

## TES TERSTANDARISASI

Nisrokha

nisrokhaabduh@yahoo.co.id

### *Abstract*

*Evaluation is very important to see the results of learning that has been done. The form of evaluation in learning is a test, which is one form of evaluation instruments to measure students' understanding of the material that has been taught. Based on the number of students, the learning achievement test is divided into group tests and individual tests. Judging from the study of psychology, the test is divided into general intelligence tests, special ability tests, learning achievement tests, and personality tests. Judging from the form of students' answers, the test is divided into written tests, oral tests, and action tests. Tests can also be divided into power test and speeds test. Judging from the way it was compiled, the test was divided into two namely standardized test and teacher-made test.*

*Keyword: assesment, test, standardized.*

### **A. Pendahuluan**

*Asesment* (penilaian) merupakan salah satu hal penting yang harus diperhatikan dalam untuk mengetahui karakteristik siswa, menentukan strategi pembelajaran dan menilai hasil belajar. Penilaian adalah proses pengumpulan, penafsiran, sintesis informasi dalam rangka mengambil keputusan.<sup>1</sup> Penilaian merupakan istilah yang umum dan mencakup semua metode yang biasa dipakai untuk mengetahui keberhasilan belajar dengan menilai unjuk kerja individu atau kelompok.<sup>2</sup> *Asesmen* tentang karakteristik atau prestasi siswa dapat bersifat informal untuk memperoleh informasi

---

<sup>1</sup> Gage, N.L., & Berliner, C. David. *Educational Psychological*. (Boston: Houghton Mifflin Company, 1992). hlm.568.

<sup>2</sup> Mimin Haryati, *Model dan Teknik Penilaian Pada Tingkat Satuan Pendidikan*, (Jakarta : Gaung Persada Pers, 2010) , hlm. 15.

umum mengenai siswa dimana sebagian besar informasi tersebut disimpan dalam pikiran guru. Asesmen dapat juga bersifat formal yang biasanya meliputi penugasan, kuis, laporan, dan tes.

Tes hanya merupakan alat dan bagian dari proses asesmen. Hal yang perlu diingat adalah bahwa tidak ada tes yang sempurna. Kegunaannya tergantung pada tujuan dan sasarnya. Secara garis besar, ada 2 (dua) jenis tes yang tercakup dalam asesmen, yakni tes terstandarisasi (*standardized test*) dan asesmen otentik (*authentic assessment*) dimana termasuk asesmen alternatif (*alternative assessment*). Satu contoh tes terstandarisasi di Indonesia adalah Ujian Nasional (UN) yang telah menjadi perdebatan di dunia pendidikan kita. Terlepas dari polemik tentang perlu atau tidaknya UN sebagai asesmen hasil belajar, ada baiknya kita memiliki pemahaman yang lebih mendalam mengenai tes terstandarisasi.

Pembahasan tentang tes terstandarisasi dalam makalah ini meliputi sejarah tes terstandarisasi, pengertian, perbedaan antara tes terstandarisasi dan tes yang dibuat oleh guru, jenis tes terstandarisasi, pengelolaan tes terstandarisasi, kegunaan dan kelemahan tes terstandarisasi. Makalah ini juga akan membahas faktor-faktor yang perlu diperhatikan dalam interpretasi tes terstandarisasi.

## **B. Pembahasan**

### **1. Sejarah**

Bukti tes terstandarisasi yang paling awal ditemukan di kekaisaran Cina. Tes tersebut diselenggarakan untuk pelamar pegawai pemerintah sekitar abad ke-6 dan ke-7 Dinasti Sui. Ujian itu meliputi musik, memanah dan berkuda, aritmetika, menulis, dan pengetahuan tentang ritual dan upacara baik yang diadakan secara pribadi maupun di publik. Kemudian, kajian tentang strategi militer, hukum negara, pendapatan dan pajak, pertanian dan geografi ditambahkan dalam ujian tersebut. Praktik ini berlangsung hingga tahun 1898.

Tes terstandarisasi diperkenalkan di Eropa pada abad ke-19 oleh konsul Inggris di Gunagzhou, Cina. Sebelum jenis tes ini diadopsi, tes terstandarisasi bukanlah bagian dari pendidikan Barat karena metode yang dipakai adalah debat yang diwarisi dari Yunani Kuno. Para akademisi Barat lebih menyukai asesmen yang tidak terstandarisasi dengan menggunakan esai tertulis. Jadi, penerapan tes terstandarisasi bukan di Eropa, melainkan di British India, yakni negara-negara yang dijajah oleh Inggris. Terinspirasi oleh penggunaan tes terstandarisasi oleh Cina, pada awal abad ke-19, perusahaan-perusahaan Inggris merekrut dan mengangkat pegawai berdasarkan ujian kompetitif untuk mencegah korupsi dan favoritisme. Praktik ini kemudian diadopsi oleh Inggris daratan pada akhir abad ke-19.

Selanjutnya, tes terstandarisasi menyebar dari Inggris ke negara-negara lain. Tidak hanya diadopsi di Persemakmuran Inggris, tetapi juga Eropa dan kemudian Amerika. Penyebarannya didorong oleh Revolusi Industri. Mengingat besarnya jumlah siswa sekolah selama dan setelah Revolusi Industri, ketika undang-undang pendidikan wajib meningkatkan populasi siswa, asesmen terbuka (*open-ended assessment*) menurun. Terlebih lagi, kurangnya proses yang terstandarisasi menyebabkan sumber kesalahan pengukuran yang cukup besar, karena para penilai menunjukkan favoritisme atau saling tidak sepakat atas jawaban yang berbeda.

Tes terstandarisasi mulai digunakan di Amerika Serikat, pada abad ke-20 yang bermuasal dari Perang Dunia I dan tes Army Alpha dan Beta yang dikembangkan oleh Robert Yerkes dan rekannya. Penggunaan tes terstandarisasi ini semakin diperkuat dengan Undang-Undang *No Child Left Behind* yang disahkan pada tahun 2002 pada pemerintahan George W. Bush. Undang-undang ini mensyaratkan ujian tahunan dalam bidang matematika, membaca, dan kemudian sains pada level kelas tertentu. UU ini menimbulkan pro dan kontra –

sebagian akan dibahas dalam bagian kelebihan dan kekurangan tes terstandarisasi. Pada bulan Maret 2010, Obama mengusulkan perubahan UU tersebut, dan menjanjikan insentif bagi negara bagian jika mereka mengembangkan asesmen yang lebih baik dan lebih mengacu pada standar negara bagian, serta menekankan indikator-indikator lain seperti kehadiran siswa, tingkat kelulusan dan iklim pembelajaran di samping skor tes. Pihak yang mendukung tes terstandarisasi tetap mengandalkan skor tes, meski Obama mengatakan bahwa: "Kita telah sering menggunakan tes ini untuk menghukum siswa dan atau, dalam beberapa kasus, menghukum sekolah."

Tes hasil belajar terstandarisasi secara nasional telah menjadi ciri dominan dalam sistem pendidikan di Indonesia. Selama periode 1965-1971 Ujian Negara telah dilakukan untuk hampir semua bidang studi bagi semua siswa di akhir tingkat sekolah, baik sekolah dasar, menengah tingkat pertama, maupun menengah tingkat atas. Meskipun kebijakan tes non terstandarisasi diterapkan selama tujuh tahun berikutnya, dimana sekolah diberikan kewenangan untuk mendesain dan menyelenggarakan ujian akhir berdasarkan pedoman dari pemerintah pusat. Pada tahun 1980 Indonesia kembali ke sistem ujian terpusat. Evaluasi Belajar Tahap Akhir Nasional, yang disingkat Ebtanas, diterapkan selama 21 tahun.

Ebtanas digunakan untuk 3 (tiga) tujuan. Pertama, Ebtanas digunakan untuk menentukan jalur pendidikan siswa selanjutnya. Jika mereka dapat mencapai skor yang tinggi, mereka berhak memilih sekolah favorit untuk tingkat pendidikan selanjutnya. Kegagalan mereka untuk memperoleh skor yang tinggi akan memaksa mereka mendapatkan sekolah dengan kualitas yang lebih rendah dan seringkali sekolah swasta yang berbiaya lebih mahal. Jika mereka berasal dari keluarga yang berada, mereka akan mendapatkan sekolah swasta yang bagus, yang biasanya mensyaratkan mereka untuk membayar sejumlah

besar sumbangan untuk pembangunan sekolah. Kedua, Ebtanas menjadi alat untuk menyaring siswa ke tingkat pendidikan selanjutnya karena jumlah sekolah makin sedikit seiring dengan tingkat pendidikan yang lebih tinggi. Ketiga, Ebtanas digunakan sebagai alat pemetaan untuk memberi informasi mengenai perbaikan kualitas pendidikan Indonesia. Pada skema Ebtanas, keputusan kelulusan siswa tergantung pada sekolah. Karena itu, siswa yang kurang bagus dalam ujian Ebtanas masih dapat lulus sekolah asalkan hasil belajar mereka di sekolah bagus. Skor Ebtanas hanya sebagian dari komponen skor total untuk kelulusan siswa, di samping ujian berskala provinsi dan nilai sekolah.

Runtuhnya pemerintahan otoriter pada tahun 1998 memberikan harapan baru akan reformasi di segala sektor pembangunan di Indonesia, termasuk pendidikan. Upaya reformasi awal dalam bidang pendidikan menyentuh sejumlah permasalahan, seperti sistem pendidikan desentralisasi, manajemen berbasis sekolah, kurikulum berbasis kompetensi, dan asesmen berdasarkan portofolio. Ada pertimbangan kuat untuk menghapus tes terstandarisasi berskala nasional. Akibatnya, Ebtanas untuk sekolah dasar dihapus pada tahun 2002 dan untuk pendidikan menengah, jumlah mata pelajaran yang diujikan dikurangi. Baik sekolah menengah tingkat pertama maupun tingkat atas hanya mempunyai 3 bidang studi yang diujikan secara nasional dibandingkan 7 mata pelajaran yang diujikan pada model tes terstandarisasi sebelumnya.

Bentuk baru dari ujian terstandarisasi nasional disebut Ujian Akhir Nasional, yang disingkat menjadi UAN. Mata pelajaran yang diujikan adalah Bahasa Indonesia, Bahasa Inggris dan Matematika. Sedangkan tes untuk bidang-bidang studi lainnya diserahkan kepada pihak sekolah dan provinsi untuk memutuskannya. Awalnya, batas lulus UAN adalah 3,01 dari 10. Karena semuanya berjalan lancar sebagaimana yang diharapkan, pada tahun 2004 Kementerian

Pendidikan memutuskan untuk meningkatkan batas kelulusan minimum 4,01. Keputusan ini menghadapi perlawanan keras dari banyak orang tua siswa dan guru karena mereka khawatir bahwa batas tersebut akan sangat sulit bagi sejumlah besar siswa untuk mencapai minimum 4,01 di ketiga mata pelajaran tersebut. Dan, kekhawatiran ini menjadi kenyataan. Kementerian Pendidikan sangat shock dengan hasil buruk yang tak terduga sehingga bereaksi cepat dengan membuat tabel konversi untuk menyeimbangkan hasil belajar siswa. Hal ini mengejutkan banyak orang karena sistem ini membuat skor siswa dengan skor jawaban benar lebih dari setengah dari jumlah total soal diturunkan untuk mensubsidi mereka yang mempunyai skor tes sangat buruk sehingga skor kelompok ini menjadi lebih tinggi.

Di bawah kabinet baru, pada tahun 2005, Kementerian Pendidikan masih melaksanakan tes dengan bentuk serupa, yang diberi nama Ujian Nasional atau disingkat UN. Meskipun UAN telah dikritik pedas, UN masih menggunakan format yang sama, menguji 3 bidang studi, yaitu Matematika, Bahasa Indonesia dan Bahasa Inggris kepada siswa di tahun terakhir sekolah menengah tingkat pertama dan sekolah menengah atas. Selain itu, UN meningkatkan batas kelulusan minimal dari 4,01 menjadi 5,01; yang segera menjadi momok bagi banyak guru, sekolah, dan orang tua yang masih dihantui oleh hasil ujian di tahun sebelumnya. Hal yang lebih menakutkan adalah bahwa UN digunakan sebagai salah satu kriteria yang menentukan kelulusan siswa. Pendek kata, kegagalan untuk mencapai batas minimum UN akan secara otomatis menyebabkan ketidaklulusan tanpa melihat hasil belajar keseluruhan siswa selama sekolah.

Menanggapi fenomena tersebut di atas, pada tanggal 31 Desember 2010, Menteri Pendidikan Nasional mengeluarkan Peraturan No. 45 Tahun 2010 tentang Kriteria Kelulusan Peserta Didik pada Sekolah Menengah Pertama/Madrasah Tsanawiyah, Sekolah Menengah

Pertama Luar Biasa, Sekolah Menengah Atas/Madrasah Aliyah, Sekolah Menengah Atas Luar Biasa, dan Sekolah Menengah Kejuruan Tahun Pelajaran 2010/2011, khususnya Pasal 6 yang mencantumkan syarat kelulusan sebagai berikut:

- a. kelulusan peserta didik dalam UN ditentukan berdasarkan NA.
- b. NA sebagaimana dimaksud pada ayat (1) diperoleh dari nilai gabungan antara Nilai S/M dari mata pelajaran yang diujinasikan dan Nilai UN, dengan pembobotan 40% (empat puluh persen) untuk Nilai S/M dari mata pelajaran yang diujinasikan dan 60% (enam puluh persen) untuk Nilai UN.
- c. peserta didik dinyatakan lulus UN apabila nilai rata-rata dari semua NA sebagaimana dimaksud pada ayat (2) mencapai paling rendah 5,5 (lima koma lima) dan nilai setiap mata pelajaran paling rendah 4,0 (empat koma nol).

## 2. Pengertian

Tes standar memiliki beberapa sinonim kata yaitu *Standardized test* dan Tes baku. Pengertian tes standar secara sempit adalah tes yang disusun oleh satu tim ahli, atau disusun oleh lembaga yang khusus menyelenggarakan tes secara profesional. Tes tersebut diketahui memenuhi syarat sebagai tes yang baik (memenuhi syarat validitas, realibilitas, dan objektivitas). Tes ini dapat digunakan dalam waktu yang relatif lama dan dapat diterapkan pada beberapa obyek mencakup wilayah yang luas. Disamping itu tes standar telah diklasifikasikan sesuai dengan tingkat usia dan kelasnya. Tes standar bertujuan untuk mengukur kemampuan peserta didik dalam tiga aspek, yaitu kedudukan belajar, kemajuan belajar, dan diagnosik.<sup>3</sup>

Secara historis, dahulu tes terstandarisasi digunakan untuk membandingkan skor tes antar siswa, kelas, sekolah, atau wilayah.

---

<sup>3</sup> Zainal Arifin, *Evaluasi Pembelajaran*, (Bandung: Remaja Rosdakarya), hlm.120

Kini, tes terstandarisasi digunakan untuk keputusan resiko tinggi (*high-stakes*), misal promosi, kelulusan, rating dan insentif guru, serta evaluasi sekolah dan wilayah. Menurut Kubiszyn & Borich, tes disebut terstandarisasi karena dikelola dan dinilai sesuai dengan prosedur yang spesifik dan seragam (standar). Tes terstandarisasi dibuat oleh ahli pembuat tes yang biasanya dibantu oleh para ahli bidang, guru, dan pengelola sekolah.<sup>4</sup> Tes, pengelolaan dan penilaian dikatakan terstandarisasi karena setiap orang yang menjalani tes tersebut memperoleh perlakuan yang sama.<sup>5</sup>

Gage & Berliner mengungkapkan bahwa tes terstandarisasi adalah tes yang diberikan ke sejumlah besar sampel representatif dari beberapa populasi sehingga skor dari tes tersebut dapat dibandingkan antar sampel yang terlibat.<sup>6</sup> Tes terstandarisasi terdiri dari tes terstandarisasi acuan norma (*norm-referenced standardized test*) dan tes terstandarisasi acuan kriteria (*criterion-referenced standardized test*), di mana keduanya saling berkaitan. *Norm-referenced standardized test* memberikan norma-norma yang memungkinkan perbandingan skor yang diperoleh antar siswa. Sedangkan *Criterion-referenced standardized test* memberikan informasi langsung mengenai seberapa baik siswa mencapai kriteria tertentu.

### 3. Tes Terstandarisasi Vs Tes Buatan Guru (*Teacher-Made Test*)

Karena tes terstandarisasi diperlukan untuk mengevaluasi siswa berdasarkan kriteria di luar yang ditetapkan untuk suatu kelas oleh guru, maka ada beberapa perbedaan antara tes terstandarisasi dengan tes buatan guru sebagaimana disajikan pada tabel berikut:<sup>7</sup>

---

<sup>4</sup> Kubiszyn, Tom & Borich, Gary. *Educational Testing and Measurement: Classroom Application and Practice*. Eighth Edition. Hoboken, (New Jersey: John Wiley & Sons, Inc., 2007), hlm. 349.

<sup>5</sup> Kubiszyn, Tom & Borich, Gary. *Educational Testing*, hlm. 354.

<sup>6</sup> Gage, N.L., and Berliner, C. David, *Educational Psychological*. hlm. 592

<sup>7</sup> Gage, N.L., and Berliner, C. David, *Educational Psychological*. hlm.593.



<b>Kriteria</b>	<b>Tes Terstandarisasi</b>	<b>Tes Buatan Guru</b>
Reliabilitas	Reliabilitas biasanya tinggi, seringkali 0.9 untuk tes acuan norma.	Reliabilitas jarang diukur; jika diukur, rata-rata sekitar 0.6.
Validitas	Validitas kriteria biasanya ditentukan; validitas konstruk didiskusikan. Validitas isi biasanya tinggi untuk tes acuan kriteria; sulit dinilai dalam kasus tes acuan norma, namun dapat diukur.	Validitas konstruk dan kriteria tidak diketahui.  Validitas isi biasanya tinggi jika prosedur sistematis dipergunakan dalam pembuatan tes.
Pengukuran Isi yang Diajarkan	Pengukuran keterampilan dasar dan hasil yang kompleks dapat disesuaikan dengan situasi setempat	Pengukuran isi unik bagi kurikulum lokal. Adaptasi tes yang berkesinambungan dimungkinkan. Cenderung menekankan pengetahuan daripada hasil pada tataran yang lebih tinggi.
Persiapan Siswa	Belajar biasanya tidak membantu siswa memperoleh skor yang lebih baik kecuali tes tersebut berkaitan erat dengan kurikulum lokal.	Belajar biasanya membantu siswa memperoleh skor yang lebih baik.
Kualitas Butir Soal	Kualitas biasanya sangat tinggi; butir soal dibuat oleh ahli tes, diujicobakan, dan direvisi sebelum digunakan dalam tes.	Kualitas tidak diketahui, beragam.

<b>Kriteria</b>	<b>Tes Terstandarisasi</b>	<b>Tes Buatan Guru</b>
Jenis Butir	Biasanya pilihan ganda.	Berbagai jenis soal dipergunakan.
Pengelolaan dan Penilaian Tes	Prosedur distandarisasi dan ajeg antar kelas. Instruksi khusus diberikan. Penghitungan skor dilakukan oleh mesin.	Prosedur tes bersifat fleksibel.  Penghitungan skor dilakukan oleh guru.
Interpretasi Skor	Skor biasanya dibandingkan dengan norma nasional. Skor biasanya dapat dipercaya ( <i>high confidence band</i> ). Panduan tes membantu dalam menafsirkan informasi dan pengambilan keputusan.	Skor dapat diinterpretasikan sesuai dengan norma dari kelompok dekat atau lokal.

Urutan langkah yang ditempuh dalam penyusunan tes buatan guru, adalah (1) menentukan tujuan mengadakan tes; (2) mengadakan pembatasan terhadap bahan yang akan diteskan; (3) merumuskan tujuan instruksional khusus dari tiap bagian bahan; (4) menderetkan semua TIK (tujuan instruksional khusus) dalam tabel persiapan yang memuat aspek tingkah laku pada TIK; (5) menyusun tabel spesifikasi yang memuat pokok materi, aspek berpikir yang diukur beserta imbalan antara kedua aspek tersebut; (6) menuliskan butir-butir soal, didasarkan pada TIK yang sudah dituliskan pada tabel TIK dan aspek tingkah laku yang dicakup.<sup>8</sup>

---

<sup>8</sup> Suharsimi Arikunto, *Dasar-dasar Evaluasi Pembelajaran*, (Jakarta: Bumi Aksara, 2011), hlm. 147

Contoh:

Tabel TIK dan Aspek Tingkah Laku yang Dicakup

<b>TIK Aspek Tingkah Laku</b>	<b>Ingatan</b>	<b>Pemahaman</b>	<b>Aplikasi</b>	<b>Keterangan</b>
1. Siswa dapat menjumlahkan 2 bilangan bersusun.		v	v	
2. Siswa dapat menerangkan hukum komulatif dan seterusnya.	v	v		

#### 4. Jenis Tes Terstandarisasi

Tes terstandarisasi dapat dibedakan menjadi dua jenis, yakni tes bakat (*aptitude test*) dan tes prestasi (*achievement test*). Tes bakat meliputi tes kecerdasan (Stanford-Binet, Wechsler), tes bakat skolastik, tes kecerdasan akademik, tes minat, tes kepribadian (*Adjective Checklist, Edwards Personal Preference Schedule/EPPS, Rorschach Inkblot Technique*), atau tes kemampuan umum. Tes ini umumnya merupakan tes acuan norma dan memberikan informasi untuk konseling siswa. Di perguruan tinggi, dunia kerja dan industri, tes tersebut digunakan untuk seleksi dan penempatan. Tes bakat yang sering digunakan adalah tes bakat verbal, matematis, spasial, dan mekanis. Contoh dari tes ini, antara lain: TPA (*Tes Potensi Akademik*), TOEFL (*Test of English as a Foreign Language*), IELTS (*International English Language Testing System*), GRE (*Graduate Record Examinations*), GMAT (*Graduate Management Admission Test*), dan lain sebagainya.

Sedangkan tes prestasi digunakan untuk mengukur prestasi atau pencapaian hasil belajar siswa pada tujuan pengajaran dalam suatu

mata pelajaran atau kurikulum. Gage dan Berliner menjelaskan bahwa jenis tes ini dibuat oleh ahli dalam bidang kurikulum dan dikelola untuk sejumlah besar siswa pada tingkat dan mata pelajaran tertentu.<sup>9</sup> Tes ini menunjukkan seberapa baik siswa, sekolah, wilayah tertentu telah mencapai hasil tertentu dibandingkan dengan kelompok norma. Contoh dari jenis tes ini ialah Ujian Nasional (UN).

Perbedaan mendasar antara tes bakat dan tes prestasi adalah fungsinya. Tes bakat digunakan untuk memprediksi prestasi atau hasil dari pengalaman belajar yang akan terjadi, sedangkan tes prestasi digunakan untuk mengukur dan mengevaluasi prestasi atau hasil dari pengalaman belajar sebelumnya. Namun, tes prestasi juga dapat digunakan untuk memprediksi prestasi di kemudian hari.

Perbedaan juga dapat diamati dari isinya. Dalam tes hasil belajar, isi harus berkaitan dengan materi yang diajarkan di sekolah, yakni mencakup apa yang telah dibaca dari buku acuan, dibahas dalam kelas, tugas yang dikerjakan di rumah, dan materi yang telah diajarkan oleh guru. Tes bakat berupa tes acuan norma, dan tidak dapat berupa tes acuan kriteria. Sedangkan tes prestasi dapat berupa tes acuan norma dan tes acuan kriteria.

## **5. Pengelolaan Tes Terstandarisasi**

Pengelolaan tes sangat krusial dalam tes terstandarisasi dan guru berperan penting. Guru harus menjelaskan kepada siswa tentang cara mengerjakan tes tersebut – mengenai aturan, format, jenis butir soal, dan waktu yang dibutuhkan – dan mencoba mengurangi kecemasan siswa sekaligus meningkatkan motivasi mereka agar mereka mengerjakannya dengan baik. Selain itu, guru pun harus menyiapkan diri dengan membaca tes, perintah pengerjaan soal, melakukan uji coba, dan terlebih dahulu mengumpulkan bahan yang diperlukan.

---

<sup>9</sup> Suharsimi Arikunto, *Dasar-dasar Evaluasi*, hlm. 598

Ketika tes terstandarisasi dianggap penting dalam evaluasi siswa, guru, dan sekolah, persiapan menghadapi tes perlu dilakukan. Terdapat beberapa cara untuk membantu siswa mempersiapkan diri menghadapi tes terstandarisasi, yaitu:<sup>10</sup>

- a. memberikan latihan dengan format soal serupa kepada siswa. Apabila tes menggunakan format pilihan ganda, maka latihan dengan format serupa dengan singkat, rutin, dan tes di kelas.
- b. menyarankan siswa untuk meninggalkan soal-soal yang sulit dan menghabiskan waktu dan kemudian kembali lagi ke soal tersebut jika selesai menjawab soal lain.
- c. jika tidak ada pengurangan nilai karena menebak dalam ujian, sarankanlah siswa untuk selalu mengisi jawaban. Apabila ada pengurangan nilai untuk tebakan yang salah, siswa seharusnya masih didorong untuk menjawab, khususnya jika mereka dapat mempersempit pilihan dengan mencoret satu pilihan atau lebih.
- d. menyarankan siswa untuk membaca semua pilihan dalam ujian pilihan ganda sebelum memilih salah satu di antaranya. Terkadang ada lebih dari satu jawaban benar, tetapi hanya satu jawaban yang menjadi jawaban yang lebih baik.
- e. menyarankan kepada siswa agar mereka menggunakan semua waktu yang tersedia. Kalau mereka selesai lebih awal, mereka sebaiknya melihat lagi jawaban mereka.

## **6. Kelebihan Tes Terstandarisasi**

Salah satu kelebihan tes terstandarisasi adalah bahwa jenis tes ini telah dipersiapkan selama beberapa tahun, biasanya butir soal dibuat

---

<sup>10</sup> Slavin, Robert E. *Psikologi Pendidikan: Teori dan Praktik*. Jilid 2. Edisi Kesembilan. (Jakarta: Indeks. 2011). hlm. 353-354.

dengan sangat hati-hati, dan penghitungan skor dilakukan oleh mesin sehingga kesalahan penghitungan dapat diminimalisir, interpretasi skor konsisten. Oleh karenanya, pengukuran tidak dipengaruhi oleh subjektivitas atay bias. Selain itu, kesalahan dalam administrasi tes juga dapat dikurangi karena dalam tes dicantumkan arahan khusus mengenai apa yang harus dikatakan dan dilakukan selama tes berlangsung. Kelebihan kedua ialah hasil tes yang terstandarisasi. Meskipun beberapa siswa mungkin memperoleh skor yang lebih rendah ketika diberikan ujian tertentu, perbedaannya akan sistematis jika menggunakan tes terstandarisasi dengan jenis yang sama.

Kelebihan ketiga adalah bahwa tes terstandarisasi berbiaya lebih murah dan lebih mudah dibanding tes alternatif. Berbiaya rendah karena pengelolaannya dapat bersifat massal atau diadakan untuk sejumlah besar orang sekaligus. Tes terstandarisasi juga lebih mudah dan cepat dalam penghitungan skornya.

Pada proses rekrutmen pegawai, tes terstandarisasi membantu memprediksi kesuksesan secara lebih akurat. Tes terstandarisasi disarankan karena wawancara tidak memberi informasi yang cukup tentang kemampuan seseorang. Dalam dunia kerja, dokter, pengacara, broker real-estate, dan pilot harus mengikuti tes terstandarisasi untuk memastikan bahwa mereka memiliki pengetahuan yang cukup bagi profesionalisme mereka.

## **7. Kelemahan Tes Terstandarisasi**

Keseragaman pengelolaan tes terstandarisasi dimana setiap orang mendapat perlakuan yang sama menimbulkan permasalahan. Individualisasi tes dengan membantu siswa yang lamban atau mendorong siswa yang lebih pandai untuk mengerjakan tes dengan lebih cepat tidaklah biasa karena hal ini dianggap melanggar prosedur tes terstandarisasi. Setiap orang mendapat perlakuan yang sama dan karenanya tes beserta administrasi dan penghitungan skor disebut

”standarisasi”. Kesamaan ini yang memungkinkan perbandingan reliabel dapat dilakukan.

Kelemahan tes terstandarisasi yang sering disoroti ialah adanya peningkatan pelanggaran prosedur penyelenggaraan tes untuk meningkatkan skor. Hal yang lebih memprihatinkan, pelanggaran-pelanggaran ini mungkin didukung secara diam-diam atau terang-terangan oleh pengelola sekolah dan wilayah. Terlepas apakah hal ini disengaja atau tidak disengaja, pelanggaran tersebut mempengaruhi hasil perbandingan prestasi siswa secara reliabel dan valid.

Selanjutnya, jenis tes ini mempunyai kelemahan dalam menilai pengetahuan yang lebih mendalam. Tes terstandarisasi terutama bukanlah alat pengukuran yang sangat baik untuk kinerja dan intelegensi siswa secara individu karena sistemnya sangat sederhana. Tes standarisasi dapat digunakan apakah seorang siswa mengetahui kapan Perang Bubat terjadi, namun tidak dapat mengetahui seberapa jauh siswa memahami faktor penyebab terjadinya perang tersebut atau seberapa dalam siswa mencerna dan memikirkan permasalahan yang lebih besar terkait dengan peristiwa bersejarah itu.

Soal pilihan ganda yang sering menjadi format untuk tes terstandarisasi mendorong siswa untuk mengambil jalan sederhana dan pemikiran bahwa hanya ada 2 kemungkinan jawaban, yakni benar atau salah. Padahal, dalam dunia nyata, konsep benar dan salah tidaklah mutlak. Karena itu, bagi yang kontra terhadap tes terstandarisasi, tes ini mendorong siswa untuk dapat mengerjakan tes dengan baik, tetapi tidak menyiapkan siswa untuk kehidupan orang dewasa yang produktif. Terlebih, bagi siswa laki-laki yang menyukai model soal pilihan ganda sebagai bagian dari permainan, maka jawaban yang dipilih bisa jadi hanya sekedar tebak-tebakan.

Selain itu, kekurangan tes terstandarisasi adalah bahwa jenis tes ini tidak dapat mengukur kreativitas, inisiatif, pemikiran konseptual,

dan imajinasi. Hal ini senada dengan pendapat yang dikemukakan oleh Gerald W. Bracey, seorang peneliti pendidikan, bahwa tes terstandarisasi tidak dapat mengukur "kreativitas, pemikiran kritis, keteguhan, motivasi, keuletan, rasa keingintahuan, ketahanan, kehandalan, antusiasme, empati, kesadaran diri, disiplin diri, kepemimpinan, kewiraan, keberanian, rasa belas kasih, kreativitas, rasa keindahan, rasa kagum, kejujuran, intergritas." Bill Ayers juga mendukung pendapat Bracey. Ia menyatakan bahwa tes terstandarisasi tidak dapat mengukur inisiatif, kreativitas, imajinasi, pemikiran konseptual, rasa ingin tahu, upaya, ironi, keputusan, komitmen, nuansa, niat baik, refleksi etis, atau sifat dan atribut berharga lainnya. Hal yang dapat diukur dan dihitung oleh tes terstandarisasi adalah fakta dan fungsi khusus, pengetahuan isi, dan aspek pembelajaran yang paling kurang menarik dan paling bermakna.

Dampak negatif lain, terutama tes hasil belajar terstandarisasi, adalah berkurangnya inisiatif dan kreativitas guru dalam pembelajaran. Hal ini disebabkan mereka harus memusatkan perhatian lebih banyak pada bahan yang akan diujikan. Mereka pun merasa bahwa apa yang mereka akan ajarkan akan sia-sia belaka jika pada akhirnya itu tidak dapat membantu siswa lulus sekolah.

#### **8. Faktor yang Harus Diperhatikan dalam Interpretasi Tes Terstandarisasi**

Menurut Kubiszyn dan Borich, untuk memperoleh keahlian akan penggunaan tes terstandarisasi secara tepat, ada beberapa faktor yang harus diperhatikan. Faktor-faktor tersebut dapat dikelompokkan menjadi dua kategori, yaitu faktor yang berhubungan dengan tes itu sendiri dan faktor yang terkait dengan siswa.<sup>11</sup> Faktor yang terkait dengan tes dapat membatasi kemampuan interpretasi hasil tes yang

---

<sup>11</sup> Kubiszyn, Tom & Borich, Gary. *Educational Testing* hlm. 360-369.



dikarenakan oleh permasalahan yang melekat pada tes itu sendiri, penggunaannya, pengelolaan, atau penghitungan skornya. Faktor-faktor tersebut meliputi:

- a. Reliabilitas. Tes terstandarisasi harus mempunyai koefisien skor sebesar 0.95 untuk konsistensi internal (*internal consistency*) dan 0.90 untuk *test-retest*. Jika tes memiliki tingkat koefisien yang tinggi, hampir dapat dipastikan bahwa skor dari tes tersebut andal atau reliabel. Namun demikian, jika koefisien reliabilitas konsistensi internal hanya 0.92, tentu saja tes masih dapat dikatakan reliabel.
- b. Akurasi. Jika skor tes reliabel, maka tes juga akurat. Dengan kata lain, hasil yang diperoleh cenderung mendekati tingkat prestasi siswa yang sesungguhnya. Tingkat signifikansi yang ditentukan dapat berkisar 95% ataupun 99%.
- c. Bukti validitas yang berhubungan dengan kriteria. Ada dua jenis bukti validitas yang berhubungan dengan kriteria, yaitu: prediktif dan konkuren. Tes hasil belajar terstandarisasi digunakan untuk mengetahui pembelajaran sebelumnya sehingga koefisien validitas prediktifnya jarang dilaporkan dan seringkali koefisien validitas konkuren yang dilaporkan. Koefisien validitas konkuren adalah perhitungan numerik tentang seberapa besar sebuah tes mempunyai korelasi dengan tes lainnya. Namun, penggunaan koefisien tersebut untuk menentukan validitas suatu tes hasil belajar merupakan hal yang sulit. Karena isi dan kekhususan dari tes yang berbeda akan beragam antara satu sama lain, tes hasil belajar yang valid mungkin menghasilkan koefisien validitas konkuren 0.6 dengan Tes X dan 0.9 dengan Tes Y, dan tes tersebut masih dapat dikatakan valid.

Sedangkan faktor yang terkait dengan siswa, antara lain sebagai berikut.

a. Bahasa dan sosiokultural

Kemampuan bahasa dapat mempengaruhi hasil tes terstandarisasi siswa, misal kurangnya kemampuan bahasa siswa terhadap bahasa yang digunakan dalam instruksi pengerjaan butir soal dapat mempengaruhi pemahamannya. Masalah bahasa terutama sekali terlihat pada tes bahasa kedua, misal tes kemampuan Bahasa Inggris yang terstandarisasi secara internasional. Interpretasi skor tes antara siswa yang menggunakan Bahasa Inggris sebagai bahasa ibu akan berbeda dengan siswa yang menggunakan bahasa tersebut sebagai bahasa kedua ataupun bahasa ketiga. Faktor sosiokultural juga dapat membawa pengaruh negatif pada kinerja siswa dalam tes terstandarisasi. Nilai-nilai budaya, seperti kurangnya daya saing, kurang seringnya mengerjakan tes terstandarisasi untuk tes masuk sekolah dan tes untuk mendapatkan pekerjaan dapat mempengaruhi kecepatan, kecermatan siswa dalam mengerjakan tes terstandarisasi.

Kesenjangan dalam sosiokultural dapat diakibatkan oleh kelaparan dan gizi buruk, kemiskinan, pandangan tentang prestasi dalam keluarga, dan berbagai perbedaan perilaku antar budaya, misal daya saing, asertif, dan kepatuhan pada tokoh yang berkuasa. Oleh karena itu, kemampuan bahasa dan perbedaan sosiokultural harus selalu menjadi pertimbangan dalam merencanakan, memilih, mengelola, dan menafsirkan hasil tes terstandarisasi.

b. Usia, Jenis Kelamin dan Tahap Perkembangan

Usia siswa mempunyai tahap perkembangan, baik fisik maupun psikis yang berbeda. Perbedaan ini juga berkaitan dengan jenis kelamin. Anak perempuan cenderung untuk

memperoleh kemampuan verbal yang lebih cepat daripada anak laki-laki. Karena itu, jangan mengabaikan faktor-faktor tersebut dalam menginterpretasikan skor tes.

c. Motivasi

Ketika motivasi siswa tinggi, ia dapat memperoleh skor yang sesuai dengan prestasi sesungguhnya. Namun, untuk beberapa jenis tes terstandarisasi yang dianggap tidak akan mempengaruhi nilai siswa di sekolah, motivasi siswa dapat berkurang atau rendah sehingga kinerja selama pengerjaan tes juga menurun.

d. Kondisi emosional saat tes

Kondisi emosi siswa seperti depresi, stres, gugup, khawatir, pada waktu tes dapat mempengaruhi skor tes. Bahkan emosi negatif siswa sebelum hari dilaksanakannya tes pun dapat membawa dampak pada skor tes.

e. Keterbatasan

Ada sebagian kecil siswa di sekolah umum yang mempunyai masalah keterbatasan fisik dan/atau emosional yang dapat menghambat prestasi mereka. Asesmen alternatif dapat saja dijadikan satu solusi, tetapi hal ini akan mengurangi komparabilitas skor dari siswa dengan keterbatasan tersebut.

f. Bakat

Bakat siswa juga merupakan salah satu faktor yang harus diperhatikan dalam menginterpretasikan skor tes terstandarisasi karena bakat, dalam ini khususnya IQ, memberikan informasi tentang perkiraan pagu pencapaian prestasi siswa. Oleh sebab itu, tes IQ perlu diberikan pada tingkatan kelas tertentu agar dapat dilihat apakah seorang siswa telah menunjukkan prestasi yang diharapkan tercapai sesuai dengan tingkat IQnya, atau malah

terjadi kesenjangan antara tes bakat dengan tes hasil belajar atau prestasi.

Seorang siswa yang mencapai skor tes hasil belajar yang lebih tinggi daripada IQnya dapat disebut *overachiever*. Sebaliknya, siswa dapat dikatakan sebagai *underachiever* jika skor tes hasil belajar lebih rendah daripada IQnya. Namun demikian, guru tidak boleh serta-merta memberi label *overachiever-underachiever* kepada siswa yang menunjukkan gejala tersebut karena interpretasi skor tes merupakan hal yang kompleks.

Selain itu untuk menyusun tes standar terdapat beberapa prosedur yang harus dilakukan dan memakan waktu yang lama. Prosedur-prosedur tersebut adalah:

- a. Penyusunan. Dalam penyusunan tes standar, dijelaskan tentang aspek-aspek yang akan diukur, misalnya kemampuan membaca, pembendaharaan pengetahuan umum, sikap, dan lain-lain. Pihak penyusun dalam tes standar juga perlu dicantumkan untuk menunjukkan jaminan mutu dan kesahihan tes standar tersebut.
- b. Uji Coba. Untuk menguji validitas dan realibilitas tes standar, perlu dilakukan percobaan-percobaan terhadap sampel yang cukup besar dan representatif.
- c. Analisa. Tes standar biasanya telah dianalisis secara statistik dan diuji secara empiris oleh para pakar agar dapat dikatakan valid untuk digunakan secara umum. Analisis soal tes bertujuan untuk mengidentifikasi soal yang baik dan soal yang jelek. Hal-hal yang harus diperhatikan dalam analisa soal tes adalah:
  - 1) Taraf Kesukaran. Suatu soal dikatakan baik jika soal tersebut tidak terlalu mudah dan tidak terlalu sukar atau soal yang terlalu mudah.
  - 2) Daya Pembeda (D). Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang

berkemampuan tinggi dengan siswa yang berkemampuan rendah.

- 3) Pola Jawaban Soal. Pola jawaban soal adalah distribusi testee dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Dari jawaban soal dapat ditentukan apakah pengecoh (distractor) berfungsi sebagai pengecoh dengan baik atau tidak.<sup>12</sup>

### C. Penutup

Tidak ada tes, baik yang terstandarisasi maupun tidak, yang sempurna dan penggunaannya memerlukan pertimbangan yang cermat dan hati-hati. Pihak-pihak yang terlibat harus memperhatikan faktor-faktor terkait, baik faktor yang berhubungan dengan tes maupun karakteristik siswa sebagai sasarannya. Terlepas dari kelemahannya, tes terstandarisasi hampir dapat dipastikan akan selalu mendapat tempat dan dibutuhkan berdampingan dengan tes berbasis kinerja (*performance-based test*) dan tes berbasis portofolio (*portfolio-based test*) sebagai bagian dari asesmen. Hal yang perlu diperhatikan adalah pemilihan jenis tes sesuai dengan tujuan dan sasaran yang hendak dicapai.

### DAFTAR PUSTAKA

- Daryanto. (2008). *Evaluasi Pendidikan*. Jakarta: Rieneka Cipta.
- Furqon. (2004). Masih perlukah ujian nasional ?. *Pikiran Rakyat*.  
<http://www.pikiran-rakyat.com/cetak/1204/23/0804.htm>, diakses pada tanggal 26 Januari 2012
- Gage, N. L., & Berliner, C. David. (1992). *Educational Psychological*. Boston: Houghton Mifflin Company.

---

<sup>12</sup> Daryanto, *Evaluasi Pendidikan*, (Jakarta: Rieneka Cipta, 2008), hlm. 179

- Kubiszyn, Tom & Borich, Gary. (2007) *Educational Testing and Measurement: Classroom Application and Practice*. Eighth Edition. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Mimin, Haryati. (2010). *Model dan Teknik Penilaian Pada Tingkat Satuan Pendidikan*. Jakarta: Gaung Persada Pers, 2010.
- Santrock, John W. (2008). *Psikologi Pendidikan*. Edisi Kedua. Jakarta: Kencana.
- Slavin, Robert E. (2011). *Psikologi Pendidikan: Teori dan Praktik*. Jilid 2. Edisi Kesembilan. Jakarta: Indeks.
- Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 45 Tahun 2010 Tentang Kriteria Kelulusan Peserta Didik pada Sekolah Menengah Pertama/Madrasah Tsanawiyah, Sekolah Menengah Pertama Luar Biasa, Sekolah Menengah Atas/Madrasah Aliyah, Sekolah Menengah Atas Luar Biasa, dan Sekolah Menengah Kejuruan Tahun Pelajaran 2010/2011.
- Arikunto, Suharsimi. (2011). *Dasar-dasar Evaluasi Pembelajaran*. Jakarta: Bumi Aksara.
- Arifin, Zainal. (2002). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- <http://standardizedtests.procon.org/> diakses pada tanggal 25 Januari 2012.
- <http://www.xtimeline.com/timeline/The-History-of-Standardized-Testing-in-Education>. diakses pada tanggal 26 Januari 2012.