3	3	0,5	0,3	0,25	0,09	0,15
3	Fatimah	2	3	-0,5	0,3	0,25	0,09	-0,2
4	Slamet	4	2	1,5	-0,7	2,25	0,49	-1,1
5	Rendra	2	3	-0,5	0,3	0,25	0,09	-0,2

6	Edy	2	4	-0,5	1,3	0,25	1,69	-0,7
7	Untung	1	1	-1,5	-1,7	2,25	2,89	2,55
8	Titut	2	3	-0,5	0,3	0,25	0,09	-0,2
9	Nina	3	3	0,5	0,3	0,25	0,09	0,15
10	Tiwuk	3	2	0,5	-0,7	0,25	0,49	-0,4
Σ		25	27			6,5	6,1	0,5
Rerata		2,5	2,7					

Tahap selanjutnya adalah memasukkan data dalam tabel tersebut ke rumus korelasi *product moment*.

$$r_{xy} = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} = \frac{0,5}{\sqrt{6,5 \cdot 6,1}} = \frac{0,5}{6,296} = 0,08$$

Hasil perhitungan tersebut pada dasarnya baru menggambarkan korelasi dua belahan tes, belum menggambarkan reliabilitas tes secara keseluruhan. Untuk mendapatkan gambaran reliabilitas tes secara keseluruhan, hasil perhitungan tersebut kemudian dimasukkan ke rumus Spearman-Brown.

$$r_{nn'} = \frac{2r_{xy}}{1 + (n-1)r_{xy}} = \frac{2 \cdot 0,08}{1 + (2-1)0,08} = \frac{0,16}{1,08} = 0,15$$

Dari hasil perhitungan dapat diketahui bahwa koefisien reliabilitas adalah 0,15. Koefisien ini lebih kecil dari harga kritis reliabilitas 0,7 ($0,15 < 0,70$), maka dapat disimpulkan tes tersebut secara keseluruhan tidak reliabel. Interpretasi hasil perhitungan koefisien reliabilitas dapat pula dilakukan dengan membandingkan hasil perhitungan tersebut dengan tabel *r product moment* pada taraf signifikansi 1% atau 5%. Apabila hasil hitung lebih besar atau sama dengan *r* tabel ($r \text{ hitung} \geq r \text{ tabel}$), maka dapat disimpulkan ada korelasi signifikan antara butir-butir soal tersebut secara keseluruhan dan tes tersebut dikategorikan reliabel. Sebaliknya, apabila hasil hitung lebih kecil daripada *r* tabel ($r \text{ hitung} < r \text{ tabel}$), maka dapat disimpulkan tidak ada korelasi signifikan antara butir-butir soal tersebut secara keseluruhan dan tes tersebut dikategorikan tidak reliabel. Untuk hasil analisis tes diatas dengan $N = 10$ (N = jumlah peserta tes) dan pada taraf signifikansi 5% diketahui harha tabel adalah 0,576. Karena 0,15 lebih kecil dari harga tabel ($0,15 < 0,576$), maka dapat disimpulkan tes tersebut secara keseluruhan tidak reliabel.

Hal yang harus diperhatikan adalah langkah-langkah perhitungan seperti diuraikan diatas adalah hanya untuk tes objektif (pilihan ganda, benar-salah) yang memiliki data atau skor diskrit. Skor diskrit adalah skor dikotomis atau skor yang hanya didasarkan pada jawaban yang dapat dikategorikan benar mutlak dan salah mutlak. Jawaban peserta didik yang benar mutlak diberi skor 1 dan jawaban salah diberi skor 0. Untuk tes yang menghasilkan skor non-diskrit, yaitu tes yang penskoranya bukan 1 dan 0, tetapi berjenjang dari skor terendah hingga tertinggi untuk setiap butir soal, rumus dan langkah-langkah perhitungan koefisien reliabilitanya akan berbeda. Skor non-diskrit biasanya terjadi pada tes bentuk uraian yang skor untuk setiap butirnya bergerak dari 0 – 10 atau model penjenjangan yang lain. Skor non-diskrit dapat juga diterapkan untuk angket, skala *Likert*, dan skala bertingkat (*rating scale*). Perhitungan koefisien reliabilitas untuk skor non-diskrit ini biasanya menggunakan rumus *Alpha* seperti berikut ini.

$$r_{nn'} = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right)$$

$$\sigma_i^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{N}}{N}$$

$$\sigma_t^2 = \frac{\sum X_t^2 - \frac{(\sum X_t)^2}{N}}{N}$$

Dimana:

$r_{nn'}$ = reliabilitas tes

k = banyaknya butir soal

σ_b^2 = jumlah varians butir

σ_t^2 = varian total

X = skor total

Contoh:

Berapakah koefisien reliabilitas tes uraian yang terdiri-dari lima butir pertanyaan dan dikerjakan oleh sepuluh peserta didik. Rentang skor setiap butir soal adalah 1 – 10. Data hasil pengukuran disajikan sebagai berikut.

No	Nama	Nomor Butir Soal					Skor Total	Kuadrat Skor Total
		1	2	3	4	5		
1	Rony	4	5	5	4	6	24	576
2	Agus	6	7	6	6	7	32	1024
3	Fatimah	7	8	7	7	8	37	1369
4	Slamet	3	9	9	9	6	36	1296
5	Rendra	5	8	7	5	9	34	1156
6	Edy	6	7	9	7	8	37	1369
7	Untung	8	10	5	8	7	38	1444
8	Titut	10	9	7	9	8	43	1849
9	Nina	9	7	8	6	9	39	1521
10	Tiwuk	5	8	9	8	8	38	1444
Jumlah		63	78	72	69	76	358	13.048
Jumlah Kuadrat Skor Per Butir		441	626	540	501	588		

$$\sigma_1^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} = \frac{441 - \frac{63^2}{10}}{10} = \frac{441 - 396,9}{10} = 4,41$$

$$\sigma_2^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} = \frac{626 - \frac{78^2}{10}}{10} = \frac{626 - 608,4}{10} = 1,76$$

$$\sigma_3^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} = \frac{540 - \frac{72^2}{10}}{10} = \frac{540 - 518,4}{10} = 2,16$$

$$\sigma_4^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} = \frac{501 - \frac{69^2}{10}}{10} = \frac{501 - 476,1}{10} = 2,49$$

$$\sigma_5^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} = \frac{588 - \frac{76^2}{10}}{10} = \frac{588 - 577,6}{10} = 1,04$$

$$\sum \sigma_b^2 = 4,41 + 1,76 + 2,16 + 2,49 + 1,04 = \mathbf{11,86}$$

$$\sigma_t^2 = \frac{\sum X_t^2 - \frac{\sum (X_t)^2}{N}}{N} = \frac{13.048 - \frac{358^2}{10}}{10} = \frac{231,6}{10} = \mathbf{23,16}$$

Selanjutnya hasil perhitungan tersebut dimasukkan ke rumus Alpha:

$$r_{nn'} = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right) = \frac{5}{4} \left(1 - \frac{11,86}{23,16} \right) = 1,25 \times 0,512 = \mathbf{0,640}$$

Untuk menginterpretasikan apakah koefisien reliabilitas hasil perhitungan tersebut menunjukkan tes yang reliabel atau tidak, hasil perhitungan tersebut selanjutnya dibandingkan dengan harga kritis reliabilitas, yaitu 0,7. Karena koefisien reliabilitas hasil perhitungan sebesar 0,640 lebih kecil dari harga kritis reliabilitas ($0,640 < 0,7$), maka tes tersebut secara keseluruhan tidak reliabel. Sama seperti tes dengan skor diskrit, interpretasi hasil perhitungan dapat juga dilakukan dengan membandingkan r hitung dengan r tabel *product moment*. Apabila hasil hitung lebih besar atau sama dengan r tabel ($r \text{ hitung} \geq r \text{ tabel}$), maka dapat disimpulkan ada korelasi signifikan antara butir-butir soal tersebut secara keseluruhan dan tes tersebut dikategorikan reliabel dan sebaliknya apabila hasil hitung lebih kecil daripada r tabel ($r \text{ hitung} < r \text{ tabel}$), maka dapat disimpulkan tidak ada korelasi signifikan antara butir-butir soal tersebut secara keseluruhan dan tes tersebut dikategorikan tidak reliabel.

Uji reliabilitas di atas adalah uji reliabilitas tes secara keseluruhan. Bukan uji reliabilitas per butir soal. Langkah-langkah uji reliabilitas tes secara keseluruhan dan uji reliabilitas butir adalah sama. Keduanya sama-sama diawali dengan menggunakan rumus korelasi *product moment*. Uji reliabilitas butir dilakukan dengan mengkorelasikan skor tiap butir soal yang

diperoleh setiap peserta didik dengan skor total dari hasil sebuah ujian. Selanjutnya, hasil perhitungan dari rumus *product moment* dimasukkan ke rumus *Spearman-Brown*.

C. Kepraktisan Tes

Tes yang praktis adalah tes yang mudah digunakan dan dikelola sebagai alat ukur terhadap hasil belajar. Dengan kata lain, tes yang praktis adalah tes yang mudah dalam proses penyiapan atau pengembangan, mudah digunakan, hasil pengukuran yang dihasilkan mudah diinterpretasikan dan mudah dalam menyimpannya. Ringkasnya, kepraktisan sebuah tes dapat dilihat sejauhmana tes itu efisien dan efektif digunakan untuk mengukur hasil belajar peserta didik. Seperangkat tes yang menghendaki peserta didik untuk menyelesaikannya dalam waktu yang cukup lama, misalnya tiga hingga lima jam, berarti tes tersebut tidak praktis. Sebuah tes yang dapat diselesaikan oleh peserta tes dalam beberapa menit saja, tetapi guru memerlukan waktu berjam-jam untuk mengoreksinya, berarti tes tersebut juga tidak praktis. Sebuah tes yang hanya bisa dikoreksi dengan komputer, sedangkan pada saat yang sama tidak tersedia komputer yang memadai, berarti tes itu pun tidak praktis. Brown (2004) menguraikan bahwa sebuah tes yang praktis harus memenuhi kriteria sebagai berikut:

1. Tes itu tidak memerlukan biaya besar dalam proses pengembangan dan pelaksanaannya.
2. Tes dapat diselesaikan dalam batasan waktu yang sesuai dan rasional.
3. Tes itu mudah dikelola dan dilaksanakan sebagai alat ukur hasil belajar. Dengan kata lain, proses pelaksanaan tes tidak memerlukan persiapan yang rumit.
4. Tes tersebut memiliki prosedur penskoran yang spesifik dan efektif.

Sebagai gambaran tentang perbedaan tes yang praktis dan tidak / kurang praktis dapat dilihat dalam tabel berikut.

Tabel 4.1

Contoh tes yang praktis dan kurang praktis

Tes yang kurang praktis	Tes yang lebih praktis
Guru menggunakan tes bentuk uraian bebas untuk mengukur hasil belajar sejumlah 250 peserta didik.	Guru menggunakan tes bentuk objektif untuk mengukur hasil belajar peserta didik yang jumlahnya banyak.
Guru memberikan tes yang menghendaki peserta didik menggunakan lembar jawab komputer (LJK), tetapi tidak tersedia <i>scanner</i> untuk mengoreksi lembar jawab.	Guru memberikan lembar jawab berupa selebar kertas untuk menjawab soal ulangan harian.
Guru memberikan tes menyimak (<i>listening test</i>) berbasis internet, tetapi jaringan internet tidak tersedia baik di wilayah dimana tes itu dilaksanakan.	Guru menggunakan <i>tape recorder</i> untuk ujian menyimak.
Guru menyiapkan tes pilihan ganda yang terdiri dari 200 butir soal dan harus dikerjakan dalam waktu tiga jam.	Guru menyiapkan 50 butir soal pilihan ganda yang harus diselesaikan oleh peserta didik dalam waktu 90 menit.

D. Dampak Tes

Tes semestinya memiliki dampak positif terhadap proses belajar dan mengajar. Dampak sebuah tes terhadap praktik belajar dan mengajar biasa juga disebut dengan *washback* (Cheng & Watanabe, 2008). Tes bisa jadi memiliki dampak negatif terhadap proses belajar dan mengajar dan dampak itu tidak diharapkan muncul dari sebuah pelaksanaan tes. Dampak negatif itu, misalnya peserta didik menjadi putus asa setelah mengikuti tes yang diberikan oleh guru. Hal ini terjadi mungkin karena tes yang harus

dikerjakan terlalu sulit atau cakupan materi yang diujikan di luar tingkat berpikir peserta didik. Guru harus menyadari hal ini. Pemilihan materi yang tepat sesuai dengan tingkat berpikir peserta didik menjadi hal yang penting untuk diperhatikan. Tes yang harus dikerjakan oleh peserta didik sekolah dasar harus berisi materi pada level sekolah dasar, demikian pula untuk peserta didik SMP/MTs dan seterusnya. Sedangkan dampak negatif suatu tes terhadap proses pembelajaran terjadi apabila selama proses pembelajaran, guru hanya fokus pada kegiatan untuk melatih peserta didik dalam menyelesaikan tes. Aktivitas pembelajaran seharusnya bukan sekedar untuk melatih peserta didik untuk menyelesaikan soal-soal tes dengan benar. Pemahaman ini terlalu mempersempit makna dari proses pembelajaran itu sendiri. Proses pembelajaran adalah aktivitas yang menghendaki adanya interaksi secara intensif antara guru, peserta didik dan sumber belajar. Tugas utama guru selama proses pembelajaran adalah mendidik, mengajar, membimbing, mengarahkan, melatih, menilai, dan mengevaluasi peserta didik. Jadi kurang tepat apabila proses pembelajaran hanya terfokus pada aktivitas untuk melatih peserta didik agar mampu mengerjakan tes dengan benar.

Sebuah tes diharapkan mampu memberikan dampak positif terhadap proses belajar dan mengajar. Contoh dampak positif dari tes terhadap proses belajar, misalnya peserta didik termotivasi untuk banyak belajar dan belajar banyak setelah mengikuti atau mengerjakan tes; peserta didik berusaha keras untuk mencari informasi bagaimana menyelesaikan sebuah butir soal secara tepat; peserta didik lebih aktif mengikuti proses pembelajaran agar lebih mampu menguasai materi yang disampaikan oleh guru, dan sebagainya. Sedangkan dampak positif sebuah tes terhadap proses mengajar guru, misalnya setelah memberikan tes kepada peserta didik, guru terdorong untuk memperbaiki metode pembelajaran yang telah dilakukan agar setiap peserta didik mampu menguasai materi pelajaran dengan baik; guru berusaha keras meng-*update* materi terkini yang harus disampaikan kepada peserta didik; guru berusaha keras untuk meningkatkan kualitas pembelajaran yang dilakukan, dan sebagainya.

E. Tingkat Kesulitan Tes

Tingkat kesulitan atau kesulitan tes semata-mata mengacu pada persentase peserta didik yang mampu menjawab butir-butir soal dengan

benar. Misalnya seorang guru ingin mengetahui berapa tingkat kesulitan butir no. 1 soal pilihan ganda yang telah ia kembangkan, maka guru cukup menghitung berapa jumlah peserta didik yang mampu menjawab benar butir soal tersebut dibagi dengan jumlah total peserta didik yang mengikuti ujian. Identifikasi terhadap tingkat kesulitan tes penting untuk dilakukan agar pada tahap selanjutnya guru mampu memilih butir-butir soal yang sesuai dengan tingkat berpikir kelompok peserta didik. Selain itu, tingkat kesulitan tes juga berguna dalam hal pencapaian tujuan tes. Tes yang digunakan untuk diagnosis sebaiknya dikembangkan butir-butir soal pada tingkat kesukaran rendah. Tes yang dimaksudkan untuk seleksi, maka tes tersebut perlu dikembangkan dengan tingkat kesulitan tinggi. Sedangkan tes sumatif di akhir semester sebaiknya dikembangkan dengan tingkat kesulitan sedang.

Tingkat kesulitan butir soal dinyatakan dalam bentuk angka dan biasa disebut dengan indeks tingkat kesulitan butir soal. Cara menghitung tingkat kesulitan tes objektif (pilihan ganda, benar-salah, dan lainnya) berbeda dengan cara untuk menghitung tingkat kesulitan untuk tes subjektif (uraian). Tingkat kesulitan tes objektif dapat dihitung dengan menggunakan rumus sebagai berikut:

$$P = \frac{N_p}{N}$$

Dimana:

- P : indeks kesulitan butir soal
 N_p : Jumlah peserta tes yang menjawab benar butir soal tertentu
 N : Jumlah total peserta didik yang mengikuti tes

Sedangkan untuk mengetahui tingkat kesulitan tes subjektif (uraian) digunakan langkah-langkah berikut.

1. Mencari rata-rata skor butir soal yang akan dicari tingkat kesulitannya dengan rumus sebagai berikut.

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

Dimana:

$\bar{X}_i$: Rata-rata skor ke-i

$\sum X_i$: Jumlah skor yang diperoleh peserta didik pada butir ke-i

N : Jumlah keseluruhan peserta didik yang mengikuti tes

2. Hasil perhitungan rata-rata tersebut kemudian dimasukkan ke rumus berikut.

$$P = \frac{\bar{X}_i}{X_{max}}$$

Dimana:

P : Tingkat kesulitan butir soal

$\bar{X}_i$: Rata-rata skor ke-i

$\bar{X}_{max}$: Skor maksimum yang ditetapkan untuk butir tertentu

Setelah butir-butir soal dianalisis untuk mengetahui indeks tingkat kesulitannya, selanjutnya ekuivalen berikut dapat digunakan sebagai panduan untuk menginterpretasikan tingkat kesulitan sebuah butir soal.

0,00 - 0,30 : butir soal tergolong sulit

0,31 - 0,70 : butir soal tergolong sedang (cukup sulit)

0,71 - 1,00 : butir soal tergolong mudah

Contoh:

Berapakah indeks tingkat kesulitan butir-butir soal pilihan ganda berikut ini. Tes terdiri-dari 10 butir dan dikerjakan oleh 10 peserta didik.

No	Nama	Nomor Butir Soal									
		1	2	3	4	5	6	7	8	9	10
1	Rony	1	1	0	0	1	1	0	1	1	0
2	Agus	0	1	1	1	1	1	1	0	0	0
3	Fatimah	0	1	0	1	1	0	0	1	1	0
4	Slamet	1	0	0	0	1	1	1	0	1	1
5	Rendra	1	0	0	1	1	1	0	0	0	1
6	Edy	0	1	1	1	1	1	0	0	0	1

7	Untung	0	0	0	0	0	0	1	0	0	1
8	Titut	1	1	0	0	0	0	0	1	1	1
9	Nina	0	1	1	1	1	0	0	1	1	0
10	Tiwuk	0	0	0	0	1	1	1	0	1	1
Jumlah		4	6	3	5	8	6	4	4	6	6

$$\text{Butir no. 1} = \frac{4}{10} = 0,4$$

$$\text{Butir no. 6} = \frac{6}{10} = 0,6$$

$$\text{Butir no. 2} = \frac{6}{10} = 0,6$$

$$\text{Butir no. 7} = \frac{4}{10} = 0,4$$

$$\text{Butir no. 3} = \frac{3}{10} = 0,3$$

$$\text{Butir no. 8} = \frac{4}{10} = 0,4$$

$$\text{Butir no. 4} = \frac{5}{10} = 0,5$$

$$\text{Butir no. 9} = \frac{6}{10} = 0,6$$

$$\text{Butir no. 5} = \frac{8}{10} = 0,8$$

$$\text{Butir no.10} = \frac{6}{10} = 0,6$$

Dengan demikian, dapat dipahami bahwa butir soal no. 1 dengan indeks kesulitan sebesar 0,4 termasuk butir soal dengan tingkat kesulitan sedang, butir no. 2 memiliki tingkat kesulitan sedang, butir no. 5 termasuk butir soal yang tergolong mudah, dan seterusnya.

Contoh analisis tingkat kesulitan soal uraian:

Berapakah tingkat kesulitan butir-butir soal uraian berikut jika skor maksimal untuk setiap butirnya adalah 10?

No	Nama	Nomor Butir Soal				
		1	2	3	4	5
1	Rony	4	5	5	4	6
2	Agus	6	7	6	6	7
3	Fatimah	7	8	7	7	8
4	Slamet	3	9	9	9	6
5	Rendra	5	8	7	5	9
6	Edy	6	7	9	7	8
7	Untung	8	10	5	8	7

8	Titut	10	9	7	9	8
9	Nina	9	7	8	6	9
10	Tiwuk	5	8	9	8	8
Jumlah		63	78	72	69	76
Rata-rata		6,3	7,8	7,2	6,9	7,6

$$\text{Butir no. 1} = \frac{6,3}{10} = 0,63$$

$$\text{Butir no. 4} = \frac{6,9}{10} = 0,69$$

$$\text{Butir no. 2} = \frac{7,8}{10} = 0,78$$

$$\text{Butir no. 5} = \frac{7,6}{10} = 0,76$$

$$\text{Butir no. 3} = \frac{7,2}{10} = 0,72$$

Dengan demikian, dapat disimpulkan bahwa butir no. 1 merupakan butir soal yang tergolong memiliki tingkat kesulitan sedang, butir no. 2 adalah butir dengan tingkat kesulitan mudah, dan seterusnya.

F. Daya Beda Tes

Daya beda tes mengacu pada sejauhmana sebuah butir soal mampu membedakan peserta didik yang telah menguasai materi yang diujikan dengan baik dan peserta didik yang belum mampu menguasai materi pelajaran. Daya beda memiliki peran penting dalam pengukuran hasil belajar peserta didik. Secara empiris, daya beda butir soal dapat dimanfaatkan oleh guru untuk memperbaiki kualitas setiap butir soal yang telah dibuat. Berdasarkan indeks daya beda, guru dapat menentukan apakah butir soal tersebut sudah dalam kategori baik, apakah butir soal tersebut perlu direvisi, atau apakah butir soal tersebut harus ditolak atau tidak dipakai sama-sekali. Indeks daya beda bergerak dari -1 hingga 1. Apabila suatu butir soal memiliki indeks daya beda minus (-), berarti ada kemungkinan bahwa peserta didik yang sebenarnya secara akademis kurang menguasai materi ajar yang diujikan, tetapi justru mereka mampu memilih jawaban yang benar. Hal ini dapat terjadi karena beberapa kemungkinan, misalnya ada kesalahan kunci jawaban yang dibuat oleh guru; butir soal itu kemungkinan memiliki lebih dari satu jawaban yang benar; kompetensi yang diuji dengan menggunakan tes tersebut tidak jelas; alternatif jawaban yang berperan sebagai pengecoh tidak berfungsi dengan baik; dan materi yang diujikan

terlalu sulit, sehingga banyak peserta didik yang menjawab butir-butir soal hanya dengan menebak.

Semakin tinggi indeks daya beda sebuah butir soal, berarti semakin mampu butir soal tersebut membedakan peserta didik yang telah menguasai materi yang diujikan dan peserta didik yang tidak atau belum mampu menguasai materi yang diujikan. Dengan kata lain, semakin tinggi indeks daya beda suatu butir soal, maka semakin baik kualitas butir soal itu. Untuk mengetahui indeks daya beda suatu butir soal pilihan ganda biasa digunakan rumus sebagai berikut:

$$DP = \frac{BA-BB}{\frac{1}{2}N} \quad \text{atau} \quad DP = \frac{2(BA-BB)}{N}$$

Dimana:

DP : Indeks daya beda

BA : Jumlah jawaban benar oleh peserta didik kelompok atas

BB : Jumlah jawaban benar oleh peserta didik kelompok bawah

N : Jumlah peserta didik yang mengikuti tes

Untuk menginterpretasikan indeks daya beda, dapat digunakan ekuivalen sebagai berikut (Crocker dan Algina,1986).

0,4 ke atas : Butir soal sangat baik dan diterima

0,30 – 0,39 : Butir soal baik, tetapi perlu sedikit revisi

0,20 – 0,29 : Butir soal cukup baik, tetapi perlu revisi perbaikan

0,00 – 0,19 : Butir soal buruk dan harus ditolak/dibuang

Contoh:

Berapakah indeks daya beda untuk butir no. 3 soal pilhan ganda berikut jika diketahui kunci jawabannya adalah C.

No	Nama	Nomor Butir Soal										Skor
		1	2	3	4	5	6	7	8	9	10	
1	Rony	1	1	1	1	1	1	0	1	1	0	8
2	Agus	0	1	1	1	1	1	1	0	0	0	6
3	Fatimah	0	1	1	0	1	0	0	0	0	0	3
4	Slamet	1	0	0	0	1	1	1	1	1	1	7
5	Rendra	1	0	0	1	0	1	0	0	0	1	4

6	Edy	0	1	1	1	1	1	0	0	1	1	7
7	Untung	0	0	0	0	0	0	1	0	0	1	2
8	Titut	1	1	0	0	0	0	0	1	1	1	5
9	Nina	0	1	1	1	1	0	0	1	1	0	6
10	Tiwuk	0	0	0	0	1	1	1	0	1	1	5

Langkah pertama yang harus dilakukan adalah mengurutkan skor dimulai dari skor tertinggi ke skor terendah guna mengelompokkan peserta didik menjadi dua kelompok, yaitu kelompok atas dan kelompok bawah.

Nama	Skor	Kelompok	Nama	Skor	Kelompok
Rony	8	Kelompok Atas	Titut	5	Kelompok Bawah
Slamet	7		Tiwuk	5	
Edy	7		Rendra	4	
Agus	6		Fatimah	3	
Nina	6		Untung	2	

Dari data tersebut dapat diketahui bahwa peserta didik kelompok atas yang dapat menjawab benar butir no. 3 ada 3 orang, sedangkan peserta didik kelompok bawah yang dapat menjawab benar butir no. 3 ada 1 orang. Selanjutnya indeks daya beda dapat dihitung:

$$DP = \frac{2(BA - BB)}{N} = \frac{2(3 - 1)}{10} = \frac{4}{10} = 0,4$$

Dengan demikian, dapat disimpulkan bahwa butir soal no. 3 termasuk butir soal yang sangat baik dan dapat diterima atau dimasukkan kedalam bank soal untuk dipakai kembali sebagai alat ukur hasil belajar peserta didik pada kesempatan yang berbeda.

Selain menggunakan rumus di atas, untuk menghitung indeks daya beda dapat juga menggunakan rumus korelasi point biserial sebagai berikut:

$$r_{pbis} = \frac{\bar{X}_b - \bar{X}_s}{SD_t} \sqrt{pq}$$

$$SD_t = \sqrt{\frac{n \sum X^2 - [\sum X]^2}{n(n-1)}}$$

Dimana:

- $\bar{X}_b$: Rata-rata skor peserta didik yang menjawab benar
 $\bar{X}_s$: Rata-rata skor peserta didik yang menjawab salah
 SD_t : Simpangan baku skor total
p : proporsi jawaban benar terhadap semua jawaban peserta didik
q : proporsi jawaban salah terhadap semua jawaban peserta didik ($q = 1 - p$)

Kelebihan dari rumus ini adalah indeks daya beda dapat memberikan refleksi terhadap kontribusi butir soal secara sesungguhnya terhadap fungsi tes sebagai alat ukur hasil belajar peserta didik; indeks daya beda langsung berhubungan dengan statistika tes; dan tidak pernah memiliki indeks daya beda 1,00 karena hanya variable-variabel yang memiliki distribusi bentuk yang sama yang berkorelasi secara tepat. Variabel kontinyu (*criterion*) dan skor dikotomis tidak selalu memiliki bentuk yang sama.

Contoh perhitungan indeks daya beda dengan rumus r_{pbis} . Berdasarkan data yang sama seperti di atas, berapakah daya beda butir soal no. 3?

Peserta didik dengan jawaban benar untuk butir no. 3	Skor keseluruhan	Peserta didik dengan jawaban salah untuk butir no. 3	Skor keseluruhan
Rony	8	Slamet	7
Agus	6	Rendra	4
Fatimah	3	Untung	2
Edy	7	Titut	6
Nina	6	Tiwuk	5
$\Sigma = 5$	$\Sigma = 30$	$\Sigma = 5$	$\Sigma = 24$
	$\bar{X} = 6$		$\bar{X} = 4,8$

Pertama-tama dicari simpangan baku dari skor untuk soal no. 3 di atas.

No	Nama	X	X ²
1.	Rony	8	64
2.	Agus	6	36
3.	Fatimah	3	9
4.	Edy	7	49
5.	Nina	6	36
6.	Slamet	7	49
7.	Rendra	4	16
8.	Untung	2	4
9.	Titut	6	36
10.	Tiwuk	5	25
Σ		54	324

$$SD_t = \sqrt{\frac{n \sum X^2 - [\sum X]^2}{n(n-1)}} = \sqrt{\frac{(10 \cdot 324) - 54^2}{10(10-1)}} = \sqrt{\frac{3240 - 2916}{90}} = \sqrt{3,6} = 1,897$$

Berdasarkan tabel dan perhitungan simpangan baku di atas, diketahui bahwa:

- Jumlah peserta didik yang menjawab benar butir no. 3 ada 5 orang.
- Jumlah peserta didik yang menjawab salah butir no. 3 ada 5 orang.
- Jumlah peserta didik secara keseluruhan adalah 10 orang.
- Skor rata-rata peserta didik yang menjawab benar 6
- Skor rata-rata peserta didik yang menjawab salah 4,8
- Simpangan baku skor total 1,897
- Jumlah skor keseluruhan 54

$$r_{pbis} = \frac{\bar{X}_b - \bar{X}_s}{SD_t} \sqrt{pq} = \frac{6 - 4,8}{1,897} \sqrt{\frac{5}{10} \cdot \frac{5}{10}} = 0,632 \times 0,5 = 0,316$$

Dari perhitungan daya beda dengan rumus korelasi poin biserial didapatkan indeks daya beda sebesar 0,316 yang berarti bahwa butir soal no. 3 tersebut baik, tetapi perlu ada sedikit revisi.

Hal yang perlu diperhatikan bahwa analisis daya beda di atas dilakukan untuk menganalisis soal pilihan ganda dengan jumlah peserta tes yang sedikit (10 orang). Bagaimana jika ujian diikuti oleh peserta didik yang jumlahnya besar? Tentu hal itu akan menyulitkan guru yang akan melakukan analisis daya beda butir soal jika harus melibatkan semua peserta didik yang mengikuti ujian yang jumlahnya mungkin sangat besar. Oleh karena itu, jika peserta didik yang mengikuti ujian jumlahnya besar (100 orang atau lebih), setelah diurutkan skornya dari yang tertinggi ke yang terendah, guru cukup mengambil 27% peserta didik dengan nilai tinggi sebagai kelompok atas dan 27% peserta didik dengan nilai rendah sebagai kelompok bawah. Pemilihan sampel peserta didik sebagai kelompok atas dimulai dari peserta didik dengan skor tertinggi dan kelompok bawah juga dimulai dari peserta didik dengan skor terendah.

Bab 5

Tes Standard dan Tes Buatan Guru

Sebelum disajikan hakikat dari tes standard (*standardized test*) dan tes buatan guru (*teacher-made test*), terlebih dahulu disajikan sebuah ilustrasi pengalaman yang dialami oleh seorang guru yang mengajar sebuah mata pelajaran di sekolah. Cerita kasus ini terjadi pada saat ujian nasional masih diberlakukan untuk mengukur hasil belajar peserta didik menjelang berakhirnya studi mereka di sekolah. Saat itu, guru tersebut telah mengajar selama hampir satu tahun di kelas XII di sebuah SMA dan setiap hari guru itu mengajar mata pelajaran kimia. Setiap selesai mengajarkan satu kompetensi dasar, guru itu kemudian memberikan ulangan (tes harian dengan bentuk tes kombinasi pilihan ganda dan uraian) kepada setiap peserta didik yang ada di kelas dimana dia biasa mengajar. Hasil ulangan kemudian diperiksa oleh guru tersebut dalam rangka mengetahui sejauhmana tingkat capaian hasil belajar setiap peserta didik. Dari hasil pemeriksaan itu, diperoleh informasi bahwa hampir semua peserta didik mendapatkan skor yang baik dengan rata-rata 80. Berdasarkan rata-rata skor ini, guru kemudian berkesimpulan bahwa hampir semua peserta didik telah menguasai materi pelajaran dengan baik. Selanjutnya, guru mengajarkan materi yang tercakup dalam kompetensi dasar yang berbeda. Seperti biasanya, setelah satu kompetensi dasar selesai diajarkan, dia kemudian memberikan ulangan harian. Hasil ulangan harian yang kedua ini pun diperoleh informasi bahwa sebagian besar peserta didik mendapatkan skor yang tinggi dengan rata-rata 85. Rata-rata skor ini menunjukkan bahwa metode mengajar guru telah berfungsi dengan baik karena semua peserta didik mampu mencapai skor tinggi saat ulangan. Hal ini berarti bahwa semua peserta didik telah menguasai materi yang telah diajarkan oleh gurunya.

Pada akhir tahun pelajaran, semua peserta didik mengikuti ujian nasional. Materi yang diujikan dalam ujian nasional didasarkan pada kurikulum dan silabus yang telah ditetapkan oleh pemerintah dan semua materi itu telah tersampaikan dengan baik kepada semua peserta didik. Setelah ujian nasional selesai dilaksanakan dan tiba saatnya peserta didik mendapatkan hasil dari ujian nasional yang telah diikuti. Hasil ujian menunjukkan bahwa sebagian besar peserta didik mendapatkan skor yang relatif rendah, tidak seperti skor yang diperoleh saat ulangan harian. Rata-rata skor ujian nasional adalah 60. Bahkan ada beberapa peserta didik yang mendapatkan skor jauh di bawah rata-rata. Ada perbedaan yang mencolok dari skor hasil ulangan harian dan skor hasil ujian nasional. Berdasarkan perbedaan tersebut, kemudian muncul sederet pertanyaan di pikiran guru tersebut. Apa yang salah dengan peserta didik yang selama ini saya ajar? Mengapa skor hasil ulangan harian berbeda jauh dengan skor hasil ujian nasional? Apakah materi yang diujikan dalam ujian nasional berbeda dengan materi yang telah saya ajarkan setiap hari? Apakah butir-butir soal ujian nasional lebih sulit daripada butir-butir soal ulangan harian? Apakah cara saya mengajar selama ini salah? Apakah metode dan teknik mengajar saya kurang tepat? Dan masih banyak pertanyaan lain yang muncul di pikiran guru tersebut. Perbedaan skor tersebut kemungkinan disebabkan oleh kurang tepatnya metode dan teknik mengajar yang digunakan oleh guru selama ini. Tetapi mungkin juga berkaitan dengan kualitas butir-butir soal yang digunakan pada saat ulangan harian dan ujian nasional. Mungkin juga disebabkan oleh cara guru dalam melakukan proses penskoran (proses penskoran yang objektif atau justru subjektif).

Kasus tersebut sebenarnya menggambarkan perbedaan dari konsep tes buatan guru dan tes standar. Secara umum terkait dengan kualitas tes, tes dapat dibedakan menjadi tes buatan guru (tes yang disiapkan dan dibuat oleh guru yang mengajar peserta didik di kelas) dan tes standar (tes yang dibuat dan disiapkan oleh para ahli, atau oleh universitas, atau oleh lembaga-lembaga pengujian formal, atau oleh lembaga lain yang kompeten). Kedua jenis tes tersebut memiliki karakteristik, cakupan materi, kualitas, tujuan, konstruksi, metode penskoran, dan teknik interpretasi skor yang berbeda satu sama lain. Ujian nasional dapat dikategorikan sebagai tes standar karena tes tersebut memiliki prosedur pelaksanaan, penskoran dan metode interpretasi skor yang seragam. Selanjutnya pada bagian ini akan dibahas

secara lebih lengkap hakikat dan perbedaan dari tes standar dan tes buatan guru.

A. Tes Standar

Pembahasan berkaitan dengan tes standar pada bagian ini dimulai dari pengertian, kriteria, karakteristik dan kegunaan tes standar dalam pendidikan. Berbagai aspek tersebut diuraikan secara lengkap dan mengacu pada pemikiran Disha (2020) dan dielaborasi berdasarkan kondisi yang terjadi dalam konteks Indonesia.

1. Pengertian Tes Standar

Tes standar biasa juga disebut dengan tes baku. Tes standar adalah sebuah tes yang disusun secara cermat oleh para ahli yang profesional atau oleh lembaga pengujian formal lain dan memiliki prosedur pelaksanaan, peskoran, serta metode interpretasi skor yang seragam. Tes standar biasanya tidak hanya digunakan secara terbatas oleh satu sekolah saja, tetapi tes ini digunakan secara luas oleh banyak sekolah untuk mengukur capaian belajar peserta didik. Dalam tes standar, materi yang diujikan telah melalui proses seleksi secara ketat sesuai dengan kriteria urgensi, keberlanjutan, relevansi, dan kepraktisan. Urgensi berarti bagian materi itu adalah sangat penting untuk dikuasai oleh setiap peserta didik. Berkelanjutan memiliki makna bahwa materi itu menjadi prasyarat untuk peserta didik mampu menguasai materi yang lain. Sedangkan relevansi bermakna bahwa materi itu memiliki korelasi dengan materi yang ada pada mata pelajaran lain. Selanjutnya, kepraktisan adalah materi itu memiliki nilai terapan yang tinggi untuk dapat mengatasi berbagai masalah yang mungkin muncul dalam kehidupan nyata.

Tes standar juga dapat dipahami sebagai sebuah tes yang disusun secara cermat oleh para ahli sesuai dengan tujuan yang telah ditentukan; prosedur pelaksanaan tes yang telah disepakati, metode penilaian dan interpretasi skor yang telah ditentukan secara rinci sehingga siapa pun yang menggunakan tes itu sebagai alat ukur dan dimana pun tes itu akan digunakan, hasilnya harus sebanding. Ringkasnya, tes standar adalah sebuah tes yang menghendaki adanya pemikiran, perencanaan, persiapan, proses penskoran, dan analisis skor yang lebih detail dan cermat. Selanjutnya, tes ini adalah tes yang disusun oleh individu atau sekelompok individu untuk dapat digunakan

dalam berbagai situasi dan tujuan. Hal ini mengandung makna bahwa tes standar dikembangkan bukan hanya untuk digunakan oleh satu sekolah secara lokal, tetapi dapat juga digunakan oleh banyak sekolah yang memiliki karakteristik yang sama. Disha (2020) lebih lanjut menguraikan acuan kriteria standar, karakteristik, dan kegunaan tes standar sebagai berikut.

2. Kriteria Tes Standar

Cakupan materi dan serangkaian butir-butir pertanyaan dalam tes standar harus memenuhi kriteria yang terstandar baik. Oleh karena itu, tes standar adalah sebuah tes yang pada saat penyusunannya telah melalui serangkaian proses standarisasi yang panjang. Proses itu diantaranya adalah:

- a. Standarisasi berkaitan dengan materi yang diujikan dan butir-butir pertanyaan.
- b. Materi dan tujuan tes adalah dua hal yang sangat diperhatikan dalam tes standar. Oleh karena itu, kisi-kisi soal memiliki peran yang sangat penting dalam rangka mewujudkan materi yang representatif untuk mencapai tujuan yang telah ditentukan. Butir-butir pertanyaan yang tidak relevan dengan tujuan tes harus dibuang dan hanya mengikutsertakan butir-butir yang relevan saja. Selain itu tingkat kesulitan dan daya beda butir soal juga menjadi pertimbangan penting dalam menyusun tes standar.
- c. Standarisasi pelaksanaan tes.
- d. Prosedur pelaksanaan tes, kondisi dimana tes akan dilaksanakan, alokasi waktu untuk menyelesaikan tes, dan sebagainya harus telah ditentukan secara jelas sebelum tes dilaksanakan.
- e. Standarisasi proses penskoran
- f. Untuk memastikan bahwa objektivitas dan keseragaman proses penskoran dapat dicapai dalam setiap situasi penilaian hasil belajar, maka kunci jawaban dan perintah untuk menyelesaikan tes harus sudah disiapkan dan dinyatakan secara jelas sebelum proses penskoran dilakukan oleh berbagai pihak yang terlibat.
- g. Standarisasi interpretasi hasil penskoran
- h. Tes standar adalah tes yang mengikuti pendekatan acuan norma, bukan tes yang mengikuti acuan kriteria. Norma-norma standar harus telah disiapkan secara jelas sebelum interpretasi skor hasil tes

dilakukan. Tes ini dilaksanakan berangkat dari sampel materi yang representatif. Oleh karena itu, skor tes kemudian diinterpretasikan mengacu pada norma atau auturan yang telah ditetapkan sebelumnya. Norma merupakan bagian integral dari keseluruhan proses interpretasi skor hasil pengujian.

3. Karakteristik Tes Standar

Berbeda dengan jenis tes lain, tes standar memiliki beberapa karakteristik yang menandai keunikan dari tes jenis ini. Berikut adalah beberapa karakteristik yang tes standar.

- a. Tes standar memuat butir-butir soal dengan kualitas tinggi.
- b. Sebelum digunakan sebagai alat ukur, butir-butir tes telah diuji-cobakan terlebih dulu untuk mengetahui sejauhmana kualitas butir-butir tes tersebut. Butir-butir tes yang digunakan hanyalah butir-butir tes yang telah diseleksi secara cermat berdasarkan tingkat kesulitan butir dan daya beda yang rasional; memiliki validitas dan reliabilitas yang baik; dan butir-butir tes dipilih berdasarkan perilaku yang diharapkan muncul dari peserta didik sesuai dengan tujuan yang telah ditentukan.
- c. Tes standar memiliki prosedur pelaksanaan tes, alokasi waktu, dan prosedur penskoran yang jelas.
- d. Instruksi atau perintah pengerjaan tes, batasan waktu untuk menyelesaikan tes, dan prosedur penilaian telah ditetapkan secara jelas sehingga setiap orang yang akan memanfaatkan tes tersebut dapat melakukan penskoran dan interpretasi hasil penskoran yang sama dengan yang dilakukan oleh pihak-pihak lain.
- e. Dalam menginterpretasikan skor, tes standar mengacu pada norma yang telah ditetapkan.
- f. Tes telah memiliki informasi berkaitan dengan reliabilitas, validitas, tingkat kesulitan, dan daya beda tes yang baik sehingga tes tersebut telah dinyatakan layak sebagai alat ukur hasil belajar peserta didik.
- g. Tes telah memiliki manual yang lengkap sebagai panduan bagi pihak-pihak yang akan menggunakan tes tersebut.
- h. Manual disediakan untuk menjelaskan tujuan dan penggunaan tes; menjelaskan secara singkat bagaimana tes itu disusun; memberikan

petunjuk khusus untuk mengelola, mencetak, dan menafsirkan hasil tes. Manual juga berisi tabel norma dan merangkum data penelitian (rekam jejak) yang tersedia pada tes tersebut.

4. Kegunaan Tes Standar

Tes standar memiliki nilai kegunaan yang lebih praktis dibandingkan dengan jenis-jenis tes yang lain, misalnya tes buatan guru. Selanjutnya tes standar dapat digunakan oleh guru berkaitan dengan hal-hal sebagai berikut.

- a. Tes terstandar dapat digunakan untuk mengukur tingkat perkembangan kemampuan peserta didik. Hasil dari pengukuran ini mampu memberikan dasar pijakan untuk mengetahui sejauhmana tingkat kekuatan dan kelemahan intelektual peserta didik secara lebih akurat.
- b. Tes ini dapat digunakan untuk menguji dan memastikan validitas tes buatan guru.
- c. Tes ini dapat digunakan untuk mendiagnosis kesulitan belajar peserta didik.
- d. Tes ini dapat digunakan untuk membantu guru dalam mengidentifikasi dan mengetahui faktor-faktor yang menjadi penyebab kesulitan belajar peserta didik.
- e. Hasil tes dapat digunakan untuk memberikan informasi sebagai dasar untuk perencanaan kurikulum dan untuk memberikan penanganan khusus bagi peserta didik yang mengalami lambat belajar.
- f. Memberikan data untuk melacak pola pertumbuhan intelektual individu selama periode tahun tertentu.
- g. Hasil tes dapat digunakan untuk membantu dalam mendesain dan mengelola program bimbingan yang lebih baik bagi peserta didik.
- h. Hasil tes dapat digunakan untuk mengevaluasi efektivitas program pembelajaran, kegiatan guru, metode pembelajaran dan faktor-faktor lain yang dianggap signifikan untuk mempengaruhi keberhasilan praktik pendidikan.

Hal yang harus dipahami adalah tidak ada dua set atau lebih tes standar yang persis sama satu dengan lainnya. Setiap tes mengukur aspek perilaku tertentu dan mungkin memiliki tujuan yang berbeda. Ada beberapa

tes dengan judul yang sama, tetapi mengukur aspek perilaku yang sangat berbeda. Sedangkan tes lain dengan judul yang berbeda, tetapi justru mengukur aspek perilaku yang hampir identik. Jadi, pihak-pihak harus lebih teliti dan cermat dalam memilih tes standar yang akan digunakan.

B. Tes Buatan Guru

Berbeda dengan tes standar, tes buatan guru memiliki pengertian, karakteristik dan kegunaan tersendiri. Dikutip dari Disha (2020) bagian berikut membahas ketiga aspek tersebut. Tes buatan guru juga memiliki cakupan materi dan kualitas yang berbeda dengan tes standar. Sesuai dengan namanya, tes buatan guru merupakan tes yang dibuat sendiri oleh guru yang mengajar mata pelajaran yang diampu. Tes tersebut dibuat untuk mengukur sejauhmana setiap peserta didik telah mampu menguasai materi pelajaran yang telah diajarkan oleh gurunya sendiri. Guru sendiri yang merumuskan setiap indikator kompetensi yang harus dicapai oleh setiap peserta didik berdasarkan materi pelajaran yang telah diajarkan. Dalam konteks ini, guru yang mengajar di kelas adalah individu yang paling banyak mengetahui tentang materi yang telah diajarkan dan bagian mana yang harus diukur sehingga tes yang dikembangkan kemungkinan besar sesuai dengan apa yang telah dipelajari oleh peserta didik.

1. Pengertian Tes Buatan Guru

Pada dasarnya, ada persamaan antara tes standard dan tes buatan guru dalam beberapa hal. Keduanya harus disusun secara cermat agar dapat digunakan untuk mengukur apa yang dimaksudkan untuk diukur. Proses penyusunannya dimulai dari membuat tabel spesifikasi yang menjadi pedoman dalam penyusunan butir-butir tes. Keduanya juga sering memiliki bentuk butir-butir soal yang sama dan petunjuk pengerjaan soal yang sama pula. Perbedaan yang menonjol dari keduanya adalah kualitas butir-butir tes (validitas, reliabilitas, daya beda, tingkat kesulitan dan sebagainya), prosedur pelaksanaan tes, dan cara menginterpretasikan skor hasil tes. Tes standar memiliki kualitas, prosedur pelaksanaan, dan metode penginterpretasian skor yang lebih baik daripada tes buatan guru.

Guru tidak serta merta dapat menggunakan tes standar yang tersedia untuk mengukur hasil belajar peserta didik, walaupun tes tersebut memiliki kualitas yang baik. Ada beberapa alasan mengapa seorang guru tidak serta

merta dapat menggunakan tes standar yang tersedia untuk mengukur hasil belajar peserta didik. Pertama, tes standar belum tentu sesuai dengan kebutuhan lokal di sekolah dimana guru mengajar. Hal ini terjadi karena tes standar mungkin dikembangkan mengacu pada kebutuhan global dalam konteks persekolahan yang lebih luas. Sekolah yang memiliki fasilitas sekolah yang terbatas belum tentu cocok untuk menggunakan tes standar yang tersedia. Kedua, pengadaan tes standar sering memerlukan biaya yang besar. Hal ini terjadi karena proses penyusunan tes standar memerlukan waktu yang lama dan proses uji-coba terbatas terlebih dulu sebelum digunakan sebagai alat ukur. Ketiga, ada kemungkinan tes standar dikembangkan dengan tujuan yang berbeda dengan tujuan yang harus dicapai dari sebuah proses pembelajaran di kelas. Oleh karena itu, guru perlu menyiapkan dan mengembangkan tes sendiri sesuai dengan tujuan dan persyaratan yang telah ditentukan.

Tes buatan guru biasanya dikembangkan dan disiapkan untuk keperluan ujian berbasis ruang kelas dan dimaksudkan untuk mengukur capaian hasil belajar di kelas itu. Tes buatan guru juga dimaksudkan untuk mengevaluasi tingkat keberhasilan metode pembelajaran yang digunakan oleh guru di kelas dan mengevaluasi sejauhmana program kurikuler telah tercapai melalui proses belajar dan mengajar yang telah dilakukan oleh guru. Tes buatan guru adalah instrumen yang paling bermakna yang dapat digunakan oleh guru untuk tujuan tersebut. Tes ini didesain lebih dimaksudkan untuk mengatasi berbagai permasalahan pembelajaran di kelas daripada untuk mengatasi pembelajaran yang bersifat umum. Tes jenis ini disiapkan untuk mengukur konten kurikulum lokal dan proses pengembangannya tidak memerlukan teknik yang canggih dan rumit. Oleh karena itu, untuk tujuan pengukuran hasil belajar dalam konteks kelas sangat direkomendasikan untuk menggunakan tes tipe objektif buatan guru ini, karena tidak harus mengikuti langkah-langkah penyusunan tes yang terstandarisasi secara lengkap dan ketat.

2. Karakteristik Tes Buatan Guru

Seperti halnya tes standar, tes buatan guru juga memiliki karakteristik yang menggambarkan keunikan dari tes jenis ini. Beberapa karakteristik tes buatan guru diuraikan dalam bagian berikut.

- a. Butir-butir tes buatan guru disusun dengan urutan berdasarkan tingkat kesulitan, dimulai dari butir-butir tes dengan kesulitan rendah ke butir-butir tes dengan tingkat kesulitan tinggi.
- b. Dikembangkan oleh guru sendiri yang mengajar di kelas dan dimaksudkan untuk tujuan prognosis dan diagnosis.
- c. Tes mencakup semua materi yang telah diajarkan oleh guru dan biasanya tes ini memiliki jumlah butir tes yang lebih banyak.
- d. Penyusunan butir-butir soal selalu diupayakan sesuai dengan kisi-kisi yang telah disiapkan.
- e. Proses penyusunan tes sering tidak harus mengikuti langkah-langkah pengembangan tes secara lengkap seperti proses pengembangan tes standar, tetapi proses pengembangannya cukup hanya mengikuti beberapa langkah saja dari keseluruhan langkah pengembangan tes.
- f. Umumnya, tes jenis ini digunakan sebagai alat evaluasi formatif.
- g. Proses persiapan dan pelaksanaan tes lebih hemat karena tidak melalui proses uji-coba terlebih dulu.
- h. Tes dikembangkan oleh guru untuk mengukur prestasi dan kecakapan peserta didik pada mata pelajaran yang diberikan/diampu oleh guru itu sendiri.
- i. Tes ini tidak memiliki norma tertentu dalam menginterpretasikan hasil, sedangkan norma menjadi hal yang sangat penting untuk tes standar.

Berdasarkan karakteristik yang telah diuraikan tersebut, tampak jelas perbedaan antara tes buatan guru dan tes standar. Tes buatan guru lebih fleksibel dalam proses pengembangannya dan digunakan dalam konteks yang lebih terbatas.

3. Kegunaan Tes Buatan Guru

Skor hasil pengukuran dengan menggunakan tes buatan guru lebih banyak dimanfaatkan oleh guru kelas untuk kepentingan perbaikan (metode pembelajaran dan cara belajar peserta didik) dan juga untuk memotivasi peserta didik untuk lebih banyak belajar dan belajar banyak. Secara rinci, tes buatan guru dapat dimanfaatkan oleh guru untuk hal-hal sebagai berikut.

- a. Membantu guru mengetahui apakah prestasi peserta didik dalam suatu kelas dalam kondisi normal, memiliki kompetensi rata-rata, diatas rata-rata, atau bahkan dibawah rata-rata.
- b. Membantu guru dalam memformalisasikan strategi baru berkaitan dengan proses belajar dan mengajar, sehingga proses belajar dan mengajar tersebut efektif dalam mentransformasikan pengetahuan, keterampilan dan sikap kepada peserta didik.
- c. Untuk mengukur prestasi akademik peserta didik dalam mata pelajaran tertentu yang telah diberikan/diampu oleh guru.
- d. Untuk menilai seberapa jauh tujuan instruksional yang ditentukan telah tercapai.
- e. Untuk mengetahui tingkat ketercapaian pengalaman belajar peserta didik.
- f. Untuk mendiagnosis kesulitan belajar yang dialami peserta didik, sehingga dapat digunakan sebagai dasar untuk menentukan langkah-langkah perbaikan yang diperlukan.
- g. Untuk keperluan klasifikasi atau menentukan tingkat kelas peserta didik berdasarkan skor yang dicapai.
- h. Tes ini dapat digunakan sebagai instrumen untuk evaluasi formatif, diagnostik dan sumatif.

Bab 6

Pelaksanaan Tes

Salah satu instrumen yang sangat penting dalam sebuah program pembelajaran adalah tes. Dalam konteks program pembelajaran, tes dapat digunakan sebagai sumber informasi dalam rangka membuat keputusan apakah program pembelajaran yang telah dilaksanakan oleh seorang guru telah secara efektif mampu mencapai tujuan yang telah ditentukan atau belum. Hasil dari aktivitas pengujian dengan memanfaatkan tes harus memberikan gambaran hasil belajar yang dicapai oleh setiap peserta didik secara akurat, adil, objektif, akuntabel, valid, dan reliabel. Gambaran berkaitan dengan hasil tes seperti tersebut di atas akan sulit dicapai apabila prosedur pelaksanaan tes tidak dilakukan dan dikelola dengan baik. Oleh karena itu, prosedur pelaksanaan tes (*test administration*) menjadi sangat krusial dalam sebuah program pembelajaran. Ringkasnya, ketika melaksanakan tes, seorang guru harus memberikan peluang kepada setiap peserta didik untuk dapat mengerjakan tes tersebut dengan baik sehingga hasil tes yang akurat, adil, objektif, akuntabel, valid, dan reliabel dapat dicapai dengan baik pula.

A. Pengertian Pelaksanaan Tes

Pelaksanaan tes (*test administration*) adalah suatu prosedur atau langkah-langkah sistematis dalam sistem pengujian yang dimaksudkan untuk meminimalisasi kesalahan pengukuran (*measurement error*) dan meningkatkan peluang untuk mendapatkan hasil tes yang akurat, adil, objektif, akuntabel, valid, dan reliabel (Bachman and Palmer, 1996). Pengertian tersebut mengandung makna bahwa untuk mendapatkan hasil tes yang mampu menggambarkan kompetensi yang sebenarnya dari hasil belajar

peserta didik, seorang guru dalam melaksanakan proses pengujian harus mengikuti prosedur sistematis dan logis. Lebih lanjut, Hughes (1989) menyatakan bahwa pelaksanaan tes mengacu pada prosedur untuk memberikan tes kepada individu atau kelompok peserta didik yang dimaksudkan untuk mengumpulkan informasi empiris berkaitan dengan kualitas tes dan juga untuk menarik kesimpulan tentang kompetensi peserta didik. Pengertian ini mengandung makna bahwa pelaksanaan tes dimaksudkan tidak hanya untuk mendapatkan informasi berkaitan hasil tes yang akurat, adil, objektif, akuntabel, valid, dan reliabel, tetapi juga dimaksudkan untuk mendapatkan gambaran berkaitan dengan kualitas butir-butir tes yang digunakan sebagai alat ukur hasil belajar. Kualitas butir-butir tes tentu saja dapat diketahui melalui proses analisis kualitatif maupun kuantitatif terhadap hasil tes yang telah dikerjakan oleh peserta didik.

Evaluasi terhadap kualitas dan kuantitas program pembelajaran telah dan akan terus menjadi bagian penting untuk dilakukan di sekolah. Bagi guru, untuk memutuskan apakah metode pembelajaran yang digunakan berfungsi dengan baik atau untuk memutuskan apakah peserta didik telah belajar dengan baik atau belum, maka guru harus melaksanakan tes dengan prosedur yang sistematis. Sebagaimana telah diketahui bahwa hasil tes dapat memberikan informasi tentang pengetahuan, kinerja, dan sikap peserta didik. Untuk mendapatkan hasil tes yang akurat hanya dapat diperoleh jika pelaksanaan tes dikelola dengan baik. Oleh karena itu, guru perlu memahami prinsip-prinsip pelaksanaan tes yang baik untuk memastikan hasil tes yang diperoleh adalah valid dan reliabel. Jika guru sebagai administrator tes tidak terlalu mahir dalam mengelola dan melaksanakan tes, maka dapat dipastikan tujuan tes tidak akan tercapai dan kemungkinan informasi yang diperoleh dari tes itu akan bias.

Sebagaimana telah dijelaskan di atas bahwa pelaksanaan tes yang baik akan memungkinkan hasil tes dapat menggambarkan kompetensi peserta didik secara akurat, adil, objektif, akuntabel, valid, dan reliabel. Secara spesifik, prosedur pelaksanaan tes yang baik akan mampu meningkatkan konsistensi dan keamanan hasil pengukuran. Konsistensi mengacu pada hasil tes yang konsisten, yaitu peserta didik berpeluang mendapatkan skor yang relatif sama atau sebanding, walaupun peserta didik itu mengikuti tes pada waktu, tempat, dan pengawas (*proctor*) yang berbeda. Sedangkan keamanan hasil pengukuran mengacu pada integritas skor yang diperoleh

oleh setiap peserta didik karena selama proses pelaksanaan tes tidak terjadi tindakan curang yang dilakukan oleh peserta didik tersebut.

B. Prinsip-prinsip Pelaksanaan Tes

Agar tes dapat menjalankan fungsinya sebagai alat ukur yang baik dan akurat terhadap hasil belajar peserta didik, pelaksanaan tes menjadi bagian dari keseluruhan proses pengujian yang sangat penting. Pelaksanaan tes tidak boleh dilaksanakan tanpa adanya perencanaan dan pengelolaan yang cermat. Perencanaan dan pengelolaan tes yang cermat memungkinkan setiap peserta didik memiliki peluang untuk mendapat hasil ujian yang adil, akuntabel, objektif, valid, dan reliabel. Perencanaan dan pengelolaan tes yang sembarangan akan memungkinkan peserta didik yang mengikuti tes merasa khawatir, tidak nyaman, dan terburu-buru dalam menyelesaikan tes. Satu hal yang harus dipahami adalah perencanaan dan pengelolaan tes yang baik hanya dapat dilakukan oleh guru sebagai administrator tes yang memiliki keterampilan dan pengalaman yang memadai. Selain itu, administrator tes harus merupakan figur yang dapat dipercaya, yaitu figur yang tidak akan memfasilitasi terjadinya kecurangan selama pelaksanaan tes berlangsung. Berkaitan dengan hal tersebut, pada bagian ini akan diuraikan prinsip-prinsip pelaksanaan tes yang baik agar akurasi, akuntabilitas, objektivitas, validitas dan reliabilitas hasil tes dapat dicapai dengan baik.

Prinsip pertama adalah setiap peserta didik harus diberi kesempatan yang sama untuk mencapai hasil belajar yang diinginkan atau hasil belajar yang dimaksudkan untuk diukur. Prinsip ini maksudnya adalah setiap peserta didik yang akan mengikuti tes harus berada dalam lingkungan tes yang kondusif dan aman sehingga lingkungan tersebut tidak akan mengganggu kondisi fisik dan psikologis setiap peserta didik. Selain itu, setiap peserta didik harus mendapatkan perangkat atau lembar tes dengan kualitas yang sama dilihat dari *lay out*, kejelasan hasil cetak soal dan kelengkapan lembar soal. Apabila tes dilaksanakan berbasis komputer, setiap peserta didik harus mendapatkan kualitas perangkat komputer yang sama dan stabilitas jaringan atau koneksi internet yang sama pula. Faktor-faktor yang mungkin mengganggu dalam mencapai validitas dan reliabilitas tes harus dikontrol secara ketat. Setiap peserta didik yang mengerjakan seperangkat tes yang sama harus berada dalam lingkungan yang sama-sama terkontrol dengan baik. Sebagai contoh, di saat satu kelompok peserta didik

yang sedang mengerjakan tes berada dalam lingkungan yang tenang dan nyaman, tidak boleh ada kelompok peserta didik lain yang terganggu deru suara kendaraan karena ruang tesnya sangat dekat dengan jalan raya yang sibuk. Apabila ada kelompok peserta didik yang mengerjakan tes dalam ruangan yang ber-AC, maka semua kelompok peserta didik yang lain pun harus mendapatkan fasilitas yang sama karena kenyamanan ruang ujian sangat berpengaruh terhadap cara peserta didik dalam menyelesaikan tes.

Dalam melaksanakan tes, guru juga harus memperhatikan kondisi psikologis dan fisik peserta didik yang sedang diuji. Kondisi psikologis dan fisik peserta didik pada saat mengerjakan soal-soal tes akan berpengaruh terhadap hasil tes yang mungkin akan mereka peroleh. Sebagai contoh, satu kelompok peserta didik mendapatkan kesempatan untuk mengikuti tes di pagi hari pada saat kondisi fisik mereka masih segar dan stamina mereka masih sangat prima. Di lain pihak, ada kelompok lain dari jenjang kelas yang sama mendapatkan kesempatan untuk mengikuti tes di siang hari dan mereka baru saja selesai mengikuti pelajaran olah-raga di lapangan. Kondisi fisik mereka tentu kelelahan akibat baru saja menyelesaikan pelajaran olah-raga yang menguras tenaga. Peluang untuk mendapatkan hasil tes yang baik tentu saja berbeda pada saat kondisi fisik mereka kelelahan dibandingkan saat kondisi fisik mereka masih segar. Kondisi yang sama terjadi apabila ada sekelompok peserta didik yang mengalami kendala psikologis. Misalnya, ada sekelompok peserta didik memiliki tingkat kecemasan yang tinggi saat mengikuti tes. Hal ini terjadi karena saat mengerjakan tes, kelompok itu juga harus memikirkan proyek dari mata pelajaran lain dan harus diselesaikan pada hari itu juga. Padahal kelompok itu belum mampu menyelesaikan proyek itu karena terbatasnya waktu yang disediakan. Padahal kelompok lain telah mampu menyelesaikan proyek itu karena memiliki waktu yang lebih longgar untuk menyelesaikannya. Dalam melaksanakan tes, guru harus memperhatikan berbagai kondisi tersebut agar setiap peserta didik memiliki peluang yang sama untuk mampu menunjukkan kompetensi mereka yang sebenarnya.

Prinsip kedua adalah peserta didik harus memiliki sikap positif terhadap tes yang sedang diikuti. Maksud dari prinsip ini adalah setiap peserta didik tidak boleh merasa cemas dan tegang selama mengikuti tes. Ketika peserta didik dalam suasana tegang, maka dipastikan peserta didik tidak akan dapat mencapai hasil tes yang optimal. Sebaliknya peserta didik

akan mampu menampilkan performa terbaiknya ketika suasana tes dalam kondisi nyaman dan pada akhirnya sikap positif peserta didik juga akan muncul saat mengikuti tes. Peserta didik akan berusaha keras dalam mencapai hasil yang optimal ketika sikap positif itu muncul saat mereka mengikuti tes. Tetapi pada umumnya guru sering kurang berhasil dalam memunculkan sikap positif peserta didik terhadap tes yang sedang mereka ikuti. Ada beberapa faktor yang menyebabkan kegagalan dalam memunculkan sikap positif peserta didik terhadap tes ini. Pertama, guru tidak memberikan kesempatan kepada peserta didik untuk membuat kesepakatan dengan guru berkaitan dengan waktu pelaksanaan tes. Biasanya guru secara sepihak menentukan waktu kapan tes akan dilaksanakan. Peserta didik tidak memiliki pilihan waktu yang sesuai dengan kesiapan mereka untuk mengikuti tes dan juga sesuai dengan kondisi psikologis dan fisik yang sedang mereka alami. Hal ini tentu akan mempengaruhi sikap mereka terhadap tes yang diberikan oleh guru. Kedua, guru sering menuntut secara verbal bahwa semua peserta didik harus berhasil mencapai nilai terbaik. Tuntutan ini justru kontra produktif karena kondisi peserta didik dalam kelas mungkin saja heterogen. Berbeda dengan peserta didik yang pandai, peserta didik yang mengalami lambat belajar menganggap tuntutan ini sebagai sebuah ancaman sehingga mereka tidak dapat berkonsentrasi selama mengikuti tes. Ketiga, bentuk tes dan jumlah butir soal yang diberikan guru tidak sesuai dengan ekspektasi peserta didik. Bentuk soal yang sangat variatif dan jumlah butir tes yang terlalu banyak juga akan berpengaruh terhadap sikap peserta didik terhadap tes yang sedang dihadapi. Keempat, tingkat kesulitan butir-butir soal yang diberikan oleh guru tidak berimbang. Tingkat kesulitan soal dengan kategori mudah (rendah), sedang dan tinggi harus berdistribusi normal. Untuk mengurangi ketegangan dan menumbuhkan sikap positif peserta didik terhadap tes yang sedang diikuti, seorang guru harus memperhatikan faktor-faktor yang menyebabkan sumber kecemasan dan ketegangan tersebut.

Prinsip ketiga adalah administrator tes (guru) harus terampil dan berpengalaman untuk melaksanakan tes dan merupakan figur yang dapat dipercaya. Prinsip ini dimaksudkan untuk memastikan bahwa pelaksanaan tes dikelola dengan baik dalam rangka mendapatkan hasil tes yang valid dan reliabel. Hal ini penting karena administrator tes bertanggung-jawab penuh terhadap tercapainya hasil tes yang akurat, akuntabel, adil, objektif, valid

dan reliabel. Administrator tes harus merupakan pihak-pihak yang dapat dipercaya dan tidak akan memfasilitasi peserta tes untuk berlaku curang. Kecurangan merupakan tindakan yang harus dihindari selama pelaksanaan tes berlangsung karena tindakan itu akan berpengaruh terhadap tidak tercapainya hasil tes yang akurat, akuntabel, adil, objektif, valid dan reliabel. Kecurangan juga mengakibatkan informasi yang diperoleh dari hasil tes menjadi bias atau tidak mampu menggambarkan kompetensi peserta didik yang sebenarnya.

Hal yang harus digarisbawai adalah persiapan yang baik akan berpengaruh terhadap kualitas pelaksanaan tes. Semakin baik persiapan yang dilakukan oleh administrator tes, maka akan semakin baik pula pelaksanaan tes itu dan juga sebaliknya. Oleh karena itu, administrator tes memiliki peran sentral terhadap akurasi tes untuk mengukur hasil belajar peserta didik. Administrator tes yang baik tidak akan memberi peluang terhadap berbagai kecurangan terjadi selama tes berlangsung. Kecurangan yang dilakukan oleh *testee* berpeluang besar terjadi jika pelaksanaan tes tidak direncanakan dengan baik. Guru sebagai administrator tes mestinya melakukan perencanaan yang baik sebelum tes dilaksanakan. Tetapi pada umumnya tahap persiapan ini sering kurang mendapat perhatian dari banyak guru sebagai administrator tes.

C. Tahapan Pelaksanaan Tes

Baik dan buruknya kualitas pelaksanaan tes sangat tergantung pada guru sebagai penguji yang sekaligus berperan sebagai administrator tes. Untuk mencapai pelaksanaan tes yang baik, guru harus melakukan berbagai aktivitas yang dibagi dalam tiga tahapan utama, yaitu tahap sebelum tes dilaksanakan, tahap saat tes dilaksanakan, dan tahap setelah tes dilaksanakan. Ketiga tahapan tersebut harus dikelola dengan baik dalam rangka mencapai hasil tes yang akurat, akuntabel, adil, objektif, valid dan reliabel.

1. Tahap sebelum tes dilaksanakan

Keamanan butir-butir soal merupakan aspek penting sebelum pelaksanaan tes. Setiap butir-butir soal, baik soal berbasis kertas (*paper-and-pencil test*) maupun soal berbasis komputer (*computer-based test*) harus disimpan di tempat yang aman dan tidak mudah diakses oleh berbagai pihak

yang tidak memiliki hak untuk mengakses soal tersebut sebelum digunakan sebagai alat ukur hasil belajar. Soal yang tidak dijaga keamanannya dengan baik memungkinkan pihak-pihak yang tidak berkepentingan, termasuk peserta didik yang akan mengikuti tes, akan dapat mengakses soal tersebut sehingga kualitas pelaksanaan tes tidak akan tercapai dan akhirnya hasil tes tidak valid dan reliabel. Untuk menghindari berbagai penyimpangan pelaksanaan tes yang tidak dikehendaki, soal harus disimpan di tempat yang aman guna mengurangi akses berbagai pihak yang tidak berkepentingan. Tes perlu disimpan di dalam ruangan yang terkunci rapat dan memiliki tingkat keamanan maksimal. Keamanan soal yang akan digunakan untuk mengukur hasil belajar juga menjadi tanggung-jawab penyusun soal. Apabila soal-soal ujian dikembangkan bukan oleh guru yang mengajar, komitmen pengembang soal tersebut untuk menjaga kerahasiaan soal sebelum digunakan juga menjadi faktor penting untuk mencapai kualitas pelaksanaan tes yang baik.

Untuk menjaga keamanan soal-soal ujian, pada saat menerima soal-soal dari percetakan atau lembaga penyedia soal, panitia tes harus memastikan bahwa soal-soal ujian dalam kondisi terbungkus rapat dan memastikan karton pembungkus soal tidak dalam kondisi rusak. Kemudian panitia harus memastikan kuantitas soal ujian sesuai kebutuhan dan soal yang diterima sesuai tingkat kelas yang akan mengikuti ujian. Setelah memastikan bahwa soal yang diterima sesuai dengan tingkat kelas yang akan mengikuti ujian, soal-soal tersebut kemudian disimpan dan segel soal dipastikan tidak rusak. Setelah memastikan soal aman sebelum digunakan, Gronlund & Linn (1990) menyarankan untuk mempertimbangkan hal-hal berikut pada saat melakukan proses persiapan pelaksanaan tes:

- a. Peserta didik harus telah diinformasikan berkaitan waktu pelaksanaan tes dan berapa lama tes harus diselesaikan.
- b. Peserta didik juga harus telah mendapatkan informasi berkaitan dengan cakupan materi yang akan diujikan.
- c. Peserta didik yang berkebutuhan khusus (misalnya mereka yang berkaca-mata minus dan memiliki kendala pendengaran) harus mendapatkan perhatian khusus sebelum tes dilaksanakan. Sebaiknya mereka ditempatkan di bagian depan atau tempat lain yang memudahkan mereka dapat mengakses soal-soal ujian dengan baik.

- d. Ruang dimana ujian akan dilaksanakan harus memiliki ventilasi, sistem pencahayaan, dan tempat duduk yang telah disusun dengan baik.
- e. Pengawas ujian telah membaca prosedur pelaksanaan tes dengan seksama berkaitan dengan waktu, peraturan ujian dan bentuk soal.
- f. Tempat duduk untuk peserta ujian harus disusun sedemikian rupa sehingga setiap peserta ujian tidak dapat melihat jawaban milik peserta yang lain.
- g. Setiap peserta ujian telah mendapatkan informasi dan penjelasan berkaitan dengan bagaimana ujian akan dilaksanakan, telah mendapatkan pelatihan tentang strategi mengerjakan soal-soal ujian, dan juga mendapatkan penjelasan berkaitan dengan manfaat ujian guna memperbaiki kualitas belajar-mengajar.
- h. Jika diperlukan, panitia ujian perlu mengamankan ruang ujian dari jangkauan orang-orang yang tidak berkepentingan dengan menempelkan tulisan yang berbunyi, misalnya “Sedang ada ujian, dilarang masuk”.

2. Tahap pada saat tes dilaksanakan

Persiapan yang baik sebelum pelaksanaan ujian belum tentu berjalan seperti yang diharapkan. Administrator tes memiliki tanggung-jawab dan tugas penting dalam mengkondisikan bahwa tes atau ujian berjalan dengan baik. Gronlund dan Linn (1990) selanjutnya memberikan rambu-rambu atau pedoman yang harus dilakukan oleh guru sebagai administrator tes agar ujian dapat berjalan dengan baik.

- a. Saat tes sedang berlangsung, pertimbangkan secara hati-hati dan pilihlah waktu yang tepat ketika ingin memberikan perintah atau klarifikasi tentang suatu hal kepada peserta tes.

Seringkali pengawas ujian menyampaikan pengumuman tentang suatu hal di saat peserta tes (peserta didik) sedang serius mengerjakan soal ujian. Misalnya, di tengah ujian yang sedang berlangsung dan peserta didik sedang serius mengerjakan tes, pengawas ujian menyampaikan pengumuman “Anak-anak, waktu untuk menyelesaikan tes adalah 120 menit. Silakan memanfaatkan waktu secara efisien agar semua soal dapat terselesaikan dengan baik”. Menyampaikan pengumuman di tengah-

tengah waktu ujian akan mengganggu konsentrasi peserta ujian. Hal ini harus diperhatikan oleh penyelenggara ujian dan sedapat mungkin dihindari. Bahkan Gronlund dan Linn (1990) menegaskan bahwa tindakan ini tidak fair dan telah merampas waktu peserta didik untuk dapat berkonsentrasi dalam mengerjakan soal-soal ujian. Pada saat mengerjakan soal-soal ujian, perhatian peserta didik secara fisik dan mental tercurah pada butir-butir soal. Pengalihan perhatian pada saat peserta tes mengerjakan butir-butir soal ujian akan mengakibatkan mereka kesulitan untuk mengingat kembali informasi yang diperlukan untuk menjawab pertanyaan. Dengan demikian, penyampaian pengumuman yang awalnya dimaksudkan baik, justru tindakan itu telah meningkatkan ketidaknyamanan peserta ujian.

- b. Jangan menginterupsi peserta tes hanya untuk menjelaskan butir soal yang tidak jelas, jika mereka tidak meminta penjelasan tentang itu.

Pada saat ada salah satu peserta ujian menanyakan hal-hal terkait dengan butir soal ujian, misalnya ada butir soal ujian yang ambigu dan tidak jelas, sebaiknya penjelasan terhadap butir tersebut tidak hanya ditujukan kepada satu peserta ujian saja. Penjelasan dapat disampaikan untuk semua peserta ujian agar tidak terjadi penjelasan yang berulang-ulang yang pada akhirnya justru mengganggu konsentrasi dan ketenangan peserta ujian.

- c. Jika ada salah satu peserta didik yang secara individu meminta penjelasan berkaitan dengan butir soal tertentu, jangan memberikan penjelasan yang dapat mengarahkan peserta didik itu mendapatkan jawaban (*clue*) terhadap butir soal tersebut.

Jika ada butir soal yang tidak jelas, maka guru atau pengawas ujian mungkin perlu memberi penjelasan berkaitan dengan apa maksud dari butir soal tersebut kepada peserta ujian. Jika hal ini terjadi, penjelasan yang disampaikan tidak boleh mengandung ‘petunjuk’ yang mengarahkan peserta tes yang pada awalnya tidak memiliki pengetahuan yang memadai dalam menjawab butir soal tertentu, pada akhirnya mendapat petunjuk untuk mampu menjawab butir soal tersebut. Penjelasan hanya diberikan berkaitan dengan kejelasan maksud butir soal, bukan memberikan ‘kata kunci’ agar peserta ujian mampu

menjawab butir soal tertentu. Apabila penjelasan yang diberikan justru memberi petunjuk kepada peserta ujian untuk dapat menjawab butir soal tertentu, hal ini akan mengakibatkan hasil ujian menjadi tidak valid dan tidak reliabel.

- d. Pastikan tidak terjadi kecurangan selama pelaksanaan ujian sedang berlangsung.

Memastikan tidak terjadi kecurangan saat ujian sedang berlangsung adalah tanggung-jawab guru yang berperan sebagai pengawas ujian. Akan tetapi, apabila ada hubungan yang akrab dan dekat antara guru dan peserta ujian kemungkinan kecurangan relatif sulit untuk dihindari. Untuk menghindari terjadinya kecurangan, pengaturan tempat duduk dan pengawasan ujian yang terorganisir dengan baik dipercaya mampu mengurangi peluang kecurangan terjadi saat ujian. Peserta tes yang berlaku curang dengan menerima bantuan jawaban illegal dari temannya telah mempengaruhi validitas dan reliabilitas hasil tes. Lebih jauh tindakan curang juga mencerminkan perilaku buruk yang terjadi pada saat ujian berlangsung. Lebih baik mengerjakan ujian berdasarkan kemampuan sendiri tetapi hasilnya valid daripada mendapatkan skor yang tinggi tetapi hasil dari tindakan curang saat ujian.

Sistem pengawasan ujian yang baik, misalnya pengawas secara periodik berkeliling ruang ujian untuk memastikan apa yang sedang dilakukan oleh peserta ujian dan bagaimana mereka mengerjakan ujian, merupakan cara yang efektif untuk mengurangi terjadinya kecurangan. Tetapi ada kendala besar yang sering terjadi saat pengawasan ujian berlangsung. Ada pengawas yang berasumsi bahwa sistem pengawasan hanya dipahami sebagai hadirnya seorang pengawas di dalam ruang ujian. Pengawas mungkin secara fisik hadir di ruang ujian, tetapi mereka tidak melakukan tugas pokok dan tanggung-jawab sebagai pengawas. Di ruang ujian, mereka menghabiskan waktunya untuk membaca koran atau novel, mengoreksi dan memberi skor hasil ujian yang telah dilaksanakan sebelumnya, dan menulis dokumen-dokumen lain yang tidak berkaitan dengan ujian. Gronlund & Linn (1990) bahkan menjelaskan bahwa sistem pengawasan ujian yang baik perlu dilakukan dengan mengamati apa yang

dilakukan peserta tes dan bagaimana mereka menyelesaikan tes, bukan hanya sekedar duduk di kursi pengawas.

Salah satu cara lain untuk mengurangi terjadinya kecurangan saat ujian adalah dengan cara membatasi penggunaan berbagai alat komunikasi di ruang ujian, baik oleh pengawas maupun peserta ujian. Bahkan pembatasan penggunaan alat komunikasi ini perlu dipastikan juga saat peserta ujian meminta izin keluar ruang ujian untuk sekedar ke kamar kecil. Peserta ujian perlu diingatkan sejak awal bahwa penggunaan alat komunikasi dilarang saat ujian sedang berlangsung dalam rangka meminimalisasi kecurangan. Lebih lanjut Rukundo & Magambo (2010) menyarankan bahwa untuk mengurangi terjadinya kecurangan selama ujian berlangsung, ada beberapa hal yang harus dilakukan oleh administrator tes, yaitu pastikan bahwa orang-orang yang masuk ruang ujian adalah mereka yang memiliki hak untuk mengikuti ujian, bukan joki yang mengambil posisi sebagai peserta ujian ilegal; tangani dan selesaikan keadaan darurat yang terjadi saat ujian secara cepat dan tepat; dan pada saat menjelang berakhirnya waktu ujian, sampaikan kepada peserta ujian berapa lama waktu yang masih tersisa untuk menyelesaikan ujian.

3. Tahap setelah tes dilaksanakan

Ketertiban saat pelaksanaan tes menjadi aspek yang sangat penting selama ujian berlangsung hingga hasil ujian ada di tangan administrator tes. Setelah ujian selesai dilaksanakan, berikut adalah hal-hal yang harus dilakukan oleh administrator tes.

- a. Semua perangkat ujian dan berbagai dokumen terkait pelaksanaan ujian, baik yang digunakan atau tidak digunakan, harus dikumpulkan kembali dan dipastikan jumlahnya sama seperti sebelum digunakan saat ujian. Semua perangkat dan dokumen itu kemudian disimpan di tempat yang aman dan terkunci.
- b. Urutkan lembar jawab ujian sesuai dengan urutan daftar dan nomor peserta ujian guna memastikan bahwa semua lembar jawaban yang terkumpul sesuai dengan apa yang tertera dalam daftar peserta ujian. Hal ini juga untuk memastikan bahwa tidak ada lembar jawab yang tertinggal di ruang ujian atau untuk memastikan bahwa semua peserta ujian telah menyerahkan lembar jawab yang telah mereka selesaikan.

Tes atau ujian adalah bagian integral dari aktivitas belajar dan mengajar dalam berbagai jenjang pendidikan, mulai dari pendidikan dasar hingga pendidikan tinggi. Agar pelaksanaan ujian dapat berjalan dengan baik dan hasil ujian dapat memberikan informasi yang akurat, adil, objektif, valid, dan reliabel, maka diperlukan prosedur dan regulasi yang cermat terhadap pelaksanaan ujian tersebut. Oleh karena itu, Rukundo & Magambo (2010) menegaskan bahwa selama pelaksanaan ujian:

- a. Semua peserta ujian harus diberi kesempatan yang adil melalui penyediaan lingkungan ujian, baik fisik dan psikologis yang kondusif.
- b. Setiap peserta didik yang sedang mengikuti ujian harus bersikap positif terhadap ujian yang sedang diikuti dan selalu mentaati peraturan yang telah ditetapkan agar hasil ujian akurat, adil, objektif, valid, dan reliabel.
- c. Guru memiliki peran sentral dan berkontribusi secara signifikan terhadap keberhasilan pelaksanaan tes atau ujian.
- d. Agar pelaksanaan tes berjalan lancar, tahapan sebelum, selama dan setelah tes harus dikelola dan diatur secara cermat dan hati-hati.

Bab 7

Teknik Penskoran

Setelah tes selesai dilaksanakan, selanjutnya tugas guru adalah memeriksa lembar-lembar jawaban tes yang telah selesai dikerjakan oleh setiap peserta didik. Selanjutnya guru membuat kesimpulan terhadap capaian belajar setiap peserta didik dalam menguasai materi yang telah diajarkan. Kesulitan yang mungkin muncul ketika seorang guru memeriksa lembar jawaban sangat dipengaruhi oleh bentuk tes yang telah diberikan dan dikerjakan oleh setiap peserta didik. Pemeriksaan tes bentuk uraian memiliki tingkat kesulitan yang lebih tinggi dibandingkan dengan tes bentuk objektif lainnya. Dengan kata lain, tes bentuk objektif merupakan tes yang paling mudah untuk diberikan skor dibandingkan dengan tes bentuk uraian. Keduanya menghendaki prosedur penskoran yang berbeda.

Proses penskoran merupakan tahapan yang sangat penting setelah seorang guru mengembangkan tes sebagai alat ukur hasil belajar peserta didik. Penskoran merupakan sebuah proses yang dilakukan oleh guru untuk menentukan angka-angka dalam bentuk skor yang mencerminkan kinerja atau kecakapan seorang peserta tes (peserta didik) yang telah mengikuti ujian tertentu. Angka-angka hasil penskoran ini pada umumnya masih berupa skor mentah (*raw score*) yang belum memberikan informasi yang bermakna terhadap hasil belajar peserta didik. Selanjutnya agar skor hasil pengukuran tersebut dapat memberikan informasi yang bermakna terhadap hasil belajar peserta didik, maka skor tersebut perlu dikonversi menjadi skor standar dengan mengikuti teknik konversi skor tertentu. Teknik penskoran soal bentuk objektif (pilihan ganda, menjodohkan, benar-salah, dan lain-lain) berbeda dengan penskoran tes bentuk subjektif (tes uraian, tes kinerja,

dan sebagainya). Penskoran dilakukan dengan melihat butir-butir soal mana yang dijawab benar oleh peserta didik.

A. Teknik Penskoran Domain Kognitif

Domain kognitif adalah ranah pendidikan yang berkaitan dengan kemampuan intelektual atau kemampuan berpikir seseorang. Domain kognitif merupakan salah satu domain yang sangat penting dalam keseluruhan proses pembelajaran karena dengan domain ini seseorang (peserta didik) akan memiliki kemampuan untuk mengingat dan memecahkan masalah tertentu. Dalam konteks pengukuran hasil belajar, ranah-ranah tersebut merupakan hal yang sangat penting untuk diidentifikasi dalam rangka memetakan perkembangan hasil belajar peserta didik. Sesuai dengan taksonomi Bloom, domain kognitif meliputi enam jenjang yang berbeda, yaitu pengetahuan (*knowledge*), pemahaman (*comprehension*), penerapan (*application*), analisis (*analysis*), evaluasi (*evaluation*) dan karya cipta (*creation*). Keenam hal tersebut dimulai dari kemampuan peserta didik yang paling sederhana (tingkat rendah) menuju kemampuan yang lebih kompleks.

Tingkat intelektualitas yang mengacu pada kemampuan peserta didik untuk mengingat kembali suatu istilah, rumusan, teori, konsep, prinsip, dan sebagainya. Jenis tes yang biasa dipakai untuk mengukur tingkat berpikir ini adalah jenis melengkapi kalimat, isian singkat, benar-salah dan sebagainya. Sedangkan pemahaman adalah kemampuan peserta didik untuk memahami suatu fakta tertentu. Kemampuan ini sedikit lebih kompleks dibandingkan dengan kemampuan untuk sekedar mengingat suatu fakta. Kemampuan ini mencakup daya intelektual seorang peserta didik untuk menangkap suatu makna dari bahan ajar yang telah dipelajari, kemudian menguraikan bahan ajar tersebut dan menyajikannya dalam bentuk yang berbeda. Selanjutnya kemampuan menerapkan mengacu pada kecakapan seorang peserta didik untuk menerapkan ide-ide umum, metode, dan prinsip dalam rangka memecahkan suatu masalah. Kemampuan analisis adalah suatu kemampuan untuk menguraikan bahan ajar menjadi bagian-bagian yang lebih spesifik dan memahami tata hubungan dari bagian-bagian tersebut. Di lain pihak, kemampuan mengevaluasi adalah kemampuan untuk menilai dari suatu hal. Terakhir, kemampuan mencipta mengacu pada sebuah kemampuan untuk menghasilkan sesuatu yang baru, baik yang benar-benar hal baru atau suatu

ide baru yang diperoleh dengan cara menghubungkan beberapa hal yang sudah ada dan menjadikannya suatu hal baru. Keenam tingkat kemampuan tersebut diukur dengan menggunakan berbagai bentuk tes yang berbeda, misalnya bentuk pilihan ganda, benar-salah, uraian dan sebagainya.

1. Teknik Penskoran Soal pilihan Ganda

Teknik penskoran tes bentuk pilihan ganda relatif lebih mudah dilakukan dibandingkan dengan bentuk tes yang lain, misalnya bentuk tes uraian. Proses penskoran bentuk pilihan ganda dapat dilakukan secara objektif karena proses penskorannya cukup membandingkan jawaban yang diberikan peserta didik dengan kunci jawaban yang telah disiapkan terlebih dulu oleh guru sebagai korektor tes. Apabila jawaban yang diberikan oleh peserta didik sesuai atau cocok dengan kunci jawaban, maka jawaban tersebut dikategorikan sebagai jawaban yang benar dan sebaliknya. Untuk menskor tes bentuk ini, ada paling tidak tiga teknik yang dapat digunakan oleh guru. Ketiga teknik tersebut adalah teknik penskoran tanpa koreksi, teknik penskoran dengan koreksi, dan teknik penskoran beda bobot. Ketiga teknik tersebut memiliki karakteristik yang berbeda satu sama lain. Ketiga teknik tersebut digunakan apabila tes yang digunakan sudah diketahui kualitasnya atau tes sudah pernah diujicobakan dan dianalisis. Sedangkan untuk tes yang belum diketahui kualitasnya, skor mentah cukup ditentukan dengan menghitung jumlah butir tes yang dijawab benar saja. Setiap butir yang dijawab benar diberi skor satu dan butir-butir yang dijawab salah diberi skor nol.

a. Teknik penskoran tanpa koreksi

Teknik penskoran tanpa koreksi adalah suatu proses penskoran yang hanya melihat berapa butir tes yang mampu dijawab benar oleh setiap peserta didik dan berapa butir yang dijawab salah oleh peserta didik tanpa adanya penalti terhadap butir-butir tes yang dijawab salah. Ini berarti bahwa setiap butir soal yang dijawab dengan benar oleh peserta didik akan diberi skor 'satu' dan untuk setiap butir soal yang dijawab salah diberi skor 'nol'. Formula yang biasa digunakan untuk memberikan skor dengan teknik ini adalah:

$$\text{Skor} = \frac{B}{N} \times 100$$

B = Jumlah butir tes yang dijawab benar

N = Jumlah total butir tes

Contoh:

Ada 50 butir untuk tes bahasa Inggris. Budi menjawab 25 butir dengan benar. Berapa skor yang diperoleh Budi?

Kemudian nilai skor yang diperoleh Budi dihitung:

$$\text{Skor} = \frac{25}{50} \times 100 = 50$$

Jadi skor yang diperoleh oleh Budi adalah 50.

b. Teknik penskoran dengan koreksi

Teknik ini mengacu pada proses penskoran dengan memberikan pertimbangan pada berapa banyak butir tes yang dijawab dengan benar, berapa banyak butir dijawab salah, dan berapa banyak butir yang tidak dijawab oleh peserta didik. Setiap butir tes yang dijawab dengan benar diberi skor 1; yang tidak dijawab diberi skor 0; dan jawaban yang salah diberikan skor -1. Formula atau rumus yang biasa digunakan adalah sebagai berikut:

$$\text{Skor} = \left[\left(B - \frac{S}{p-1} \right) / N \right] \times 100$$

B = Jumlah butir yang dijawab dengan benar

S = Jumlah butir yang dijawab salah

p = Jumlah alternatif jawaban untuk setiap butir tes

N = Total butir tes

Contoh:

Tes pilihan ganda terdiri dari 40 butir dengan empat alternatif jawaban untuk setiap butirnya. Amir mampu menjawab dengan benar 20 butir dari 40 butir tes tersebut, 12 butir dijawab salah dan ada 8 butir yang tidak dijawab oleh Amir. Berapa skor yang diperoleh oleh Amir?

Skor yang diperoleh Amir dihitung sebagai berikut:

$$\begin{aligned}\text{Skor} &= \left[\left(\frac{12}{4 - 1} \right) / 40 \right] \times 100 \\ &= 40\end{aligned}$$

Jadi skor yang diperoleh oleh Amir adalah 40.

c. Teknik penskoran dengan beda bobot

Teknik ini mengacu pada proses pemberian skor dengan mempertimbangkan bobot atau poin yang berbeda untuk setiap kelompok butir tes. Bobot untuk butir-butir tertentu biasanya disesuaikan dengan tingkat kognitif peserta didik (pengetahuan, pemahaman, penerapan, analisis, evaluasi, kreasi atau daya cipta). Rumus yang biasa digunakan untuk menghitung skor dengan teknik adalah sebagai berikut:

$$\text{Skor} = \sum \frac{B_i \times b_i}{S_t} \times 100 \%$$

B_i = Jumlah butir yang dijawab benar

b_i = Bobot untuk kelompok butir soal tertentu

S_t = Skor teoritis atau skor maksimum yang mungkin diperoleh ketika semua butir dijawab dengan benar

Contoh:

Sebuah tes mata pelajaran biologi terdiri dari 40 butir tes dan memiliki enam ranah kognitif yang berbeda dan setiap ranah diberi bobot sebagai berikut: 1

poin untuk ranah pengetahuan, 2 poin untuk ranah pemahaman, 3 untuk ranah aplikasi, 4 poin untuk ranah analisis, 5 poin untuk ranah evaluasi, dan 6 poin untuk ranah kreasi. Badu dapat menjawab dengan benar 8 butir dari 12 butir untuk ranah pengetahuan; 12 dari 20 butir untuk ranah pemahaman benar; 2 butir dari 4 butir untuk ranah aplikasi benar; 1 butir dari 2 butir untuk ranah analisis benar; 1 butir dari ranah evaluasi benar; dan 1 butir untuk ranah kreasi juga benar. Berapa skor yang didapat Badu?

Untuk memudahkan proses penskoran hasil belajar yang dicapai oleh Badu tersebut, perlu dibuat tabel bantuan sebagai berikut:

Domain kognitif	Jumlah butir	Bobot butir (b _i)	Jum. butir x b _i	B _i
Pengetahuan	12	1	12	8
Pemahaman	20	2	40	12
Aplikasi	4	3	12	2
Analisis	2	4	8	1
Evaluasi	1	5	5	1
Kreasi	1	6	6	1
Total	40		S _t = 83	25

$$\text{Skor} = \sum \frac{(8 \times 1) + (12 \times 2) + (2 \times 3) + (1 \times 4) + (1 \times 5) + (1 \times 6)}{83} \times 100 \%$$

$$= 63,9\%$$

Jadi skor yang diperoleh oleh Badu untuk mata pelajaran biologi dengan data tersebut adalah sebesar 63,9% yang berarti penguasaan Badu terhadap materi yang diujikan sebesar 63,9%.

2. Teknik Penskoran Soal Benar-salah

Teknik penskoran butir tes bentuk benar-salah dilakukan dengan rumus yang sangat sederhana. Karena tes bentuk benar-salah merupakan salah satu dari tes bentuk objektif, maka objektivitas dalam pemberian skor

juga sangat tinggi dibandingkan dengan tes bentuk uraian. Penskoran tes bentuk benar-salah bahkan dapat dilakukan oleh orang-orang yang tidak banyak memahami substansi materi yang diujikan. Ada paling tidak dua teknik penskoran butir tes bentuk benar-salah, yaitu penskoran dengan denda dan penskoran tanpa denda. Keduanya memiliki kelebihan dan kekurangan masing-masing.

a. Penskoran dengan denda

Teknik ini digunakan apabila skor yang diperoleh peserta didik sebanyak jawaban yang cocok dengan kunci jawaban. Teknik ini digunakan karena guru menganggap adanya unsur tebakan (*guessing*) yang dilakukan oleh peserta didik dalam menjawab butir-butir tes. Penskoran tes benar-salah dengan teknik ini sangat dipengaruhi oleh berapa jumlah butir tes yang dijawab salah oleh peserta didik. Butir-butir soal yang tidak dijawab oleh peserta didik juga dianggap salah. Salah satu kelemahan dari teknik ini adalah ada kemungkinan peserta didik yang mengikuti tes akan mendapatkan skor negatif. Rumus yang biasa digunakan untuk teknik ini adalah sebagai berikut:

$$S = \Sigma R - \Sigma W$$

S	= Skor yang diperoleh
ΣR	= Jumlah butir yang dijawab benar
ΣW	= Jumlah butir yang dijawab salah

atau dengan rumus yang lain sebagai berikut:

$$S = T - 2W$$

S	= Skor yang diperoleh
T	= Total butir tes yang diujikan
W	= Butir yang dijawab salah

Contoh:

Guru memberikan tes bentuk benar-salah dengan jumlah butir sebanyak 40 butir tes untuk dikerjakan oleh setiap peserta didik. Seorang peserta didik

mampu mengerjakan 30 butir dengan benar dan sejumlah 10 butir dijawab salah. Berapakah skor yang diperoleh peserta didik tersebut?

$$S = 30 - 10 = 20$$

Dengan rumus yang kedua:

$$S = 40 - 2(10) = 20$$

Jadi skor yang diperoleh peserta didik tersebut adalah 20.

b. Penskoran tanpa denda

Teknik ini sering digunakan oleh guru untuk melakukan penskoran tes bentuk benar-salah. Teknik ini tidak melihat berapa butir tes yang dijawab salah oleh seorang peserta didik. Skor yang diperoleh oleh seorang peserta didik hanya didasarkan pada berapa butir tes yang dijawab benar. Dengan demikian, rumus yang biasa dipakai untuk memberi skor dengan teknik ini adalah sebagai berikut:

$$S = \sum R$$

Contoh:

Guru memberikan tes bentuk benar-salah dengan jumlah butir sebanyak 40 butir tes untuk dikerjakan oleh setiap peserta didik. Seorang peserta didik mampu mengerjakan 30 butir dengan benar dan sejumlah 10 butir dijawab salah. Berapakah skor yang diperoleh peserta didik tersebut?

$$S = 30$$

Jadi, skor yang diperoleh peserta didik tersebut adalah 30. Dengan demikian, penskoran dengan teknik ini tidak memungkinkan seorang peserta didik mendapatkan skor negatif.

3. Teknik Penskoran Soal Bentuk Menjodohkan

Penskoran tes bentuk menjodohkan (*matching type*) hanya didasarkan pada berapa banyak butir tes yang dijawab dengan benar oleh seorang peserta didik. Kelemahan yang sering terjadi dengan teknik penskoran ini adalah pemberi skor (guru) sering mengabaikan tingkat kesukaran butir-

butir tes yang dikerjakan oleh peserta didik. Rumus yang biasa digunakan adalah:

$$S = \Sigma R$$

Dengan rumus penskoran di atas, butir-butir yang dijawab salah dan tidak dijawab atau dikosongkan oleh peserta didik, keduanya dianggap salah karena yang dihitung sebagai skor hanya butir-butir yang dijawab benar.

Contoh:

Seorang guru matematika telah mengembangkan 10 butir tes berbentuk menjodohkan. Peserta didik bernama A mampu menjawab dengan benar 8 butir dan sisanya dijawab salah. Maka skor yang diperoleh oleh siswa A tersebut adalah:

$$10 - 2 = 8$$

Jadi skor yang diperoleh oleh peserta didik bernama A tersebut adalah 8.

4. Teknik Penskoran tes Bentuk Uraian

Dalam konteks penskoran, tes uraian dapat dibedakan menjadi dua jenis, yaitu tes uraian objektif dan non-objektif. Perbedaan keduanya terletak pada kepastian dalam memberikan skor. Pada bentuk tes uraian objektif, ada kepastian dalam hal jawaban dan pedoman penskorannya. Jawaban dan pedoman penskorannya sudah diuraikan secara jelas dan detail sehingga penskoran dapat dilakukan dengan lebih objektif. Sedangkan tes uraian non-objektif, pedoman penskoran biasanya dinyatakan dalam bentuk rentang, misalnya 0 – 4 atau 0 – 10. Proses penskoran untuk tes bentuk ini seringkali banyak dipengaruhi oleh subjektivitas guru yang memberi skor. Guru harus mampu meminimalisir subjektivitasnya sehingga skor yang diperoleh oleh peserta didik benar-benar mencerminkan kompetensi sebenarnya. Ada beberapa strategi yang dapat dilakukan oleh guru untuk mengurangi subjektivitasnya dalam memberikan skor untuk tes bentuk ini, yaitu:

- a. Guru harus menyiapkan garis-garis besar jawaban yang diharapkan muncul dan diberikan oleh peserta didik sebelum melakukan proses penskoran.
- b. Sedapat mungkin menghindari untuk melihat identitas peserta didik yang lembar jawabnya sedang dilakukan proses penskoran.
- c. Guru harus mampu menghilangkan atau menghindari faktor-faktor yang mengakibatkan munculnya subjektivitas ketika sedang memberikan skor. Faktor-faktor itu misalnya, rapi dan tidaknya tulisan peserta didik, gaya bahasa yang digunakan oleh peserta didik dalam menjawab butir-butir tes, dan sebagainya.
- d. Terlebih dulu guru harus menetapkan metode penskoran tertentu sebelum melakukan proses penskoran.

Untuk mendapatkan skor yang mampu mencerminkan kompetensi peserta didik yang sebenarnya, ada dua metode penskoran yang dapat diterapkan terhadap butir-butir tes bentuk uraian. Dua metode itu adalah *point method* dan *rating method*. *Point method* adalah metode penskoran yang dilakukan dengan membandingkan setiap jawaban yang diberikan oleh setiap peserta didik dengan jawaban ideal yang ditetapkan sebagai kunci jawaban dan besar kecilnya skor yang diberikan sangat tergantung pada tingkat kesamaan atau kecocokan antara jawaban dan kunci jawaban. Sedangkan *rating method* adalah metode penskoran dimana setiap jawaban yang diberikan oleh peserta didik ditetapkan dalam kelompok-kelompok tertentu menurut kualitas jawaban tersebut. Kelompok-kelompok tersebut sangat menentukan berapa skor yang akan diberikan untuk setiap jawaban. Sebagai contoh, sebuah butir tes akan diberi skor maksimal 5, maka butir tes tersebut dapat dibuat 6 kelompok jawaban dengan rentang 5 hingga 0. Selanjutnya, sebuah jawaban dikategorikan masuk kelompok 0, jika jawaban tersebut sama sekali tidak cocok dengan kunci jawaban yang telah ditentukan. Sedangkan sebuah jawaban dikategorikan masuk kelompok 5, apabila jawaban tersebut secara keseluruhan cocok dengan kunci jawaban yang telah ditentukan, dan seterusnya.

- e. Guru sebaiknya terlebih dahulu melakukan penskoran terhadap satu nomor butir tes yang sama untuk setiap peserta didik sebelum berpindah untuk melakukan penskoran untuk nomor butir yang

lain. Misalnya guru terlebih dahulu menyelesaikan penskoran butir no. 1 untuk setiap peserta didik sebelum memberikan skor butir no. 2. Hal ini dilakukan untuk menghindari berubah-ubahnya skor untuk satu butir tes akibat dipengaruhi perubahan standar penskoran dari satu butir tes ke butir yang lain.

- f. Guru sebaiknya segera menyusun pedoman penskoran setelah butir-butir tes selesai dikembangkan.
- g. Guru sebaiknya menggunakan teknik *inter-rater*, yaitu suatu proses penskoran terhadap hasil belajar peserta didik yang dilakukan oleh lebih dari satu orang (pemberi skor). Skor yang diberikan oleh setiap *rater* kemudian saling dibandingkan untuk mengetahui sejauhmana konsistensi skor yang diberikan oleh beberapa *rater* tersebut. Semakin konsisten skor yang diberikan oleh beberapa *rater*, semakin objektif pula skor tersebut.

Beberapa strategi tersebut dapat diterapkan secara konsisten oleh guru yang sedang melakukan proses penskoran terhadap lembar jawaban. Hal ini dimaksudkan agar skor yang diberikan tidak bias dan mencerminkan kemampuan sebenarnya yang telah dicapai setiap peserta didik.

a. Penskoran Tes Uraian Objektif

Skor yang diberikan untuk tes bentuk uraian objektif lebih bersifat dikotomis, yaitu proses pemberian skor yang didasarkan pada dua kategori (benar dan salah) yang berlaku secara mutlak. Setiap jawaban yang dikategorikan benar diberi skor 1, dan setiap butir tes yang dijawab salah atau tidak dijawab sama sekali oleh peserta didik diberi skor 0. Peserta didik kemungkinan akan memberikan jawaban yang tidak sama persis dengan kunci jawaban yang telah ditentukan sebelumnya. Tetapi tingkat benar dan salah dari jawaban itu didasarkan pada kecocokan antara kata-kata kunci yang ada di setiap jawaban yang diberikan oleh peserta didik dengan kata-kata kunci yang ada dalam kunci jawaban. Untuk mendapatkan skor yang objektif, maka perlu diperhatikan langkah-langkah penskoran sebagai berikut:

- 1) Guru perlu menuliskan kata-kata kunci sebagai pedoman untuk menentukan benar dan salahnya sebuah jawaban yang diberikan oleh setiap peserta didik.
- 2) Setiap jawaban yang benar diberi skor 1, sedangkan jawaban yang salah atau butir tes yang tidak dijawab diberi skor 0. Tidak ada skor di luar dua angka tersebut.
- 3) Apabila ada satu butir tes yang memiliki sub-butir tes, maka guru harus merinci kata-kata kunci dari jawaban tersebut menjadi beberapa kata kunci sub-jawaban dan kemudian tentukan skor untuk sub-sub jawaban tersebut.
- 4) Hitung skor dari semua kata kunci yang telah ditentukan pada tes tersebut sebagai skor maksimum.

Contoh:

Indikator: Peserta didik mampu menentukan kecepatan rata-rata dalam satuan meter sebuah sepeda motor yang melaju dari kota A ke kota B.

Butir tes : Seseorang memacu sepeda motornya dari kota A ke kota B yang berjarak 10 km dengan waktu tempuh 15 menit. Berapa meter rata-rata kecepatan motor tersebut selama melaju dari kota A dan kota B?

Pedoman penskoran:

Langkah	Kunci Jawaban	Skor
1	Rumus kecepatan = Jarak tempuh / waktu tempuh	1
2	= $10 / 15$	1
3	= 0,67 km	1
4	Kecepatan rata-rata dalam meter = $0,67 \times 1000$ meter	1
5	= 670 meter	1
Skor maksimum		5

b. Penskoran Tes Uraian Non-Objektif

Hal yang mula-mula harus dilakukan untuk menskor tes uraian bentuk non-objektif adalah membuat rentang penskoran, misalnya 0 – 5, 0 – 8, 0 –

10, dan sebagainya. Dalam hal ini skor minimal untuk setiap butir tes adalah 0, sedangkan skor maksimum ditentukan oleh guru sebagai pengembang tes dengan memperhatikan kompleksitas materi yang diujikan. Langkah-langkah penskoran tes uraian bentuk non-objektif adalah sebagai berikut:

- 1) Guru perlu merumuskan garis-garis besar sebagai kriteria jawaban yang benar untuk digunakan sebagai pedoman dalam memberikan skor tiap butir tes.
- 2) Guru menentukan rentang skor untuk setiap butir tes sesuai dengan kompleksitas materi yang diujikan.
- 3) Skor yang diberikan untuk setiap jawaban untuk butir tertentu ditentukan sesuai dengan kualitas jawaban yang diberikan oleh peserta didik.
- 4) Hitung skor yang diperoleh seorang peserta didik untuk butir tertentu dengan rumus sebagai berikut:

$$S_b = \frac{S_p}{S_m} \times B$$

S_b = Skor untuk butir tes tertentu

S_p = Skor perolehan oleh peserta didik untuk butir tes tertentu

S_m = Skor maksimum untuk butir tes tertentu

B = Bobot butir tes tertentu

- 5) Hitung skor akhir yang diperoleh seorang peserta didik dengan cara menjumlahkan semua skor yang diperoleh dari semua butir tes.

Contoh:

Indikator : Jelaskan mengapa pencemaran lingkungan di Indonesia menjadi semakin parah dalam satu dekade terakhir!

Kriteria jawaban: Jawaban yang diberikan oleh peserta didik kemungkinan dapat beraneka ragam, tetapi jawaban tersebut dapat dikelompokkan dalam kriteria sebagai berikut.

Kriteria jawaban	Rentang skor
Pencemaran yang diakibatkan oleh kebiasaan masyarakat untuk membuang sampah sembarangan beserta penjelasannya.	0 – 3
Pencemaran udara yang diakibatkan oleh asap industri beserta penjelasannya.	0 – 3
Pencemaran udara sebagai akibat ulah masyarakat yang membakar hutan dan lahan secara besar-besaran beserta penjelasannya.	0 – 3
Pencemaran air sebagai akibat dari buangan limbah industri, limbah rumah tangga, dan limbah pertanian beserta penjelasannya.	0 – 4
Skor maksimum	13

5. Penskoran Bentuk Tes campuran

Ada kalanya seperangkat tes terdiri dari campuran antara bentuk objektif (pilihan ganda, menjodohkan, benar salah, dll) dan benk tes non-objektif (uraian). Dalam sebuah perangkat tes, jumlah butir-butir tes bentuk objektif biasanya lebih banyak dibandingkan dengan jumlah butir-butir uraian. Bobot skor untuk butir soal objektif pun berbeda dengan bobot butir bentuk uraian. Misalnya, seperangkat tes terdiri dari n_1 soal pilihan ganda dan n_2 soal uraian. Bobot soal pilihan ganda b_1 dan bobot soal esai b_2 . Jika peserta didik menjawab dengan benar n_1 untuk tes pilihan ganda, dan menjawab dengan benar n_2 untuk tes uraian, maka skor peserta didik adalah:

$$\text{Skor} = b_1 \left[\frac{n_1}{n_1} \times 100 \right] + b_2 \left[\frac{n_2}{n_2} \times 100 \right]$$

b_1 = Bobot untuk tes 1 (pilihan ganda)

b_2 = Bobot untuk tes 2 (uraian)

n_1 = Tes 1 (pilihan ganda)

n_2 = Tes 2 (uraian)

Contoh:

Suatu perangkat tes terdiri dari 50 butir bentuk pilihan ganda dengan 5 pilihan, dan 5 buah butir bentuk uraian. Amir dapat menjawab benar soal pilihan ganda 40 butir dan salah 10 butir, sedang bentuk uraian bisa dijawab benar 20 dari skor maksimum 25. Apabila bobot pilihan ganda adalah 0,45 dan bentuk uraian 0,55, maka skor yang diperoleh Amir dapat dihitung sebagai berikut.

- ✓ Skor pilihan ganda tanpa koreksi jawaban adalah $\frac{40}{50} \times 100 = 80$
- ✓ Skor bentuk uraian adalah $\frac{20}{25} \times 100 = 80$
- ✓ Skor akhir adalah $(0,45 \times 80) + (0,55 \times 80) = 80$

6. Penskoran Tugas

Untuk melakukan penskoran terhadap tugas yang telah dikerjakan oleh peserta didik, diperlukan beberapa kriteria tertentu yang akan digunakan sebagai dasar atau panduan untuk memberikan skor terhadap tugas tersebut. Semakin rinci kriteria yang ditentukan oleh guru, semakin informatif makna dari skor tersebut terhadap kompetensi yang dimiliki oleh peserta didik. Sebagai contoh, kriteria yang dapat digunakan oleh guru untuk mengukur kompetensi peserta didik dalam menyelesaikan tugas adalah:

- a. Ketepatan waktu dalam menyelesaikan dan menyerahkan tugas kepada guru.
- b. Keseriusan peserta didik dalam menyelesaikan tugas dilihat dari bentuk fisik, kerapian, artistik, dan sebagainya.
- c. Sistematika tugas yang menunjukkan alur berpikir peserta didik dalam menyelesaikan tugas tersebut.
- d. Kelengkapan isi tugas yang menyangkut kepadatan dan ketuntasan dari isi tugas tersebut.
- e. Kualitas tugas yang telah diselesaikan oleh peserta didik dilihat dari kesesuaian antara tugas yang dihasilkan dengan garis-garis besar tugas yang telah ditentukan oleh guru.

Berbagai kriteria tersebut kemudian diberi bobot tertentu sebagai pedoman untuk menentukan nilai akhir. Sebagai contoh, berbagai kriteria tersebut diberi bobot sebagai berikut:

- A1 : Ketepatan waktu diberi bobot 2
- A2 : Bentuk fisik diberi bobot 1
- A3 : Sistematika tugas diberi bobot 3
- A4 : Kelengkapan isi tugas diberi bobot 3
- A4 : Kualitas tugas diberi bobot 3

Dengan demikian, nilai akhir dari tugas yang telah diselesaikan oleh seorang peserta didik dapat dihitung dengan formula sebagai berikut:

$$NAT = \frac{(2 \times A1) + (1 \times A2) + (3 \times A3) + (3 \times A4) + (3 \times A4)}{12}$$

B. Teknik Penskoran Domain Afektif

Kemampuan afektif yang dimiliki oleh peserta didik pada dasarnya bukan merupakan hasil belajar secara langsung melalui proses pembelajaran di kelas, tetapi kemampuan tersebut diperoleh melalui pengalaman peserta didik melalui proses interaksi dengan lingkungan sekitar. Kemampuan afektif meliputi berbagai hal berkaitan dengan sikap, minat, nilai, dan kosep diri. Kemampuan afektif sangat menentukan kesuksesan belajar peserta didik. Seorang peserta didik yang memiliki minat yang tinggi dan sikap positif terhadap mata pelajaran tertentu, maka peluang untuk sukses dalam belajar untuk mata pelajaran tersebut juga semakin tinggi dan sebaliknya.

Untuk mengukur domain afektif yang dimiliki oleh peserta didik, guru harus memahami terlebih dahulu definisi aspek yang akan diukur. Misalnya seorang guru akan mengukur aspek sikap dan minat terhadap mata pelajaran yang sedang diampu, maka guru tersebut harus paham betul apa itu sikap dan minat terlebih dahulu. Mardapi (2008) menyatakan bahwa sikap adalah kecenderungan seseorang untuk merespon secara positif atau negatif terhadap suatu objek, situasi, konsep, atau orang. Dalam konteks sekolah, sikap dapat berupa sikap peserta didik terhadap mata pelajaran tertentu. Sikap juga mengacu pada pernyataan evaluatif terhadap segala sesuatu, dapat berupa objek, orang atau peristiwa. Sikap mempunyai tiga komponen utama, yaitu kesadaran, perasaan, dan perilaku. Sikap seseorang terhadap sesuatu dapat berupa sikap positif atau sikap negatif. Selanjutnya minat adalah dorongan atau keinginan seseorang pada suatu objek tertentu,

misalnya minat terhadap mata pelajaran, hobi, nonton film, dan sebagainya. Minat seseorang terhadap suatu objek tertentu akan selalu didasari oleh keinginan dan kebutuhan dari orang tersebut.

Untuk mengukur domain afektif, diperlukan instrumen yang baik agar hasil pengukuran mampu memberikan informasi yang akurat. Mardapi (2008) menjelaskan bahwa untuk mengembangkan instrumen yang baik untuk mengukur domain afektif, ada beberapa langkah yang perlu diperhatikan, yaitu:

1. Memilih domain afektif yang akan dinilai, misalnya sikap dan minat.
2. Menentukan spesifikasi instrumen yang memuat tujuan dan kisi-kisi instrumen.
3. Menuliskan instrumen.
4. Menentukan skala instrumen, misalnya skala Likert, skala Thurstone, dan sebagainya.
5. Menentukan sistem penskoran.
6. Menelaah instrumen.
7. Melakukan uji coba instrumen.
8. Merakit instrumen.
9. Melaksanakan pengukuran.
10. Menafsirkan hasil pengukuran.

Setelah instrumen untuk mengukur domain afektif telah dikembangkan, maka instrumen tersebut dapat digunakan untuk mengukur aspek yang ingin diukur.

Contoh:

Seorang guru bahasa Inggris ingin mengukur minat peserta didik untuk belajar bahasa Inggris. Guru kemudian mengembangkan instrumen yang terdiri dari 10 butir dengan rentang respon peserta didik 1 – 5 (dimulai dari sangat tidak berminat hingga sangat berminat). Bagaimana kategori minat peserta didik terhadap mata pelajaran bahasa Inggris tersebut?

Jika rentang yang digunakan dalam instrumen adalah 1 - 5, maka skor terendah seorang peserta didik adalah 10 (hasil perkalian dari 10 x 1) dan skor tertinggi sebesar 50 (hasil perkalian dari 10 x 5). Dengan demikian,

median dari skor tersebut adalah $(10 + 50)/2$ atau sebesar 30. Apabila dibagi menjadi 4 kategori, maka peserta didik dengan skor:

- 10 – 20 : kategori tidak berminat untuk belajar bahasa Inggris,
- 21 – 30 : kategori kurang berminat untuk belajar bahasa Inggris,
- 31 – 40 : kategori berminat untuk belajar bahasa Inggris,
- 41 – 50 : kategori sangat berminat untuk belajar bahasa Inggris.

Apabila seorang peserta didik berdasarkan hasil pengukuran mendapat skor 42, berarti peserta didik tersebut dalam kategori memiliki minat yang sangat baik dalam belajar bahasa Inggris karena skor 42 berada pada rentang 41 - 50.

C. Teknik Penskoran Domain Psikomotor

Tes ranah psikomotor adalah tes yang digunakan untuk mengukur keterampilan atau kinerja yang telah dikuasai oleh peserta didik. Untuk mengukur kemampuan ini, peserta didik diminta untuk mempraktikkan keterampilan tertentu dengan mengikuti prosedur dan indikator yang telah ditentukan sebelum tes dilaksanakan. Keterampilan yang ditunjukkan oleh peserta didik yang semakin mendekati atau sesuai dengan prosedur dan indikator yang telah ditentukan, maka semakin baik skor yang akan didapatkan oleh peserta didik tersebut. Keterampilan peserta didik kemudian diukur dengan menggunakan instrumen dengan skala tertentu, misalnya skala 5, dengan kriteria terendah 1 yang berarti keterampilannya sangat tidak sempurna dan kriteria tertinggi 5 yang berarti keterampilan peserta didik sangat sempurna.

Contoh:

Pada mata pelajaran fisika, guru ingin mengukur sejauhmana keterampilan peserta didik dalam menggunakan termometer untuk mengukur suhu badan seseorang. Untuk mengukur keterampilan tersebut, guru menentukan indikator-indikator keterampilan yang akan diukur sebagai berikut:

1. Cara peserta didik mengeluarkan termometer dari wadah dimana termometer itu ditempatkan.
2. Cara peserta didik menurunkan air raksa pada posisi terendah pada termometer tersebut.

3. Cara memasang termometer pada posisi badan seseorang yang akan diukur suhu tubuhnya.
4. Waktu yang diperlukan oleh peserta didik untuk memasang termometer pada badan seseorang yang akan diukur suhu badanya.
5. Cara mengambil kembali termometer dari badan seseorang yang telah selesai diukur suhu badanya.
6. Cara membaca hasil pengukuran suhu badan berdasarkan tinggi dan rendahnya air raksa pada pipa kapiler termometer.

Sejauhmana keterampilan peserta didik dalam menggunakan termometer untuk mengukur suhu tubuh seseorang didasarkan pada enam indikator yang telah ditentukan tersebut. Apabila seorang peserta didik mendapatkan skor 5 pada indikator nomor 1 yang berarti keterampilan peserta didik sangat baik dalam mengeluarkan termometer dari wadah dimana termometer itu ditempatkan; mendapatkan skor 4 pada indikator nomor 2; mendapatkan skor 4 pada indikator nomor 3; mendapatkan skor 3 pada indikator nomor 4; mendapatkan skor 5 pada indikator nomor 5; dan mendapatkan skor 1 pada indikator nomor 6 yang berarti peserta didik dalam membaca hasil pengukuran sangat tidak baik. Dengan demikian, total skor yang diperoleh oleh peserta didik tersebut adalah $5 + 4 + 4 + 3 + 5 + 1 = 22$. Skor terendah yang mungkin akan diperoleh peserta didik adalah 6 dan skor tertinggi yang mungkin diperoleh oleh peserta didik adalah 30. Median dari skor tersebut adalah $(6 + 30)/2 = 18$. Apabila skor tersebut dibagi menjadi 4 kategori, maka skor:

- 6 – 12 : peserta didik dinyatakan gagal,
- 13 – 18 : peserta didik dinyatakan kurang berhasil,
- 19 – 24 : peserta didik dinyatakan berhasil,
- 25 – 30 : peserta didik dinyatakan sangat berhasil.

Dengan demikian, berdasarkan capaian skor oleh peserta didik sebesar 22 dapat disimpulkan peserta didik tersebut memiliki keterampilan yang baik dalam menggunakan termometer sebagai alat ukur suhu badan seseorang.

Bab 8

Konversi Skor

Setelah pengukuran hasil belajar peserta didik dengan menggunakan alat ukur berupa tes selesai dilaksanakan, guru akan mendapatkan data hasil pengukuran berupa skor. Skor merupakan data mentah berupa angka-angka yang mencerminkan capaian kompetensi yang diperoleh setiap peserta didik setelah mengikuti program pembelajaran. Pada dasarnya skor yang diperoleh setiap peserta didik tersebut masih berupa skor mentah yang belum bermakna dalam menggambarkan kompetensi peserta didik. Oleh karena itu, skor mentah selanjutnya perlu diolah menjadi skor standar agar skor tersebut lebih bermakna dalam menggambarkan kompetensi peserta didik. Kegiatan mengolah skor mentah menjadi skor standar ini biasa disebut dengan konversi skor. Hasil konversi skor ini kemudian akan dihasilkan nilai yang lebih bermakna dalam menggambarkan kompetensi peserta didik setelah mengikuti kegiatan belajar dan mengajar.

Pada umumnya, banyak orang menggunakan istilah skor dan nilai secara bergantian untuk merujuk hal yang ekuivalen. Sebenarnya dua istilah tersebut berbeda satu sama lain, tetapi keduanya berkaitan erat dan merujuk pada representasi hasil belajar peserta didik. Skor mengacu pada hasil proses penskoran yang dilakukan dengan menjumlahkan angka-angka dari setiap butir-butir tes yang dijawab benar oleh peserta didik. Angka-angka ini masih berupa skor mentah yang perlu dilakukan pengolahan lebih lanjut untuk menjadi sebuah nilai. Sedangkan nilai adalah angka ubahan dari skor yang diperoleh dengan cara mengolah atau menkonversi skor tersebut dengan menggunakan acuan tertentu, yaitu acuan norma dan acuan kriteria. Pada dasarnya, kegiatan menskor adalah proses pengubahan jawaban terhadap butir-butir tes yang diberikan oleh peserta didik menjadi angka-angka yang bersifat kuantitatif dan angka-angka hasil pengukuran tersebut selanjutnya

perlu diproses atau dikonversi menjadi nilai (*grade*). Ada paling tidak dua cara yang dapat digunakan untuk menkonversi skor mentah menjadi skor standar (baku), yaitu dengan menggunakan penilaian acuan patokan (*criterion-referenced assessment*) dan penilaian acuan norma (*norm-referenced assessment*). Kedua pendekatan yang digunakan untuk mengkonversi skor mentah menjadi skor standar tersebut memiliki karakteristik dan asumsi masing-masing yang harus dipenuhi sebelum proses konversi dilakukan.

A. Penilaian Acuan Patokan

Penilaian acuan patokan (PAP) merupakan pendekatan penilaian yang dilakukan dengan cara membandingkan hasil pengukuran kompetensi peserta didik dengan kriteria (patokan) yang telah ditentukan sebagai batas lulus atau batas tuntas terhadap capaian hasil belajar peserta didik tersebut. Pendekatan ini menghendaki guru telah menentukan terlebih dulu kriteria atau patokan yang akan digunakan sebagai pembanding untuk memutuskan apakah seorang peserta didik telah lulus / tidak lulus atau telah tuntas / tidak tuntas dalam menguasai kompetensi tertentu. Hal ini dapat dipahami bahwa jika peserta didik telah mampu mendapatkan skor yang sama atau lebih tinggi daripada kriteria yang telah ditentukan, maka peserta didik tersebut dinyatakan lulus atau telah tuntas dalam menguasai materi yang telah diajarkan oleh guru. Sebaliknya, jika peserta didik belum mampu mencapai skor sesuai dengan patokan yang telah ditentukan, maka peserta didik tersebut dinyatakan belum lulus atau belum tuntas dalam menguasai materi yang telah diajarkan oleh guru.

1. Kegunaan Penilaian Acuan Patokan

Penilaian acuan patokan (PAP) biasa juga disebut dengan pendekatan penilaian absolut atau standar mutu mutlak. Ini terjadi karena tingkat keberhasilan belajar peserta didik akan didasarkan pada kriteria atau patokan yang telah terlebih dulu ditentukan oleh guru. Apabila peserta didik tidak mampu mencapai skor sesuai kriteria atau patokan yang telah ditentukan, maka peserta didik tersebut dianggap belum berhasil dalam belajarnya, dan juga sebaliknya. Setiap peserta didik bertanggung-jawab sendiri-sendiri atas capaian hasil belajarnya tanpa dipengaruhi oleh capaian belajar kelompoknya (teman-teman sekelasnya). Seorang peserta didik yang mendapatkan skor, misalnya 90, skor tersebut diperoleh karena peserta didik

telah bekerja keras secara mandiri untuk menguasai pengetahuan atau keterampilan yang telah ditentukan oleh guru. Skor tersebut tidak dipengaruhi oleh skor rata-rata yang diperoleh oleh teman-teman sekelasnya. Oleh karena itu, PAP memiliki karakteristik sebagai berikut:

1. PAP dilakukan dengan membandingkan skor peserta didik sebagai hasil pengukuran dengan patokan yang telah ditetapkan oleh guru;
2. Pendekatan penilaian ini bersifat absolut dan objektif; dan
3. Digunakan untuk melakukan penilaian penguasaan peserta didik pada pokok bahasan atau kompetensi dasar tertentu.

PAP sangat cocok diterapkan untuk penilaian formatif, yaitu penilaian yang dilakukan pada setiap akhir suatu kompetensi dasar atau akhir suatu pokok bahasan tertentu yang dimaksudkan untuk mengetahui sejauhmana suatu proses pembelajaran telah berjalan sesuai yang direncanakan. Apabila banyak peserta didik telah mampu mencapai kompetensi tertentu yang direpresentasikan dengan capaian skor yang sama atau lebih tinggi dengan patokan yang telah ditentukan, guru dapat melanjutkan untuk menyampaikan materi ajar yang berbeda. Sebaliknya, apabila masih banyak peserta didik yang belum mampu mencapai kompetensi yang direpresentasikan dengan skor yang masih dibawah patokan yang telah ditentukan, maka guru harus melakukan berbagai upaya terlebih dahulu agar peserta didik mampu mencapai skor sesuai patokan. Upaya itu, misalnya dengan melakukan *remedial teaching*; memperbaiki metode dan teknik pembelajaran yang telah dilakukan, dan sebagainya. PAP tidak sepenuhnya cocok diterapkan untuk penilaian sumatif. Hal ini disebabkan PAP tidak mempertimbangkan kompetensi peserta didik dalam kelompoknya (rata-rata skor yang diperoleh teman-teman sekelasnya), sehingga penentuan skor akhir kadang-kadang menjadi “tidak manusiawi”. Apabila butir-butir tes yang harus dikerjakan oleh setiap peserta didik ada dalam kategori butir-butir tes yang sulit, maka kemungkinan besar banyak peserta didik yang mendapatkan skor yang rendah. Walaupun peserta didik tersebut sebenarnya adalah peserta didik dalam kategori pandai. Jika ini terjadi, skor akhir yang akan diisikan ke dalam laporan hasil belajar (rapor) juga akan rendah. Sebaliknya apabila butir-butir tes yang dikerjakan peserta didik dalam kategori mudah, akan memungkinkan semua peserta didik mendapatkan skor tinggi. Dengan demikian, skor-skor tersebut tidak dapat

menggambarkan tingkat kompetensi peserta didik secara faktual. Dengan demikian, PAP hanya akan dapat diterapkan secara efektif apabila butir-butir tes sudah dikategorikan standar atau sudah diuji-cobakan untuk mengetahui kualitas tes tersebut.

Ketika guru akan menerapkan PAP sebagai pendekatan dalam kegiatan pembelajaran dan penilaian, guru harus memperhatikan dua jenis kompetensi yang harus dikuasai oleh peserta didik. Pertama, kompetensi yang menjadi prasyarat untuk menguasai kompetensi yang lain. Kompetensi ini ada di hampir semua mata pelajaran, terutama mata pelajaran eksakta. Kedua, Kompetensi yang tidak menjadi prasyarat untuk menguasai kompetensi yang lain. Apabila peserta didik belum menguasai kompetensi jenis ini, maka tidak dikuasainya kompetensi tersebut tidak “berdampak” terhadap penguasaan materi yang lain. Pemahaman terhadap dua jenis kompetensi ini menjadi suatu hal yang sangat penting karena pemahaman tersebut dapat digunakan sebagai fondasi bagi guru untuk menentukan bahan-bahan ajar mana yang menjadi prasyarat untuk menguasai bahan ajar yang lain. Bahkan untuk menentukan bahan ajar sebagai prasyarat untuk menguasai bahan ajar pada mata pelajaran yang lain.

2. Konversi Skor dengan Pendekatan Penilaian Acuan Patokan

Untuk mengkonversi (mengolah atau mengubah) skor mentah menjadi skor standar dengan menggunakan pendekatan penilaian acuan patokan, guru terlebih dulu harus menentukan panduan konversi. Panduan konversi itu akan digunakan sebagai pedoman untuk memutuskan apakah seorang peserta didik telah mencapai batas tuntas (*passing grade*) yang telah ditentukan atau belum. Batas tuntas juga biasa disebut dengan kriteria ketuntasan minimal (KKM), yaitu kriteria batas nilai paling rendah yang harus dicapai oleh peserta didik agar dapat dikategorikan tuntas dalam menguasai kompetensi tertentu. Dalam menentukan batas tuntas suatu mata pelajaran, ada tiga hal yang harus dipertimbangkan, yaitu kompleksitas materi, sumber daya pendukung (sarana), dan *intake*. Semakin kompleks kompetensi yang harus dikuasai peserta didik, maka semakin rendah skor batas tuntasnya. Sebaliknya semakin mudah kompetensi yang harus dikuasai oleh peserta didik, maka semakin tinggi skor batas tuntasnya. Semakin tinggi daya pendukung proses pembelajaran terhadap mata pelajaran tertentu, semakin tinggi pula skor batas tuntas mata pelajaran tersebut dan

demikian pula sebaliknya. *Intake* mengacu pada kompetensi awal yang dimiliki oleh peserta didik. Semakin baik kompetensi awal yang dimiliki oleh peserta didik, semakin tinggi pula batas tuntasnya. Batas tuntas akan menjadi target pencapaian kompetensi sesuai dengan kompetensi dasar tertentu. Batas tuntas juga dapat digunakan sebagai kontrak pedagogis antara guru, peserta didik, dan orang tua sehingga proses pembelajaran selalu diarahkan agar mampu memfasilitasi peserta didik untuk mencapai kompetensi yang telah ditetapkan.

Proses konversi skor dengan menggunakan pendekatan penilaian acuan patokan ini dilakukan dengan mengikuti beberapa langkah, yaitu:

- 1) Menentukan skor mentah.
- 2) Menghitung skor menjadi nilai menggunakan rumus PAP dengan mengalikan skor ideal.
- 3) Membuat pedoman konversi hasil perhitungan.
- 4) Mengubah skor menjadi nilai.

Langkah-langkah tersebut memberikan panduan sistematis bagi guru dalam menkonversi skor mentah menjadi skor standar yang lebih bermakna bagi peserta didik. Selanjutnya, pendekatan penilaian acuan patokan dapat digunakan untuk menkonversi skor pada skala-5 ataupun skala-10. Pada bagian berikut disajikan cara untuk mengkonversi skor dalam dua skala tersebut.

a. Konversi skor skala-5

Proses mengkonversi skor skala-5 dimulai dari menentukan panduan konversi, menentukan tabel konversi, dan mengubah skor menjadi nilai.

Contoh:

Diketahui skor mentah mata pelajaran Bahasa Inggris dari 36 peserta didik kelas IX SMP adalah sebagai berikut:

35	46	45	45	37	44
48	43	39	44	30	49
50	42	46	46	45	47
45	45	41	42	39	43
45	37	39	41	45	45
37	44	48	36	37	47

Tentukan skor standar dari skor-skor mentah tersebut jika skor maksimumnya ditetapkan 50.

1) Menentukan panduan konversi

Konversi skor skala-5 biasanya menggunakan panduan konversi sebagai berikut:

Tingkat Penguasaan	Skor Standar
90% - 100%	A
80% - 89%	B
70% - 79%	C
60% - 69%	D
$\leq 59\%$	E

2) Menentukan tabel konversi

Untuk membuat tabel konversi skala-5, guru harus menentukan skor maksimum (skor teoretis) terlebih dahulu. Skor maksimum adalah skor tertinggi apabila semua butir tes dijawab dengan benar oleh peserta didik. Menurut data skor di atas, skor maksimumnya adalah 50. Selanjutnya, tabel konversi dapat dibuat mengikuti panduan konversi skala-5 seperti di atas. Jika skor maksimum adalah 50, maka penguasaan 90% dihitung $0,9 \times 50 = 45$; penguasaan 80% dihitung $0,8 \times 50 = 40$; dan penguasaan 70% dihitung $0,7 \times 50 = 35$; penguasaan 60% dihitung $0,6 \times 50 = 30$. Dengan demikian, tabel konversi skala-5 dapat ditentukan sebagai berikut:

Skor mentah	Skor standar
45 – 50	A
40 – 44	B
35 – 39	C
30 – 34	D
≤ 29	E

3) Mengubah skor menjadi nilai

Berdasarkan tabel konversi tersebut, selanjutnya skor standar (nilai) dapat ditentukan. Seorang peserta didik dengan skor mentah 35, maka peserta didik tersebut mendapat nilai C karena skor 35 berada pada interval

35 – 39; seorang peserta didik dengan skor mentah 46, maka peserta didik tersebut mendapatkan nilai A karena skor mentah 46 berada pada interval 45 – 50; seorang peserta didik dengan skor mentah 30, maka peserta didik tersebut mendapatkan nilai D karena skor tersebut berada interval 30 – 34; dan seterusnya. Apabila batas lulus telah ditentukan bagi mereka yang mendapat nilai minimal C, maka peserta didik dengan skor mentah kurang dari 35 dinyatakan belum lulus atau belum tuntas dalam menguasai kompetensi yang telah ditentukan.

b. Konversi skor skala-10

Proses konversi skor mentah menjadi skor standar skala-10 pada dasarnya memiliki tahapan-tahapan yang sama dengan proses konversi skala-5, yaitu dimulai dari menentukan panduan konversi, menentukan tabel konversi, dan mengubah skor menjadi nilai. Dengan menggunakan data skor di atas konversi skor ke skala-10 dapat dilakukan sebagai berikut.

1) Menentukan panduan konversi

Konversi skor skala-10 biasanya menggunakan panduan konversi sebagai berikut:

Tingkat Penguasaan	Skor Standar
95% - 100%	10
85% - 94%	9
75% - 84%	8
65% - 74%	7
55% - 64%	6
45% - 54 %	5
35% - 44%	4
25% - 34%	3
15% - 24%	2
05% - 14%	1

2) Menentukan tabel konversi

Tabel konversi sekala-10 dengan skor teoretis 50 dapat dibuat sebagai berikut. Penguasaan 95% dihitung $0,95 \times 50 = 47,5$ dan dibulatkan menjadi 48; penguasaan 85% dihitung $0,85 \times 50 = 42,5$ dan dibulatkan

menjadi 43; penguasaan 75% dihitung $0,75 \times 50 = 37,5$ dan dibulatkan menjadi 38; penguasaan 65% dihitung $0,65 \times 50 = 32,5$ dan dibulatkan menjadi 33; penguasaan 55% dihitung $0,55 \times 50 = 27,5$ dan dibulatkan menjadi 28; penguasaan 45% dihitung $0,45 \times 50 = 22,5$ dan dibulatkan menjadi 23; penguasaan 35% dihitung $0,35 \times 50 = 17,5$ dan dibulatkan menjadi 18; penguasaan 25% dihitung $0,25 \times 50 = 12,5$ dan dibulatkan menjadi 13; penguasaan 15% dihitung $0,15 \times 50 = 7,5$ dan dibulatkan menjadi 8; dan penguasaan 05% dihitung $0,05 \times 50 = 2,5$ dan dibulatkan menjadi 3. Dengan demikian, tabel konversi skala-10 dapat ditentukan sebagai berikut:

Skor Mentah	Skor Standar
48 – 50	10
43 – 47	9
38 – 42	8
33 – 37	7
28 – 32	6
23 – 27	5
18 – 22	4
13 – 17	3
8 – 12	2
3 – 7	1

3) Mengubah skor menjadi nilai.

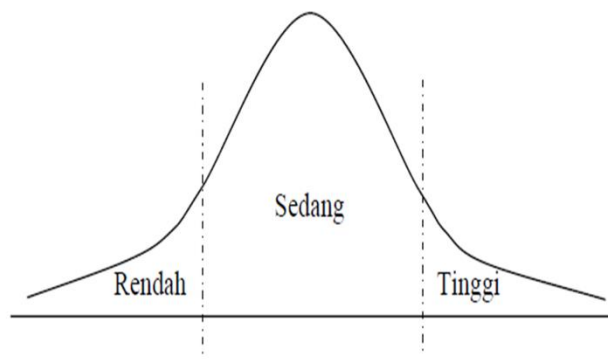
Berdasarkan tabel konversi tersebut, selanjutnya skor standar (nilai) skala-10 dapat ditentukan. Seorang peserta didik dengan skor mentah 35, maka peserta didik tersebut mendapat nilai 7 karena skor 35 berada pada interval 33 – 37; seorang peserta didik dengan skor mentah 46, maka peserta didik tersebut mendapatkan nilai 9 karena skor mentah 46 berada pada interval 43 – 47; seorang peserta didik dengan skor mentah 30, maka peserta didik tersebut mendapatkan nilai 6 karena skor tersebut berada interval 28 – 32; dan seterusnya. Apabila batas lulus telah ditentukan bagi mereka yang mendapat nilai minimal 7, maka peserta didik dengan skor mentah kurang dari 33 dinyatakan belum lulus atau belum tuntas dalam menguasai kompetensi yang telah ditentukan.

B. Penilaian Acuan Norma

Berbeda dengan penilaian acuan patokan yang merupakan model pendekatan penilaian yang mengacu pada suatu kriteria atau patokan yang bersifat absolut, penilaian acuan norma (PAN) merupakan suatu pendekatan penilaian yang mengacu pada norma kelompok. Hal ini berarti bahwa PAN sangat mempertimbangkan rata-rata skor dan standar deviasi yang diperoleh kelompok peserta didik. PAN merupakan sistem penilaian yang didasarkan pada skor yang diperoleh sekelompok peserta didik setelah mengikuti suatu proses pembelajaran. Dua peserta didik yang mendapatkan skor yang sama, tetapi mereka berada dalam kelompok yang berbeda, maka skor standar yang diperoleh dua peserta didik tersebut kemungkinan akan berbeda. Ini terjadi karena skor akhir (skor standar) yang diperoleh seorang peserta didik sangat dipengaruhi oleh rata-rata skor dan standar deviasi skor yang dicapai oleh kelompok peserta didik tersebut. Dalam pendekatan PAN, norma dapat dipahami sebagai prestasi belajar kelompok. Sedangkan kelompok disini dapat dipahami sebagai sejumlah peserta didik yang mengikuti tes. Kelompok peserta didik dapat mengacu pada sejumlah peserta didik dalam suatu kelas, sekolah, kabupaten, provinsi, dan sebagainya yang mengikuti tes dalam jenis, jenjang dan tingkatan yang sama.

1. Kegunaan Penilaian Acuan Norma

PAN diterapkan mengikuti prinsip-prinsip distribusi normal. Artinya PAN dapat diterapkan apabila skor-skor yang didapatkan oleh sekelompok peserta didik tersebut berdistribusi normal atau distribusinya membentuk kurva normal. Seperti tampak pada gambar 8.1 bahwa untuk dapat diterapkan



Gambar 8.1 Skor berdistribusi normal

PAN, sebaran skor dalam kategori rendah, sedang, dan tinggi yang diperoleh oleh kelompok peserta didik harus berdistribusi normal karena pendekatan ini mengikuti asumsi normalitas data. Untuk mendapatkan gambaran normalitas data tersebut tentu saja menghendaki data skor yang dikumpulkan dari sekelompok peserta didik yang jumlahnya banyak. Oleh karena itu, pendekatan PAN dapat diterapkan jika memenuhi dua syarat utama, yaitu sebaran skor-skor yang diperoleh oleh peserta didik terpengaruh seperti kurva normal; dan jumlah peserta didik yang mengikuti tes harus banyak atau besar (50 – 100 orang peserta didik).

Jika dalam pendekatan PAP menganut kriteria atau patokan yang bersifat absolut, PAN menganut metode penghitungan skor akhir yang bersifat relative, yang berarti bahwa skor akhir atau skor standar yang diperoleh oleh peserta didik sangat dipengaruhi oleh naik turunnya rata-rata skor dan standar deviasi yang dihasilkan oleh kelompok peserta didik tersebut pada saat mengikuti suatu tes. Misalnya, apabila seorang guru telah menentukan pedoman konversi berdasarkan hasil tes yang diikuti dan dilakukan oleh kelompok peserta didik tertentu, maka pedoman tersebut hanya dapat digunakan untuk kelompok tersebut dan kemungkinan tidak dapat digunakan lagi sebagai pedoman konversi untuk kelompok yang berbeda, karena distribusi skor yang diperoleh oleh peserta didik dalam kelompok tersebut mungkin saja berbeda dengan kelompok lain. Hal inilah yang mengakibatkan bahwa pendekatan PAN bersifat relatif, bukan bersifat absolut seperti pada PAP. Oleh karena itu, pendekatan PAP dan PAN dapat dibedakan paling tidak berdasarkan tiga kriteria yang berbeda, yaitu perbedaan ditinjau berdasarkan proses pengembangan butir-butir tes, standar penilaian kinerja peserta didik, dan tujuan tes.

2. Konversi Skor dengan Pendekatan Penilaian Acuan Norma

Untuk mengkonversi skor mentah dengan menggunakan pendekatan penilaian acuan norma, ada beberapa langkah yang harus dilakukan. Ketiga langkah itu adalah menghitung rata-rata skor, menghitung standar deviasi skor, dan menyusun panduan konversi untuk mengubah skor mentah menjadi skor standar. Hal yang harus digarisbawahi adalah untuk menghitung rata-rata skor dan standar deviasi untuk data tunggal dan data kelompok dilakukan dengan rumus yang berbeda. Data tunggal adalah data yang disusun sendiri menurut nilai dan besarnya masing - masing. Disebut

data tunggal karena jumlah data (skor) ditaksir tidak terlalu banyak sehingga tidak perlu menggunakan tabel distribusi frekuensi. Sedangkan data kelompok adalah data yang jumlahnya banyak (biasanya lebih dari 30 data) sehingga memerlukan penyajian data dalam tabel distribusi frekuensi di mana data - data tersebut akan dikelompokkan dalam beberapa kelas, dan setiap kelas mempunyai interval nilai tertentu.

Adapun langkah-langkah untuk menkonversi skor dengan pendekatan acuan norma adalah sebagai berikut.

1) Menghitung rata-rata skor

Rata-rata skor biasanya diberikan simbol $\bar{X}$. Untuk menghitung rata-rata skor data tunggal digunakan rumus sebagai berikut:

$$\bar{X} = \frac{\sum x_i}{n}$$

Dimana:

$\bar{X}$ = rata-rata skor

$\sum x_i$ = skor peserta didik ke- i

n = jumlah keseluruhan peserta didik

Sedangkan cara menghitung rata-rata skor untuk data kelompok digunakan rumus sebagai berikut:

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i}$$

Dimana:

$\bar{X}$ = rata-rata skor

x_i = nilai tengah kelas ke- i

f_i = frekuensi kelas ke- i

2) Menghitung standar deviasi skor

Standar deviasi adalah nilai statistik yang digunakan untuk menentukan seberapa dekat elemen data-data (skor) yang ada dengan nilai rata-rata (mean) dari suatu sampel (peserta didik) dan bagaimana data-data tersebar di dalam sampel tersebut. Standar deviasi adalah ukuran

penyebaran yang paling banyak digunakan. Semua gugus data (skor) dipertimbangkan sehingga lebih stabil dibandingkan dengan ukuran lainnya. Biasanya standar deviasi dimanfaatkan oleh para ahli statistik atau orang yang berkecimpung dalam dunia tersebut untuk mengetahui apakah sampel data yang diambil mewakili seluruh populasi. Rumus untuk mencari standar deviasi dibedakan pula untuk data tunggal dan data kelompok. Rumus untuk mencari standar deviasi untuk data tunggal adalah sebagai berikut:

$$S = \sqrt{\frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}}$$

Dimana:

S = standar deviasi
 x_i = skor x ke-i
 n = jumlah peserta didik

Sedangkan rumus untuk mencari standar deviasi untuk data berkelompok adalah sebagai berikut:

$$s = i \sqrt{\frac{\sum f_i f_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$$

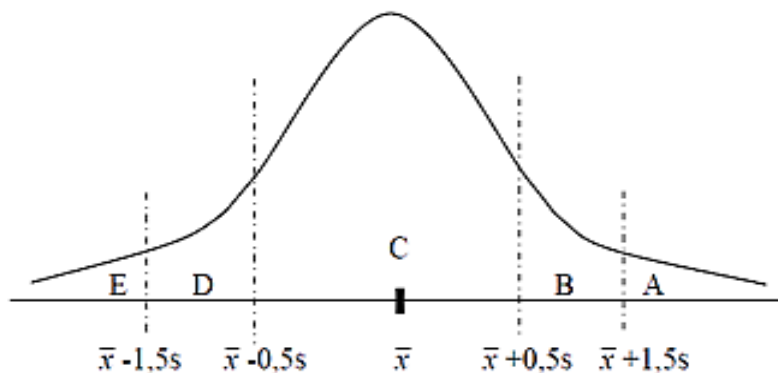
Dimana:

S = standar deviasi
 i = panjang kelas
 f_i = frekuensi skor ke-i
 x_i = skor ke-i
 n = jumlah peserta didik

3) Menyusun panduan konversi untuk mengubah skor mentah menjadi skor standar.

Panduan konversi dapat dibedakan untuk skor standar lima (*stanfive*), standar sembilan (*stanine*), standar sebelas (*stanel*), nilai standar z (*z-score*), dan standar t (*t score*).

a) Konversi standar lima (*stanfive*) secara distributif dapat digambarkan seperti kurva berikut ini:



Catatan:

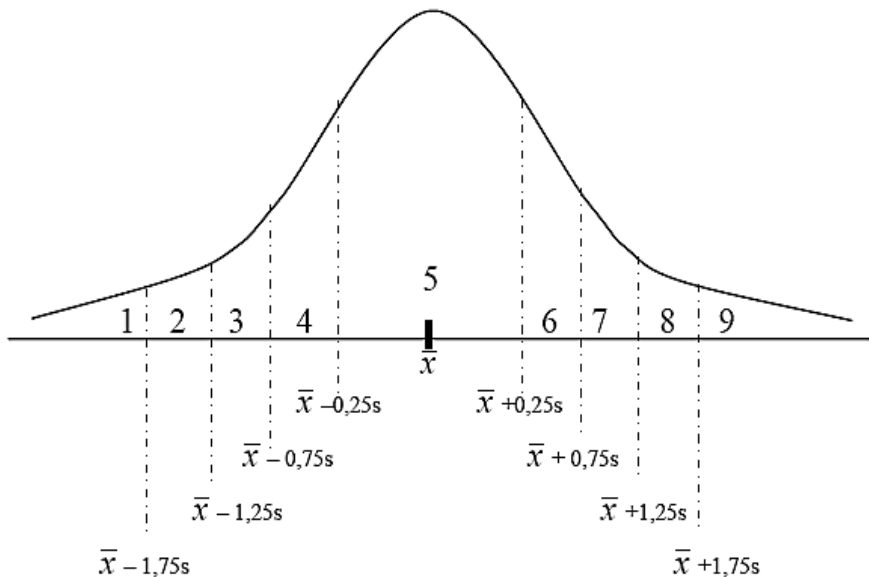
s = standar deviasi

Gambar 8.2 Konversi skor skala-5

Sedangkan konversi skor mentah hasil tes menjadi nilai standar lima (*stanfive*) dapat digunakan panduan sebagai berikut:

$\bar{X} + 1,5s$	A
$\bar{X} + 0,5s$	B
$\bar{X}$	C
$\bar{X} - 0,5s$	D
$\bar{X} - 1,5s$	E

- b) Konversi standar sembilan (*stanine*) secara distributif dapat digambarkan seperti kurva berikut ini:

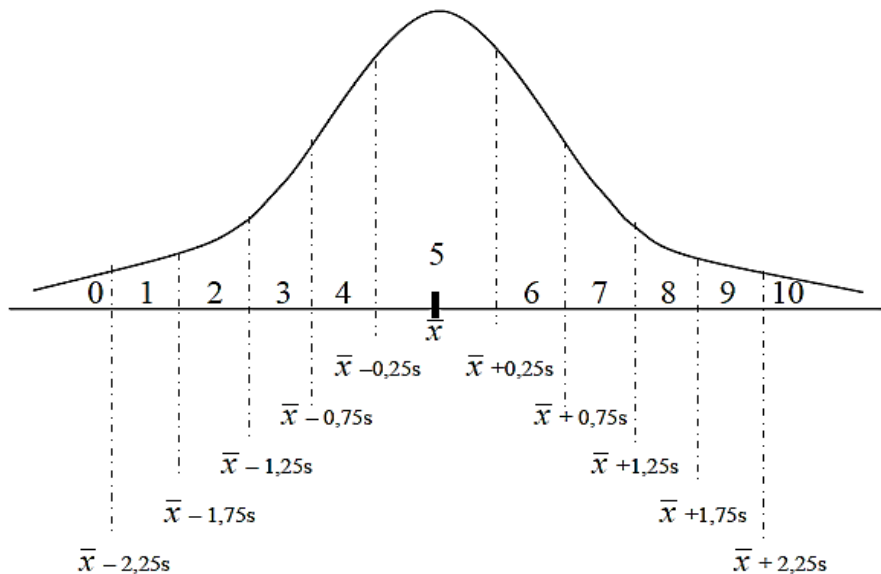


Gambar 8.3 Konversi skor skala-9

Sedangkan konversi skor mentah hasil tes menjadi nilai standar sembilan (*stanine*) dapat digunakan panduan sebagai berikut:

$\bar{X} + 1,7s$	→ 9
$\bar{X} + 1,25s$	→ 8
$\bar{X} + 0,75s$	→ 7
$\bar{X} + 0,25s$	→ 6
$\bar{X}$	→ 5
$\bar{X} - 0,25s$	→ 4
$\bar{X} - 0,75s$	→ 3
$\bar{X} - 1,25s$	→ 2
$\bar{X} - 1,75s$	→ 1

- c) Konversi standar sebelas (*stanel*) secara distributif dapat digambarkan seperti kurva berikut ini:



Gambar 8.4 Konversi skor skala-11

Sedangkan konversi skor mentah hasil tes menjadi nilai standar sebelas (*stanel*) dapat digunakan panduan sebagai berikut:

$\bar{X} + 2,25s$	→ 10
$\bar{X} + 1,75s$	→ 9
$\bar{X} + 1,25s$	→ 8
$\bar{X} + 0,75s$	→ 7
$\bar{X} + 0,25s$	→ 6
$\bar{X} - 0,25s$	→ 5
$\bar{X} - 0,75s$	→ 4
$\bar{X} - 1,25s$	→ 3
$\bar{X} - 1,75s$	→ 2
$\bar{X} - 2,25s$	→ 1
	→ 0

Contoh konversi skor dengan pendekatan PAN untuk data tunggal:

Seorang guru Bahasa Inggris mengajar sepuluh peserta didik di kelas VII SMP. Setelah menyelesaikan satu kompetensi dasar, guru tersebut memberikan soal ulangan bentuk pilihan ganda dengan jumlah 50 butir untuk dikerjakan oleh sepuluh orang peserta didik tersebut. Perolehan skor masing-masing peserta didik adalah sebagai berikut.

No.	Nama Peserta Didik	Skor
1.	Dona Agnovani	43
2.	Raka Dewi	36
3.	Rita Octaviana	32
4.	Rony Danovan	40
5.	Tiwuk Nurkamti	41
6.	Siti Sukaesih	50
7.	Nurul Khasanah	34
8.	Roberto Manchini	35
9.	Rumelu Adaumi	35
10.	Junaedi Rahman	41

1) Pedoman konversi skala-5

Langkah pertama, menghitung rata-rata skor dan standar deviasi:

No.	Nama Peserta Didik	Skor (X_i)	X_i^2
1.	Dona Agnovani	43	1849
2.	Raka Dewi	36	1296
3.	Rita Octaviana	32	1024
4.	Rony Danovan	40	1600
5.	Tiwuk Nurkamti	41	1681
6.	Siti Sukaesih	50	2500
7.	Nurul Khasanah	34	1156
8.	Roberto Manchini	35	1225
9.	Rumelu Adaumi	35	1225
10.	Junaedi Rahman	41	1681
Jumlah		387	15237

$$\bar{X} = \frac{\sum x_i}{n} = \frac{387}{10} = 38,7$$

Rata-rata skor ditemukan sebesar 38,7 dan rata-rata tersebut dibulatkan menjadi 39.

$$s = \sqrt{\frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}} = \sqrt{\frac{10 \sum 15237 - 387^2}{10(10-1)}} = 5,38$$

Standar deviasi dari skor-skor tersebut ditemukan sebesar 5,38.

Langkah kedua adalah membuat panduan konversi untuk mengubah skor mentah menjadi skor standar

$\bar{X} + 1,5s$	$39 + (1,5 \times 5,38) = 37,07$ dibulatkan 37	→ A
$\bar{X} + 0,5s$	$39 + (0,5 \times 5,38) = 31,69$ dibulatkan 32	→ B
$\bar{X} - 0,5s$	$39 - (0,5 \times 5,38) = 26,31$ dibulatkan 26	→ C
$\bar{X} - 1,5s$	$39 - (1,5 \times 5,38) = 20,93$ dibulatkan 21	→ D
		→ E

Selanjutnya panduan konversi dapat dibuat sebagai berikut:

Interval Skor	Nilai
37 keatas	A
32 – 36	B
26 – 31	C
21 – 25	D
20 ke bawah	E

Dengan demikian, peserta didik dengan skor mentah 43, maka peserta didik tersebut mendapat nilai A karena skor 43 berada pada interval 37 – keatas; peserta didik dengan skor mentah 36, maka peserta didik tersebut mendapat nilai B karena skor 36 berada pada interval 32 – 36; peserta didik dengan skor mentah 32, maka peserta didik tersebut mendapat nilai B karena skor 32 berada pada interval 32 – 36, dan seterusnya.

2) Pedoman konversi skala-9

Pada dasarnya, langkah-langkah untuk mengkonversi skor skala-9 sama seperti langkah-langkah pada konversi skor skala-5. Yang membedakan adalah panduan konversi yang akan digunakan sebagai dasar mengkonversi skor mentah menjadi skor standar. Untuk mengkonversi data (skor) diatas, karena rata-rata skor dan standar deviasi telah diketahui, maka langkah tersebut dapat tinggalkan dan langsung melangkah untuk membuat panduan konversi dengan pedekatan PAN untuk skala-9.

$\bar{X} + 1,75s$	$39 + (1,75 \times 5,38) = 48,42$ dibulatkan 48	9
$\bar{X} + 1,25s$	$39 + (1,25 \times 5,38) = 45,73$ dibulatkan 46	8
$\bar{X} + 0,75s$	$39 + (0,75 \times 5,38) = 43,04$ dibulatkan 43	7
$\bar{X} + 0,25s$	$39 + (0,25 \times 5,38) = 40,35$ dibulatkan 40	6
$\bar{X} - 0,25s$	$39 - (0,25 \times 5,38) = 37,66$ dibulatkan 38	5
$\bar{X} - 0,75s$	$39 - (0,75 \times 5,38) = 34,97$ dibulatkan 35	4
$\bar{X} - 1,25s$	$39 - (1,25 \times 5,38) = 32,28$ dibulatkan 32	3
$\bar{X} - 1,75s$	$39 - (1,75 \times 5,38) = 29,59$ dibulatkan 30	2
		1

Selanjutnya panduan konversi dapat disusun sebagai berikut:

Interval Skor	Nilai
48 ke atas	9
46 – 47	8
43 – 45	7
40 – 42	6
38 – 39	5
35 – 37	4
32 – 34	3
30 – 31	2
29 ke bawah	1

Dengan demikian, peserta didik dengan skor mentah 43, maka peserta didik tersebut mendapat nilai 7 karena skor 43 berada pada interval 43 – 45; peserta didik dengan skor mentah 36, maka peserta didik tersebut mendapat nilai 4 karena skor 36 berada pada interval 35 – 37; peserta didik dengan skor mentah 32, maka peserta didik tersebut mendapat nilai 3 karena skor 32 berada pada interval 32 – 34, dan seterusnya.

3) Pedoman konversi skala-11

Rata-rata skor dan standar deviasi dari skor-skor tersebut telah diketahui, maka panduan konversi dengan pedekatan PAN untuk skala-11 adalah sebagai berikut.

$\bar{X} + 2,25s$	$39 + (2,25 \times 5,38) = 51,12$ dibulatkan 51	10
$\bar{X} + 1,75s$	$39 + (1,75 \times 5,38) = 48,42$ dibulatkan 48	9
$\bar{X} + 1,25s$	$39 + (1,25 \times 5,38) = 45,73$ dibulatkan 46	8
$\bar{X} + 0,75s$	$39 + (0,75 \times 5,38) = 43,04$ dibulatkan 43	7
$\bar{X} + 0,25s$	$39 + (0,25 \times 5,38) = 40,35$ dibulatkan 40	6
		5

$\bar{X} - 0,25s$	$39 - (0,25 \times 5,38) = 37,66$ dibulatkan 38	→ 4
$\bar{X} - 0,75s$	$39 - (0,75 \times 5,38) = 34,97$ dibulatkan 35	→ 3
$\bar{X} - 1,25s$	$39 - (1,25 \times 5,38) = 32,28$ dibulatkan 32	→ 2
$\bar{X} - 1,75s$	$39 - (1,75 \times 5,38) = 29,59$ dibulatkan 30	→ 1
$\bar{X} - 2,25s$	$39 - (2,25 \times 5,38) = 26,87$ dibulatkan 27	→ 0

Salanjutnya berdasarkan hitungan di atas, panduan konversi dapat disusun sebagai berikut:

Interval Skor	Nilai
51 ke atas	10
48 – 50	9
46 – 47	8
43 – 45	7
40 – 42	6
38 – 39	5
35 – 37	4
32 – 34	3
30 – 31	2
27 – 29	1
26 ke bawah	0

Dengan demikian, peserta didik dengan skor mentah 43, maka peserta didik tersebut mendapat nilai 7 karena skor 43 berada pada interval 43 – 45; peserta didik dengan skor mentah 36, maka peserta didik tersebut mendapat nilai 4 karena skor 36 berada pada interval 35 – 37; peserta didik dengan skor mentah 32, maka peserta didik tersebut mendapat nilai 3 karena skor 32 berada pada interval 32 – 34, dan seterusnya.

Konversi skor seperti contoh di atas merupakan konversi skor untuk data (skor) tunggal yang jumlahnya tidak terlalu banyak. Untuk mengkonversi skor yang jumlahnya banyak, maka data

tersebut perlu diubah menjadi data kelompok sehingga data akan lebih mudah disajikan.

Contoh konversi untuk data kelompok:

Seorang guru Bahasa Inggris mengajar 50 peserta didik di kelas VII SMP. Setelah menyelesaikan satu kompetensi dasar, guru tersebut memberikan soal ulangan bentuk pilihan ganda dengan jumlah 50 butir untuk dikerjakan oleh 50 orang peserta didik tersebut, dan perolehan skor masing-masing peserta didik sebagai berikut.

43	45	35	35	35
34	35	34	37	37
35	37	43	45	45
34	35	43	37	45
37	37	37	45	35
43	43	34	36	47
43	45	45	36	45
37	47	45	35	35
35	45	35	37	35
35	44	35	37	35

Konversikan data berupa skor tersebut ke dalam skala-5 dan skala-9 dengan menggunakan pendekatan penilaian acuan norma.

Untuk mengkonversi skor (data kelompok) tersebut, ada langkah-langkah yang harus dilakukan, yaitu:

1) Menghitung rata-rata skor ($\bar{X}$) dan standar deviasi (s)

Untuk menghitung rata-rata skor dan standar deviasi dari skor-skor di atas, perlu dibuat tabel frekuensi terlebih dulu. Agar dapat dibuat tabel frekuensi, ada beberapa tahapan yang harus dilakukan, yaitu:

(a) Menentukan rentang skor

Untuk menentukan rentang skor (r), dilakukan dengan mengurangkan skor tertinggi dengan skor terendah atau dengan rumus sebagai berikut:

Rentang skor (r) = skor tertinggi – skor terendah

Dari data di atas, skor tertinggi adalah 47 dan skor terendah adalah 34. Maka rentang skornya adalah:

$$r = 47 - 34 = 13$$

(b) Menentukan banyak kelas interval

Setelah rentang kelas diketahui, selanjutnya dicari banyaknya kelas interval dengan rumus:

$$k = 1 + 3,3 \log n$$

dimana:

k = banyak kelas interval

1 dan 3,3 = konstanta

n = banyaknya data

dengan demikian banyaknya kelas interval data diatas dapat dihitung seperti berikut:

$$\begin{aligned} k &= 1 + 3,3 \cdot \log 50 \\ &= 1 + (3,3 \times 1,6990) \\ &= 6,6067 \text{ (dibulatkan menjadi 7)} \end{aligned}$$

Jadi banyaknya interval kelas adalah 7

(c) Menentukan panjang kelas

Panjang kelas atau luas kelas adalah jarak antara tepi atas kelas dan tepi bawah kelas. Untuk mencari panjang kelas dapat digunakan rumus sebagai berikut:

$$\text{Panjang kelas (i)} = \frac{r}{k}$$

Dimana:

r = rentang skor

k = banyaknya kelas interval

Dengan demikian, panjang kelas dari data tersebut diatas dapat dihitung sebagai berikut:

$$\text{Panjang kelas} = \frac{r}{k} = \frac{13}{7} = 1,85 \text{ (dibulatkan menjadi 2)}$$

Catatan: pembulatan panjang kelas boleh tidak mengikuti kaidah matematik. Artinya pembulatan boleh ke atas atau ke bawah dengan pertimbangan agar semua skor dapat masuk.

(d) Membuat tabel distribusi frekuensi kelompok

Kelas Interval	Tally	Frekuensi
34 – 35	III III III III	19
36 – 37	III III II	12
38 – 39		0
40 – 41		0
42 – 43	III I	6
44 – 45	III III I	11
46 – 47	II	2

(e) Menentukan $(\bar{X})$ dan (s)

Kelas interval	f_i	x_i	$f_i x_i$	x'_i	$f_i x'_i$	$f_i x'_i (x'_i)$
34 – 35	19	34,5	655,5	+3	57	171
36 – 37	12	36,5	438	+2	24	48
38 – 39	0	38,5	0	+1	0	0
40 – 41	0	40,5	0	0	0	0
42 – 43	6	42,5	255	-1	-6	6
44 – 45	11	44,5	489,5	-2	-22	44
46 – 47	2	46,5	93	-3	-6	18
Jumlah	50		1931	0	47	287

Berdasarkan tabel di atas, selanjutnya dapat ditentukan ($\bar{X}$) dan (s) sebagai berikut:

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1931}{50} = 38,62 \text{ (dibulatkan menjadi 39)}$$

Selanjutnya standar deviasi (s) adalah sebagai berikut:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2} = \sqrt{\frac{287}{50} - \left(\frac{47}{50}\right)^2} = 4,41$$

2) Membuat dan mengkonversi skor skala-5

$\bar{X} + 1,5s$	$39 + (1,5 \times 4,41) = 45,62$ dibulatkan 46	A
$\bar{X} + 0,5s$	$39 + (0,5 \times 4,41) = 41,21$ dibulatkan 41	B
$\bar{X} - 0,5s$	$39 - (0,5 \times 4,41) = 36,79$ dibulatkan 37	C
$\bar{X} - 1,5s$	$39 - (1,5 \times 4,41) = 32,39$ dibulatkan 32	D
		E

Selanjutnya panduan konversi dapat dibuat sebagai berikut:

Interval Skor	Nilai
46 keatas	A
41 – 45	B
37 – 40	C
32 – 36	D
31 ke bawah	E

Dengan demikian, peserta didik dengan skor mentah 43, maka peserta didik tersebut mendapat nilai B karena skor 43 berada pada interval 41 – 45; peserta didik dengan skor mentah 34, maka peserta didik tersebut mendapat nilai D karena skor 34

berada pada interval 32 – 36; peserta didik dengan skor mentah 35, maka peserta didik tersebut mendapat nilai D karena skor 35 berada pada interval 32 – 36, dan seterusnya.

3) Membuat dan mengkonversi skor skala-9

$\bar{X} + 1,75s$	$39 + (1,75 \times 4,41) = 46,72$ dibulatkan 47	9
$\bar{X} + 1,25s$	$39 + (1,25 \times 4,41) = 44,51$ dibulatkan 45	8
$\bar{X} + 0,75s$	$39 + (0,75 \times 4,41) = 42,31$ dibulatkan 42	7
$\bar{X} + 0,25s$	$39 + (0,25 \times 4,41) = 40,10$ dibulatkan 40	6
$\bar{X} - 0,25s$	$39 - (0,25 \times 4,41) = 37,89$ dibulatkan 38	5
$\bar{X} - 0,75s$	$39 - (0,75 \times 4,41) = 35,69$ dibulatkan 36	4
$\bar{X} - 1,25s$	$39 - (1,25 \times 4,41) = 33,49$ dibulatkan 33	3
$\bar{X} - 1,75s$	$39 - (1,75 \times 4,41) = 31,28$ dibulatkan 31	2
		1

Selanjutnya panduan konversi dapat disusun sebagai berikut:

Interval Skor	Nilai
47 ke atas	9
45 – 46	8
42 – 44	7
40 – 41	6
38 – 39	5
36 – 37	4
33 – 35	3
31 – 32	2
30 ke bawah	1

Dengan demikian, peserta didik dengan skor mentah 43, maka peserta didik tersebut mendapat nilai 7 karena skor 43 berada pada interval 42 – 44; peserta didik dengan skor mentah 34, maka peserta didik tersebut mendapat nilai 3 karena skor 34 berada pada interval 33 – 35; peserta didik dengan skor mentah 35, maka peserta didik tersebut mendapat nilai 3 karena skor 35 berada pada interval 33 – 35, dan seterusnya.

d) Skor standar T (*T-score*)

Nilai standar T pada dasarnya adalah pengolahan atau konversi skor mentah kedalam skor standar skala-100. Rumus yang biasa digunakan adalah:

$$T - \text{Skor} = 50 + \left(\frac{X - \bar{X}}{s} \right) 10$$

Dimana:

50 dan 10 = bilangan tetap (konstanta)

X = skor mentah yang diperoleh oleh peserta didik

$\bar{X}$ = rata-rata skor

s = standar deviasi

Contoh:

Dengan menggunakan data tunggal (skor yang diperoleh oleh 10 peserta didik pada halaman sebelumnya), konversikan skor-skor mentah tersebut ke dalam skor standar T.

Penyelesaian:

Berdasarkan penghitungan yang telah dilakukan sebelumnya, rata-rata skor tersebut adalah 38,7 dan standar deviasinya adalah 5,38.

Dengan demikian, seorang peserta didik dengan skor mentah 43, maka skor standar T adalah:

$$T - \text{Skor} = 50 + \left(\frac{X - \bar{X}}{s} \right) 10 = 50 + \left(\frac{43 - 38,7}{5,38} \right) 10 = 57,99 \quad \text{dan}$$

(dibulatkan menjadi 58). Dengan demikian seorang peserta didik

dengan skor mentah 43, skor standar (skor skala-100) yang diperoleh adalah 58.

Seorang peserta didik dengan skor mentah 36, maka skor standar T yang diperoleh adalah:

$$T - \text{Skor} = 50 + \left(\frac{x - \bar{x}}{s} \right) 10 = 50 + \left(\frac{36 - 38,7}{5,38} \right) 10 = 44,98 \quad \text{dan}$$

(dibulatkan menjadi 45). Dengan demikian, seorang peserta didik dengan skor mentah 36, skor standar T (skor skala -100) adalah 45.

Seorang peserta didik dengan skor mentah 32, maka skor standar T yang diperoleh adalah:

$$T - \text{Skor} = 50 + \left(\frac{x - \bar{x}}{s} \right) 10 = 50 + \left(\frac{32 - 38,7}{5,38} \right) 10 = 37,55 \quad \text{dan}$$

(dibulatkan menjadi 38). Dengan demikian, seorang peserta didik dengan skor mentah 32, skor standat T (skor skala -100) adalah 38; dan seterusnya.

e) Skor standar Z (*Z-score*)

Skor standar Z pada dasarnya digunakan untuk mengkonversi beberapa skor yang didapatkan dari berbagai hasil pengukuran yang berbeda. Sebagai contoh Tes Potensi Skolastik yang biasa digunakan untuk menyeleksi calon mahasiswa yang akan masuk ke perguruan tinggi tertentu. Tes ini bertujuan untuk mengukur kemampuan kognitif calon mahasiswa yang dianggap penting untuk menunjang keberhasilan calon mahasiswa tersebut di perguruan nanti. Tes ini mengukur berbagai kemampuan yang berbeda, misalnya kemampuan penalaran umum; kemampuan kuantitatif berupa pengetahuan dan penguasaan matematika dasar; pengetahuan dan pemahaman umum; dan kemampuan memahami bacaan dan menulis. Skor yang diperoleh calon mahasiswa dari berbagai kemampuan tersebut kemungkinan akan berbeda satu dengan lainnya. Untuk menentukan, misalnya 5 calon mahasiswa yang akan diterima di perguruan tinggi tersebut, maka diperlukan skor standar yang mencerminkan kedudukan relatif 5 calon mahasiswa tersebut. Untuk memudahkan penghitungan skor standar, berbagai kemampuan tersebut perlu diberikan simbol tertentu. Misalnya

kemampuan penalaran umum diberi simbol (X_1); kemampuan kuantitatif berupa pengetahuan dan penguasaan matematika dasar diberi simbol (X_2); pengetahuan dan pemahaman umum diberi simbol (X_3); dan kemampuan memahami bacaan dan menulis diberi simbol (X_4). Rumus yang biasa digunakan adalah:

$$Z = \frac{x}{SD_x}$$

Dimana:

Z = Z-score

x = deviasi skor x , yaitu selisih antara skor X dengan M_x

SD_x = standar deviasi dari skor-skor X

Selanjutnya, langkah-langkah untuk mengkonversi skor mentah menjadi skor standar z adalah sebagai berikut:

- 1) Menjumlahkan skor-skor kemampuan X_1, X_2, X_3 , dan X_4 .
- 2) Menghitung rata-rata hitung (mean) dari skor kemampuan X_1, X_2, X_3 , dan X_4 dengan menggunakan rumus:

$$M_1 = \frac{\sum X_1}{N}; M_2 = \frac{\sum X_2}{N}; M_3 = \frac{\sum X_3}{N}; \text{ dan seterusnya.}$$

- 3) Mencari deviasi X_1, X_2, X_3 , dan X_4 dengan rumus sebagai berikut:

$$x_1 = X_1 - M_{x1}; x_2 = X_2 - M_{x2}; \text{ dan seterusnya.}$$

- 4) Mengudaratkan $x_1; x_2; x_3$, dan x_4 dan selanjutnya menjumlahkan hasil-hasil pengkuadratan tersebut.
- 5) Mencari standar deviasi skor-skor dari keempat kemampuan tersebut dengan rumus:

$$SD_{x1} = \sqrt{\frac{\sum x_1^2}{N}}; \text{ dan seterusnya.}$$

- 6) Mencari Z-score dengan rumus:

$$Z_1 = \frac{X_1}{SD_{x1}}; Z_2 = \frac{X_2}{SD_{x2}}; \text{ dan seterusnya}$$

Selanjutnya hasil penghitunga tersebut dijumlahkan dari atas ke bawah sehingga dapat diperoleh $\sum Z_1$; $\sum Z_2$; $\sum Z_3$; dan $\sum Z_4$

- 7) Z-score yang dimiliki oleh setiap calon mahasiswa dijumlahkan dari kiri ke kanan sehingga akan terlihat calon mahasiswa mana yang mendapatkan total z-score positif dan z-score negatif.

Contoh penghitungan:

Ada 10 orang calon mahasiswa mendaftar untuk dapat diterima di sebuah perguruan tinggi. Skor yang diperoleh 10 calon mahasiswa untuk empat kemampuan tersebut ditampilkan dalam tabel berikut. Tentukan calon mahasiswa mana saja yang dinyatakan diterima di perguruan tinggi tersebut.

No.	Calon Mhs.	Skor Mentah (X)				Deviasi (x)			
		X ₁	X ₂	X ₃	X ₄	x_1	x_2	x_3	x_4
1.	Yuni	121	85	100	82	13,6	1	-11,5	14,5
2.	Rony	120	78	122	67	12,6	-6	10,5	-0,5
3.	Rossy	98	85	131	56	-9,4	1	19,5	-11,5
4.	Vina	100	76	102	67	-7,4	-8	-9,5	-0,5
5.	Dhilla	101	78	107	65	-6,4	-6	-4,5	-2,5
6.	Mirza	122	90	102	67	14,6	6	-9,5	-0,5
7.	Wirta	120	92	121	62	12,6	8	9,5	-5,5
8.	Cahyo	97	87	122	67	-10,4	3	10,5	-0,5
9.	Kunil	98	89	109	70	-9,4	5	-2,5	2,5
10.	Hadi	97	80	99	72	-10,4	-4	-12,5	4,5
Σ		1074	840	1115	675	0	0	0	0
Rata-rata (M)		107,4	84,0	111,5	67,5				

No	Calon Mhs.	Kuadrat deviasi (x^2)				Z-score				Σ
		x_1^2	x_2^2	x_3^2	x_4^2	Z_1	Z_2	Z_3	Z_4	
1.	Yuni	184.96	1	132.25	210.25	1,24	0,19	-1,06	2,27	2,64
2.	Rony	158.76	36	110.25	0.25	1,15	-1,12	0,96	-0,08	0,91
3.	Rossy	88.36	1	380.25	132.2	-0,86	0,19	1,79	-1,80	-0,68

					5					
4.	Vina	54.76	64	90.25	0.25	-0,67	-1,49	-0,87	-0,08	-3,11
5.	Dhilla	40.96	36	20.25	6.25	-0,58	-1,12	-0,41	-0,39	-2,5
6.	Mirza	213.16	36	90.25	0.25	1,33	1,12	-0,87	-0,08	1,5
7.	Wirta	158.76	64	90.25	30.25	1,15	1,49	0,87	-0,86	2,65
8.	Cahyo	108.16	9	110.25	0.25	-0,95	0,56	0,96	0,08	0,65
9.	Kunil	88.36	25	6.25	6.25	-0,86	0,93	-0,23	0,39	0,23
10.	Hadi	108.16	16	156.25	20.25	-0,95	-0,74	-1,15	0,71	-2,13
Σ		1204,4	288	1186,5	406,5	0	0	0	0	
SD		10.97	5,4	10,89	6,38					

Dari tabel di atas, nilai Z selanjutnya diurutkan dimulai dari yang bernilai positif tertinggi kemudian diikuti dengan rendah rendah di bawahnya, dan seterusnya. Jika dalam tes tersebut hanya ingin meluluskan lima calon mahasiswa saja, maka yang diambil adalah lima calon mahasiswa yang memiliki nilai positif tertinggi, yaitu Wirta (2,65); Yuni (2,64); Mirza (1,5); Rony (0,91); dan Cahyo (0,65).

Bab 9

Penilaian Autentik

Penilaian merupakan proses pengumpulan dan pengolahan informasi untuk menentukan pencapaian hasil belajar peserta didik. Penilaian hasil belajar adalah sebuah proses untuk mengumpulkan informasi tentang capaian hasil belajar peserta didik dalam berbagai kompetensi, yaitu sikap spiritual dan sikap sosial, pengetahuan, dan keterampilan yang dilakukan secara sistematis pada saat dan setelah proses pembelajaran. Melalui proses penilaian, guru akan mendapatkan gambaran sejauhmana peserta didik telah menguasai berbagai kompetensi sebagai hasil dari proses belajar mereka. Hasil dari sebuah proses penilaian harus mampu memberikan gambaran yang sebenarnya mengenai sikap, pengetahuan, dan keterampilan peserta didik. Pertanyaan yang muncul kemudian adalah bentuk penilaian seperti apa yang hasilnya mampu memberikan gambaran nyata atau sebenarnya dari kompetensi peserta didik? Apakah penilaian konvensional yang dilakukan oleh guru dalam bentuk tes pilihan ganda, benar-salah, menjodohkan, uraian dan sebagainya telah mampu memberikan gambaran kompetensi peserta didik yang sebenarnya terkait dengan pengetahuan, keterampilan dan sikap? Apakah penilaian konvensional tersebut telah mampu memberikan gambaran pengetahuan, keterampilan dan sikap secara simultan? Berbagai pertanyaan tersebut mestinya memberikan panduan bagi guru untuk mampu melakukan proses penilaian yang holistik agar hasil penilaian lebih bermakna bagi peserta didik.

Salah satu bentuk penilaian yang dapat dilakukan oleh guru untuk mendapatkan gambaran secara nyata terhadap kompetensi peserta didik adalah penilaian autentik. Hasil dari penilaian autentik dipercaya mampu memberikan gambaran yang lebih bermakna berkaitan dengan kompetensi peserta didik dibandingkan dengan penilaian konvensional dalam bentuk

pilihan ganda yang telah tersandar sekali pun. Ketika guru akan melakukan penilaian terhadap hasil belajar peserta didik dengan menerapkan konsep penilaian autentik, guru harus terlebih dahulu memahami konsep penilaian tersebut dan bagaimana konsep penilaian tersebut diterapkan di kelas. Penilaian autentik menuntut guru untuk menerapkan kriteria yang berkaitan dengan konstruksi pengetahuan, aktivitas mengamati dan mencoba, dan nilai prestasi, baik prestasi di dalam maupun luar sekolah.

A. Penilaian Non-autentik dan Penilaian Autentik

Untuk menilai hasil belajar peserta didik, ada dua jenis penilaian yang dapat digunakan oleh guru, yaitu penilaian non-autentik dan penilaian autentik. Penilaian non-autentik biasa juga disebut dengan penilaian konvensional atau penilaian tradisional. Pada penilaian jenis ini, guru sejak awal sudah menentukan kunci jawaban sebagai acuan untuk menentukan apakah jawaban yang diberikan oleh peserta didik dapat dikategorikan sebagai jawaban benar atau salah. Selain itu, butir-butir pertanyaan yang akan digunakan oleh guru sebagai alat ukur telah disiapkan dengan tahapan-tahapan yang sistematis sehingga butir-butir pertanyaan memenuhi standar tertentu. Bahkan sebelum digunakan sebagai alat ukur, butir-butir pertanyaan itu perlu diujicobakan terlebih dahulu untuk mengetahui kualitasnya. Bentuk pilihan ganda, menjodohkan, benar-salah dan isian singkat adalah contoh-contoh penilaian non-autentik yang sering dipakai oleh guru untuk mengukur hasil belajar peserta didik. Salah satu kelemahan dari penilaian bentuk ini adalah hasil penilaian sering tidak mencerminkan kemampuan peserta didik yang sebenarnya karena ada peluang peserta didik untuk menjawab butir-butir pertanyaan dengan sekedar menebak berbagai pilihan jawaban yang tersedia. Tetapi, penilaian konvensional tidak berarti harus dihindari dalam proses pembelajaran. Penilaian non-autentik masih relevan untuk digunakan karena penilaian ini lazim digunakan oleh guru, efektif, efisien dan penilaian ini memiliki legitimasi akademis.

Di lain pihak, penilaian autentik merupakan bentuk penilaian yang mengharuskan peserta didik untuk merekonstruksi dan menentukan jawaban berdasarkan kemampuan yang mereka miliki sendiri. Penilaian autentik biasa juga disebut dengan penilaian alternatif, penilaian kinerja, penilaian informal, dan penilaian portofolio. Dalam penilaian autentik, peserta didik tidak akan disediakan pilihan jawaban yang harus mereka pilih untuk

menjawab berbagai pertanyaan yang diajukan. Penilaian autentik menghendaki peserta didik menrekonstruksi sendiri jawaban atas pertanyaan yang diajukan melalui proses bernalar dan berpikir kritis. Lebih lanjut, Suyanto (2020) menjelaskan bahwa kemampuan berpikir ini digunakan untuk menentukan strategi menjawab, menentukan pengetahuan yang diperlukan untuk menjawab, memilih dan menentukan alternatif prosedur untuk menyelesaikan masalah; mengevaluasi kekuatan sekaligus kelemahan prosedur dan pendekatan serta memilih yang sesuai untuk digunakan menyelesaikan masalah; mengumpulkan informasi tambahan jika diperlukan; mengonstruksi jawaban dengan alur berpikir dan struktur yang jelas atau dalam suatu bentuk karya baru seni, budaya, dan gerakan kinestetik. Melalui suatu permasalahan autentik, peserta didik dapat mengembangkan kesungguhan, ketelitian, serta ketangguhan dalam menyelesaikan masalah.

Ada beberapa pengertian berkaitan dengan apa yang dimaksud dengan penilaian autentik yang dikemukakan oleh para ahli. Seperti, Mueller (2014) yang menyatakan bahwa penilaian autentik adalah salah bentuk penilaian dimana peserta didik diminta untuk melakukan tugas yang mencerminkan dunia nyata dengan cara menerapkan pengetahuan dan keterampilan yang dimilikinya. Pengertian ini mengisyaratkan bahwa penilaian autentik dilakukan tidak sekedar hanya mengukur kemampuan peserta didik untuk mengingat kembali pengetahuan dan keterampilan yang telah dipelajari, tetapi penilaian ini diarahkan bagaimana peserta didik mampu menerapkan pengetahuan dan keterampilan tersebut untuk mengatasi berbagai permasalahan di dunia nyata. Pengertian serupa disampaikan oleh Wiggins (1993) yang menyatakan bahwa ukuran autentik mengacu pada sejauhmana pertanyaan-pertanyaan yang digunakan untuk proses penilaian mampu melibatkan peserta didik untuk menggunakan pengetahuannya dalam rangka menunjukkan kinerja secara efektif dan kreatif. Dengan demikian, penilaian autentik dapat dipahami sebagai sebuah penilaian yang berfokus pada berbagai tugas yang kompleks dan kontekstual dengan dunia nyata. Penilaian ini memungkinkan setiap peserta didik untuk menunjukkan kompetensinya dalam situasi nyata, bukan sekedar mengingat dan memahami berbagai pengetahuan dan keterampilan yang bersifat abstrak. Hal ini sangat kontras dengan penilaian konvensional terstandar dalam bentuk pilihan ganda, benar-salah, menjodohkan, isian singkat, dan

sebagainya yang cenderung fokus pada kemampuan untuk mengingat dan memahami suatu konsep, prosedur, dan fakta.

Tugas yang digunakan sebagai penilaian autentik dapat berupa proyek yang bersifat kompleks dan membutuhkan waktu untuk menyelesaikan hingga beberapa minggu atau tugas yang dapat diselesaikan dalam waktu yang singkat. Tetapi, satu hal yang harus digarisbawahi bahwa tugas-tugas itu harus mereplikasikan berbagai tantangan dalam dunia nyata dan berbagai tahapan yang dilakukan oleh peserta didik untuk menyelesaikan tugas. Dalam praktiknya, penerapan penilaian autentik tidak mengikuti kriteria-kriteria yang biasa diterapkan dalam penilaian konvensional, yaitu proses penilaian yang terstandarisasi, objektif dan efisien. Proses penskoran dalam penilaian autentik juga tidak dapat dilakukan dengan memanfaatkan mesin seperti halnya dalam penilaian konvensional.

Penilaian autentik menghendaki peserta didik untuk menggunakan daya nalar, pikiran kritis, dan kemampuan untuk mencari solusi yang kreatif untuk dapat menyelesaikan tugas yang diberikan oleh guru. Karena alasan ini, penilaian autentik tidak lagi sekedar menekankan peserta didik untuk menerapkan kemampuan untuk menghafal, tetapi penilaian autentik menghendaki peserta didik untuk menerapkan kemampuan tingkat tinggi atau lebih dikenal dengan istilah *higher order thinking skills* (HOTS). Sedangkan penilaian yang hanya sekedar mengarahkan peserta didik untuk menghafal dan memahami suatu konsep, prosedur, dan fakta dikenal dengan istilah *lower order thinking skills* (LOTS). Kemampuan berpikir tingkat tinggi akan mampu mengarahkan peserta didik untuk mencari solusi terhadap permasalahan nyata dan krusial yang dihadapi dalam kehidupannya. Oleh karena itu, Suyanto (2020) menyatakan bahwa orientasi penilaian autentik adalah mengarahkan peserta didik mempunyai sikap solutif terhadap permasalahan yang ada di kenyataan hidup, membangun kesadaran untuk memberi kepedulian terhadap lingkungan kehidupan sosial, ekonomi, budaya, seni, fisik, dan teknologi yang merupakan suatu yang tak terpisah dengan kehidupan peserta didik. Lebih lanjut, Suyanto menegaskan bahwa kosekuensi penilaian autentik adalah guru harus bekerja lebih keras, memperhatikan masalah kehidupan di sekitarnya, mengembangkan soal untuk penilaian, dan memeriksa jawaban peserta didik yang belum tersedia melalui program komputer atau pun dapat dilakukan orang lain yang tidak paham materi yang dipersoalkan dalam penilaian autentik. Kesabaran dan

kepedulian terhadap hasil belajar peserta didik yang lebih baik merupakan kepuasan dan kebanggaan guru. Dalam penilaian autentik ini, kedudukan guru menjadi semakin penting.

Mueller (2014) secara tegas membedakan antara penilaian autentik dan penilaian non-autentik (penilaian tradisional atau konvensional) seperti disajikan dalam tabel 9.1.

Tabel 9.1

Perbedaan penilaian autentik dan penilaian konvensional

Penilaian non-autentik	Penilaian autentik
• Memilih/Merespon: Peserta didik memilih jawaban, menentukan pilihan, dan menjawab dengan uraian • Dikondisikan: Aktivitas peserta didik dikondisikan sesuai dengan keinginan penguji, seperti memilih jawaban yang dikondisikan guru • Mengingat/ Menyatakan: Peserta didik mengingat atau menyatakan informasi yang mereka kuasai. • Struktur Dirancang Guru: Peserta didik perlu berhati-hati untuk mengembangkan struktur yang guru harapkan, memenuhi target seperti yang guru inginkan • Bukti Tidak Langsung: Dalam	• Melaksanakan kegiatan: Peserta didik melakukan aktivitas yang sesungguhnya sehingga memperoleh pengalaman belajar. • Kenyataan Hidup: Guru menilai kenyataan yang sesungguhnya peserta didik lakukan pada kehidupan nyata dalam waktu pendek • Konstruksi/Aplikasi: Penilaian Autentik memperhatikan peserta didik menganalisis atau mengaplikasikan ilmu dalam proses berkreasi, berinovasi atau mencipta. • Struktur Prilaku Dikembangkan Peserta didik: Penilaian autentik memberi ruang kepada peserta didik mengembangkan konstruksi sesuai dengan keinginannya • Bukti Langsung: Dalam

penilaian tradisional melalui tes pilihan ganda, misalnya, memperoleh bukti kompetensi peserta didik tidak langsung

penilaian autentik guru memperoleh bukti langsung tentang perkembangan kompetensi yang ditunjukkan peserta didik secara langsung

B. Kelebihan dan Kekurangan Penilaian Non-autentik dan Autentik

Seperti telah dijelaskan di bagian awal bahwa penilaian non-autentik atau penilaian tradisional mengacu pada metode penilaian konvensional yang terstandar, berbasis kertas dan pensil dalam bentuk pilihan ganda, benar-salah, dan menjodohkan. Sedangkan penilaian autentik adalah sebuah proses penilaian dimana peserta didik diminta untuk melakukan tugas-tugas yang mengacu pada dunia nyata guna mendemonstrasikan terapan pengetahuan dan keterampilan yang telah dipelajarinya. Penilaian non-autentik dimaksudkan untuk mengevaluasi apakah peserta didik telah menguasai materi yang diajarkan oleh guru; untuk menentukan apakah peserta didik telah berhasil memperoleh pengetahuan yang telah ditentukan; dan untuk memberikan peringkat terhadap hasil belajar peserta didik dan membandingkannya dengan standar yang telah ditentukan. Sedangkan penilaian autentik lebih dimaksudkan untuk mengukur kecakapan peserta didik dengan meminta mereka untuk melakukan tugas-tugas yang mengacu dunia nyata; untuk memberikan jalan kepada peserta didik untuk belajar dan mendemonstrasikan kemampuan terbaiknya berkaitan dengan apa yang telah mereka pelajari; untuk memberikan umpan balik dan membantu peserta didik mengelola cara belajar mereka sendiri; dan untuk mengevaluasi kompetensi yang dicapai oleh peserta didik.

Dalam implementasinya, penilaian non-autentik dan penilaian autentik memiliki kelebihan dan kekurangan masing-masing. Berangkat dari kelebihan dan kekurangan dari dua jenis penilaian tersebut, guru diharapkan mampu menentukan pilihannya dalam rangka mengumpulkan berbagai informasi terkait proses dan hasil belajar setiap peserta didik. Lebih lanjut, Abdao (2015) menjelaskan secara detail kelebihan dan kekurangan dari penilaian non-autentik dan penilaian autentik yang diuraikan pada tabel 9.2 dan tabel 9.3.

Tabel 9.2

Kelebihan penilaian non-autentik dibandingkan dengan penilaian autentik

Penilaian non-autentik	Penilaian autentik
Kelebihan:	Kekurangan:
• Mudah dilakukan penskoran. Guru dapat melakukan evaluasi terhadap hasil belajar peserta didik dengan mudah dan cepat. • Proses persiapan penilaian lebih mudah dan cepat; proses penilaian juga mudah dilaksanakan. • Hasil penilaian lebih objektif, reliabel dan valid • Pelaksanaan proses penilaian lebih ekonomis	• Proses evaluasi lebih sulit dan kompleks. Guru memerlukan lebih banyak waktu untuk menyelesaikan proses evaluasi hasil belajar peserta didik. • Proses persiapan penilaian banyak menyita waktu • Hasil penilaian kurang objektif dan cenderung subjektif sehingga rentan terhadap ketidakadilan. Reliabilitas dan validitas menjadi rendah jika proses penilaian tidak dipandu dengan rubrik penilaian yang disusun/dipersiapkan dengan baik. • Pelaksanaan proses penilaian kurang ekonomis karena memungkinkan peserta didik memerlukan dukungan banyak biaya untuk menyelesaikan tugas.

Walupun penilaian autentik diidentifikasi memiliki beberapa kekurangan, tetapi penilaian autentik pun juga memiliki banyak kelebihan. Pemahaman yang baik terhadap kekurangan dan kelebihan dari kedua jenis penilaian tersebut akan mengarahkan guru untuk mampu mendesain proses pembelajaran yang lebih bermakna bagi peserta didik. Selanjutnya,

kelebihan penilaian autentik dibandingkan dengan penilaian non-autentik dapat dilihat pada tabel 9.3.

Tabel 9.3

Kelebihan penilaian autentik dibandingkan dengan penilaian non-autentik

Penilaian non-autentik	Penilaian autentik
Kekurangan:	Kelebihan:
• Hanya memberikan gambaran singkat kepada guru tentang apa yang benar-benar telah dipelajari peserta didik.	• Mampu memberikan kepada guru gambaran nyata tentang bagaimana dan di mana peserta didik harus belajar; memberikan lebih banyak informasi tentang kekuatan, kelemahan, kebutuhan, dan preferensi peserta didik sehingga guru dapat membantu mereka dalam menyesuaikan proses instruksional untuk meningkatkan kualitas belajar dan mengajar.
• Hanya memberikan pilihan yang sangat terbatas kepada setiap peserta didik untuk mendemonstrasikan pengetahuan dan keterampilan yang telah mereka pelajari.	• Memberikan banyak alternatif/cara kepada peserta didik untuk mendemonstrasikan apa yang mereka telah pelajari; dan menawarkan berbagai aktivitas penilaian yang menarik dan menantang.
• Proses penilaian terpisah dengan proses pembelajaran	• Proses penilaian terintegrasi dengan proses pembelajaran karena saat proses pembelajaran sedang berlangsung pun guru dapat melakukan penilaian dengan teknik observasi.
• Lebih banyak mengungkap keterampilan kognitif tingkat	• Mengungkap keterampilan kognitif tingkat tinggi yang dimiliki oleh

rendah yang dimiliki peserta didik: pengetahuan dan pemahaman

peserta didik: tidak hanya sebatas pada pengetahuan dan pemahaman, tetapi juga hingga analisis, aplikasi, evaluasi, dan kreasi.

- Lebih banyak menekankan pada kemampuan peserta didik untuk menghafal dan mengingat fakta, prinsip, dan prosedur tertentu.
 - Proses penilaian terstruktur pada guru (*teacher-structured*): guru mengarahkan dan bertindak sebagai evaluator; peserta didik hanya menjawab berbagai pertanyaan yang diberikan oleh guru.
 - Peserta didik lebih banyak bekerja sendiri-sendiri yang mengakibatkan munculnya sikap kompetitif yang tinggi diantara peserta didik.
 - Memunculkan perasaan cemas bagi peserta didik yang dapat mengganggu proses pembelajaran.
 - Waktu ditentukan secara pasti dan terbatas, peserta didik dipaksa untuk menyelesaikan tes dalam
- Meningkatkan kemampuan peserta didik untuk menerapkan keterampilan dan pengetahuan dalam situasi kehidupan nyata; dan juga memanfaatkan keterampilan kognitif untuk pemecahan masalah tingkat tinggi
 - Proses penilaian terstruktur pada peserta didik (*student-structured*): Peserta didik lebih banyak terlibat dalam pembelajaran dan hasil penilaian akan digunakan sebagai panduan untuk memperbaiki proses pembelajaran.
 - Seringkali melibatkan peserta didik untuk saling bekerja-sama dalam kelompok sehingga hal ini dapat mempromosikan sikap kerja-sama dalam tim, dan juga meningkatkan keterampilan kolaboratif dan interpersonal.
 - Mengurangi perasaan cemas dan menciptakan suasana rileks bagi peserta didik sehingga dapat meningkatkan kualitas pembelajaran.
 - Waktu penilaian lebih fleksibel

waktu yang telah ditentukan.

- Berfokus pada kecerdasan peserta didik.
 - Berfokus pada perkembangan belajar peserta didik.
 - Peserta didik diminta mengekspresikan pemahaman mereka tentang materi pelajaran dengan menggunakan berbagai bentuk kecerdasan yang mereka sukai.
 - Memberikan peluang kepada berbagai pihak (orang tua dan masyarakat umum) untuk dapat mengamati produk hasil belajar peserta didik, sehingga mereka dapat memberikan dukungan yang optimal terhadap proses belajar peserta didik.
-

Berdasarkan dua tabel tersebut, dapat dipahami bahwa penilaian non-autentik maupun penilaian autentik memiliki kelebihan dan kekurangan masing-masing. Hal tersebut mengisyaratkan bahwa guru semestinya memanfaatkan kedua jenis penilaian tersebut secara simultan untuk mengumpulkan berbagai informasi tentang kemajuan belajar peserta didik. Selain itu, pemanfaatan kedua penilaian tersebut juga dapat mengarahkan guru untuk mendapatkan informasi yang lebih akurat dan menggambarkan kompetensi faktual yang dimiliki oleh setiap peserta didik.

C. Jenis-jenis Penilaian Autentik

Penilaian autentik menjadi sebuah metode penilaian yang dilakukan oleh guru untuk mendapatkan gambaran kompetensi peserta didik yang lebih bermakna, baik untuk guru sendiri maupun untuk peserta didik. Bermakna bagi guru karena berdasarkan penilaian autentik, guru dapat mengetahui sejauhmana efektivitas program pembelajaran yang telah dilakukan untuk mencapai tujuan pembelajaran yang telah ditentukan. Guru kemudian dapat

melakukan modifikasi metode pembelajaran yang telah dilakukan selama ini atau bahkan mengganti dengan metode yang baru apabila diidentifikasi metode pembelajaran yang digunakan selama ini tidak efektif untuk mencapai tujuan pembelajaran yang telah ditentukan. Di pihak lain, penilaian autentik akan lebih bermakna bagi peserta didik karena hasil penilaian autentik mampu memberikan gambaran mengenai kompetensi yang sebenarnya dimiliki oleh peserta didik. Kompetensi ini dapat diketahui secara akurat karena penilaian autentik meghendaki peserta didik harus melakukan tugas yang mencerminkan kehidupan nyata (*real world*) melalui aktivitas mendemonstrasikan, mempresentasikan, memperagakan dan sebagainya. Sebagai metode penilaian berbasis kinerja, penilaian autentik terdiri-dari beberapa jenis penilaian yang diuraikan pada bagian berikut.

1. Penilaian Berbasis Kinerja (*Performance-based Assessments*)

Guru pada umumnya mash belum banyak menerapkan penilaian berbasis kinerja untuk mengumpulkan infromasi berkaitan dengan hasil belajar peserta didik. Guru pada umumnya masih banyak mengandalkan penilaian tradisional sebagai sebuah pendekatan untuk tujuan membuat keputusan terhadap hasil belajar. Masalah yang sering muncul berkaitan dengan penerapan pendekatan ini adalah guru hanya menekankan pada sejauhmana peserta didik telah menguasai bahan ajar yang telah disampaikan di dalam kelas. Dengan demikian, guru tidak banyak mendapatkan informasi tentang bagaimana peserta didik menggunakan atau memanfaatkan pengetahuannya untuk melakukan tugas yang mencerminkan dunia nyata. Misalnya, pada saat guru bahasa Inggris menggunakan tes bentuk pilihan ganda untuk mengukur hasil belajar peserta didik, tes ini lebih banyak mengukur kemampuan peserta didik untuk mengingat dan memahami materi yang telah diajarkan dan tes bentuk ini tidak banyak mengungkap kompetensi peserta didik untuk menggunakan bahasa Inggris untuk berkomunikasi, baik secara lisan mapun tulisan. Oleh karena itu, guru perlu melakukan perubahan pola pikir berkaitan dengan bagaimana melakukan penilaian yang tidak hanya mengukur kemampuan peserta didik untuk mengingat kembali materi yang telah diajarkan, tetapi bagaimana sebuah proses penilaian mampu memberikan informasi akurat tentang kemampuan peserta didik untuk menerapkan pengetahuannya untuk menyelesaikan tugas yang kompleks.

Sudah sering dijumpai bahwa ketika peserta didik mampu mengerjakan tes tradisional dengan baik, tidak ada jaminan bagi mereka mampu menerapkan pengetahuannya dalam situasi autentik. Pengukuran kemampuan peserta didik untuk menggunakan pengetahuannya dalam situasi dunia nyata memerlukan kreativitas guru untuk menempatkan mereka ke dalam situasi dimana peserta didik itu dapat menggunakan pengetahuannya untuk menyelesaikan tugas otentik (Lim & Griffith, 2011). Oleh karena itu, penilaian berbasis kinerja merupakan pendekatan penilaian yang sangat berguna untuk tujuan ini. Penilaian berbasis kinerja merupakan strategi penilaian yang menghendaki peserta didik untuk menerapkan pengetahuan dan keterampilan yang dimiliki dalam kinerja nyata agar lebih bermakna bagi kehidupan mereka (Hibbard, 1996). Model penilaian seperti ini memberikan informasi yang lebih akurat kepada guru tentang seberapa baik memahami suatu pengetahuan tertentu sekaligus menerapkannya pengetahuan itu dalam dunia nyata. Model penilaian ini mampu memberikan informasi kepada guru tidak hanya sekedar kemampuan peserta didik untuk menghafal dan mengingat fakta, prinsip, prosedur, dan konsep semata.

Sebagai salah model penilaian autentik, aktivitas dalam proses penilaian berbasis kinerja harus memenuhi berbagai kriteria tertentu. The National Capital Language Resource Center (NCLRC, 2011) menyatakan bahwa ada beberapa kriteria yang harus dipenuhi agar penilaian berbasis kinerja memenuhi karakteristik sebagai penilaian autentik, yaitu:

- a. Penilaian dikembangkan berkaitan dengan topik atau masalah yang menarik bagi peserta didik;
- b. Penilaian dikembangkan dengan mereplikasi konteks dan situasi dunia nyata;
- c. Penilaian harus melibatkan berbagai tugas multi-tahap dan mengacu pada masalah nyata yang membutuhkan penerapan pengetahuan dan keterampilan yang dimiliki peserta didik;
- d. Penilaian perlu mewajibkan peserta didik untuk menghasilkan produk atau kinerja yang berkualitas;
- e. Kriteria dan standar penilaian harus diketahui oleh peserta didik;
- f. Penilaian harus melibatkan adanya interaksi antara asesor (guru, teman sebaya, diri sendiri) dan peserta didik yang dinilai; dan

- g. Penilaian perlu memberi peluang bagi peserta didik untuk melakukan penilaian diri dan koreksi diri saat mereka sedang menyelesaikan tugas yang diberikan.

Penilaian berbasis kinerja diidentifikasi memiliki beberapa manfaat yang dapat diambil oleh peserta didik. Blaz (2001) menguraikan beberapa manfaat penilaian berbasis kinerja sebagai berikut:

- a. Tugas kinerja banyak melibatkan dan menarik minat peserta didik;
- b. Penilaian berbasis kinerja adalah indikator yang akurat dan bermakna tidak hanya tentang apa yang ketahui oleh peserta didik, tetapi juga tentang apa yang peserta didik dapat lakukan;
- c. Penilaian berbasis kinerja dapat meningkatkan kepercayaan diri peserta didik karena mereka mampu mengetahui dan memahami sesuatu sekaligus menerapkannya sesuai standar tertentu;
- d. Penilaian berbasis kinerja dapat meningkatkan kejelasan berkaitan dengan arah penilaian karena berdasarkan rubrik yang diberikan oleh guru, peserta didik akan mengetahui persis apa yang diharapkan untuk mendapatkan nilai atau skor tertentu;
- e. Penilaian berbasis kinerja mampu meningkatkan kepercayaan diri guru dalam menilai hasil belajar peserta didik; dan
- f. Tugas kinerja yang telah diselesaikan oleh peserta didik mampu memberikan informasi kepada guru tentang seberapa baik metode pembelajaran yang digunakan oleh guru, dan juga memberi informasi kepada peserta didik tentang seberapa baik mereka telah belajar.

Model penilaian berbasis kinerja, seperti diuraikan oleh McTighe & Ferrara (1998), dapat dibedakan menjadi tiga jenis, yaitu penilaian produk, penilaian kinerja, dan penilaian yang berorientasi pada proses. Produk mengacu pada sesuatu yang dihasilkan oleh peserta didik dan merupakan contoh konkret dari penerapan pengetahuan yang dimilikinya. Produk yang dihasilkan oleh peserta didik dapat berupa, misalnya halaman website, batik, laporan, brosur dan lain sebagainya. Produk-produk ini biasanya diselesaikan di luar kelas. Sedangkan kinerja mengacu pada bagaimana peserta didik dapat menerapkan pengetahuan dan keterampilan yang dimilikinya di bawah pengamatan guru. Kegiatan ini biasanya dilaksanakan

di dalam kelas karena harus melibatkan guru untuk mengamati kinerja setiap peserta didik, walaupun proses persiapan dapat dilakukan di luar kelas. Selanjutnya, penilaian yang berorientasi pada proses dapat memberikan wawasan tentang pemikiran, penalaran, dan motivasi peserta didik. Peserta didik dapat memberikan informasi diagnostik tentang bagaimana harus merefleksikan pembelajaran yang telah dilakukan dan selanjutnya menentukan strategi untuk memperbaikinya.

Guna menghindari subjektivitas hasil penilaian berbasis kinerja, guru harus menentukan terlebih dulu rubrik penilaian. Rubrik penilaian mencakup spesifikasi keterampilan yang dinilai dan berbagai hal yang mempengaruhi terbentuknya tingkat keberhasilan kinerja peserta didik, misalnya motivasi, kepercayaan diri, kerjasama, dan sebagainya. Oleh karena itu, agar rubrik penilaian dapat digunakan sebagai panduan yang baik untuk melakukan proses penilaian terhadap kinerja peserta didik, rubrik penilaian harus dikembangkan secara hati-hati. Griffith & Lim (2012) menjelaskan bahwa untuk mengembangkan rubrik yang baik, ada beberapa tahapan yang harus dilakukan, yaitu:

1. Menentukan perilaku peserta didik yang akan dinilai.

Langkah pertama adalah menentukan perilaku apa yang harus dilakukan peserta didik. Dalam menentukan perilaku yang akan dinilai ini, guru dapat dipandu dengan menjawab beberapa pertanyaan, misalnya: konsep, keterampilan, dan pengetahuan apa yang akan dinilai? Bagaimana peserta didik dapat memahami konsep, keterampilan dan pengetahuan itu, pada tingkatan seperti apa kinerja yang harus ditunjukkan oleh peserta didik; jenis pengetahuan apa yang akan dinilai: kemampuan bernalar, mengingat, atau proses? Dengan menjawab pertanyaan-pertanyaan ini, guru akan dapat menentukan aktivitas apa yang harus dilakukan oleh peserta didik selama proses penilaian berbasis kinerja.

2. Memilih jenis aktivitas yang harus dilakukan oleh peserta didik.

Setelah menentukan tujuan asesmen, guru harus segera memutuskan kegiatan apa yang harus dilakukan oleh peserta didik selama proses penilaian berbasis kinerja ini. Keputusan harus mempertimbangkan masalah-masalah yang berkaitan dengan

batasan waktu, sumber daya, dan berapa banyak data/informasi yang diperlukan selama proses penilaian berlangsung.

3. Menentukan kriteria penilaian yang akan digunakan sebagai dasar untuk menentukan kualitas kinerja peserta didik;
Setelah menentukan aktivitas dan tugas yang harus dilakukan oleh peserta didik, guru harus segera menentukan kriteria yang akan digunakan untuk menentukan keberhasilan kinerja peserta didik.

Hal yang harus dipahami oleh guru adalah penilaian berbasis kinerja tidak akan dapat dilakukan oleh guru dengan baik tanpa didukung adanya penguasaan pengetahuan dan keterampilan yang baik pula oleh peserta didik. Oleh karena itu, sebelum penilaian ini dilakukan, guru harus membekali terlebih dahulu pengetahuan dan keterampilan melalui proses pembelajaran.

2. Penilaian Berbasis Proyek (*Project-based Assessments*)

Penilaian berbasis proyek adalah pendekatan penilaian yang digunakan oleh guru untuk menilai kinerja peserta didik. Peserta didik harus melakukan serangkaian aktivitas dan menghasilkan solusi terhadap suatu permasalahan tertentu berdasarkan informasi proyek yang diberikan kepada mereka. Penilaian berbasis proyek merupakan alternatif dari proses penilaian yang memungkinkan peserta didik untuk terlibat dalam keseluruhan pembelajaran dengan cara yang lebih konkret. Peserta didik tidak hanya sekedar mempelajari teori, tetapi melalui penilaian ini, peserta didik langsung diminta untuk menerapkan topik atau materi yang telah dipelajari dengan mengeksplorasi topik tersebut secara mendalam melalui proyek yang harus diselesaikan dalam waktu tertentu. Guru dapat menggunakan proyek sebagai bagian dari proses pembelajaran yang sedang berlangsung atau sebagai pengganti ujian akhir tradisional.

Penilaian berbasis proyek seringkali merupakan komponen penting dari pembelajaran berbasis proyek (*Project-Based Learning* atau *PBL*), dimana keseluruhan proses pembelajaran difokuskan untuk melibatkan peserta didik dalam pemecahan masalah tertentu. Seperti PBL, penilaian berbasis proyek adalah sebuah proses penilaian yang berpusat pada peserta didik (*student-centered assessment*) dan memerlukan proses refleksi pada proses dan materi/isi pembelajaran agar lebih bermakna.

Agar penilaian berbasis proyek berjalan dengan baik, dibutuhkan perencanaan yang cermat dan guru harus bersikap fleksibel, karena tidak ada dua proyek yang harus dihasilkan oleh peserta didik yang benar-benar sama. Apabila guru akan menerapkan penilaian berbasis proyek di kelas, mungkin perlu mempertimbangkan dan mengikuti praktik terbaik (*best practice*) yang pernah dilakukan oleh *Buck Institute for Education (BIE)* berikut ini untuk memastikan agar penilaian lebih bermakna dan efektif:

1. Biarkan peserta didik untuk memilih topik, masalah, dan arah proyeknya sendiri agar penilaian itu lebih bermakna bagi mereka secara personal.
2. Guru harus menentukan parameter proyek dan rubrik penilaian secara jelas sejak awal sehingga peserta didik — dan mungkin juga orang tua peserta didik — mengetahui apa yang harus dilakukan dan produk apa yang akan dihasilkan dari proyek tersebut.
3. Jika peserta didik harus bekerja dalam kelompok, guru harus memastikan telah memiliki strategi untuk menilai kinerja peserta didik secara individu, bukan hanya sekedar memberikan nilai bersama pada produk akhir yang dihasilkan oleh kelompok tersebut. Hal ini penting untuk diperhatikan karena kualitas kinerja masing-masing peserta didik selama menyelesaikan proyek mungkin berbeda satu sama lain.
4. Guru harus sering-sering mengecek pekerjaan peserta didik, baik secara informal maupun dengan menggunakan tolok ukur yang telah diberitahukan kepada peserta didik dengan maksud untuk membantu proses belajar mereka dan untuk memastikan mereka sukses dalam belajarnya.

Penilaian berbasis proyek banyak berguna bagi peserta didik jika diterapkan secara konsisten dan benar, baik di dalam kelas maupun di luar kelas. Trach (2018) menjelaskan bahwa penilaian berbasis proyek dapat berguna bagi peserta didik berkaitan dengan hal-hal sebagai berikut:

1. Autentik

Penilaian berbasis proyek mampu memberikan pengalaman dan kesempatan kepada peserta didik untuk menerapkan pengetahuan dan keterampilan yang dimiliki dalam kehidupan nyata.

2. Motivasi

Penilaian berbasis proyek yang dirancang dengan baik dapat mengarahkan peserta didik mampu menentukan pilihannya dengan baik pula. Artinya, peserta didik dapat memilih tugas/proyek yang paling berarti bagi mereka. Hal ini dapat meningkatkan keterlibatan dan motivasi peserta didik untuk melakukan hal yang terbaik dalam proses belajarnya sehingga pembelajaran akan lebih bermakna dan melekat pada peserta didik di dalam hidupnya.

3. Metakognitif

Penilaian berbasis proyek yang dirancang dengan baik dapat memberikan umpan balik yang signifikan kepada peserta didik tentang kemajuan belajarnya. Untuk mencapai tujuan ini, guru perlu merancang rubrik yang jelas dan terperinci agar dapat digunakan peserta didik untuk mengukur kemajuan dan keberhasilannya saat bekerja menyelesaikan proyek.

4. Kolaborasi

Proses menyelesaikan proyek adalah kesempatan yang sangat baik bagi peserta didik bekerja-sama untuk mencapai tujuan bersama.

5. Kreativitas

Karena pembelajaran berbasis proyek seringkali melibatkan pemecahan masalah, peserta didik dapat mengembangkan kreativitas dan keterampilan berpikir kritis yang berguna untuk memberi solusi terhadap permasalahan hidup yang mungkin dihadapinya. Proyek berbasis kelompok diidentifikasi sangat efektif untuk mencapai tujuan ini. Penelitian telah membuktikan bahwa ketika peserta didik fokus pada mengatasi konflik bersama-sama dan tidak berusaha menghindarinya, keterampilan akademiknya akan meningkat lebih baik.

Pembelajaran dan penilaian berbasis proyek dapat diterapkan pada proses pembelajaran dalam mata pelajaran apapun dan pada usia peserta didik berapa pun. Tetapi pembelajaran dan penilaian jenis ini memerlukan persiapan yang matang agar lebih bermakna bagi guru sendiri dan peserta didik.

3. Penilaian Berbasis Portofolio (*Portfolio-based Assessments*)

Penilaian berbasis portofolio merupakan salah satu jenis penilaian autentik yang mampu merekam aktivitas belajar dan kompetensi peserta didik dalam rentang waktu tertentu. Perekaman aktivitas belajar dan prestasi ini mampu memberikan gambaran tentang perkembangan kemajuan belajar peserta didik yang dapat digunakan oleh guru untuk mengukur efektivitas proses pembelajaran yang telah dilakukan. Penilaian berbasis portofolio bukan hanya sekedar mengumpulkan kinerja peserta didik dalam periode waktu tertentu, tetapi portofolio harus direncanakan secara matang agar mampu menggambarkan secara akurat usaha belajar, kemajuan belajar dan prestasi belajar yang telah dilakukan dan dicapai oleh peserta didik. Penilaian berbasis portofolio terdiri dari berbagai karya pilihan peserta didik, dan karya-karya itu mencakup paling tidak dua hal, yaitu refleksi peserta didik terhadap kinerjanya selama periode tertentu dan juga komentar atau feedback yang diberikan oleh guru berkaitan dengan refleksi dan kinerja peserta didik tersebut. Portofolio dapat digunakan untuk menunjukkan apa yang telah mampu dilakukan oleh peserta didik dengan pengetahuan dan keterampilan yang dimilikinya. Portofolio adalah salah satu strategi yang dimanfaatkan oleh guru untuk melacak kemajuan belajar peserta didik sepanjang tahun.

Portofolio dapat dimaknai sebagai kumpulan karya peserta didik yang dapat menunjukkan upaya, kemajuan, dan prestasi belajar peserta didik sesuai dengan kurikulum dalam periode tertentu. Portofolio dapat digunakan oleh guru sebagai salah satu model ujian yang mampu memberikan gambaran tentang pengalaman belajar peserta didik yang terpilih dan portofolio mampu mendokumentasikan hasil belajar peserta didik yang dinilai. Hal ini bermakna bahwa tidak semua karya peserta didik dapat dikumpulkan sebagai bagian dari portofolio, hanya karya-karya peserta didik yang terpilih dan terkait dengan kurikulum dan aspek yang dinilai saja yang harus dikumpulkan. Sebagai sebuah penilaian autentik, portofolio dapat mengarahkan guru untuk memenuhi beberapa tujuan, misalnya untuk:

1. meningkatkan kinerja peserta didik;
2. meningkatkan keterampilan peserta didik untuk melakukan penilaian diri dan menentukan tujuan belajar yang ingin dicapai;

3. menghadirkan potret atau gambaran secara jelas tentang profil peserta didik sebagai seorang pembelajar;
4. menyediakan media untuk mengkomunikasikan kemajuan belajar peserta didik bagi orang tua, guru, dan anggota masyarakat;
5. membangun rasa tanggung jawab peserta didik atas proses pembelajaran mereka sendiri;
6. membangun sikap memiliki dan rasa bangga bagi peserta didik atas prestasi yang telah dicapai;
7. membangun rasa percaya diri peserta didik terhadap kemampuan dan keterampilan yang telah dicapai sebagai seorang pembelajar.

Lebih lanjut, Arifin (2010) menjelaskan bahwa tujuan portofolio ditetapkan oleh apa yang harus dikerjakan dan siapa yang akan menggunakan penilaian portofolio tersebut. Hal ini dapat dimaknai bahwa pemanfaatan portofolio sebagai model penilaian harus direncanakan secara matang berkaitan dengan kompetensi apa yang akan diukur. Sejak awal guru dan peserta didik secara bersama-sama harus menentukan kompetensi yang akan diukur. Selajutnya hasil pengukuran itu akan dapat digunakan oleh guru untuk memberikan feedback yang tepat terhadap proses dan hasil belajar peserta didik. Hasil penilaian juga dapat digunakan oleh peserta didik untuk memperbaiki proses belajar yang telah dilakukan. Selain itu, orang tua juga dapat menggunakan hasil penilaian portofolio untuk memberikan bimbingan dan motivasi terhadap aktivitas belajar di rumah.

Penilaian berbasis portofolio juga memiliki fungsi yang sangat penting bagi peserta didik, guru, dan orang tua. Arifin (2010) selanjutnya menguraikan bahwa penilaian portofolio berfungsi sebagai berikut:

1. Portofolio sebagai sumber informasi bagi guru dan orang tua untuk mengetahui pertumbuhan dan perkembangan kemampuan peserta didik, tanggung jawab dalam belajar, perluasan dimensi belajar, dan pembaharuan proses pembelajaran.
2. Portofolio sebagai alat pengajaran merupakan komponen kurikulum, karena portofolio mengharuskan peserta didik untuk mengoleksi dan menunjukkan hasil kerja mereka.
3. Portofolio sebagai alat penilaian autentik (*authentic assessment*).
4. Portofolio sebagai sumber informasi bagi peserta didik untuk melakukan *self-assessment*.

Mencermati fungsi penting portofolio seperti diuraikan di atas, tampaknya guru perlu secara konsisten untuk menerapkan penilaian berbasis-portofolio ini sebagai upaya untuk mendapatkan informasi akurat dan faktual tentang kompetensi peserta didik.

Bab 10

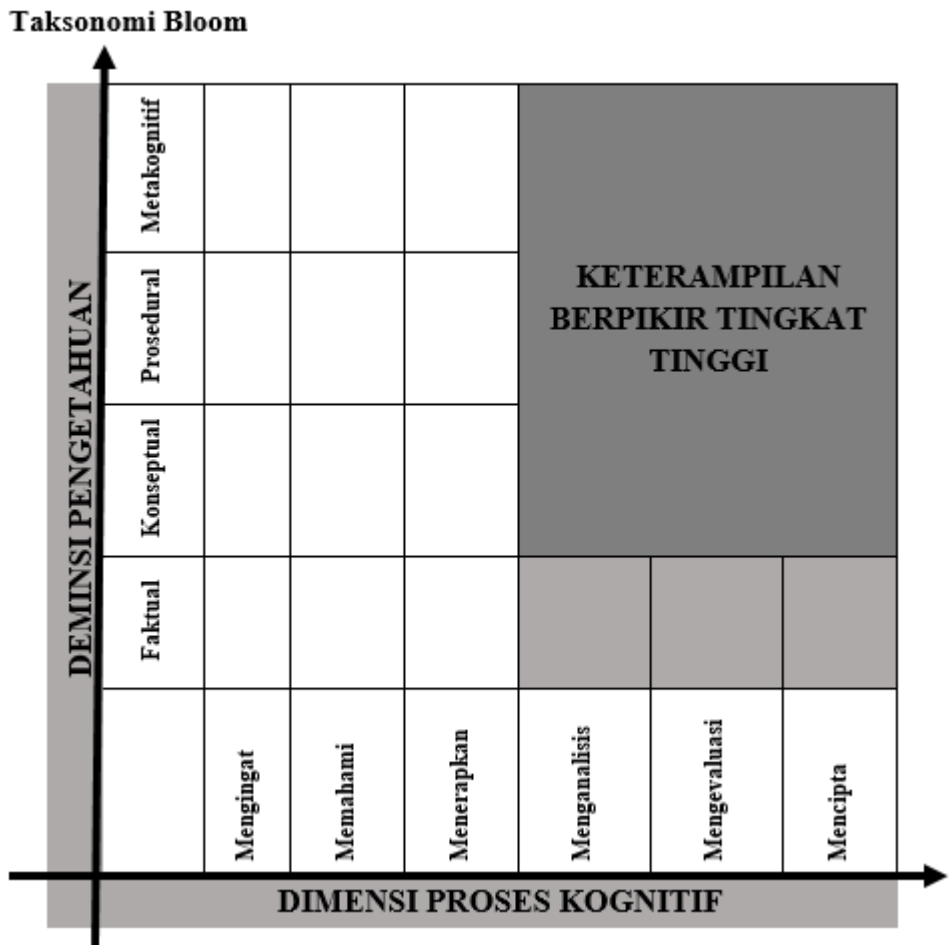
Penilaian Berbasis HOTS (Higher Order Thinking Skills)

Keterampilan berpikir tingkat tinggi (*higher-order thinking skills/HOTS*) yang harus dimiliki oleh peserta didik telah menjadi sebuah *trend* dan mendapatkan banyak perhatian oleh pendidik dan pemerhati pendidikan dalam dekade terakhir. Hal ini mengisyaratkan bahwa dalam dekade terakhir ini telah terjadi pergeseran paradigma proses pembelajaran tradisional yang banyak mengarahkan peserta didik untuk sekedar mengingat dan memahami suatu fakta, konsep, prinsip, dan prosedur. Kemudian bergeser menjadi proses pembelajaran modern yang menghendaki peserta didik agar mampu mentransfer pengetahuan, berpikir kritis, dan memecahkan berbagai permasalahan yang dihadapinya. Tiga kemampuan ini oleh Brookhart (2010) biasa disebut dengan kemampuan berpikir tingkat tinggi. Lalu, apa yang membedakan antara pembelajaran tradisional dan modern? Bagaimana guru harus mendesain proses pembelajaran modern? Bagaimana pendekatan penilaian yang tepat untuk diterapkan pada pembelajaran modern? Beberapa pertanyaan ini mengemuka menyertai terjadinya pergeseran paradigma proses pembelajaran dalam dekade terakhir ini.

A. Pembelajaran Berbasis HOTS

Pembelajaran berbasis HOTS adalah salah satu model pembelajaran modern yang banyak menjadi perhatian pada saat ini. Tetapi pada praktiknya, masih banyak yang memiliki miskonsepsi antara pembelajaran berbasis HOTS dan penilaian berbasis HOTS. Keduanya adalah dua hal yang berbeda, tetapi keduanya memiliki hubungan yang erat dan kuat. Penilaian berbasis HOTS tidak dapat dilaksanakan dengan baik, apabila

proses pembelajaran yang telah dilaksanakan oleh guru belum mencerminkan sebuah proses pembelajaran yang mengarahkan peserta didik untuk melakukan transfer pengetahuan, berpikir kritis, dan memecahkan masalah yang dihadapi dalam kehidupan faktual. Penilaian terhadap hasil pembelajaran modern tidak akan



Gambar 10.1: Gradasi tingkat berpikir

mampu memberikan informasi yang akurat terhadap kompetensi peserta didik apabila proses penilaian masih dilakukan dengan menerapkan prinsip-prinsip penilaian tradisional yang menekankan pada kemampuan peserta didik untuk sekedar mengingat dan memahami fakta, konsep, prinsip, dan

prosedur. Fokus pembelajaran dan penilaian HOTS adalah mengarahkan peserta didik untuk mampu menganalisis, mengevaluasi, dan mencipta suatu konsep, prosedur, dan metakognitif. Sejalan dengan taksonomi Bloom, tingkat berpikir yang diharapkan muncul dari proses pembelajaran dan penilaian HOTS diilustrasikan dalam gambar 10.1.

Fakta, konsep, prinsip, prosedur, dan metakognitif mengacu pada materi pembelajaran. Fakta adalah materi pembelajaran yang mengacu berbagai hal yang merujuk pada kenyataan dan kebenaran, meliputi nama-nama objek, peristiwa sejarah, lambang, nama tempat, nama orang, nama bagian atau komponen suatu benda, dan sebagainya. Sedangkan konsep merupakan materi pembelajaran yang meliputi definisi, pengertian, ciri khusus, hakekat, inti/isi dan sebagainya. Selanjutnya prinsip adalah materi pembelajaran yang berupa hal-hal utama, pokok, dan memiliki posisi terpenting, meliputi dalil, rumus, adagium, postulat, paradigma, teorema, serta hubungan antar konsep yang menggambarkan implikasi sebab akibat. Prosedur adalah materi pembelajaran berupa langkah-langkah sistematis atau berurutan dalam mengerjakan suatu aktivitas dan kronologi suatu sistem. Metakognitif merupakan kesadaran tentang apa yang diketahui dan apa yang belum diketahui. Strategi Metakognitif merujuk kepada cara untuk meningkatkan kesadaran mengenai proses berfikir dan pembelajaran yang berlaku. Penilaian yang hanya menekankan pada kemampuan peserta didik untuk sekedar mengingat dan memahami pokok-pokok materi tersebut belum cukup. Penguasaan yang baik terhadap pokok materi tersebut harus diikuti kemampuan untuk benalar, berfikir kritis, dan mampu memecahkan berbagai permasalahan tertentu. Tiga kemampuan ini biasa disebut dengan *higher-order thinking skills* (HOTS).

Sejalan dengan yang telah diuraikan di atas, pembelajaran HOTS bertujuan untuk mendorong peserta didik untuk memiliki keterampilan berpikir kritis, kreatif, inovatif, dan mampu menyelesaikan masalah. Saat ini, kreativitas dan inovasi menjadi modal penting dalam kesuksesan proses belajar peserta didik. Kreativitas dan inovasi juga menjadikan sebuah bangsa memiliki daya saing dibandingkan dengan bangsa yang lainnya. Agar proses pembelajaran HOTS berjalan efektif dalam mencapai tujuan yang telah ditentukan, guru harus mendesain skenario pembelajaran yang sistematis dan tercermin dalam rencana pelaksanaan pembelajaran (RPP). Guru harus menyiapkan bahan ajar, sumber belajar, media pembelajaran/alat peraga,

dan instrumen tes untuk mengukur hasil belajar sesuai kriteria HOTS. Pembelajaran HOTS tercermin dalam penerapan pendekatan saintifik yang meliputi 5M, yaitu: mengamati, menanya, mengumpulkan informasi, menalar/mengasosiasikan, dan mengomunikasikan. Oleh karena itu, skenario pembelajaran yang didesain oleh guru harus mengikuti tahapan-tahapan pendekatan saintifik tersebut agar tujuan pembelajaran HOTS dapat dicapai secara optimal.

1. Strategi Pembelajaran HOTS

Salah satu ciri pembelajaran modern abad 21 adalah guru mengarahkan peserta didik untuk menggunakan proses berpikir tingkat tinggi. Tidak sekedar mengingat dan memahami materi ajar, peserta didik diarahkan mampu menghubungkan pengetahuan dan keterampilan yang dimiliki dengan berbagai konsep sehingga hasil belajar lebih bermakna dalam kehidupan sehari-harinya. Untuk mencapai tujuan ini, strategi perlu dirancang dengan baik sehingga proses pembelajaran benar-benar berbasis HOTS. Cox (2019) menyatakan bahwa ada sepuluh strategi pembelajaran yang dapat digunakan guru untuk meningkatkan kualitas pembelajaran HOTS, yaitu:

- a. Guru perlu memastikan terlebih dahulu bahwa peserta didik telah memahami hakikat dari pembelajaran HOTS.
Pemahaman yang baik terhadap hakikat pembelajaran HOTS akan sangat membantu peserta didik mampu melakukan berbagai aktivitas belajar yang mencerminkan prinsip-prinsip pembelajaran HOTS. Pemahaman ini juga dapat menghindarkan peserta didik terjebak dalam berbagai aktivitas yang tercermin dalam proses pembelajaran tradisional.
- b. Pembelajaran harus diarahkan agar peserta didik menghubungkan suatu konsep dengan konsep yang lain.
Guru perlu membimbing peserta didik untuk mampu menghubungkan satu konsep dengan konsep lainnya. Proses pembelajaran yang mengarahkan peserta didik untuk mampu menghubungkan satu konsep dengan konsep yang lain ini akan membantu mereka lebih memahami materi yang sedang dipelajari.
- c. Peserta didik diarahkan untuk membuat kesimpulan.
Selama proses pembelajaran, guru mestinya mengarahkan peserta didik untuk membuat kesimpulan dari suatu peristiwa yang bersifat faktual.

Sebagai contoh, guru dapat mulai dengan memberikan gambar kepada peserta didik tentang orang-orang yang sedang mengantri di dapur umum. Guru selanjutnya meminta peserta didik untuk melihat gambar dan fokus pada detailnya. Kemudian, peserta didik diminta membuat kesimpulan berdasarkan apa yang mereka lihat dalam gambar.

- d. Dorong peserta didik untuk bertanya.

Apabila peserta didik memiliki kesempatan bertanya tanpa ada perasaan khawatir bahwa pertanyaan yang diberikan akan dianggap sebagai pertanyaan yang tidak berkualitas oleh guru dan teman-teman sebayanya, kreativitas peserta didik akan berkembang lebih baik.

- e. Guru dapat menggunakan *graphic organizer* selama proses pembelajaran. *Graphic organizer* adalah tampilan visual dan grafik yang menggambarkan hubungan antara fakta, istilah, dan/atau ide dalam suatu tugas pembelajaran. *Graphic organizer* mampu memberikan peserta didik jalan untuk membingkai pikiran mereka secara terorganisir.

- f. Arahkan peserta didik untuk memahami strategi untuk memecahkan suatu masalah.

Guru perlu membimbing peserta didik untuk menentukan langkah-langkah untuk memecahkan suatu masalah. Guru perlu juga mengarahkan peserta didik untuk menggunakan berbagai metode pemecahan masalah yang berbeda-beda sehingga mereka akan terbantu untuk mendapatkan solusi terhadap permasalahan tertentu dengan lebih cepat dan mudah.

- g. Proses pembelajaran harus mendorong peserta didik untuk berpikir kreatif.

Berpikir kreatif adalah ketika peserta didik menemukan, berimajinasi, dan merancang sendiri apa yang mereka pikirkan dan lakukan. Dengan menggunakan indra kreatif dapat membantu peserta didik memproses dan memahami informasi dengan lebih baik.

- h. Dorong peserta didik menggunakan *mind movies* pada saat mengalami kesulitan belajar.

Ketika konsep yang sedang dipelajari dirasa sulit, dorong peserta didik untuk membuat film dalam pikiran mereka. *Mind Movies* mampu mengubah pembelajaran yang membosankan menjadi lebih menyenangkan.

- i. Arahkan proses pembelajaran untuk membantu peserta didik mengelaborasi berbagai respon yang mereka berikan.

Berpikir tingkat tinggi menuntut peserta didik untuk benar-benar memahami suatu konsep, bukan sekedar mengulang atau menghafalnya. Guru perlu meminta peserta didik untuk menguraikan jawaban yang mereka berikan atas suatu pertanyaan yang diberikan guru dengan mengajukan pertanyaan lanjutan yang relevan sehingga peserta didik mampu menjelaskan gagasan/jawabannya secara lebih rinci.

j. Biasakan peserta didik melakukan aktivitas tanya-jawab.

Strategi ini terbukti efektif untuk menciptakan kondisi bagi peserta didik untuk berpikir tingkat tinggi.

Mendesain proses pembelajaran HOTS bukan suatu aktivitas instan yang mudah untuk dilakukan oleh guru. Guru perlu membiasakan diri untuk mengasah keterampilannya sehingga mampu mendesain pembelajaran HOTS secara berkesinambungan. Oleh karena itu praktik baik seperti diuraikan oleh Cox tersebut dapat digunakan sebagai panduan dalam rangka menciptakan pembelajaran HOTS yang efektif dan bermakna bagi peserta didik.

2. Model-model Pembelajaran HOTS

Hal terpenting dari sebuah pembelajaran HOTS ada pada proses pembelajaran yang tergambar pada kegiatan inti. Pada kegiatan inti ini, guru harus mempraktikkan desain, model, strategi, dan metode pembelajaran tertentu. Ada beberapa model pembelajaran yang dapat dimanfaatkan oleh guru untuk menerapkan pembelajaran HOTS, misalnya: pembelajaran berbasis proyek (*project-based learning*), pembelajaran berbasis masalah (*problem-based learning*), dan pembelajaran berbasis penemuan (*discovery/inquiry-based learning*). Ketiga model pembelajaran tersebut memiliki karakteristik masing-masing, tetapi ketiganya menghendaki aktivitas pembelajaran yang berpusat pada peserta didik (*student-centered teaching and learning activities*).

a. Pembelajaran Berbasis Proyek

Pembelajaran berbasis proyek (*Project-based learning*) adalah sebuah model pembelajaran dimana peserta didik mendapatkan pengetahuan dan keterampilan melalui aktivitas bekerja dalam jangka waktu yang relatif lama untuk menyelidiki dan merespon pertanyaan, masalah, dan tantangan yang autentik, menarik, dan menantang. Pembelajaran ini menghendaki aktivitas

pedagogis yang berpusat pada peserta didik dan melibatkan pendekatan pembelajaran yang dinamis. Melalui pendekatan ini, peserta didik dipercaya akan mendapatkan pengetahuan dan keterampilan yang lebih mendalam melalui proses eksplorasi aktif terhadap tantangan dan permasalahan yang faktual atau nyata. Proses pembelajaran berbasis proyek sangat berbeda dengan pembelajaran tradisional yang lebih menekankan pada aspek ingatan dan berpusat pada arahan guru (*teacher-led instruction*). Metode pembelajaran ini dapat diterapkan dengan melibatkan peserta didik secara kelompok maupun individu.

Pembelajaran berbasis proyek memiliki beberapa karakteristik unik yang berbeda dengan pembelajaran tradisional berbasis ceramah. Diantara karakteristik pembelajaran berbasis proyek ini adalah:

1. Tugas atau proyek diselesaikan oleh peserta didik, baik secara individu maupun kelompok, dalam jangka waktu tertentu dimulai dari tahap perencanaan, penyusunan hingga pelaporan produk.
2. Peserta didik bertanggung-jawab penuh terhadap pelaksanaan proyek dan produk yang akan dihasilkan.
3. Selama proses penyelesaian proyek, peserta didik dapat melibatkan banyak pihak, seperti guru, tema sejawat, dan masyarakat.
4. Proses pembelajaran ini dapat melatih peserta didik untuk mampu berpikir kritis dan kreatif.
5. Proses pembelajaran ini sangat mengakomodasi adanya ketidaklengkapan gagasan yang disampaikan oleh peserta didik.

Sebagai sebuah metode pembelajaran, pembelajaran berbasis proyek memiliki langkah-langkah sistematis yang harus dilakukan agar proyek efektif meningkatkan kompetensi peserta didik. Langkah-langkah ini juga memberikan panduan bagi guru dan peserta didik tentang apa yang harus dilakukan dan bagaimana menyelesaikan proyeknya. Adapun langkah-langkah (sintaks) pembelajaran berbasis proyek adalah sebagai berikut:

1. Memberikan pertanyaan esensial

Proses pembelajaran diawali dengan guru menyampaikan topik dan memberikan pertanyaan esensial, yaitu pertanyaan yang dapat mengarahkan peserta didik melakukan aktivitas tertentu dalam rangka menyelesaikan proyek.

2. Mendesaian perencanaan proyek

Guru memastikan bahwa setiap peserta didik mengetahui prosedur dalam menyelesaikan proyek atau produk yang akan dihasilkan. Sedangkan peserta didik berdiskusi dengan kelompoknya untuk menyusun rencana penyelesaian proyek dan memecahkan masalah. Termasuk melakukan pembagian tugas, mempersiapkan alat, bahan, media dan sumber daya yang akan dibutuhkan selama menyelesaikan proyek.

3. Menyusun jadwal

Guru dan peserta didik membuat kesepakatan berkaitan dengan durasi untuk menyelesaikan proyek. Selanjutnya peserta didik membuat jadwal secara rinci dengan memperhatikan waktu yang telah disepakati bersama.

4. Memonitor perkembangan dan kemajuan proyek

Guru memantau aktivitas peserta didik selama menyelesaikan proyek, memantau perkembangan penyelesaian proyek dan memberikan feedback guna membantu peserta didik yang mengalami kesulitan. Peserta didik bekerja melaksanakan proyek sesuai jadwal yang telah dibuat, mencatat poin-poin penting terkait penyelesaian proyek, dan melakukan diskusi tentang berbagai permasalahan yang muncul selama penyelesaian proyek, baik dengan teman kelompoknya maupun dengan guru.

5. Menguji hasil

Guru mengukur tingkat ketercapaian peserta didik dalam menyelesaikan proyek, dan memberikan feedback tentang tingkat pemahaman yang sudah dicapai peserta didik. Peserta didik menilai kelayakan proyek yang telah diselesaikan dan membuat laporan tentang produk yang dihasilkan dengan memaparkannya kepada guru dan teman sejawat dari kelompok yang berbeda.

6. Megevaluasi pengalaman belajar peserta didika

Guru membimbing peserta didik dalam memaparkan proyek yang telah diselesaikan atau produk yang dihasilkan dan memberi tanggapan terhadap proyek dan produk tersebut, dan selanjutnya bersama-sama dengan peserta didik membuat kesimpulan terhadap keberhasilan proyek yang telah dikerjakan oleh peserta didik. Peserta didik baik secara individu atau kelompok memaparkan proyek yang telah diselesaikan, peserta didik alain memberi tanggapan, dan selanjutnya bersama-sama gurur membuat kesimpulan.

Hal yang harus digarabawahi adalah topik proyek yang harus diselesaikan oleh peserta didik merupakan topik yang kontekstual dan menarik; peserta didik cukup menyelesaikan satu proyek saja untuk satu topik dan harus diselesaikan dalam waktu yang telah ditentukan; penyelesaian proyek memanfaatkan berbagai bahan yang tersedia di lingkungan sekitar; dan penilaian autentik menekankan pada kemampuan peserta didik untuk merancang, menerapkan, menemukan dan menyampaikan produknya kepada orang lain.

b. Pembelajaran Berbasis masalah

Pembelajaran berbasis masalah (*problem-based learning*) adalah sebuah model pembelajaran yang mengedepankan kemampuan berpikir kritis peserta didik, baik secara individu atau kelompok untuk menyelesaikan permasalahan kontekstual sehingga proses pembelajaran menjadi lebih bermakna. Tujuan dari pembelajaran berbasis masalah adalah untuk meningkatkan kompetensi peserta didik untuk mengatasi berbagai permasalahan kontekstual dengan mengintegrasikan keterampilan berpikir tingkat tinggi (*higher-order thinking skills*) yang dimilikinya.

Model pembelajaran ini, seperti model pembelajaran berbasis proyek, juga memiliki karakteristik unik dan spesifik. Beberapa karakteristik pembelajaran ini adalah:

1. Pembelajaran berangkat dari permasalahan kontekstual yang dipilih oleh guru dan peserta didik.
2. Masalah kontekstual yang diangkat masih berupa masalah yang mengambang (*ill-structured*), yaitu masalah yang belum memiliki solusi dan tujuan yang jelas.
3. Masalah yang diangkat adalah masalah yang multi-perspektif.
4. Masalah yang diangkat merupakan masalah yang menantang, yaitu masalah yang dapat mengarahkan peserta didik untuk bekerja-keras untuk dapat menyelesaikan sehingga usaha dan kerja keras itu mampu membentuk pengetahuan baru.
5. Model pembelajaran berbasis masalah lebih mengutamakan belajar mandiri dengan memanfaatkan berbagai sumber pengetahuan yang tidak tunggal.
6. Proses pembelajaran bersifat kolaboratif, komunikatif dan kooperatif.

Model pembelajaran berbasis masalah ini menghendaki peserta didik mampu memecahkan masalah kontekstual dan melalui proses pemecahan masalah itu peserta didik dapat merekonstruksi pengetahuan baru. Untuk mencapainya tujuan ini, guru dan peserta didik perlu mengikuti langkah-langkah sistematis dari model pembelajaran tersebut, yaitu:

1. Orientasi peserta didik pada masalah tertentu

Guru menyampaikan masalah tertentu yang harus dicari solusinya oleh peserta didik. Peserta didik mencermati dan memahami masalah yang telah ditentukan oleh guru untuk segera dicari pemecahannya.

2. Mengorganisasikan peserta didik untuk belajar secara efektif

Guru memastikan setiap peserta didik memahami tugas masing-masing dalam kelompoknya dan peserta didik selanjutnya melakukan berbagai diskusi untuk mencari data, bahan, dan alat yang diperlukan untuk memecahkan masalah.

3. Membimbing peserta didik untuk melakukan penyelidikan, baik secara individu atau kelompok

Jika diperlukan, guru dapat membantu peserta didik untuk mencari data dan bahan yang diperlukan untuk pemecahan masalah. Sedangkan peserta didik melakukan berbagai penyelidikan untuk mencari data dan referensi sebagai bahan diskusi dalam rangka pemecahan masalah.

4. Mengembangkan dan menyajikan hasil karya

Guru memantau proses diskusi peserta didik dan membimbing peserta didik untuk membuat laporan yang harus dipresentasikan. Setelah menemukan solusi terhadap permasalahan yang ditentukan, peserta didik selanjutnya mempresentasikan solusi tersebut kepada guru dan teman-temannya.

5. Menganalisis dan mengevaluasi proses pemecahan masalah

Guru memberikan evaluasi terhadap solusi yang telah dipresentasikan oleh peserta didik dan memberikan kesempatan kepada peserta didik lain untuk memberikan tanggapan terhadap solusi yang telah dipresentasikan.

Dengan mencermati berbagai karakteristik dan langkah-langkah pembelajaran tersebut, pembelajaran berbasis masalah dipercaya mampu mengarahkan peserta didik, baik secara individu maupun kelompok, untuk belajar bagaimana belajar (*learning how to learn*) melalui proses kerja kolaboratif, komunikatif, dan kooperatif. Sebagai akibatnya, pengetahuan dan

keterampilan yang diperoleh peserta didik melalui model pembelajaran ini menjadi lebih mendalam dan melekat dalam kehidupan keseharian peserta didik tersebut.

c. Pembelajaran Berbasis Penemuan

Pembelajaran berbasis penemuan atau inkuiri (*discovery/inquiry-based learning*) adalah model pembelajaran yang mengutamakan aktivitas peserta didik untuk bertanya, mengembangkan gagasan, dan menganalisis permasalahan tersebut untuk mendapatkan bukti-bukti yang logis. Secara pedagogis, model pembelajaran ini dapat didefinisikan dari dua perspektif yang berbeda, yaitu menurut perspektif peserta didik dan perspektif guru. Dari sudut pandang peserta didik, pembelajaran berbasis inkuiri berfokus pada penyelidikan pertanyaan atau masalah terbuka. Peserta didik harus menggunakan penalaran berbasis bukti dan pemecahan masalah yang kreatif untuk mencapai kesimpulan yang harus mereka pertahankan atau presentasikan. Sedangkan dari sudut pandang guru, pembelajaran berbasis inkuiri berfokus pada menggerakkan peserta didik untuk memiliki rasa ingin tahu yang tinggi dengan menerapkan kemampuan berpikir kritis yang mereka miliki untuk menemukan pengetahuan baru. Guru harus mendorong peserta didik untuk mengajukan pertanyaan; membimbing mereka melakukan proses penyelidikan; mengarahkan mereka kapan harus memulai penyelidikan; dan bagaimana mereka merencanakan dan melakukan kegiatan penyelidikan. Apapun jenis aktivitas yang digunakan guru untuk menerapkan model pembelajaran ini, aktivitas itu harus memberikan kesempatan peserta didik mengembangkan strategi unik untuk memecahkan pertanyaan terbuka.

Langkah-langkah (sintaks) yang dapat digunakan oleh guru dan peserta didik untuk menerapkan model pembelajaran berbasis inkuiri ini adalah sebagai berikut:

1. Pemberian rangsangan

Guru mengawali proses pembelajaran dengan mengajukan pertanyaan, merekomendasikan buku yang harus dibaca peserta didik, dan menentukan aktivitas yang harus dilakukan peserta didik untuk memecahkan masalah. Peserta didik dihadapkan pada ketidaktahuannya berkaitan masalah yang diberikan dan kemudian

didorong untuk memunculkan rasa ingin tahu sebagai awal untuk melakukan penyelidikan. Stimulasi ini diberikan untuk menciptakan kondisi belajar agar peserta didik terdorong untuk melakukan eksplorasi berbagai bahan untuk melakukan penyelidikan.

2. Pernyataan masalah

Guru memberikan kesempatan kepada peserta didik untuk mengidentifikasi berbagai permasalahan yang relevan dengan bahan pelajaran, dan kemudian meminta mereka memilih salah satu untuk dicarikan jawaban sementara (hipotesis). Pada tahap ini peserta didik mencari jawaban sementara atas pertanyaan yang telah dipilih.

3. Pengumpulan data

Guru memberi kesempatan kepada peserta didik untuk mengumpulkan berbagai informasi atau data yang relevan untuk membuktikan hipotesis yang telah ditentukan sebelumnya dan peserta didik berusaha untuk membuktikan kebenaran hipotesis yang telah ditentukan melalui teknik wawancara, observasi, membaca referensi, melakukan eksperimen dan sebagainya.

4. Pengolahan data

Guru membantu peserta didik pada saat mereka melakukan pengolahan data dengan memberikan bimbingan dan feedback yang diperlukan. Pada tahap ini, peserta didik mengolah atau menganalisis data yang diperoleh melalui teknik wawancara, observasi, membaca referensi dan melakukan eksperimen dengan cara membuat tabulasi dan klasifikasi dari data tersebut.

5. Verifikasi atau pembuktian temuan

Guru mengarahkan peserta didik untuk menemukan suatu konsep, aturan dan pemahaman tertentu berangkat dari masalah yang telah dipilih. Sedangkan peserta didik melakukan verifikasi secara sistematis dan holistik untuk membuktikan kebenaran hipotesis yang telah dibuat sebelumnya.

6. Menarik simpulan/generalisasi

Peserta didik mengakhiri penyelidikan yang telah dilakukan dengan menarik sebuah simpulan dan memberikan justifikasi atas kinerja dan simpulan yang dibuat oleh peserta didik.

Selain ketiga model pembelajaran tersebut, pembelajaran HOTS dapat juga dilaksanakan dengan model-model pembelajaran berbasis kooperatif yang lain, misalnya *jigsaw*, *Student Team Achievement Divison* (STAD), *Think Pair Share* (TPS), dan sebagainya.

B. Penilaian Berbasis HOTS

Berpikir tingkat tinggi adalah kemampuan berpikir kritis, logis, reflektif, metakognitif, dan kreatif. Dalam kaitanya dengan pengukuran hasil belajar, HOTS dapat dipahami sebagai instrumen pengukuran yang digunakan untuk mengukur keterampilan berpikir tingkat tinggi, yaitu keterampilan berpikir yang tidak sekadar mengingat (*remember*), memahami (*undestand*), atau menerapkan (*apply*). Secara spesifik, Brookhart (2010) menyatakan bahwa tes berbasis HOTS adalah tes yang diarahkan agar peserta didik pada saat merespon atau menjawab butir-butir tes mampu:

1. mentransfer satu konsep ke konsep lainnya;
2. memproses dan menerapkan informasi;
3. mencari kaitan dari berbagai informasi yang berbeda-beda;
4. menggunakan informasi untuk menyelesaikan masalah; dan
5. menelaah ide dan informasi secara kritis.

Karakteristik tes seperti tersebut mengisyaratkan bahwa pengembangan tes berbasis HOTS tidak dapat dilakukan dengan serampangan, tetapi memerlukan perencanaan yang matang dengan memperhatikan indikator-indikator penting dari kemampuan berpikir tingkat tinggi.

Ada paling tidak empat indikator kemampuan berpikir tingkat tinggi yaitu, transfer pembelajaran (*transfer of learning*), berpikir kritis (*critical thinking*), pemecahan masalah (*problem solving*), dan berpikir logis (*logical thinking*). Transfer pembelajaran merupakan kemampuan peserta didik untuk memanfaatkan pengetahuan dan keterampilan yang telah dipelajari sebelumnya untuk diterapkan untuk memecahkan masalah pada konteks atau situasi pembelajaran baru. Sedangkan berpikir kritis mengacu pada proses intelektual yang dimiliki oleh peserta didik untuk secara aktif dan terampil dalam menkonseptualisasikan, menerapkan, menganalisis, mensintesis, dan mengevaluasi berbagai informasi yang dikumpulkan melalui observasi, pengalaman, refleksi, penalaran, dan komunikasi. Selanjutnya keterampilan pemecahan masalah mengacu pada kemampuan peserta didik untuk

memecahkan masalah secara efektif, tepat waktu dan lancar tanpa ada rintangan yang berarti. Keterampilan ini melibatkan kemampuan peserta didik untuk mengidentifikasi dan mendefinisikan masalah, menghasilkan solusi alternatif, mengevaluasi dan memilih alternatif solusi terbaik, dan menerapkan solusi terpilih untuk permasalahan kontekstual. Di lain pihak, berpikir logis adalah kemampuan untuk menganalisis situasi dan menghasilkan solusi yang masuk akal terhadap suatu permasalahan tertentu. Sama seperti berpikir kritis, berpikir logis juga membutuhkan keterampilan bernalar untuk mempelajari suatu masalah secara objektif, yang memungkinkan peserta didik mampu membuat kesimpulan rasional.

Satu hal yang harus dipahami bahwa soal-soal yang sulit untuk dikerjakan oleh peserta didik tidak selalu identik dengan soal HOTS. Ini berarti bahwa butir-butir tes yang mudah dikerjakan oleh peserta didik pun dapat dikategorikan sebagai soal HOTS apabila butir-butir tes tersebut menghendaki peserta didik untuk memanfaatkan kemampuan untuk mentransfer berbagai informasi, memecahkan masalah, berpikir kritis dan berpikir logis ketika mereka menjawab atau merespon butir-butir tes itu. Sebagai contoh dalam mata pelajaran bahasa Inggris, untuk mengetahui arti dari kata tertentu yang jarang digunakan mungkin sulit, tetapi ini bukanlah *higher-order thinking*, kecuali untuk mengetahui arti kata tersebut melibatkan proses bernalar (seperti misalnya mencari arti dari kata tersebut berdasarkan konteks dimana kalimat tersebut digunakan dan/atau menebak makna kata tersebut berdasarkan stimulus tertentu).

1. Level Kognitif

Ketika mengembangkan tes untuk mengukur hasil belajar peserta didik, guru harus memperhatikan level kognitif peserta didik, karena level kognitif tersebut akan mempengaruhi indikator soal sebagai tolok ukur pencapaian kompetensi oleh peserta didik. Level kognitif merupakan tingkat kemampuan peserta didik secara individual maupun kelompok yang dapat dijabarkan ke dalam tiga tingkatan kognitif, yaitu level 1, level 2, dan level 3. Ketiga level tersebut memiliki karakteristik dan keunikan masing-masing yang mengacu pada tingkat berpikir peserta didik.

a) Level 1 (pengetahuan dan pemahaman)

Pada level ini, tes dimaksudkan untuk mengukur kompetensi peserta didik dengan tingkat kesulitan rendah karena pada level ini butir-butir tes

hanya menghendaki peserta didik untuk mengingat dan memahami suatu fakta, konsep, prinsip, dan prosedur. Secara rinci, butir-butir tes pada level 1 dimaksudkan mengukur kompetensi peserta didik untuk:

1. menunjukkan kemampuan untuk mengingat dan memahami materi yang telah dipelajari dan membuat generalisasi sederhana;
2. menunjukkan kemampuan pemecahan masalah pada tingkat dasar, paling tidak pemecahan masalah tersebut dilakukan dengan satu cara;
3. menunjukkan kemampuan dasar dalam membaca grafik, tabel, dan materi visual lainnya;
4. mengkomunikasikan fakta-fakta dasar dengan menggunakan beberapa terminology sederhana.

Contoh butir tes pada level 1:

Di antara *eubacteria* berikut yang dapat menimbulkan sakit perut (diare) pada manusia adalah....

- A. *Pseudomonas sp*
- B. *Thiobacillus ferrooxidans*
- C. *Clostridium botulinum*
- D. *Escherichia coli*
- E. *Acetobacter xylinum*

(Butir tes diadopsi dari Dit. PSMA, 2018)

b) Level 2 (penerapan)

Tes pada level 2 menghendaki peserta didik untuk menggunakan pengetahuan faktual, konsep, dan prosedural tertentu pada konsep lain dalam mata pelajaran yang sama atau mata pelajaran lain dan juga menggunakan pengetahuan faktual, konsep, dan prosedural tertentu untuk menyelesaikan masalah kontekstual. Butir-butir tes level 1 diarahkan agar peserta didik mampu:

1. menunjukkan pengetahuan dan keterampilan yang telah dikuasai dan mengaplikasikannya dalam konteks tertentu;
2. menginterpretasikan dan menganalisis suatu data dan informasi yang digunakan dalam sebuah tes secara mendalam;
3. memecahkan suatu permasalahan sederhana dalam suatu materi pelajaran;

4. memahami dan mengentepretasikan grafik, tabel, gambar, dan materi visual lainnya;
5. mengkomunikasikan suatu fakta dengan cara menjelaskannya dengan menggunakan analogi atau terminology lain.

Contoh butir tes level 2:

Jumlah uang yang beredar di masyarakat sebesar Rp 100 milyar, tingkat harga umum yang berlaku Rp 200.000,00 dan jumlah barang yang diperdagangkan 5.000.000 unit, maka kecepatan uang yang beredar menurut teori kuantitas Irving Fisher adalah

- A. 5 kali
- B. 10 kali
- C. 50 kali
- D. 100 kali
- E. 1000 kali

(Butir tes diadopsi dari Dit. PSMA, 2018)

c) Level 3 (penalaran)

Tes padal level ini dimaksudkan untuk mengukur kompetensi peserta didik dalam menggunakan kemampuan bernalar dan berlogika untuk mengambil keputusan, meprediksi dan merefleksi, dan menyusun strategi memecahkan masalah tertentu. Secara rinci, butir-butir tes level 3 dimaksudkan untuk mengukur kompetensi peserta didik untuk mampu:

1. menunjukkan penguasaan dan pemahaman yang luas terhadap materi pelajaran dan merapkannya dalam berbagai konteks situasi yang berbeda.
2. menganalisis, mensintesis, dan mengevaluasi berbagai berbagai gagasan dan informasi faktual.
3. menjabarkan hubungan antara informasi dan konsep ke dalam situasi faktual.
4. menafsirkan berbagai informasi kompleks pada materi pembelajaran.
5. menemukan solusi dari berbagai masalah yang ada dan mengaitkan dengan banyak variabel.
6. mendemonstrasikan berbagai ide dan pemikiran orisinal.

Contoh butir tes level 3:

BOTOL MINUMAN



Gambar disamping menunjukkan 2 buah botol minuman. Kapasitas botol A 1500 ml dan kapasitas botol B 1000 ml. Seseorang mempunyai sebuah ember yang berisi air sebanyak 6 liter. Air tersebut akan dimasukkan ke dalam dua jenis botol seperti pada gambar di samping tanpa sisa.

Manakah kemungkinan banyak masing-masing botol yang dapat diisi air tersebut.

- A. 2 Botol A dan 4 Botol B
- B. 2 Botol A dan 3 Botol B
- C. 4 Botol A dan 0 Botol B
- D. 3 Botol A dan 2 Botol B
- E. 0 Botol A dan 6 Botol B

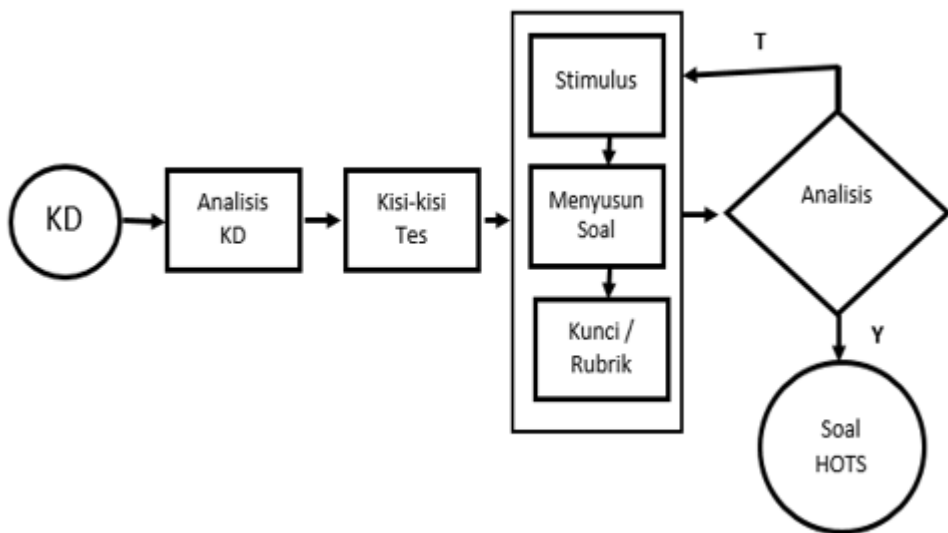
(Butir tes diadopsi dari Dit. PSMA, 2018)

Dari ketiga level tersebut, level 3 merupakan model penilaian berbasis HOTS karena butir-butir tes pada level 3 menghendaki peserta didik untuk mampu menganalisis, membandingkan, menginterpretasi, dan memecahkan masalah. Dengan demikian, ketika mengembangkan tes berbasis HOTS, guru harus memperhatikan berbagai karakteristik tes berbasis HOTS agar setiap butir yang dikembangkan benar-benar dapat mengukur kemampuan berpikir tingkat tinggi. Secara umum, karakteristik tes berbasis HOTS adalah tes tersebut menghendaki peserta didik untuk mentransfer satu konsep ke konsep lainnya; memproses dan menerapkan informasi; mencari kaitan dari berbagai informasi yang berbeda-beda; menggunakan informasi untuk menyelesaikan masalah; dan menelaah ide dan informasi secara kritis.

2. Langkah-langkah Pengembangan Tes Berbasis HOTS

Ada beberapa langkah atau tahapan untuk mengembangkan tes berbasis HOTS. Langkah-langkah ini memberikan panduan praktis bagi guru sebagai pengembang soal dalam mengembangkan tes berbasis HOTS untuk mengukur kemampuan berpikir tingkat tinggi. Langkah-langkah tersebut dilukiskan seperti dalam gambar 10.2. Gambar tersebut melukiskan tahapan-tahapan pengembangan butir-butir tes yang bersifat hirarkis dan

saling berkaitan. Aktivitas yang dilakukan pada tahap sebelumnya akan mempengaruhi kualitas pada tahap berikutnya. Oleh karena itu, pengembang atau penyusun butir-butir tes berbasis HOTS harus melakukan perencanaan secara cermat pada setiap tahapan sehingga pada akhirnya akan dihasilkan butir-butir tes yang berkualitas dan benar-benar mampu mengukur kemampuan berpikir tingkat tinggi yang dimiliki oleh peserta didik.



Gambar 10.2: Alur pengembangan tes berbasis HOTS

a. Menganalisis Kompetensi Dasar (KD)

Kompetensi dasar (KD) merupakan kemampuan dan materi pembelajaran minimal yang harus dicapai peserta didik untuk suatu mata pelajaran pada masing-masing satuan pendidikan yang mengacu pada kompetensi inti. Setiap KD memiliki kata kerja operasional tersendiri dan cakupan materi yang berbeda-beda pula. Setiap KD paling tidak memuat dua komponen utama, yaitu kemampuan dan materi pokok. Kemampuan mengacu pada taksonomi Bloom yang terdiri dari enam dimensi, yaitu pengetahuan (C1), pemahaman (C2), menerapkan (C3), menganalisis (C4), mengevaluasi (C5) dan mencipta (C6). Selanjutnya, semua materi kemudian dikelompokkan menjadi beberapa lingkup materi. Lingkup materi disusun berdasarkan hasil analisis KD pada mata pelajaran tertentu sesuai kurikulum

yang berlaku. Materi yang tercakup dalam setiap lingkup materi dipetakan ke dalam tiga level kognitif yang berbeda, yaitu pengetahuan dan pemahaman (C1 dan C2), aplikasi (C3), dan penalaran (C4, C5, atau C6). Selanjutnya, untuk mengembangkan indikator tes berbasis HOTS, dipilih level kognitif penalaran (C4, C5, atau C6).

b. Menyusun Kisi-Kisi

Kisi-kisi soal merupakan sebuah format berupa matriks yang memuat pedoman dan kriteria untuk menulis tes atau merakit tes menjadi suatu instrumen penilaian. Sebuah kisi-kisi yang baik paling tidak memuat kompetensi dasar, lingkup materi yang akan diujikan, level kognitif, indikator butir-butir tes, bentuk tes, dan nomor tes. Pengembangan tes berbasis HOTS menekankan pada level kognitif penalaran (level 3) yang terdiri tiga dimensi kemampuan utama yang harus dikuasai oleh peserta didik, yaitu menganalisis, mengevaluasi dan mencipta. Kemampuan menganalisis mengacu pada kemampuan peserta didik untuk menguraikan suatu masalah menjadi bagian-bagian yang spesifik, mengelompokkan informasi, membandingkan, menentukan hubungan antara satu kelompok/informasi dengan kelompok/informasi lainnya. Sedangkan kemampuan mengevaluasi merupakan kemampuan untuk menyimpulkan informasi berdasarkan suatu kriteria, memprediksi, hipotesa. Selanjutnya, kemampuan mencipta adalah keterampilan untuk membuat sesuatu yang baru dari apa yang sudah ada sehingga hasil tersebut merupakan satu kesatuan utuh dan berbeda dari komponen yang digunakan untuk membentuknya.

Indikator butir tes menggambarkan kompetensi peserta didik yang akan diuji sesuai dengan level kognitif dan materi. Dari satu indikator dapat disusun beberapa butir tes yang paralel. Sebuah indikator yang baik harus memuat kompetensi peserta didik yang dapat diamati (*observable*) dan diukur (*measurable*). Selain itu, sebuah indikator harus memuat subjek (peserta didik), perilaku yang akan diukur (misalnya membandingkan, menganalisis, memecahkan masalah, memprediksi, menginterpretasi, menyimpulkan, merumuskan, dll). Berikut adalah contoh indikator yang dapat digunakan untuk mengembangkan tes berbasis HOTS:

- Disajikan sebuah teks tentang pencemaran lingkungan, peserta didik dapat memprediksi dampak pencemaran lingkungan yang dapat mempengaruhi kualitas hidup masyarakat sekitar.
- Disajikan lima pernyataan tentang kerawanan sosial, peserta didik dapat merancang strategi yang tepat untuk mengatasi dampak kerawanan sosial terhadap kehidupan sosial kemasyarakatan.

Kedua indikator tersebut secara eksplisit memuat kompetensi yang dapat diamati dan terukur, memuat subjek (peserta didik) yang akan diukur kompetensinya, memuat perilaku yang akan diukur (memprediksi dan merancang).

c. Menentukan Stimulus

Stimulus adalah pernyataan yang menimbulkan masalah dan merangsang pemikiran dan pencarian informasi. Selain berupa pernyataan, stimulus dapat juga berupa gambar, grafik, tabel, teks, dan sebagainya. Dalam memilih dan menentukan stimulus untuk tes berbasis HOTS, guru perlu mempertimbangkan beberapa kriteria sebagai berikut:

1. Memuat satu atau beberapa informasi, dapat berupa gambar, grafik, tabel, wacana, dll yang memiliki keterkaitan dalam sebuah kasus sesuai dengan lingkup materi yang diujikan.
2. Menuntut kemampuan menginterpretasi, mencari hubungan, menganalisis, menyimpulkan, memprediksi, atau menciptakan.
3. Bersifat kontekstual dan menarik (terkini) untuk memotivasi peserta didik untuk membaca. Pengecualian untuk mata pelajaran bahasa dan sejarah boleh tidak kontekstual.
4. Terkait langsung dengan pertanyaan (pokok soal).
5. Stimulus berisi informasi yang bermanfaat untuk menjawab soal dan sebagai dasar menyusun pertanyaan.

d. Menulis Butir-butir Tes

Setelah selesai menganalisis KD, menyusun kisi-kisi dan menentukan stimulus, tahap selanjutnya adalah menuliskan butir-butir tes. Ada beberapa tips untuk menulis butir-butir tes berbasis HOTS, yaitu:

1. Gunakan konteks faktual yang mudah dipahami oleh peserta didik.
Tulislah butir-butir tes yang mengacu pada fenomena faktual dan

ada di sekitar peserta didik untuk memudahkan mereka dalam memahami konteks yang menaungi butir-butir tes tersebut.

2. Tulislah butir-butir pertanyaan berdasarkan grafik, tabel, gambar, bagan dan sebagainya yang dapat merangsang peserta didik untuk berpikir kritis dan analitis.
3. Apabila butir-butir pertanyaan menghendaki jawaban singkat dari peserta didik, tanyakan pula alasan peserta didik mengapa menjawab demikian. Misalnya ada pertanyaan yang menghendaki peserta didik untuk merespon setuju atau tidak setuju, maka tanyakan mengapa mereka menyatakan setuju atau tidak setuju.
4. Gunakan variasi butir-butir tes yang variatif agar sikap kritis dan kemampuan bernalar peserta didik dapat diungkap yang luas dan mendalam.

Berikut adalah beberapa contoh butir-butir tes berbasis HOTS untuk beberapa mata pelajaran, baik bentuk tes pilihan ganda maupun uraian.

- Soal mata pelajaran matematika: uraian

Ahmad adalah peserta didik lulusan SMA yang ingin melanjutkan ke Fakultas Pertanian IPB. Syarat untuk dapat diterima di Fakultas Pertanian IPB harus lulus tes Matematika dengan nilai tidak kurang dari 70 dan tes Biologi dengan nilai tidak kurang dari 50, serta jumlah nilai Matematika dan Biologi tidak boleh kurang dari 130.

Ternyata jumlah nilai Ahmad untuk dua kali nilai Matematika dan tiga kali nilai Biologi sama dengan 300.

Apakah Ahmad diterima atau ditolak di Fakultas Pertanian IPB? Jelaskan jawaban Anda!

(Butir tes diadopsi dari Dit. PSMA, 2018)

- Soal mata pelajaran olah-raga: pilihan ganda



Pada perlombaan seni pencak silat, empat orang juri memberikan tiga kriteria penilaian pada peserta lomba dengan interval skor 1-10. Hasil penilaian tampak pada tabel berikut

Nama Atlit	Juri 1			Juri 2			Juri 3		
	K.1	K.2	K.3	K.1	K.2	K.3	K.1	K.2	K.3
Toni	6	8	9	7	7	8	8	8	4
Rizki	6	8	8	6	7	7	6	7	5
Yanto	7	7	8	8	8	7	6	7	7
Bowo	7	8	8	8	8	7	9	9	6

Kriteria Penilaian :

- K.1 = Serangan
- K.2 = Pukulan
- K.3 = Tendangan

Skor akhir adalah rata-rata tertinggi dari ketiga kriteria. Disyaratkan juga pemenang tidak memiliki skor kurang dari 6.

Berdasarkan aturan tersebut, siapakah yang menjadi pemenang?

- A. Toni
- B. Rizki
- C. Yanto
- D. Bowo

(Butir tes diadopsi dari Dit. PSMA, 2018)

- Soal mata pelajaran bahasa Inggris: uraian

Kancil and Crocodile

Kancil was a clever mousedeer. He had many enemies. One of them was Crocodile. Crocodile lived in a river in the forest. Now, one day, Kancil went to the river. It was a very hot day, and he wanted to have a bath. Kancil bathed and splashed about in the water.

Crocodile saw Kancil. "A nice meal," he thought. Then, he crawled behind Kancil and grabbed him. He caught one of Kancil's legs.

Kancil was terrified. Then, he had an idea. He saw a twig floating near him. He picked it up and said, "You stupid fool! So you think you've got me. You're biting a twig - not my leg. Here, this is my leg."

And with that, he showed Crocodile the twig. Crocodile could not see well. He was a very stupid creature, too. He believed the cunning mousedeer. He freed the mousedeer's leg and snapped upon the twig. Kancil ran out of the water immediately.

"Ha! Ha!" he laughed. "I tricked you!"

1. In what ways are Kancil and Crocodile different?
2. Do you think Kancil has done the right thing? Why?
3. Compose a letter of apology from Kancil to Crocodile.

(Butir tes diadopsi dari Dit. PSMA, 2018)

Ketiga butir pertanyaan tersebut merupakan tes berbasis HOTS dan digunakan untuk mengukur tiga dimensi yang berbeda. Pertanyaan no. 1 adalah butir tes untuk mengukur kemampuan 'menganalisis' pesan yang ada dalam teks. Sedangkan pertanyaan no. 2 adalah butir tes untuk mengukur kemampuan 'mengevaluasi' pesan dalam teks dan pertanyaan no. 3 merupakan butir tes yang dimaksudkan untuk mengukur kemampuan peserta didik dalam 'menciptakan' suatu teks baru.

- Soal bahasa Indonesia: uraian

KAKEK DAN PENCURI PEPAYA



Cerita ini tentang seorang kakek yang sederhana, hidup sebagai orang kampung yang bersahaja. Suatu sore, ia mendapati pohon pepaya di depan rumahnya telah berbuah. Walaupun hanya dua buah namun telah menguning dan siap dipanen. Ia berencana memetik buah itu di keesokan hari. Namun, tatkala pagi tiba, ia mendapati satu buah pepayanya hilang dicuri orang.

Kakek itu begitu bersedih, hingga istrinya merasa heran. “Suamiku, jangan hanya karena sebuah pepaya saja engkau demikian murung” ujar sang istri. “Bukan itu yang aku sedihkan,” jawab sang kakek. “Aku berpikir, betapa sulitnya orang itu mengambil pepaya kita. Ia harus sembunyi-sembunyi di tengah malam agar tidak ketahuan orang. Belum lagi mesti memanjatnya dengan susah payah untuk bisa memetik pepaya.”

“Oleh karena itu istriku...,” lanjut sang kakek. “Saya akan meminjam tangga dan saya taruh di bawah pohon pepaya kita. Mudah-mudahan ia datang kembali malam ini dan tidak akan kesulitan lagi mengambil pepaya yang satunya.” Namun saat pagi kembali hadir, ia mendapati pepaya yang tinggal sebuah itu tetap ada beserta tangganya tanpa bergeser sedikitpun. Sang Kakek tetap menunggu. Namun di pagi berikutnya, tetap saja buah pepaya itu masih di tempatnya.

Di sore harinya, sang kakek kedatangan seorang tamu yang menenteng dua buah pepaya besar di tangannya. Sang kakek belum pernah mengenal si tamu tersebut. Singkat cerita, setelah berbincang lama, sang tamu dengan amat menyesal mengaku bahwa dialah yang telah mencuri pepayanya.

“Sebenarnya, di malam berikutnya saya ingin mencuri buah pepaya yang tersisa. Namun saat saya menemukan ada tangga di sana, saya tersadarkan dan sejak itu saya bertekad untuk tidak mencuri lagi. Untuk itu, saya kembalikan pepaya Anda dan untuk menebus kesalahan saya, saya hadiahkan pepaya yang baru saya beli di pasar untuk Anda.”

1. Ani berpendapat bahwa sifat Kakek tersebut dermawan. Setujukah kamu dengan pendapat Ani tersebut? Jelaskan alasanmu!
2. Apakah pendapatmu jika pada cerita tersebut si pencuri tetap mengambil pepaya milik Kakek yang kedua?
3. Apakah yang membuat perasaan Kakek sedih setelah menyadari satu buah pepaya miliknya hilang?
4. Apakah yang dilakukan sang pencuri untuk menebus kesalahannya?

Pertanyaan no. 1 dan no. 2 merupakan butir-butir tes berbasis HOTS karena untuk menjawab dua pertanyaan tersebut peserta didik perlu menggunakan kemampuan berpikir kritis, mentransfer informasi, dan bernalar. Sedangkan pertanyaan no. 3 dan no. 4 bukan merupakan butir-butir tes berbasis HOTS karena untuk mampu menjawab dua butir tersebut peserta didik cukup mengandalkan kemampuan untuk mengingat dan memahami isi teks.

e. Membuat Kunci Jawaban dan Rubrik Penskoran

Segera setelah butir-butir tes selesai dibuat, guru sebagai pengembang tes harus segera membuat kunci jawaban dan rubrik penskoran. Kunci jawaban memiliki peran penting karena kunci jawaban dapat digunakan oleh guru sebagai pedoman untuk memutuskan apakah jawaban yang diberikan oleh peserta didik dikategorikan benar atau salah. Rubrik merupakan panduan penilaian yang memuat berbagai kriteria yang diinginkan oleh pengembang tes untuk melakukan penilaian terhadap hasil pekerjaan peserta didik dan biasanya digunakan untuk tes bentuk subjektif (uraian, proyek, tugas, dan sebagainya). Setiawan (2020) menjelaskan bahwa rubrik memiliki manfaat yang penting dalam keseluruhan proses penilaian kinerja peserta didik. Adapun manfaat atau kegunaan dari penilaian rubrik ialah sebagai berikut:

- a. Rubrik menjelaskan deskripsi tugas.
- b. Rubrik memberikan informasi bobot penilaian.
- c. Peserta didik memperoleh umpan balik yang cepat dan akurat.
- d. Penilaian lebih objektif dan konsisten.
- e. Para peserta didik jadi pembelajaran aktif.
- f. Para peserta didik memperoleh “*content knowledge*” dan “*procedural knowledge*”.
- g. Para peserta didik dapat menilai kinerja kelompoknya sendiri.
- h. Baik pendidik maupun peserta didik memperoleh alat refleksi yang efektif tentang proses pembelajaran yang telah berlangsung.
- i. Sebagai alat atau pedoman penilaian kinerja atau hasil kerja mahasiswa didik.

Rubrik biasanya dibuat dalam bentuk tabel dua lajur, yaitu baris yang berisi kriteria dan kolom yang berisi mutu. Kriteria dapat dinyatakan secara garis

besar. Kemudian dirinci menjadi komponen-komponen penting atau dapat pula komponen-komponen ditulis langsung tanpa dikelompokkan dalam garis besar.

f. Menganalisis Butir-butir Tes

Analisis butir-butir tes dimaksudkan untuk mengetahui kualitas butir-butir tes yang telah dikembangkan. Hasil analisis butir tes akan memberi gambaran kepada guru sebagai pengembang tes apakah pertanyaan-pertanyaan telah ditulis dengan baik atau masih ada beberapa kesalahan; juga untuk mengetahui apakah ada masalah dalam memahami instruksi. Apabila hasil analisis telah menunjukkan bahwa butir-butir tes telah dikonstruksi dengan baik dan instruksi telah mudah dipahami, maka tes tersebut telah siap digunakan untuk mengukur hasil belajar peserta didik. Tetapi ketika ketika hasil analisis masih ditemukan beberapa kesalahan dan instruksi masih belum jelas, maka pengembang tes berbasis HOTS harus kembali ke tahapan-tahapan sebelumnya untuk melakukan revisi stimulus yang ada pada butir-butir tertentu, kemudian butir-butir tes ditulis kembali untuk perbaikan. Termasuk pada akhirnya memperbaiki kunci jawaban dan rubrik penilaian hingga akhirnya dihasilkan butir-butir tes berbasis HOTS yang berkualitas.

Bab 11

Pelaporan Hasil Penilaian

Setelah tes dilaksanakan dan dikerjakan oleh peserta didik, tahap selanjutnya adalah guru memeriksa atau mengoreksi lembar jawab yang dikumpulkan oleh peserta didik dalam rangka menentukan skor dan nilai yang dicapai oleh peserta didik itu. Skor dan nilai itu kemudian oleh guru dicatat sebagai dokumen hasil belajar peserta didik. Dokumen itu kemudian dilaporkan kepada semua pihak yang berkepentingan yang berisi skor dan nilai sebagai representasi hasil belajar peserta didik tersebut. Dengan kata lain, hasil penilaian harus dilaporkan kepada pihak-pihak yang berkepentingan (peserta didik, orang tua, guru, masyarakat) dengan tujuan untuk memantau (*monitoring*) dan memberikan umpan balik (*feedback*) terhadap proses belajar peserta didik. Laporan hasil penilaian juga dapat digunakan oleh guru untuk mendapatkan informasi dalam rangka meningkatkan proses pembelajaran dan pada gilirannya laporan penilaian menjadi bagian integral untuk menetapkan harapan berkaitan dengan prestasi belajar yang ingin dicapai oleh peserta didik.

A. Hakikat dan Tujuan Pelaporan Hasil Penilaian

Pelaporan hasil penilaian dapat dipahami sebagai sebuah proses untuk mengkomunikasikan pengetahuan dan keterampilan peserta didik yang didapatkan dari penilaian hasil belajar. Pelaporan ini dimaksudkan untuk memberikan informasi yang relevan berkaitan dengan kemajuan belajar peserta didik agar diketahui oleh peserta didik itu sendiri, staf administrasi, orang tua, dan *stakeholders* lainnya. Pelaporan hasil penilaian dapat pula dipahami sebagai pemanfaatan hasil penilaian yang relevan untuk memastikan bahwa aktivitas belajar-mengajar selama ini telah sesuai secara

maksimal dengan pengetahuan pengetahuan dan tingkat pembelajaran peserta didik saat ini.

Laporan hasil belajar peserta didik harus dapat menggambarkan pencapaian kompetensi peserta didik pada semua mata pelajaran. Peraturan Pemerintah No. 19 tahun 2005 pasal 25 ayat (4) menjelaskan bahwa laporan hasil penilaian mencakup pengetahuan, keterampilan, dan sikap. Oleh karena itu, penilaian hasil belajar harus mencerminkan ketiga aspek kompetensi dimaksud dengan mempertimbangkan karakteristik masing-masing mata pelajaran. Hasil penilaian oleh pendidik dan satuan pendidikan disampaikan dalam bentuk satu nilai pencapaian kompetensi mata pelajaran untuk masing-masing nilai pengetahuan dan nilai praktik sesuai dengan karakteristik kompetensi mata pelajaran yang bersangkutan, serta kualifikasi untuk kondisi afektif/sikap, disertai dengan deskripsi kemajuan belajar/ketercapaian kompetensi peserta didik sebagai pencerminan kompetensi secara utuh. Setiap akhir semester, satuan pendidikan perlu melaporkan hasil belajar peserta didik kepada orangtua/wali peserta didik sehingga mereka dapat memantau kemajuan belajar anak-anaknya yang pada akhirnya orang tua dapat membantu dan mendorong proses belajar mereka dan pada gilirannya putra-putrinya dapat mencapai hasil belajar yang optimal. Laporan hasil penilaian yang biasa juga disebut laporan hasil belajar (LHB) peserta didik dapat berbentuk buku atau lembaran dan dapat ditulis secara manual atau komputerisasi.

Boggs (2020) menjelaskan bahwa laporan hasil penilaian proses dan hasil belajar meliputi aspek kognitif (pengetahuan), psikomotorik (keterampilan) dan afektif (sikap). Tetapi, tidak semua mata pelajaran dinilai aspek psikomotoriknya, yang dinilai aspek psikomotoriknya hanya mata pelajaran yang melakukan kegiatan praktik. Informasi ranah kognitif dan psikomotor diperoleh dari sistem penilaian yang digunakan untuk mata pelajaran yang sesuai dengan tuntutan kompetensi dasar. Informasi ranah afektif diperoleh melalui kuesioner, inventori, dan pengamatan yang sistematis. Aspek kognitif dan afektif dinilai untuk seluruh mata ajar. Informasi aspek kognitif dan psikomotor diperoleh melalui sistem penilaian sesuai dengan tuntutan indikator-indikator dari kompetensi dasar yang telah ditetapkan. Aspek afektif diperoleh melalui lembar pengamatan yang sistematis, kuesioner dan inventori. Hasil penilaian kognitif dan psikomotorik dapat berupa nilai angka maupun deskripsi kualitatif mengenai

kompetensi dasar tertentu. Sedangkan deskripsi kualitatif dapat dilaporkan dalam bentuk deskripsi mengenai kompetensi dasar tertentu dari mata pelajaran tersebut. Pelaporan hasil inventori afektif ini akan sangat bermanfaat khususnya untuk mengetahui sikap dan minat peserta didik terhadap pelajaran tertentu dan hasilnya dapat dimanfaatkan untuk memperbaiki sikap serta minat peserta didik terhadap pembelajaran tersebut. Sedangkan pelaporan ranah afektif dilakukan secara kualitatif.

Setelah laporan hasil penilaian didokumentasikan secara lengkap, selanjutnya hasil penilaian tersebut disampaikan kepada peserta didik, orang tua, sekolah, dan masyarakat. Laporan yang berisi catatan tentang peserta didik diusahakan selengkap mungkin agar dapat memberikan informasi yang lengkap pula. Akan tetapi, setiap saat membuat laporan yang lengkap merupakan beban yang berat bagi seorang guru. Oleh karena itu, pembuatan laporan dapat pula bersifat singkat, disesuaikan dengan kebutuhan. Laporan yang dibuat guru untuk peserta didik dan orang tua berisi catatan prestasi belajar peserta didik. Catatan itu dapat dibedakan atas dua cara, yaitu lulus atau belum lulus dan berhasil atau belum berhasil. Prestasi yang dilaporkan guru kepada peserta didik dan orang tua dapat dilihat dalam buku rapor yang diisi pada setiap semester. Guru juga harus membuat laporan untuk sekolah, sebagai lembaga yang bertanggung jawab terhadap berlangsungnya proses belajar-mengajar. Oleh karena itu, pihak sekolah berkepentingan untuk mengetahui catatan perkembangan peserta didik yang ada di dalamnya. Dengan demikian hasil belajar peserta didik akan diperhatikan dan dipikirkan oleh pihak sekolah. Laporan yang dibuat guru untuk pihak sekolah sebaiknya lebih lengkap. Guru tidak semata-mata melaporkan prestasi peserta didik tetapi juga menyinggung problem kepribadian mereka. Laporan tidak hanya dalam bentuk angka tapi juga dalam bentuk deskripsi tentang peserta didik. Pada umumnya laporan untuk masyarakat berkaitan dengan jumlah lulusan sekolah. Setiap peserta didik yang telah lulus membawa bukti bahwa mereka memiliki suatu pengetahuan dan keterampilan tertentu. Tetapi pengetahuan dan keterampilan yang diperoleh peserta didik dari suatu sekolah tidaklah sama. Tingkat keberhasilan ini dinyatakan secara lengkap dalam laporan prestasi.

B. Manfaat Pelaporan Hasil Penilaian

Pelaporan hasil penilaian dapat dimanfaatkan oleh peserta didik, orang tua, guru, dan masyarakat untuk memperkuat motivasi belajar peserta didik; memperkuat daya ingat dan meningkatkan kemampuan transfer hasil belajarnya; memperbesar pemahaman peserta didik terhadap dirinya; dan memberikan umpan balik terhadap efektivitas program pembelajaran. Selanjutnya Boggs (2020) menyatakan bahwa manfaat hasil penilaian untuk pihak-pihak yang berkepentingan diuraikan dalam bagian berikut.

1. Manfaat Pelaporan Hasil Penilaian untuk Peserta Didik

Melaporkan hasil belajar kepada peserta didik dapat dilakukan oleh guru setiap saat (pada saat proses pembelajaran sedang berlangsung atau setelah proses pembelajaran selesai dilaksanakan). Proses pelaporan ini dapat dilakukan oleh guru, baik secara lisan maupun tertulis. Laporan kepada peserta didik meliputi laporan kemajuan belajar peserta didik, kelebihan dan kekurangan proses dan hasil belajar mereka dan juga tindakan apa yang akan diambil untuk proses pembelajaran berikutnya. Untuk mendapatkan informasi mengenai hasil belajar peserta didik pada ranah kognitif dan psikomotor dilakukan melalui tes, sedangkan ranah afektif dilakukan melalui angket, inventori, dan pengamatan Selanjutnya, pelaporan hasil penilaian ini dapat dimanfaatkan oleh peserta didik untuk:

- a) mengetahui kemajuan hasil belajar diri;
- b) mengetahui konsep-konsep atau teori yang belum dikuasai;
- c) memotivasi diri untuk belajar lebih baik; dan
- d) memperbaiki strategi belajar.

Guna memberikan informasi yang akurat berkaitan dengan proses dan hasil belajar peserta didik, laporan hasil penilaian yang harus disampaikan kepada peserta didik harus mencakup berbagai hal yang berisi tentang:

- a) hasil pencapaian belajar peserta didik;
- b) kekuatan dan kelemahan peserta didik dalam semua mata pelajaran, dan
- c) minat peserta didik pada masing-masing mata pelajaran.

2. Manfaat Pelaporan Hasil Penilaian untuk Orang Tua

Selain dimanfaatkan oleh peserta didik sendiri, laporan hasil belajar juga dimanfaatkan sepenuhnya oleh orang tua dari peserta didik. Orang tua dari peserta didik tersebut sangat memerlukan laporan hasil belajar putra-

putrinya sehingga orang tuanya dapat memantau dan mengetahui sejauhmana kemajuan belajar putra-putrinya. Dengan mengetahui laporan hasil belajar yang diperoleh oleh putra-putrinya, orang tua dari peserta didik tersebut diharapkan dapat bekerja-sama secara sinergis dengan pihak sekolah dalam rangka memotivasi, mengawasi dan memberikan bantuan terhadap proses belajar peserta didik dan pada akhirnya prestasi belajar peserta didik dapat dicapai secara optimal. Secara konkrit, informasi hasil belajar dimanfaatkan oleh orang tua untuk memotivasi anak agar belajar lebih baik. Untuk itu diperlukan informasi yang akurat tentang hasil belajar peserta didik, yang meliputi ranah kognitif, psikomotor, dan afektif. Informasi ini digunakan orang tua untuk:

- a) membantu anaknya belajar;
- b) memotivasi anaknya belajar;
- c) membantu sekolah meningkatkan hasil belajar peserta didik, dan
- d) membantu sekolah melengkapi fasilitas belajar.

Agar orang tua mampu secara sinergis membantu proses belajar putra-putrinya, bentuk laporan hasil belajar harus dibuat selengkap mungkin dan rinci berkaitan dengan kekurangan dan kelebihan proses belajar peserta didik; capaian pengetahuan, keterampilan dan sikap peserta didik; dan bakat-minat yang dimiliki oleh peserta didik tersebut.

3. Manfaat Pelaporan Hasil Penilaian untuk Guru dan Kepala Sekolah

Guru merupakan pihak yang secara langsung dapat memanfaatkan laporan penilaian hasil belajar peserta didik. Dengan laporan belajar, guru dapat menentukan apakah metode pembelajaran yang digunakan selama ini perlu dipertahankan, diperbaiki, atau bahkan diganti dengan metode pembelajaran yang baru. Apabila metode tersebut sama sekali tidak efektif untuk memfasilitasi proses belajar peserta didik, maka guru harus memutuskan untuk mencari dan mengganti metode pembelajaran yang lain. Tetapi apabila metode yang digunakan selama ini telah efektif untuk memfasilitasi proses belajar peserta didik, maka guru dapat mempertahankan metode tersebut. Laporan hasil belajar juga dapat digunakan oleh guru pengganti apabila guru yang pertama pindah tugas atau peserta didik naik kelas dan diajar oleh guru yang baru. Guru pengganti dapat memanfaatkan laporan hasil belajar untuk menentukan strategi pembelajaran yang tepat dalam rangka mempertahankan dan/atau

meningkatkan prestasi belajar yang telah dicapai oleh peserta didik selama ini.

Laporan hasil belajar untuk guru dan kepala sekolah harus mencakup hasil belajar dalam semua ranah untuk semua mata pelajaran. Informasi yang diperlukan adalah kompetensi dasar yang telah dikuasai dan yang belum dikuasai oleh peserta didik. Guru memerlukan informasi yang spesifik untuk masing-masing kelas yang diajar, sedangkan kepala sekolah memerlukan informasi yang umum untuk semua kelas dalam satu sekolah. Secara rinci, laporan hasil belajar peserta didik dapat dimanfaatkan oleh guru untuk:

- a) Mengetahui kekuatan dan kelemahan peserta didik dalam satu rombongan belajar dan sekolah yang mencakup semua mata ajar;
- b) Mendorong para guru untuk lebih baik lagi dalam memberikan pelayanan belajar kepada peserta didik;
- c) Membantu guru dalam mencari strategi yang lebih tepat;
- d) Mendorong sekolah untuk memberikan fasilitas belajar yang lebih baik lagi.

4. Manfaat Pelaporan Hasil Penilaian untuk Masyarakat

Pada umumnya, laporan hasil penilaian berkaitan dengan jumlah dan kualitas lulusan dari suatu jenjang sekolah tertentu. Setiap peserta didik yang telah lulus membawa bukti bahwa mereka memiliki suatu pengetahuan dan keterampilan tertentu. Namun pengetahuan dan keterampilan yang diperoleh peserta didik dari suatu sekolah tidaklah sama. Tingkat keberhasilan ini dinyatakan secara lengkap dalam laporan prestasi. Dengan laporan ini, masyarakat pada akhirnya memiliki gambaran berkaitan dengan kualitas pendidikan di suatu sekolah tertentu. Berdasarkan gambaran itu, masyarakat luas dapat menentukan pilihan apakah mereka harus menyekolahkan putra-putrinya di sekolah tersebut atau tidak. Apabila laporan hasil belajar menunjukkan bahwa sekolah tersebut memiliki kualitas lulusan yang baik, maka mereka dapat menentukan pilihannya untuk menyekolahkan putra-putrinya di sekolah tersebut dan sebaliknya. Dengan demikian, laporan hasil penilaian menjadi suatu hal yang krusial bagi peserta didik, guru, orang tua, dan juga masyarakat luas.

C. Model Laporan Hasil Penilaian

Pada dasarnya ada berbagai variasi model laporan hasil penilaian yang dapat digunakan oleh guru untuk menyampaikan informasi berkaitan dengan kemajuan hasil belajar peserta didik. Berdasarkan Keputusan Dirjen Mandikdasmen tahun 2008, laporan hasil penilaian berisi paling tidak hal-hal sebagai berikut:

1. Identitas Peserta Didik

Identitas peserta didik memuat berbagai informasi mengenai latar-belakang peserta didik sebagai berikut:

1. Nama Peserta Didik (lengkap) :
2. Nomor Induk :
3. Tempat dan Tanggal Lahir :
4. Jenis Kelamin :
5. Agama :
6. Anak ke :
7. Status dalam Keluarga :
8. Alamat Peserta Didik :
- Telepon :
9. Diterima di Sekolah ini
 - a. Di Kelas :
 - b. Pada Tanggal :
 - c. Semester :
10. Sekolah Asal
 - a. Nama Sekolah :
 - b. Alamat :
11. Ijazah SMP/MTs/Paket B
 - a. Tahun :
 - b. Nomor :
12. Surat Keterangan Hasil Ujian Nasional (SKHUN)
SMP/MTs/Paket B :
 - a. Tahun :
 - b. Nomor :
13. Orang Tua :
 - a. Ayah :

- b. Ibu :
14. Alamat Orang Tua :
- Telepon :
15. Pekerjaan Orang Tua :
- a. Ayah :
- b. Ibu :
16. Nama Wali :
17. Alamat Wali :
- Telepon :
18. Pekerjaan Wali :

.....,

Kepala Sekolah

(.....)

2. Format Nilai Hasil Belajar Peserta Didik

Nama Peserta Didik : ...

Kelas/Semester : ...

Nomor Induk : ...

Tahun Pelajaran: ...

Nama Sekolah : ...

No	Komponen	Kriteria Ketuntasan Minimal (KKM)	Nilai Hasil Belajar				
			Pengetahuan		Praktik		Sikap/ Afektif
A	Mata Pelajaran		Angka	Huruf	Angka	Huruf	Predikat
1	Pendidikan Agama						
2	Pendidikan Kewarganegaraan						
3	Bahasa Indonesia						

4	Bahasa Inggris						
5	Matematika						
6	Fisika						
7	Biologi						
8	Kimia						
9	Sejarah						
10	Geografi						
11	Ekonomik						
12	Sosiologi						
13	Seni Budaya						
14	Pendidikan Jasmani, Olahraga, dan Kesehatan						
15	Teknologi Informasi dan Komunikasi						
16	Keterampilan/ Bahasa Asing*)						
B	Muatan Lokal **)						

.....,

Orang Tua/Wali

Wali Kelas

Kepala Sekolah

.....

.....

.....

Kolom Kriteria Ketuntasan Belajar (KKM) untuk setiap mata pelajaran diisi dengan nilai KKM yang telah ditentukan oleh masing-masing satuan pendidikan, dalam bentuk bilangan bulat, dengan rentang 0 – 100. Kolom Pengetahuan diisi dengan nilai kumulatif dari hasil pencapaian SK dan

KD setiap mata pelajaran dan muatan lokal per semester, ditulis secara kuantitatif dalam bentuk bilangan bulat dan huruf. Kolom Praktik diisi dengan nilai kumulatif dari hasil pencapaian SK dan KD untuk aspek praktik pada mata pelajaran dan muatan lokal tertentu per semester, ditulis secara kuantitatif dalam bentuk bilangan bulat dan huruf. Kolom sikap/afektif diisi dengan hasil penilaian aspek sikap/afektif pada setiap mata pelajaran dan muatan lokal melalui pengamatan selama proses pembelajaran berlangsung per semester. Nilai sikap/afektif dicantumkan dalam bentuk predikat, dengan klasifikasi tinggi, sedang, rendah, atau amat baik, baik, cukup, dan kurang.

3. Format ketercapaian kompetensi peserta didik

Nama Peserta Didik : ...

Kelas/Semester : ...

Nomor Induk : ...

Tahun Pelajaran: ...

Nama Sekolah : ...

Ketercapaian Kompetensi Peserta Didik

No.	Komponen	Ketercapaian Kompetensi
A	Mata Pelajaran	
1	Pendidikan Agama	
2	Pendidikan Kewarganegaraan	
3	Bahasa Indonesia	
4	Bahasa Inggris	
5	Matematika	
6	Fisika	
7	Biologi	
8	Kimia	
9	Sejarah	

10	Geografi	
11	Ekonomi	
12	Sosiologi	
13	Seni Budaya	
14	Pendidikan Jasmani, Olahraga, dan Kesehatan	
15	Teknologi Informasi dan Komunikasi	
16	Keterampilan/ Bahasa Asing*)	
B	Muatan Lokal**)	

Kolom ketercapaian kompetensi diisi dengan uraian singkat/deskripsi yang menggambarkan tingkat pencapaian kompetensi secara utuh (baik KD yang telah tuntas maupun yang belum tuntas) untuk setiap mata pelajaran.

4. Pengembangan diri

Nama Peserta Didik : ...

Kelas/Semester : ...

Nomor Induk : ...

Tahun Pelajaran : ...

Nama Sekolah : ...

Pengembangan Diri :

No.	Jenis Kegiatan		Kegiatan
A.	Kegiatan Ekstrakurikuler		

B.	Keikutsertaan Dalam Organisasi/Kegiatan di Sekolah		

Tabel pengembangan diri diisi dengan jenis pengembangan diri (kegiatan kreativitas) yang diikuti oleh peserta didik dan keikutsertaan dalam organisasi/kegiatan sekolah.

5. Akhlak mulia dan kepribadian

No.	Aspek Yang Dinilai	Keterangan
1	Kedisiplinan	
2	Kebersihan	
3	Kesehatan	
4	Tanggung Jawab	
5	Span Santun	
6	Percaya Diri	
7	Kompetitif	
8	Hubunga Sosial	
9	Kejujuran	
10	Pelaksanaan Ibadah Ritual	

Kolom keterangan diisi dengan deskripsi singkat tentang predikat prestasi dan ketercapaian kemampuan baik keterampilan maupun pengetahuan, aktivitas/kegiatan sekolah yang diikuti peserta didik, serta sikap yang

ditunjukkan oleh peserta didik selama mengikuti kegiatan dan setelah mengikuti kegiatan pengembangan diri. Kolom keterangan pada Tabel Akhlak Mulia dan Kepribadian diisi dengan kategori penilaian Sangat Baik, Baik, atau Kurang Baik dan deskripsi tentang sikap/kebiasaan peserta didik yang paling dominan (baik positif maupun negatif), dalam kehidupan sehari-hari di sekolah untuk setiap aspek yang dinilai.

6. Ketidakhadiran

No	Alasan Ketidakhadiran	Keterangan
1	Sakit	
2	Izin	
3	Tanpa Keterangan	

Catatan Wali Kelas:

Orangtua/Wali

Wali Kelas

Kepala Sekolah

Kolom keterangan pada tabel ketidakhadiran diisi dengan lama waktu (hari, jam atau satuan waktu lainnya).

7. Catatan wali kelas

Catatan Wali Kelas:

Keterangan Kenaikan Kelas: Naik/Tidak Naik *)
Program : IPA / IPS / Bahasa *)

*) Coret yang tidak perlu

..... ,

Orang Tua/ Wali

.....

Wali Kelas

.....

Kepala Sekolah

.....

8. Catatan prestasi peserta didik

Catatan Prestasi Yang Telah Dicapai

Nama Peserta Didik: ...

Nomor Induk : ...

Nama Sekolah : ...

Kelas/Semester : ...

Tahun Pelajaran: ...

Program : ...

No.	Prestasi Yang Pernah Tercapai	Bukti Sertifikat / Piagam / Trophy Yang Diperoleh (Sebutkan)
1	Kurikuler	
2	Ekstrakurikuler	
3	Lain-lain	
4	Catatan Khusus Lainnya	

9. Keterangan pindah sekolah

Keterangan Pindah Sekolah

(Diisi oleh sekolah yang ditinggalkan/lama).

Nama Peserta Didik : ...

Kelas/Semester : ...

Nomor Induk : ...

Tahun Pelajaran : ...

Nama Sekolah : ...

Program : ...

KELUAR			
Tanggal	Kelas dan Semester yang ditinggalkan	Sebab-sebab keluar, dan atas permintaan (tertulis) dari :	Tanda Tangan Kepala Sekolah, Stempel Sekolah dan Tanda Tangan Orang Tua/Wali
			Kepala Sekolah, (.....) Orang Tua/Wali (.....)
			Kepala Sekolah, (.....) Orang Tua/Wali (.....)

Keterangan Pindah Sekolah

(Diisi oleh sekolah yang baru)

Nama Peserta Didik : ...

Kelas/Semester : ...

Nomor Induk : ...

Tahun Pelajaran: ...

Nama Sekolah : ...

Program : ...

MASUK

No.	Identitas Peserta Didik		 Kepala Sekolah, (.....)
1.	Nama Peserta Didik		
2.	Nomor Induk		
3.	Nama Sekolah		
4.	Masuk :		
5.	a. Tanggal		
	b. Di Kelas		
	c. Semester		

	Tahun Pelajaran	_____	
No.	Identitas Peserta Didik		
1.	Nama Peserta Didik	_____	
2.	Nomor Induk	_____	Kepala Sekolah,
3.	Nama Sekolah	_____	(.....)
4.	Masuk :	_____	
5.	a. Tanggal	_____	
	b. Di Kelas	_____	
	c. Semester	_____	
	Tahun Pelajaran	_____	
No.	Identitas Peserta Didik		
1.	Nama Peserta Didik	_____	
2.	Nomor Induk	_____	Kepala Sekolah,
3.	Nama Sekolah	_____	(.....)
4.	Masuk :	_____	
5.	a. Tanggal	_____	
	b. Di Kelas	_____	
	c. Semester	_____	
	Tahun Pelajaran	_____	
No.	Identitas Peserta Didik		
1.	Nama Peserta Didik	_____	
2.	Nomor Induk	_____	Kepala Sekolah,
3.	Nama Sekolah	_____	(.....)
4.	Masuk :	_____	
5.	a. Tanggal	_____	
	b. Di Kelas	_____	
	c. Semester	_____	
	Tahun Pelajaran	_____	

Pelaksanaan pindah sekolah lintas provinsi/kabupaten/kota disesuaikan dengan peraturan yang berlaku pada masing-masing Dinas Pendidikan Provinsi dan atau Kabupaten/Kota. Satuan pendidikan dapat menentukan persyaratan pindah/mutasi peserta didik sesuai prinsip manajemen berbasis sekolah, antara lain mencakup hal-hal sebagai berikut:

- Menyesuaikan bentuk LHB dari sekolah asal sesuai dengan LHB yang digunakan di sekolah tujuan
- Melakukan tes atau program matrikulasi bagi peserta didik pindahan

Bab 12

Evaluasi Program Pembelajaran

Pembelajaran dapat dipahami sebagai sebuah proses interaksi antara peserta didik dengan pendidik dan sumber belajar pada suatu lingkungan belajar. Berdasarkan pengertian di atas dapat dipahami bahwa sebuah proses pembelajaran melibatkan banyak komponen, yaitu pendidik (guru), peserta didik, materi ajar dan lingkungan. Karena sebuah proses pembelajaran melibatkan banyak komponen yang berkaitan secara sistemik, maka sebuah proses pembelajaran pada hakikatnya adalah sebuah program yang memiliki tujuan yang ingin dicapai. Program adalah sebuah terminologi yang mengacu pada tata hubungan dari suatu makna yang dirancang dan diterapkan dengan tujuan tertentu. Di lain pihak, istilah program pembelajaran pada dasarnya terdiri-dari dua kata, yaitu program dan pembelajaran dan keduanya saling berhubungan secara magnetis dan tidak terpisahkan. Berangkat dari dua kata tersebut, program pembelajaran dapat dimaknai sebagai sebuah rancangan kegiatan yang berkesinambungan dalam proses pembelajaran yang melibatkan berbagai komponen pembelajaran untuk mencapai tujuan yang telah ditetapkan. Komponen pembelajaran ini meliputi mata pelajaran, kompetensi inti, kompetensi dasar, indikator pencapaian kompetensi, tujuan pembelajaran, materi ajar, metode pembelajaran, kegiatan pembelajaran, media, sumber belajar dan alat penilaian. Selanjutnya, untuk mengidentifikasi apakah sebuah program pembelajaran benar-benar telah mampu secara efektif mencapai tujuan yang telah ditentukan atau belum, maka program pembelajaran tersebut perlu dievaluasi.

A. Hakaikat Evaluasi

Evaluasi merupakan sebuah tindakan atau tahapan-tahapan sistematis dari prosedur penelitian sosial yang dimaksudkan untuk mengukur konseptualisasi, desain, implementasi, dan kebermanfaatan suatu program (Rossi & Freeman, 1993). Pengertian ini mengandung makna bahwa sebuah evaluasi selalu melibatkan kegiatan untuk mengukur kekuatan dan kelemahan suatu program (misalnya program pembelajaran), kebijakan, produk, dan organisasi dalam rangka meningkatkan efektivitasnya. Dengan kata lain, Cronbach (1982) menyatakan bahwa kata kunci dari sebuah proses evaluasi adalah ditemukan bukti-bukti bahwa sebuah program berjalan efektif sesuai dengan rencana. Di lain sisi, Worthen & Sanders (1987) menjelaskan bahwa sebuah evaluasi yang baik selalu memanfaatkan pengukuran (*measurement*) dan pengamatan (*observation*) yang akurat, *reliable* dan valid guna mengumpulkan bukti-bukti empiris dan sistematis dan kemudian bukti-bukti itu digunakan untuk menganalisis hasil secara objektif.

Tujuan pokok dari sebuah evaluasi adalah sedikit berbeda dengan tujuan penelitian evaluasi (Cronbach, 1982). Senada dengan ini Ogle (2002) menyatakan bahwa tujuan penelitian evaluasi adalah untuk mendapatkan informasi khusus yang diperlukan untuk membuat sebuah keputusan, tidak sekedar mendapatkan informasi mengenai efektivitas program. Berdasarkan pengertian tersebut dapat dipahami bahwa penelitian evaluasi merupakan tindakan untuk mengumpulkan dan menganalisis data secara sistematis yang diperlukan untuk membuat keputusan. Dengan demikian, penelitian evaluasi harus dilakukan melalui prosedur terencana dan logis sehingga didapatkan data yang akurat guna memutuskan apakah sebuah program itu perlu dilanjutkan; dilanjutkan, tetapi perlu diadakan modifikasi program; atau bahkan program itu perlu dihentikan sama sekali.

McLaughlin & Jordan (1999) menyatakan bahwa penelitian evaluasi merupakan sebuah metode untuk mempelajari berbagai hal yang bersifat kompleks dari implementasi suatu program berdasarkan pemahaman yang komprehensif yang diperoleh dari hasil deskripsi dan analisis. Penelitian evaluasi dimaksudkan untuk mengilustrasikan berbagai temuan dari implementasi program dan mendeskripsikan hal-hal kritis dari implementasi program itu. Penelitian evaluasi juga dimaksudkan untuk mengeksplorasi berbagai gagasan baru untuk memperbaiki suatu program, menginvestigasi

jalannya suatu program, dan juga menguji sebab akibat adanya kekurangan dan kelebihan dari implementasi program secara mendalam. Selanjutnya Beijaard et al (2005) menyatakan bahwa berkaitan dengan tujuan penelitian evaluasi yang dimaksudkan untuk memperbaiki program, justifikasi apakah suatu program itu berhasil atau tidak dalam mencapai tujuan yang diharapkan, diperlukan standar baku (*benchmark*) yang dapat digunakan sebagai acuan dari justifikasi tersebut.

B. Evaluasi Program Pembelajaran

Valades & Bamberger (1994) menyatakan bahwa evaluasi program merupakan aktivitas manajemen internal maupun eksternal yang dimaksudkan untuk mengukur kelayakan suatu desain program dan metode implementasinya dalam rangka mencapai tujuan yang telah ditentukan, baik tujuan khusus maupun tujuan desain program dalam skala luas. Sedangkan Worthen, et al (1996) menyatakan bahwa evaluasi program merupakan metode sistematis yang dimaksudkan untuk mengukur dan memperbaiki suatu program melalui prosedur yang bersifat etis, akurat, dan fisibel. Pendapat serupa diberikan oleh Mclaughlin & Jordan (1999) yang menyatakan bahwa evaluasi program merupakan proses pengumpulan informasi secara sistematis mengenai aktivitas, karakteristik, dan *outcomes* dari implementasi suatu program dengan maksud untuk melakukan perbaikan program dan mengetahui efektivitas program tersebut.

Berangkat dari beberapa pengertian di atas, evaluasi program pembelajaran dapat dipahami sebagai aktivitas pengumpulan informasi secara sistematis dari implementasi suatu program pembelajaran yang dimaksudkan untuk mengetahui kelayakan program tersebut dalam rangka mencapai tujuan yang diharapkan (misalnya, tercapainya suatu kompetensi tertentu yang harus dikuasai oleh peserta didik). Dalam pengertian yang lebih luas, evaluasi program pembelajaran dimaksudkan untuk mengukur dampak yang dihasilkan dari sebuah program pembelajaran, baik dampak yang diinginkan maupun yang tidak diinginkan, dan juga untuk mengukur faktor-faktor yang mempengaruhi tingkat dan distribusi manfaat yang dihasilkan dari program tersebut. Pada akhirnya evaluasi program pembelajaran dapat digunakan oleh guru sebagai panduan (*guidelines*) untuk meningkatkan kualitas proses pembelajaran yang telah berlangsung

sekaligus untuk mengestimasi dampak yang mungkin didapat untuk meningkatkan kompetensi peserta didik.

Valadez & Bamberger (1994) selanjutnya menegaskan pula bahwa informasi yang diperoleh melalui kegiatan evaluasi program pembelajaran dapat dimanfaatkan oleh berbagai pihak, seperti guru, kepala sekolah, dan juga dinas pendidikan dalam beberapa hal, yaitu (1) memperbaiki metode dan teknik pembelajaran yang digunakan oleh guru agar selalu konsisten dengan tujuan yang telah ditentukan, dan juga untuk meyakinkan tingkat keberhasilan yang mungkin diperoleh dari program tersebut dalam rangka mencapai tujuan yang telah ditentukan; (2) evaluasi program pembelajaran dapat dimanfaatkan untuk menentukan apakah implementasi program tersebut dapat menjangkau orang-orang (peserta didik, orang tua, masyarakat umum, dan lain-lain) yang menjadi sasaran (*beneficiary*) program atau tidak; (3) mengukur apakah implementasi program pembelajaran itu mampu mencapai tujuan yang telah ditentukan; dan (4) mengukur dampak program pembelajaran terhadap pembangunan (misalnya kualitas sumber daya manusia) dalam skala luas. Informasi ini merupakan bagian sentral dan krusial dari evaluasi program pembelajaran yang harus dieksplorasi oleh seorang evaluator, baik evaluator eksternal maupun internal.

Seperti halnya dengan program-program yang lain, pelaksanaan program pembelajaran telah berjalan perlu dievaluasi untuk mendapatkan informasi mengenai kekuatan dan kelemahan program tersebut; apakah program berjalan sesuai yang direncanakan; dan apakah program tersebut telah mampu secara efektif mencapai tujuan yang diharapkan, misalnya bertujuan untuk meningkatkan kompetensi peserta didik pada mata pelajaran tertentu. Selanjutnya evaluasi program pembelajaran ini perlu dilakukan mengikuti model evaluasi yang sesuai dengan karakteristik program itu, sehingga kegiatan evaluasi mampu mendapatkan informasi yang akurat terhadap implementasi program pembelajaran tersebut. Model evaluasi menurut Valadez & Bamberger (1994) merupakan seperangkat hipotesis yang dapat diukur mengenai implementasi suatu program yang mana seperangkat hipotesis itu dapat diubah atau dipertahankan. Pengertian ini mengandung makna bahwa model evaluasi bukan merupakan jalinan antarvariabel yang mutlak dan bersifat permanen, tetapi sebuah model

evaluasi dapat dimodifikasi sesuai dengan konteks dimana model itu digunakan untuk mengevaluasi suatu program.

C. Model Evaluasi Program Pembelajaran

Valadez dan Bamberger (1994) menyatakan bahwa ada beberapa model evaluasi yang dapat diaplikasikan dalam mengukur efektivitas suatu program. Dalam kerangka evaluasi program, model-model evaluasi itu muncul karena adanya perbedaan tujuan, struktur dan jenis analisis yang digunakan untuk mengukur kinerja dan efektivitas program itu. Model evaluasi ini diharapkan dapat membantu dalam merencanakan dan mendesain suatu proses pembelajaran, membantu dalam mengamati dan mengukur proses pelaksanaan proses pembelajaran, membantu dalam membuat estimasi kuantitatif terhadap dampak dari suatu proses pembelajaran, dan juga membantu mengidentifikasi faktor-faktor yang mempengaruhi keberhasilan implementasi suatu program pembelajaran serta dampak yang mengikutinya.

Model evaluasi merupakan seperangkat aturan, ketentuan, larangan, dan panduan yang digunakan untuk melakukan evaluasi guna mengukur efektivitas implementasi suatu program. Ada beberapa model evaluasi yang berkembang dan digunakan oleh para evaluator dalam mengevaluasi suatu program. Beberapa model itu mempunyai asumsi dan karakteristik yang berbeda satu sama lain. Di antara model evaluasi itu adalah:

1. Model Evaluasi Kirkpatrick

Model Kirkpatrick pertama kali dikembangkan oleh Donald Kirkpatrick (1959) dan hingga saat ini digunakan secara luas oleh para evaluator untuk mengevaluasi program-program pembelajaran. Model evaluasi Kirkpatrick terdiri dari empat tahapan evaluasi, yaitu:

a) Reaksi (*reaction*).

Evaluasi pada tahap ini dilakukan untuk mengetahui kesan atau reaksi setiap peserta didik apakah merasa senang dan tertarik dengan program pembelajaran yang telah diimplementasikan oleh guru.

b) Pembelajaran (*learning*)

Evaluasi pada tahap ini dilakukan untuk mengidentifikasi sejauhmana capaian pengetahuan dan keterampilan setiap peserta didik setelah mengikuti program pembelajaran itu.

c) Perilaku (*behaviour*)

Evaluasi perilaku dilakukan untuk mengetahui apakah setiap peserta didik yang mengikuti program pembelajaran itu telah mengubah perilakunya dengan mengaplikasikan pengetahuan dan keterampilan yang telah dipelajari dari sebuah program pembelajaran.

d) Hasil (*results*)

Fokus pada tahap ini adalah untuk mengetahui dampak suatu program pembelajaran terhadap kompetensi peserta didik.

2. Model Evaluasi Stake

Worthen and Sanders (1987) menyatakan bahwa model evaluasi ini harus dilakukan melalui beberapa tahapan sebagai berikut:

- e) Mengidentifikasi latar belakang, justifikasi, dan diskripsi program pembelajaran secara rasional.
- f) Menentukan *antecedents* yang diinginkan dari implementasi program itu (input, sumber daya, dan kondisi yang ada), *transactions* (aktivitas dan proses) dan *outcomes*.
- g) Mencatat *antecedents*, *transactions*, dan *outcomes* yang teramati (termasuk mencatat kejadian-kejadian yang kemunculannya sebenarnya tidak diharapkan dari implementasi program itu).
- h) Menyatakan secara eksplisit standar (kriteria, harapan, atau kinerja program sejenis) untuk menilai *antecedents*, *transactions*, dan *outcomes* dari program itu.
- i) Mencatat segala bentuk penilaian (*judgement*) yang telah dilakukan terhadap *antecedents*, *transactions*, dan *outcomes* dari program pembelajaran itu.

3. Model Evaluasi Alkin

Menurut model evaluasi Alkin, ada lima tahapan yang harus dilakukan untuk mengevaluasi suatu program (Pace & Friedlander, 1978), yaitu:

- b) Melakukan penilaian terhadap system (*system assessment*);
- c) Perencanaan program (*program planning*);
- d) Implementasi program (*program implementation*);
- e) Perbaikan program (*program improvement*);
- f) Sertifikasi program (*program certification*).

4. Model CIPP Stufflebeam

Model evaluasi CIPP (*context, input, process, product*) seperti yang dikemukakan oleh Ogle (2002) dapat digunakan untuk melakukan evaluasi formatif (pada saat program pembelajaran sedang berlangsung) maupun sumatif (pada saat suatu program pembelajaran telah berlangsung). Salah satu karakteristik terpenting dari model CIPP Stufflebeam ini adalah model ini menerapkan langkah-langkah evaluasi secara holistik. Seluruh elemen-elemen evaluasi selalu berorientasi pada sistem dan terstruktur untuk mengakomodasi kebutuhan (*needs*) evaluasi secara universal.

Selanjutnya Braden (1992:) mencatat pula bahwa tidak banyak model evaluasi yang menawarkan tahapan evaluasi proses pelaksanaan program secara lengkap, tetapi hal itu dapat ditemukan secara komprehensif dalam model CIPP Stufflebeam. Selanjutnya tahapan-tahapan model evaluasi ini dapat diuraikan sebagai berikut:

- a) Evaluasi konteks (*context evaluation*) dimaksudkan untuk mendeskripsikan status, konteks atau *setting* dengan maksud untuk mengidentifikasi kebutuhan-kebutuhan yang belum dan telah terpenuhi, peluang yang potensial, masalah-masalah yang muncul selama program pembelajaran sedang berlangsung dan juga tujuan suatu program yang akan dievaluasi.
- b) Evaluasi input (*input evaluation*) dimaksudkan untuk mengetahui berbagai informasi yang diperlukan untuk membantu guru untuk memilih prosedur dan sumber-sumber untuk mendesain metode dan materi program pembelajaran yang sesuai.
- c) Evaluasi proses (*proses evaluation*) diperlukan untuk memastikan bahwa program pembelajaran akan berjalan sesuai yang diinginkan dan juga untuk mengidentifikasi berbagai kekurangan dan kelebihan prosedur pelaksanaan program.
- d) Evaluasi produk (*product evaluation*) dilakukan untuk mengukur sejauhmana suatu program telah mencapai tujuan yang telah ditentukan.

D. Pentingnya Evaluasi Program Pembelajaran

Program pembelajaran pada dasarnya merupakan sebuah program yang mencakup rangkaian aktivitas terencana yang dimaksudkan untuk mengembangkan pengetahuan, keterampilan, dan sikap peserta didik.

Pengetahuan mengacu pada pemahaman bahan ajar secara luas dan mendalam. Sedangkan keterampilan mengacu pada kemampuan untuk menggunakan atau menerapkan pengetahuan yang dimiliki untuk melakukan hal-hal praktis dalam kehidupan peserta didik. selanjutnya sikap mengacu pada suatu pikiran, kecenderungan dan perasaan seseorang untuk mengenal aspek-aspek tertentu pada lingkungan yang seringnya bersifat permanen karena sulit diubah. Perhatian berbagai kalangan mengenai implementasi program pembelajaran saat ini menunjukkan tren yang terus meningkat. Hal itu terjadi karena ada asumsi krusial bahwa apabila seorang guru tidak melaksanakan program pembelajaran yang terencana secara sistematis, maka efektivitas program pembelajaran tersebut dalam mencapai tujuan yang diharapkan tidak akan optimal. Program pembelajaran utamanya dilakukan untuk memperluas dan mengembangkan pengetahuan, keterampilan, dan sikap peserta didik, sehingga peserta didik mampu memiliki kecakapan yang berguna dalam hidupnya. Selanjutnya evaluasi dilakukan untuk mengetahui sejauhmana tujuan program itu telah tercapai.

Guskey (2000) menjelaskan alasan mengapa evaluasi program pembelajaran sangat penting untuk dilakukan. Pertama, evaluasi program pembelajaran sangat penting untuk dilakukan agar guru mempunyai pemahaman yang lebih baik mengenai hakekat dan dinamika dari program pembelajaran yang dilakukan. Paling tidak guru memahami bahwa program pembelajaran merupakan sebuah proses yang berkelanjutan dan berkesinambungan, bukan hanya aktivitas yang terjadi pada suatu saat saja. Program pembelajaran merupakan aktivitas yang panjang yang dimaksudkan untuk menciptakan pengalaman nyata dalam rangka meningkatkan kompetensi peserta didik. Oleh karena itu evaluasi menjadi suatu yang sangat penting dalam implementasi sebuah program pembelajaran.

Kedua, program pembelajaran merupakan sebuah proses yang disengaja yang dimaksudkan untuk menghasilkan perubahan kompetensi peserta didik. Program pembelajaran didesain untuk menghasilkan perubahan positif dan perbaikan sikap peserta didik Oleh karena itu, informasi mengenai efektivitas program dalam mencapai tingkat perubahan itu perlu dikumpulkan dan dianalisis. Aktivitas mengumpulkan informasi, proses memaknai informasi, dan melaporkan hasil analisis informasi ini pada dasarnya merupakan rangkaian aktivitas yang merujuk pada proses evaluasi.

Ketiga, evaluasi program pembelajaran perlu dilakukan karena adanya kebutuhan informasi yang digunakan sebagai dasar untuk perbaikan, khususnya perbaikan kualitas program pembelajaran itu sendiri dan umumnya perbaikan mutu pendidikan. Melalui proses evaluasi, dampak (baik dampak yang diinginkan maupun yang tidak diinginkan) dari implementasi program, tingkat keberhasilan, sumber daya yang diperlukan, dan berbagai kekurangan/kelemahan implementasi program pembelajaran dapat diidentifikasi dan akhirnya berdasarkan informasi itu perbaikan implementasi program dapat dilakukan guna meningkatkan efektivitas program dalam mencapai tujuan yang telah ditetapkan sebelumnya.

Keempat, evaluasi program pembelajaran perlu dilakukan untuk keperluan akuntabilitas. Evaluasi program pembelajaran diperlukan untuk membuktikan bahwa program itu telah membawa perubahan yang diinginkan. Apabila program tersebut secara empiris telah mampu memperbaiki kinerja guru, berarti akuntabilitas program adalah baik. Sebaliknya apabila program tersebut secara empiris tidak mampu memperbaiki kinerja guru, berarti akuntabilitas program adalah tidak baik. Dengan demikian, secara sistemis dapat dipahami bahwa tanpa dilakukan evaluasi, maka akuntabilitas program dalam mencapai dampak yang diinginkan tidak dapat dibuktikan secara akurat.

Daftar Pustaka

- Abdao, D. (2015). *Traditional vs. authentic assessment: reflection, insight, and realization*. <https://abdao.wordpress.com/2015/07/18/traditional-vs-authentic-assessment/>
- Allen, M.J. & Yan, W.M. (1979). *Introduction to measurement theory*. California: Brooks/Cole Publishing Company
- Anderson, L.W. dan Krathwohl, D.R. (2001). *A Taxonomy for Learning, Teaching, and Asuming: A Revision of Bloom's Taxonomy of Educational Objectives*. A Bridged Edition. New York: Addison Wesley Longman, Inc.
- Arifin, Z. (2010). *Evaluasi pembelajaran: Prinsip, teknik, prosedur*. Bandung: PT Remaja Rosdakarya
- Arreola, R.A. (1998). *Writing learning objectives*. Memphis: The University of Tennessee
- Bachman, L., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Barahal, S.L. (2008). Thinking about thinking. *Phi Delta Kappan*, 90(4), 298-302.
- Beijaard, D; Meijer, P.C.; Morine-Dersheimer, G; & Tilleme, H. (2005). *Teacher professional development in changing conditions*. Netherlands: Springer.
- Blaz, D. (2001). *A Collection of Performance Tasks and Rubrics: Foreign Languages*. Larchmont, NY: Eye on Education.
- Briggs, M., Woodfield, A., Martin, C, & Swatton, P. (2008). *Assessment for learning and teaching in primary school*. London: Learning Matters Ltd.
- Boggs, O. (2020). *Laporan penilaian hasil belajar dan manfaatnya*. <https://www.slideserve.com/orsen/laporan-penilaian-hasil-belajar-dan-manfaatnya>
- Brookhart, S.M. (2010). *How to assess higher-order thinking skills in your classroom*. Alexandria: ASCD

- Brown, H.D. (2004). *Language assessment: Principles and classroom practices*. New York: Longman Publishing
- Cheng, L. & Watanabe, Y. (2008). *Washback in language testing*. Mahwah, NJ: Taylor & Francis.
- Cox, J. (2019). *Teaching strategies that enhance higher-order thinking skills*. <https://www.teachhub.com/teaching-strategies/2019/10/teaching-strategies-that-enhance-higher-order-thinking/>
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Fransico: Jossey-Bass Publishers
- Direktorat Pendidikan SMA (2018). *Pengembangan Soal Higher Order Thinking Skills*. Materi disampaikan dalam Workshop Sekolah Rujukan di Yogyakarta
- Dirjen Mandikdasmen (2008). *Bentuk dan tatacara penyusunan laporan hasil belajar peserta didik satuan pendidikan dasar dan menengah (SD/MI/SDLB,SMP/MTs./SMPLB, DAN SMA/MA/SMK/SMALB)*
- Disha, M. (2019). *Essay Test: Types, Advantages and Limitations*. <https://www.yourarticlelibrary.com/statistics-2/essay-test-types-advantages-and-limitations-statistics/92656>
- Dorobat, D. (2007). *The methodology of evaluation and testing*. Ministerul Educatiei si Certarii
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment*. New York: Routledge
- Griffith, W.I., & Lim, H. (2011). Performance-Based Assessment: Rubrics, Web 2.0 Tools and Language Competencies. *MEXTESOL Journal*, 36(1), 1 – 12.
- Gronlund, N.E. (1985). *Measurement and evaluation in teaching*, fifth edition, New York: McMillan Publishing co., Inc.
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan Publishing Company.
- Guskey, T.R. (2000). *Evaluating professional development*. California: Corwin Press.

- Harris, M. & McCann, P. (1994). *Assessment: Handbooks for the English classroom*. Macmillan Oxford: Macmillan Publishers Ltd
- Heaton, J.B. (1990). *Classroom testing*. New York: Longman Publishing.
- Hibbard, K. M., Van Wagenen, L., Lewbel, S., Waterbury-Wyatt, S., Shaw, S. & Pelletier, K. (1996). *A Teacher's Guide to Performance-based Learning and Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Izard, J. (2005). *Trial testing and item analysis in test construction*. Paris: International Institute for Educational Planning/UNESCO
- Johnson, R.L., Penny, J.A., & Gordon, B. (2009). *Assessing performance: Designing, scoring, and validating performance task*. New York: The Guilford Press
- Kaplan, R.M., & Saccuzzo, D.P. (1982). *Psychological testing: Principles, application, and issues*. Monterey: Brooks/Cole Publishig Company.
- Lim, H-Y, & Griffith, W.I. (2011). Practice doesn't make perfect. *MEXTESOL Journal*, 34(3).
- Magno, C. & Ouano, J. (2000). *Designing written assessment of student learning*. New York: Sringer
- Mardapi, J. (2008). *Teknik penyusunan instrumen tes dan non-tes*. Yogyakarta: Mitra Cendekia Press
- Mclaughlin, J.A., & Jordan, G.B. (1999). Logic models: a tool for telling your program's performance story. *Journal of Evaluation and Program Planning*. 22 (1), 65-72.
- McTighe, J., & Ferrara, S. (1998). *Assessing Learning in the Classroom*. Washington, DC: National Education Association.
- Mueller, J. (2014). *Authentic assessment toolbox*. <http://jfmueeller.faculty.noctrl.edu/toolbox/whatisit.htm>
- National Capital Language Resource Center (NCLRC). (2011). *The Essentials of Language Teaching*. <http://nclrc.org/essentials/assessing/alternative.htm>
- Ogle, G.J. (2002). *Towards a formative evaluation tool*. Dissertation: Virginia Polytechnic Institute and State University – USA
- O'neill, P., Moore, C., & Huot, B. (2009). *A guide to college writing assessment*. Utah: Utah University Press

- Osterlind, S.J. (2002). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats*. New York: Kluwer Academic Publishers.
- Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach* (5th Ed.). Newbury Park: Sage Publications.
- Rukundo, A., & Magambo, J. (2010). Effective test administration in schools: Principles and good practices for test administrators in Uganda. *African Journal of Teacher Education*, 1(1), 166 – 173.
- Setiawan, S. (2020). *Pengertian Rubrik – Tujuan, Jenis, Isi, Macam, Langkah, Manfaat, Contoh, Para Ahli*.
<https://www.gurupendidikan.co.id/pengertian-rubrik/>
- Suyanto (2020). *Penilaian autentik*. <https://suyanto.id/penilaian-autentik/>
- Trach, E. (2018). *How Students Benefit From Project-Based Assessments*.
<https://www.schoolology.com/blog/how-students-benefit-project-based-assessments-examples>
- Van der Linden, W.J. (2005). *Linear models for optimal test design*. New York: Springer
- Valadez, J., & Bamberger, M. (1994). *Monitoring and evaluating social programs in developing countries*. Washington: the World Bank
- Widoyoko, S.E.P. (2010). *Evaluasi program pembelajaran: Panduan praktis bagi pendidik dan calon pendidik*. Yogyakarta: Pustaka Pelajar
- Wiggins, G.P. (1993). *Assessing Student Performance*. San Francisco: Jossey-Bass Publishers.
- Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.
- Worthen, B.R., Sanders, J.R., & Fitzpatrick, J.L. (1996). *Program evaluation: Alternative approaches and practical guidelines 2nd Edition*. New York: Longman

Lampiran 1

Tabel r Product Moment Pada Sig.0,05 (Two Tail)

N	r	N	r	N	r	N	r	N	r	N	r
1	0.997	41	0.301	81	0.216	121	0.177	161	0.154	201	0.138
2	0.95	42	0.297	82	0.215	122	0.176	162	0.153	202	0.137
3	0.878	43	0.294	83	0.213	123	0.176	163	0.153	203	0.137
4	0.811	44	0.291	84	0.212	124	0.175	164	0.152	204	0.137
5	0.754	45	0.288	85	0.211	125	0.174	165	0.152	205	0.136
6	0.707	46	0.285	86	0.21	126	0.174	166	0.151	206	0.136
7	0.666	47	0.282	87	0.208	127	0.173	167	0.151	207	0.136
8	0.632	48	0.279	88	0.207	128	0.172	168	0.151	208	0.135
9	0.602	49	0.276	89	0.206	129	0.172	169	0.15	209	0.135
10	0.576	50	0.273	90	0.205	130	0.171	170	0.15	210	0.135
11	0.553	51	0.271	91	0.204	131	0.17	171	0.149	211	0.134
12	0.532	52	0.268	92	0.203	132	0.17	172	0.149	212	0.134
13	0.514	53	0.266	93	0.202	133	0.169	173	0.148	213	0.134
14	0.497	54	0.263	94	0.201	134	0.168	174	0.148	214	0.134
15	0.482	55	0.261	95	0.2	135	0.168	175	0.148	215	0.133
16	0.468	56	0.259	96	0.199	136	0.167	176	0.147	216	0.133
17	0.456	57	0.256	97	0.198	137	0.167	177	0.147	217	0.133
18	0.444	58	0.254	98	0.197	138	0.166	178	0.146	218	0.132
19	0.433	59	0.252	99	0.196	139	0.165	179	0.146	219	0.132
20	0.423	60	0.25	100	0.195	140	0.165	180	0.146	220	0.132
21	0.413	61	0.248	101	0.194	141	0.164	181	0.145	221	0.131
22	0.404	62	0.246	102	0.193	142	0.164	182	0.145	222	0.131
23	0.396	63	0.244	103	0.192	143	0.163	183	0.144	223	0.131
24	0.388	64	0.242	104	0.191	144	0.163	184	0.144	224	0.131
25	0.381	65	0.24	105	0.19	145	0.162	185	0.144	225	0.13
26	0.374	66	0.239	106	0.189	146	0.161	186	0.143	226	0.13
27	0.367	67	0.237	107	0.188	147	0.161	187	0.143	227	0.13
28	0.361	68	0.235	108	0.187	148	0.16	188	0.142	228	0.129
29	0.355	69	0.234	109	0.187	149	0.16	189	0.142	229	0.129
30	0.349	70	0.232	110	0.186	150	0.159	190	0.142	230	0.129
31	0.344	71	0.23	111	0.185	151	0.159	191	0.141	231	0.129
32	0.339	72	0.229	112	0.184	152	0.158	192	0.141	232	0.128
33	0.334	73	0.227	113	0.183	153	0.158	193	0.141	233	0.128
34	0.329	74	0.226	114	0.182	154	0.157	194	0.14	234	0.128
35	0.325	75	0.224	115	0.182	155	0.157	195	0.14	235	0.127
36	0.32	76	0.223	116	0.181	156	0.156	196	0.139	236	0.127
37	0.316	77	0.221	117	0.18	157	0.156	197	0.139	237	0.127
38	0.312	78	0.22	118	0.179	158	0.155	198	0.139	238	0.127
39	0.308	79	0.219	119	0.179	159	0.155	199	0.138	239	0.126
40	0.304	80	0.217	120	0.178	160	0.154	200	0.138	240	0.126

Lampiran 2 Tabel Uji-t

Pr df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.050	0.01 0.02	0.005 0.010	0.001 0.002
1	1.00000	3.07768	6.31375	12.70620	31.82052	63.65674	318.30884
2	0.81650	1.88562	2.91999	4.30265	6.96456	9.92484	22.32712
3	0.76489	1.63774	2.35336	3.18245	4.54070	5.84091	10.21453
4	0.74070	1.53321	2.13185	2.77645	3.74695	4.60409	7.17318
5	0.72669	1.47588	2.01505	2.57058	3.36493	4.03214	5.89343
6	0.71756	1.43976	1.94318	2.44691	3.14267	3.70743	5.20763
7	0.71114	1.41492	1.89458	2.36462	2.99795	3.49948	4.78529
8	0.70639	1.39682	1.85955	2.30600	2.89646	3.35539	4.50079
9	0.70272	1.38303	1.83311	2.26216	2.82144	3.24984	4.29681
10	0.69981	1.37218	1.81246	2.22814	2.76377	3.16927	4.14370
11	0.69745	1.36343	1.79588	2.20099	2.71808	3.10581	4.02470
12	0.69548	1.35622	1.78229	2.17881	2.68100	3.05454	3.92963
13	0.69383	1.35017	1.77093	2.16037	2.65031	3.01228	3.85198
14	0.69242	1.34503	1.76131	2.14479	2.62449	2.97684	3.78739
15	0.69120	1.34061	1.75305	2.13145	2.60248	2.94671	3.73283
16	0.69013	1.33676	1.74588	2.11991	2.58349	2.92078	3.68615
17	0.68920	1.33338	1.73961	2.10982	2.56693	2.89823	3.64577
18	0.68836	1.33039	1.73406	2.10092	2.55238	2.87844	3.61048
19	0.68762	1.32773	1.72913	2.09302	2.53948	2.86093	3.57940
20	0.68695	1.32534	1.72472	2.08596	2.52798	2.84534	3.55181
21	0.68635	1.32319	1.72074	2.07961	2.51765	2.83136	3.52715
22	0.68581	1.32124	1.71714	2.07387	2.50832	2.81876	3.50499
23	0.68531	1.31946	1.71387	2.06866	2.49987	2.80734	3.48496
24	0.68485	1.31784	1.71088	2.06390	2.49216	2.79694	3.46678
25	0.68443	1.31635	1.70814	2.05954	2.48511	2.78744	3.45019
26	0.68404	1.31497	1.70562	2.05553	2.47863	2.77871	3.43500
27	0.68368	1.31370	1.70329	2.05183	2.47266	2.77068	3.42103
28	0.68335	1.31253	1.70113	2.04841	2.46714	2.76326	3.40816
29	0.68304	1.31143	1.69913	2.04523	2.46202	2.75639	3.39624
30	0.68276	1.31042	1.69726	2.04227	2.45726	2.75000	3.38518
31	0.68249	1.30946	1.69552	2.03951	2.45282	2.74404	3.37490
32	0.68223	1.30857	1.69389	2.03693	2.44868	2.73848	3.36531
33	0.68200	1.30774	1.69236	2.03452	2.44479	2.73328	3.35634
34	0.68177	1.30695	1.69092	2.03224	2.44115	2.72839	3.34793
35	0.68156	1.30621	1.68957	2.03011	2.43772	2.72381	3.34005
36	0.68137	1.30551	1.68830	2.02809	2.43449	2.71948	3.33262
37	0.68118	1.30485	1.68709	2.02619	2.43145	2.71541	3.32563
38	0.68100	1.30423	1.68595	2.02439	2.42857	2.71156	3.31903
39	0.68083	1.30364	1.68488	2.02269	2.42584	2.70791	3.31279

40	0.68067	1.30308	1.68385	2.02108	2.42326	2.70446	3.30688
41	0.68052	1.30254	1.68288	2.01954	2.42080	2.70118	3.30127
42	0.68038	1.30204	1.68195	2.01808	2.41847	2.69807	3.29595
43	0.68024	1.30155	1.68107	2.01669	2.41625	2.69510	3.29089
44	0.68011	1.30109	1.68023	2.01537	2.41413	2.69228	3.28607
45	0.67998	1.30065	1.67943	2.01410	2.41212	2.68959	3.28148
46	0.67986	1.30023	1.67866	2.01290	2.41019	2.68701	3.27710
47	0.67975	1.29982	1.67793	2.01174	2.40835	2.68456	3.27291
48	0.67964	1.29944	1.67722	2.01063	2.40658	2.68220	3.26891
49	0.67953	1.29907	1.67655	2.00958	2.40489	2.67995	3.26508
50	0.67943	1.29871	1.67591	2.00856	2.40327	2.67779	3.26141
51	0.67933	1.29837	1.67528	2.00758	2.40172	2.67572	3.25789
52	0.67924	1.29805	1.67469	2.00665	2.40022	2.67373	3.25451
53	0.67915	1.29773	1.67412	2.00575	2.39879	2.67182	3.25127
54	0.67906	1.29743	1.67356	2.00488	2.39741	2.66998	3.24815
55	0.67898	1.29713	1.67303	2.00404	2.39608	2.66822	3.24515
56	0.67890	1.29685	1.67252	2.00324	2.39480	2.66651	3.24226
57	0.67882	1.29658	1.67203	2.00247	2.39357	2.66487	3.23948
58	0.67874	1.29632	1.67155	2.00172	2.39238	2.66329	3.23680
59	0.67867	1.29607	1.67109	2.00100	2.39123	2.66176	3.23421
60	0.67860	1.29582	1.67065	2.00030	2.39012	2.66028	3.23171
61	0.67853	1.29558	1.67022	1.99962	2.38905	2.65886	3.22930
62	0.67847	1.29536	1.66980	1.99897	2.38801	2.65748	3.22696
63	0.67840	1.29513	1.66940	1.99834	2.38701	2.65615	3.22471
64	0.67834	1.29492	1.66901	1.99773	2.38604	2.65485	3.22253
65	0.67828	1.29471	1.66864	1.99714	2.38510	2.65360	3.22041
66	0.67823	1.29451	1.66827	1.99656	2.38419	2.65239	3.21837
67	0.67817	1.29432	1.66792	1.99601	2.38330	2.65122	3.21639
68	0.67811	1.29413	1.66757	1.99547	2.38245	2.65008	3.21446
69	0.67806	1.29394	1.66724	1.99495	2.38161	2.64898	3.21260
70	0.67801	1.29376	1.66691	1.99444	2.38081	2.64790	3.21079
71	0.67796	1.29359	1.66660	1.99394	2.38002	2.64686	3.20903
72	0.67791	1.29342	1.66629	1.99346	2.37926	2.64585	3.20733
73	0.67787	1.29326	1.66600	1.99300	2.37852	2.64487	3.20567
74	0.67782	1.29310	1.66571	1.99254	2.37780	2.64391	3.20406
75	0.67778	1.29294	1.66543	1.99210	2.37710	2.64298	3.20249
76	0.67773	1.29279	1.66515	1.99167	2.37642	2.64208	3.20096
77	0.67769	1.29264	1.66488	1.99125	2.37576	2.64120	3.19948
78	0.67765	1.29250	1.66462	1.99085	2.37511	2.64034	3.19804
79	0.67761	1.29236	1.66437	1.99045	2.37448	2.63950	3.19663
80	0.67757	1.29222	1.66412	1.99006	2.37387	2.63869	3.19526
81	0.67753	1.29209	1.66388	1.98969	2.37327	2.63790	3.19392
82	0.67749	1.29196	1.66365	1.98932	2.37269	2.63712	3.19262

83	0.67746	1.29183	1.66342	1.98896	2.37212	2.63637	3.19135
84	0.67742	1.29171	1.66320	1.98861	2.37156	2.63563	3.19011
85	0.67739	1.29159	1.66298	1.98827	2.37102	2.63491	3.18890
86	0.67735	1.29147	1.66277	1.98793	2.37049	2.63421	3.18772
87	0.67732	1.29136	1.66256	1.98761	2.36998	2.63353	3.18657
88	0.67729	1.29125	1.66235	1.98729	2.36947	2.63286	3.18544
89	0.67726	1.29114	1.66216	1.98698	2.36898	2.63220	3.18434
90	0.67723	1.29103	1.66196	1.98667	2.36850	2.63157	3.18327
91	0.67720	1.29092	1.66177	1.98638	2.36803	2.63094	3.18222
92	0.67717	1.29082	1.66159	1.98609	2.36757	2.63033	3.18119
93	0.67714	1.29072	1.66140	1.98580	2.36712	2.62973	3.18019
94	0.67711	1.29062	1.66123	1.98552	2.36667	2.62915	3.17921
95	0.67708	1.29053	1.66105	1.98525	2.36624	2.62858	3.17825
96	0.67705	1.29043	1.66088	1.98498	2.36582	2.62802	3.17731
97	0.67703	1.29034	1.66071	1.98472	2.36541	2.62747	3.17639
98	0.67700	1.29025	1.66055	1.98447	2.36500	2.62693	3.17549
99	0.67698	1.29016	1.66039	1.98422	2.36461	2.62641	3.17460
100	0.67695	1.29007	1.66023	1.98397	2.36422	2.62589	3.17374
101	0.67693	1.28999	1.66008	1.98373	2.36384	2.62539	3.17289
102	0.67690	1.28991	1.65993	1.98350	2.36346	2.62489	3.17206
103	0.67688	1.28982	1.65978	1.98326	2.36310	2.62441	3.17125
104	0.67686	1.28974	1.65964	1.98304	2.36274	2.62393	3.17045
105	0.67683	1.28967	1.65950	1.98282	2.36239	2.62347	3.16967
106	0.67681	1.28959	1.65936	1.98260	2.36204	2.62301	3.16890
107	0.67679	1.28951	1.65922	1.98238	2.36170	2.62256	3.16815
108	0.67677	1.28944	1.65909	1.98217	2.36137	2.62212	3.16741
109	0.67675	1.28937	1.65895	1.98197	2.36105	2.62169	3.16669
110	0.67673	1.28930	1.65882	1.98177	2.36073	2.62126	3.16598
111	0.67671	1.28922	1.65870	1.98157	2.36041	2.62085	3.16528
112	0.67669	1.28916	1.65857	1.98137	2.36010	2.62044	3.16460
113	0.67667	1.28909	1.65845	1.98118	2.35980	2.62004	3.16392
114	0.67665	1.28902	1.65833	1.98099	2.35950	2.61964	3.16326
115	0.67663	1.28896	1.65821	1.98081	2.35921	2.61926	3.16262
116	0.67661	1.28889	1.65810	1.98063	2.35892	2.61888	3.16198
117	0.67659	1.28883	1.65798	1.98045	2.35864	2.61850	3.16135
118	0.67657	1.28877	1.65787	1.98027	2.35837	2.61814	3.16074
119	0.67656	1.28871	1.65776	1.98010	2.35809	2.61778	3.16013
120	0.67654	1.28865	1.65765	1.97993	2.35782	2.61742	3.15954
121	0.67652	1.28859	1.65754	1.97976	2.35756	2.61707	3.15895
122	0.67651	1.28853	1.65744	1.97960	2.35730	2.61673	3.15838
123	0.67649	1.28847	1.65734	1.97944	2.35705	2.61639	3.15781
124	0.67647	1.28842	1.65723	1.97928	2.35680	2.61606	3.15726
125	0.67646	1.28836	1.65714	1.97912	2.35655	2.61573	3.15671

126	0.67644	1.28831	1.65704	1.97897	2.35631	2.61541	3.15617
127	0.67643	1.28825	1.65694	1.97882	2.35607	2.61510	3.15565
128	0.67641	1.28820	1.65685	1.97867	2.35583	2.61478	3.15512
129	0.67640	1.28815	1.65675	1.97852	2.35560	2.61448	3.15461
130	0.67638	1.28810	1.65666	1.97838	2.35537	2.61418	3.15411
131	0.67637	1.28805	1.65657	1.97824	2.35515	2.61388	3.15361
132	0.67635	1.28800	1.65648	1.97810	2.35493	2.61359	3.15312
133	0.67634	1.28795	1.65639	1.97796	2.35471	2.61330	3.15264
134	0.67633	1.28790	1.65630	1.97783	2.35450	2.61302	3.15217
135	0.67631	1.28785	1.65622	1.97769	2.35429	2.61274	3.15170
136	0.67630	1.28781	1.65613	1.97756	2.35408	2.61246	3.15124
137	0.67628	1.28776	1.65605	1.97743	2.35387	2.61219	3.15079
138	0.67627	1.28772	1.65597	1.97730	2.35367	2.61193	3.15034
139	0.67626	1.28767	1.65589	1.97718	2.35347	2.61166	3.14990
140	0.67625	1.28763	1.65581	1.97705	2.35328	2.61140	3.14947
141	0.67623	1.28758	1.65573	1.97693	2.35309	2.61115	3.14904
142	0.67622	1.28754	1.65566	1.97681	2.35289	2.61090	3.14862
143	0.67621	1.28750	1.65558	1.97669	2.35271	2.61065	3.14820
144	0.67620	1.28746	1.65550	1.97658	2.35252	2.61040	3.14779
145	0.67619	1.28742	1.65543	1.97646	2.35234	2.61016	3.14739
146	0.67617	1.28738	1.65536	1.97635	2.35216	2.60992	3.14699
147	0.67616	1.28734	1.65529	1.97623	2.35198	2.60969	3.14660
148	0.67615	1.28730	1.65521	1.97612	2.35181	2.60946	3.14621
149	0.67614	1.28726	1.65514	1.97601	2.35163	2.60923	3.14583
150	0.67613	1.28722	1.65508	1.97591	2.35146	2.60900	3.14545
151	0.67612	1.28718	1.65501	1.97580	2.35130	2.60878	3.14508
152	0.67611	1.28715	1.65494	1.97569	2.35113	2.60856	3.14471
153	0.67610	1.28711	1.65487	1.97559	2.35097	2.60834	3.14435
154	0.67609	1.28707	1.65481	1.97549	2.35081	2.60813	3.14400
155	0.67608	1.28704	1.65474	1.97539	2.35065	2.60792	3.14364
156	0.67607	1.28700	1.65468	1.97529	2.35049	2.60771	3.14330
157	0.67606	1.28697	1.65462	1.97519	2.35033	2.60751	3.14295
158	0.67605	1.28693	1.65455	1.97509	2.35018	2.60730	3.14261
159	0.67604	1.28690	1.65449	1.97500	2.35003	2.60710	3.14228
160	0.67603	1.28687	1.65443	1.97490	2.34988	2.60691	3.14195

Pr df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.050	0.01 0.02	0.005 0.010	0.001 0.002
161	0.67602	1.28683	1.65437	1.97481	2.34973	2.60671	3.14162
162	0.67601	1.28680	1.65431	1.97472	2.34959	2.60652	3.14130
163	0.67600	1.28677	1.65426	1.97462	2.34944	2.60633	3.14098
164	0.67599	1.28673	1.65420	1.97453	2.34930	2.60614	3.14067
165	0.67598	1.28670	1.65414	1.97445	2.34916	2.60595	3.14036
166	0.67597	1.28667	1.65408	1.97436	2.34902	2.60577	3.14005
167	0.67596	1.28664	1.65403	1.97427	2.34888	2.60559	3.13975
168	0.67595	1.28661	1.65397	1.97419	2.34875	2.60541	3.13945
169	0.67594	1.28658	1.65392	1.97410	2.34862	2.60523	3.13915
170	0.67594	1.28655	1.65387	1.97402	2.34848	2.60506	3.13886
171	0.67593	1.28652	1.65381	1.97393	2.34835	2.60489	3.13857
172	0.67592	1.28649	1.65376	1.97385	2.34822	2.60471	3.13829
173	0.67591	1.28646	1.65371	1.97377	2.34810	2.60455	3.13801
174	0.67590	1.28644	1.65366	1.97369	2.34797	2.60438	3.13773
175	0.67589	1.28641	1.65361	1.97361	2.34784	2.60421	3.13745
176	0.67589	1.28638	1.65356	1.97353	2.34772	2.60405	3.13718
177	0.67588	1.28635	1.65351	1.97346	2.34760	2.60389	3.13691
178	0.67587	1.28633	1.65346	1.97338	2.34748	2.60373	3.13665
179	0.67586	1.28630	1.65341	1.97331	2.34736	2.60357	3.13638
180	0.67586	1.28627	1.65336	1.97323	2.34724	2.60342	3.13612
181	0.67585	1.28625	1.65332	1.97316	2.34713	2.60326	3.13587
182	0.67584	1.28622	1.65327	1.97308	2.34701	2.60311	3.13561
183	0.67583	1.28619	1.65322	1.97301	2.34690	2.60296	3.13536
184	0.67583	1.28617	1.65318	1.97294	2.34678	2.60281	3.13511
185	0.67582	1.28614	1.65313	1.97287	2.34667	2.60267	3.13487
186	0.67581	1.28612	1.65309	1.97280	2.34656	2.60252	3.13463
187	0.67580	1.28610	1.65304	1.97273	2.34645	2.60238	3.13438
188	0.67580	1.28607	1.65300	1.97266	2.34635	2.60223	3.13415
189	0.67579	1.28605	1.65296	1.97260	2.34624	2.60209	3.13391
190	0.67578	1.28602	1.65291	1.97253	2.34613	2.60195	3.13368
191	0.67578	1.28600	1.65287	1.97246	2.34603	2.60181	3.13345
192	0.67577	1.28598	1.65283	1.97240	2.34593	2.60168	3.13322
193	0.67576	1.28595	1.65279	1.97233	2.34582	2.60154	3.13299
194	0.67576	1.28593	1.65275	1.97227	2.34572	2.60141	3.13277
195	0.67575	1.28591	1.65271	1.97220	2.34562	2.60128	3.13255
196	0.67574	1.28589	1.65267	1.97214	2.34552	2.60115	3.13233
197	0.67574	1.28586	1.65263	1.97208	2.34543	2.60102	3.13212
198	0.67573	1.28584	1.65259	1.97202	2.34533	2.60089	3.13190
199	0.67572	1.28582	1.65255	1.97196	2.34523	2.60076	3.13169
200	0.67572	1.28580	1.65251	1.97190	2.34514	2.60063	3.13148

Daftar Pustaka

- Abdao, D. (2015). *Traditional vs. authentic assessment: reflection, insight, and realization*. <https://abdao.wordpress.com/2015/07/18/traditional-vs-authentic-assessment/>
- Allen, M.J. & Yan, W.M. (1979). *Introduction to measurement theory*. California: Brooks/Cole Publishing Company
- Anderson, L.W. dan Krathwohl, D.R. (2001). *A Taxonomy for Learning, Teaching, and Asuming: A Revision of Bloom's Taxonomy of Educational Objectives*. A Bridged Edition. New York: Addison Wesley Longman, Inc.
- Arifin, Z. (2010). *Evaluasi pembelajaran: Prinsip, teknik, prosedur*. Bandung: PT Remaja Rosdakarya
- Arreola, R.A. (1998). *Writing learning objectives*. Memphis: The University of Tennessee
- Bachman, L., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Barahal, S.L. (2008). Thinking about thinking. *Phi Delta Kappan*, 90(4), 298-302.
- Beijaard, D; Meijer, P.C.; Morine-Dershimer, G; & Tilleme, H. (2005). *Teacher professional development in changing conditions*. Netherlands: Springer.
- Blaz, D. (2001). *A Collection of Performance Tasks and Rubrics: Foreign Languages*. Larchmont, NY: Eye on Education.
- Briggs, M., Woodfield, A., Martin, C, & Swatton, P. (2008). *Assessment for learning and teaching in primary school*. London: Learning Matters Ltd.
- Boggs, O. (2020). *Laporan penilaian hasil belajar dan manfaatnya*. <https://www.slideserve.com/orsen/laporan-penilaian-hasil-belajar-dan-manfaatnya>
- Brookhart, S.M. (2010). *How to assess higher-order thinking skills in your classroom*. Alexandria: ASCD

- Brown, H.D. (2004). *Language assessment: Principles and classroom practices*. New York: Longman Publishing
- Cheng, L. & Watanabe, Y. (2008). *Washback in language testing*. Mahwah, NJ: Taylor & Francis.
- Cox, J. (2019). *Teaching strategies that enhance higher-order thinking skills*. <https://www.teachhub.com/teaching-strategies/2019/10/teaching-strategies-that-enhance-higher-order-thinking/>
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Fransico: Jossey-Bass Publishers
- Direktorat Pendidikan SMA (2018). *Pengembangan Soal Higher Order Thinking Skills*. Materi disampaikan dalam Workshop Sekolah Rujukan di Yogyakarta
- Dirjen Mandikdasmen (2008). *Bentuk dan tatacara penyusunan laporan hasil belajar peserta didik satuan pendidikan dasar dan menengah (SD/MI/SDLB,SMP/MTs./SMPLB, DAN SMA/MA/SMK/SMALB)*
- Disha, M. (2019). *Essay Test: Types, Advantages and Limitations*. <https://www.yourarticlelibrary.com/statistics-2/essay-test-types-advantages-and-limitations-statistics/92656>
- Dorobat, D. (2007). *The methodology of evaluation and testing*. Ministerul Educatiei si Certarii
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment*. New York: Routledge
- Griffith, W.I., & Lim, H. (2011). Performance-Based Assessment: Rubrics, Web 2.0 Tools and Language Competencies. *MEXTESOL Journal*, 36(1), 1 – 12.
- Gronlund, N.E. (1985). *Measurement and evaluation in teaching*, fifth edition, New York: McMillan Publishing co., Inc.
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan Publishing Company.
- Guskey, T.R. (2000). *Evaluating professional development*. California: Corwin Press.

- Harris, M. & McCann, P. (1994). *Assessment: Handbooks for the English classroom*. Macmillan Oxford: Macmillan Publishers Ltd
- Heaton, J.B. (1990). *Classroom testing*. New York: Longman Publishing.
- Hibbard, K. M., Van Wagenen, L., Lewbel, S., Waterbury-Wyatt, S., Shaw, S. & Pelletier, K. (1996). *A Teacher's Guide to Performance-based Learning and Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Izard, J. (2005). *Trial testing and item analysis in test construction*. Paris: International Institute for Educational Planning/UNESCO
- Johnson, R.L., Penny, J.A., & Gordon, B. (2009). *Assessing performance: Designing, scoring, and validating performance task*. New York: The Guilford Press
- Kaplan, R.M., & Saccuzzo, D.P. (1982). *Psychological testing: Principles, application, and issues*. Monterey: Brooks/Cole Publishig Company.
- Lim, H-Y, & Griffith, W.I. (2011). Practice doesn't make perfect. *MEXTESOL Journal*, 34(3).
- Magno, C. & Ouano, J. (2000). *Designing written assessment of student learning*. New York: Sringer
- Mardapi, J. (2008). *Teknik penyusunan instrumen tes dan non-tes*. Yogyakarta: Mitra Cendekia Press
- Mclaughlin, J.A., & Jordan, G.B. (1999). Logic models: a tool for telling your program's performance story. *Journal of Evaluation and Program Planning*. 22 (1), 65-72.
- McTighe, J., & Ferrara, S. (1998). *Assessing Learning in the Classroom*. Washington, DC: National Education Association.
- Mueller, J. (2014). *Authentic assessment toolbox*. <http://jfmueeller.faculty.noctrl.edu/toolbox/whatisit.htm>
- National Capital Language Resource Center (NCLRC). (2011). *The Essentials of Language Teaching*. <http://nclrc.org/essentials/assessing/alternative.htm>
- Ogle, G.J. (2002). *Towards a formative evaluation tool*. Dissertation: Virginia Polytechnic Institute and State University – USA
- O'Neill, P., Moore, C., & Huot, B. (2009). *A guide to college writing assessment*. Utah: Utah University Press

- Osterlind, S.J. (2002). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats*. New York: Kluwer Academic Publishers.
- Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach* (5th Ed.). Newbury Park: Sage Publications.
- Rukundo, A., & Magambo, J. (2010). Effective test administration in schools: Principles and good practices for test administrators in Uganda. *African Journal of Teacher Education*, 1(1), 166 – 173.
- Setiawan, S. (2020). *Pengertian Rubrik – Tujuan, Jenis, Isi, Macam, Langkah, Manfaat, Contoh, Para Ahli*.
<https://www.gurupendidikan.co.id/pengertian-rubrik/>
- Suyanto (2020). *Penilaian autentik*. <https://suyanto.id/penilaian-autentik/>
- Trach, E. (2018). *How Students Benefit From Project-Based Assessments*.
<https://www.schoology.com/blog/how-students-benefit-project-based-assessments-examples>
- Van der Linden, W.J. (2005). *Linear models for optimal test design*. New York: Springer
- Valadez, J., & Bamberger, M. (1994). *Monitoring and evaluating social programs in developing countries*. Washington: the World Bank
- Widoyoko, S.E.P. (2010). *Evaluasi program pembelajaran: Panduan praktis bagi pendidik dan calon pendidik*. Yogyakarta: Pustaka Pelajar
- Wiggins, G.P. (1993). *Assessing Student Performance*. San Francisco: Jossey-Bass Publishers.
- Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.
- Worthen, B.R., Sanders, J.R., & Fitzpatrick, J.L. (1996). *Program evaluation: Alternative approaches and practical guidelines 2nd Edition*. New York: Longman

