

## **KUALITAS INSTRUMEN EVALUASI PEMBELAJARAN**

Oleh:

**Buhori, M.Pd.I**

STAI - Sulthan Syarif Hasyim Siak Sri Indrapura Riau

**Buhoriri.m.pd.i@gmail.com**

### **ABSTRAK**

Tujuan penelitian ini adalah untuk mengetahui kualitas instrumen evaluasi pembelajaran. Setelah penulis melakukan analisis dari berbagai referensi maka disimpulkan bahwa baik buruknya suatu tes atau suatu alat evaluasi dapat kita tinjau dari beberapa segi, yaitu: validitas, reliabilitas, tingkat kesukaran dan daya beda. Ketika mengkonstruksi atau memilih tes atau instrumen evaluasi, pertanyaan yang sering muncul dari pemikiran seorang guru adalah pada kondisi apakah interpretasi skor yang dihasilkan dari penggunaan instrumen tepat, bermanfaat, dan dapat digunakan pada keadaan sekolah yang ada. Tanpa melihat jenis instrumen dan kemampuan guru dalam memilih jenis tes dalam melakukan evaluasi terhadap siswa, semua instrumen seperti yang disebut di atas agar memiliki kemampuan mengevaluasi perlu memenuhi karakteristik, yaitu valid, reliabel, dan dapat digunakan.

**Kata Kunci:** Kualitas Instrumen Evaluasi

### **1. PENDAHULUAN**

Ketika mengkonstruksi atau memilih tes atau instrumen evaluasi, pertanyaan yang sering muncul dari pemikiran seorang guru adalah pada kondisi apakah interpretasi skor yang dihasilkan dari penggunaan instrumen tepat, bermanfaat, dan dapat digunakan pada keadaan sekolah yang ada? Ada banyak macam tes dan bervariasi pula kegunaannya tergantung dari tujuan yang hendak dicapai. Tes dan instrumen evaluasi dalam kategori *paper and pencil* dapat digunakan untuk tes pencapaian hasil belajar, yang di dalamnya tercakup fokus-fokus penting tujuan pilihan (*selection*), penempatan, diagnosis dan sertifikasi profesi. Tes kecakapan (*aptitude test*) digunakan untuk memprediksi keberhasilan siswa dalam prospek belajar, atau dalam prospek bekerja. Tes penghargaan (*appraisal test*) digunakan untuk mengestimasi keberhasilan belajar siswa, perkembangan sosial, dan juga untuk mengetahui problem siswa dan pengaruh evaluasi penerapan program terhadap keadaan anak.

Tanpa melihat jenis instrumen dan kemampuan guru dalam memilih jenis tes dalam melakukan evaluasi terhadap siswa, semua instrumen seperti yang disebut di atas agar memiliki kemampuan mengevaluasi perlu memenuhi karakteristik tertentu yang hendak diuraikan secara ringkas dan jelas pada bab ini. Ketiga karakteristik penting tersebut, yaitu valid, reliabel, dan dapat digunakan (*usability*).<sup>1</sup> Kegiatan evaluasi juga diatur dalam Undang-Undang Republik Indonesia No. 20 Tahun 2003 tentang Sistem Pendidikan

---

<sup>1</sup> Sukardi, *Evaluasi Pendidikan: Prinsip dan Operasionalnya*, (Jakarta: PT. Bumi Aksara, 2010), hlm. 29

Nasional Bab XVI Pasal 58 ayat 1 yang berbunyi: evaluasi hasil belajar siswa dilakukan untuk memantau proses, kemajuan, dan peningkatan hasil belajar siswa. hasil belajar secara berkelanjutan Supriyansah dan Miranda. Oleh karena itu, tercapai atau tidaknya tujuan pembelajaran dapat dilihat dari hasil evaluasi. Oleh karena itu proses pembelajaran tidak dapat dipisahkan dari evaluasi. Ada beberapa teknik yang dapat digunakan dalam evaluasi yaitu teknik tes dan non tes.<sup>2</sup> Dalam hal ini, evaluasi pendidikan adalah salah satu bagian dari kegiatan yang dilakukan oleh seorang guru untuk mendukung agar tercapainya tujuan pendidikan tersebut, dan diantara evaluasi yang dilakukan oleh guru yaitu evaluasi hasil belajar, dimana evaluasi ini dilakukan untuk mengukur sejauh mana pengetahuan dan keterampilan siswa setelah menerima.<sup>3</sup>

## 2. METODE PENELITIAN

Penelitian yang digunakan adalah studi literatur, dengan objek penelitian Kualitas Instrumen Evaluasi Pembelajaran dengan mengambil referensi dari berbagai sumber termasuk jurnal. Dengan cara menganalisis dari berbagai sumber tersebut lalu disimpulkan tentang Instrumen Evaluasi Pembelajaran tersebut.

## 3. PEMBAHASAN

Pada dasarnya instrumen penilaian merupakan alat yang dipakai guna mengukur ketercapaian tujuan pembelajaran yang ditargetkan. instrumen penilaian dikenal juga dengan alat evaluasi (instrumen evaluasi), yang mana berguna untuk mengukur pencapaian dari hasil program pembelajaran.<sup>4</sup>

Baik buruknya suatu tes atau suatu alat evaluasi dapat kita tinjau dari beberapa segi, yaitu:

1. Validitas.
2. Reliabilitas.
3. Tingkat kesukaran.
4. Daya beda.

Di bawah ini akan kami uraikan lebih lanjut tentang keempat faktor tersebut.

---

<sup>2</sup> Jurnal, Hidayatul Maulidah, Analisis Kualitas Instrumen Tes Fisika Kelas X Menggunakan Software Anates, Volume 2 Nomor 3, Desember 2022, hlm, 154.

<sup>3</sup> Sawaluddin, Langkah-Langkah Dan Teknik Evaluasi Hasil Belajar Pendidikan Agama Islam, Vol. 6, No. 1, Januari – Juni 2020 (13-24), hlm, 13

<sup>4</sup> Sahrul Dahri, Pengembangan Instrumen Evaluasi Belajar Via Ponsel Pada Mata Kuliah Pendidikan Agama Islam, Jurnal Lentera Vol. 21 No. 1 (Maret, 2022), hlm, 128

## 1. Validasi Tes

Suatu alat pengukur dapat dikatakan alat pengukur yang valid apabila alat pengukur tersebut dapat mengukur apa yang hendak diukur secara tepat. Misalnya barometer adalah suatu alat yang valid untuk mengukur tekanan udara. Tetapi alat ini tidak valid untuk mengukur suhu. Demikian pula dalam alat-alat evaluasi. Suatu tes hasil belajar dapat dikatakan tes yang valid apabila tes tersebut betul-betul dapat mengukur hasil belajar. Jadi bukan sekedar mengukur daya ingatan atau kemampuan bahasa saja misalnya.

Validitas, suatu tes dapat ditinjau dari beberapa segi, seperti yang akan kami uraikan di bawah ini.

### a. Validitas ramalan (Predictive validity).

Validitas ramalan artinya ketepatan (kejituan) daripada suatu alat pengukur ditinjau dari kemampuan tes tersebut untuk meramalkan prestasi yang dicapainya kemudian. Misalnya suatu tes hasil belajar dapat dikatakan mempunyai validitas ramalan yang tinggi, apabila hasil yang dicapai oleh anak dalam tes tersebut betul-betul dapat meramalkan sukses tidaknya anak-anak dalam pelajaran-pelajaran yang akan datang. Cara yang dipergunakan untuk menilai tinggi rendahnya validitas ramalan ini ialah dengan jalan mencari korelasi antara nilai-nilai yang dicapai oleh anak-anak dalam tes tersebut dengan nilai-nilai yang dicapainya kemudian.

Apabila koefisien korelasi yang diperoleh cukup tinggi, maka berarti validitas ramalan tes tersebut tinggi. Sebaliknya pula apabila koefisien korelasi yang diperoleh rendah, maka berarti pula validitas ramalan tes tersebut rendah.

### b. Validitas bandingan (Concurrent validity).

Validitas bandingan artinya kejituan daripada suatu tes dilihat dari korelasinya, terhadap kecakapan yang telah dimiliki saat ini secara riil. Perbedaan antara validitas, ramalan dengan validitas bandingan ialah dilihat dari segi waktunya. Validitas ramalan melihat hubungannya dengan masa yang akan datang, sedangkan validitas bandingan melihat hubungannya dengan masa sekarang. Seperti juga dijelaskan oleh Wrightstone bahwa: "The difference between concurrent, validity and predictive validity is solely a matter of time. Predictive validity requires correspondence with the criterion at the time of testing. (Wrightstone, 1961 hal. 44).

Cara yang dipergunakan untuk menilai validitas bandingan ialah dengan jalan mengkorelasikan hasil-hasil yang dicapai dalam tes tersebut dengan hasil-hasil yang dicapai dalam tes yang sejenis yang telah diketahui mempunyai validitas yang tinggi (misalnya tes standar). Tinggi rendahnya koefisien korelasi yang diperoleh menunjukkan tinggi rendahnya validitas tes yang akan kita nilai kualitasnya.

c. Validitas isi (*Content validity*).

Validitas isi artinya kejituan daripada suatu tes ditinjau dari isi tes tersebut. Suatu tes hasil belajar dapat dikatakan valid, apabila materi tes tersebut betul-betul merupakan bahan-bahan yang representatif terhadap bahan-bahan pelajaran yang diberikan. Misalnya apabila kita ingin memberikan tes bahasa Inggris kepada anak-anak kelas II, maka item-itemnya harus diambil dari bahan-bahan pelajaran kelas II. Kalau di dalamnya kita selipkan item-item yang diambil dari bahan-bahan pelajaran kelas III maka tes tersebut sudah tidak valid lagi.

Untuk menilai apakah suatu tes memiliki validitas isi atau tidak dapat kita lakukan dengan jalan membandingkan materi tes tersebut dengan analisa rasional yang kita lakukan terhadap bahan-bahan yang seharusnya dipergunakan dalam menyusun tes tersebut.

Apabila materi tes tersebut telah cocok dengan analisa rasional yang kita lakukan, berarti tes yang kita nilai itu mempunyai validitas isi. Sebaliknya apabila materi tes tersebut menyimpang dari analisa rasional kita, berarti tes tersebut tidak valid ditinjau dari validitas isinya.

d. Validitas susunan (*Construct Validity*).

Validitas susunan artinya kejituan daripada suatu tes ditinjau dari susunan tes tersebut. Misalnya kalau kita ingin memberikan tes kecakapan ilmu pasti, kita harus membuat soal yang ringkas dan jelas yang benar-benar akan mengukur kecakapan ilmu pasti, bukan mengukur kemampuan bahasa karena soal itu ditulis secara berkepanjangan dengan bahasa yang sukar dimengerti.

Untuk mengetahui apakah suatu tes memenuhi syarat-syarat validitas susunan atau tidak maka kita harus membandingkan susunan tes tersebut dengan syarat-syarat penyusunan tes yang baik. Apabila susunan tes tersebut telah memenuhi syarat-syarat penyusunan tes maka berarti tes tersebut memenuhi syarat validitas susunan. Begitu pula sebaliknya, apabila tidak memenuhi syarat-syarat penyusunan tes berarti tidak memenuhi validitas susunan.

Validitas ramalan dan validitas bandingan disebut pula validitas empiris (*empirical validity*), sebab pengujian daripada validitas tersebut didasarkan atas perhitungan-perhitungan secara empiris. Sebaliknya validitas isi dan validitas susunan disebut pula validitas rasional (*logical validity*) sebab pengujian terhadap, validitas tersebut didasarkan atas analisa rasional.<sup>5</sup>

Dalam penilaian ada empat unsur pokok yaitu objek yang dinilai, kriteria sebagai tolak ukur, data tentang objek yang dinilai, dan pertimbangan keputusan (*judgement*). Sehingga bisa disimpulkan bahwa penilaian tidak sekedar pengumpulan data peserta

---

<sup>5</sup> Wayan Nurkencana & Sunartana, *Evaluasi Pendidikan*, (Surabaya: Usana Offset Printing, 1986), hlm. 127

didik, tetapi juga pengolahannya untuk memperoleh gambaran proses dan hasil belajar peserta didik. Penilaian tidak sekedar memberi soal peserta didik kemudian selesai, tetapi guru harus menindaklanjutinya untuk kepentingan pembelajaran.<sup>6</sup>

Berdasarkan penjelasan di atas maka guru tak harus sekedar menguasai berbagai materi pembelajaran saja tetapi guru dituntut untuk mengevaluasi siswa yang diajar berdasarkan hasil penilaian yang diberikan dengan prosedur yang benar dan relevan sesuai dengan acuan yang ada..

Banyak faktor yang dapat mempengaruhi hasil tes evaluasi tidak valid. Beberapa faktor tersebut secara garis besar dapat dibedakan menurut sumbernya, yaitu faktor internal dari tes, faktor eksternal tes, dan faktor yang berasal dari siswa yang bersangkutan.

a. Faktor yang berasal dari dalam tes

Beberapa sumber yang pada umumnya berasal dari faktor internal tes evaluasi di antaranya sebagai berikut:

- 1) Arahan tes yang disusun dengan makna tidak jelas sehingga dapat mengurangi validitas tes.
- 2) Kata-kata yang digunakan dalam struktur instrumen evaluasi, terlalu sulit.
- 3) Item-item tes dikonstruksi dengan jelek.
- 4) Tingkat kesulitan item tes tidak tepat dengan materi pembelajaran yang diterima siswa.
- 5) Waktu yang dialokasikan tidak tepat, hal ini termasuk kemungkinan terlalu kurang atau terlalu longgar
- 6) Jumlah item tes terlalu sedikit sehingga tidak mewakili sampel materi pembelajaran.
- 7) Jawaban masing-masing item evaluasi bisa diprediksi siswa.

b. Faktor yang berasal dari administrasi dan skor. Faktor ini dapat mengurangi validitas interpretasi tes evaluasi, khususnya tes evaluasi yang dibuat oleh guru. Berikut beberapa contoh faktor yang sumbernya berasal dari proses administrasi dan skor.

- 1) Waktu pengerjaan tidak cukup sehingga siswa dalam memberikan jawaban dalam situasi yang tergesa-gesa.
- 2) Adanya kecurangan dalam tes sehingga tidak bisa membedakan antara siswa yang belajar dengan yang melakukan kecurangan.
- 3) Pemberian petunjuk dari pengawas yang tidak dapat dilakukan pada semua siswa.

---

<sup>6</sup> Jurnal Geografi Media Informasi Pengembangan Ilmu dan Profesi Kegeografian, Pradika Adi Wijayanto, Evaluasi Kualitas Instrumen Tes Dalam Pembelajaran Geografi Di Man 2 Kota Batu, Jurnal Geografi Volume 13 No 2 (102 dari 224)

- 4) Teknik pemberian skor yang tidak konsisten, misalnya pada tes esai, juga dapat mengurangi validitas tes evaluasi.
  - 5) Siswa tidak dapat mengikuti arahan yang diberikan dalam tes baku.
  - 6) Adanya joki (orang lain bukan siswa) yang masuk dan menjawab item tes yang diberikan.
- c. Faktor-faktor yang berasal dari siswa

Seringkali terjadi bahwa interpretasi terhadap terhadap item- item tes evaluasi tidak valid, karena dipengaruhi oleh jawaban siswa daripada interpretasi item-item pada tes evaluasi. Sebagai contoh, sebelum tes para siswa menjadi tegang karena guru pengampu mata pelajaran dikenal “killer”, galak, dan sebagainya sehingga siswa yang ikut tes banyak yang gagal. Contoh lain, ketika siswa melakukan tes penampilan keterampilan, ruangan terlalu ramai atau gaduh sehingga para siswa tidak dapat konsentrasi dengan baik. Ini semua dapat mengurangi nilai validitas instrumen evaluasi.<sup>7</sup>

## 2. Realibilitas Suatu Tes

Suatu tes dapat dikatakan tes yang reliable apabila tes tersebut menunjukkan hasil-hasil yang mantap. Ada beberapa cara yang dapat dipergunakan untuk mencari taraf reliabilitas daripada suatu tes;

### a. Teknik ulangan.

Mencari reliabilitas suatu tes dengan teknik ulangan ialah dengan jalan memberikan tes tersebut kepada sekelompok anak-anak dalam dua kesempatan yang berlainan. Misalnya suatu tes diberikan kepada group A. Selang tiga hari atau seminggu kemudian tes tersebut diberikan lagi kepada group A dengan syarat-syarat tertentu. (misalnya soal-soal dalam tes tidak dibicarakan selama waktu antara itu, situasi tempat dibuat sama dan sebagainya). Skor yang diperoleh oleh anak-anak dalam periode pertama dikorelasikan dengan skor yang mereka peroleh dalam periode kedua. Besar kecilnya koefisien korelasi yang diperoleh menunjukkan reliabilitas dari tes tersebut.

### b. Teknik bentuk pastel.

Dalam teknik ini dipergunakan dua buah tes yang sejenis (tetapi tidak identik), mengenai isinya; proses mental yang diukur, tingkat kesukaran jumlah item dan aspek-aspek yang lain.

Kedua tes ini diberikan kepada kelompok subyek tanpa adanya tenggang waktu. Skor yang diperoleh dari kedua tes tersebut dikorelasikan besar kecilnya koefisien korelasi yang diperoleh menunjukkan reliabilitas daripada tes tersebut.

---

<sup>7</sup> Sukardi, *op.cit.*, hlm. 39

Jika dibandingkan dengan teknik ulangan, teknik bentuk paralel ini lebih menguntungkan karena:

- 1) Item-item yang dipergunakan tidak sama maka pengaruh daripada hasil latihan dapat dihindarkan.
  - 2) Tidak adanya tenggang waktu maka perbedaan faktor-faktor yang mempengaruhi pelaksanaan tes boleh dikatakan tidak ada. Misalnya faktor situasi tes, administrasi, pengawasan dan sebagainya. Kelemahan daripada teknik ini ialah adanya kesukaran untuk menyusun item-item yang betul-betul paralel.
- c. Teknik belah dua

Dalam teknik ini, tes yang telah diberikan kepada sekelompok subyek dibelah menjadi dua bagian. Kemudian tiap-tiap bagian diberikan skor secara terpisah. Ada dua prosedur yang dapat dipergunakan untuk membelah dua suatu tes yaitu

- 1) Prosedur ganjil genap, artinya seluruh item yang bernomor ganjil dikumpulkan menjadi satu kelompok, dan seluruh item yang bernomor genap menjadi kelompok yang lain.
- 2) Prosedur secara random, misalnya dengan jalan lotre, atau dengan jalan menggunakan tabel bilangan random.

Koefisien korelasi yang diperoleh dari kedua belahan itu menunjukkan reliabilitas dari setengah tes.<sup>8</sup>

### 3. Tingkat Kesukaran dan Daya Beda Suatu Tes

Suatu tes tidak boleh terlalu mudah, dan juga tidak boleh terlalu sukar. Sebuah item yang terlalu mudah sehingga dapat dijawab dengan benar oleh semua anak bukanlah merupakan item yang baik. Begitu pula item yang terlalu sukar sehingga tidak dapat dijawab oleh semua anak juga bukan merupakan item yang baik. Jadi item yang baik adalah item yang mempunyai derajat kesukaran tertentu.

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

Untuk mencari Derajat Kesukaran (DK) dan Daya Beda (DB) suatu item dapat dilakukan dengan jalan mengadakan analisis item-item, (items analysis). Dalam analisis item di samping mencari DK dan DB-nya, juga dapat dicari efektivitas setiap option yang digunakan dalam item tersebut.<sup>9</sup>

---

<sup>8</sup> Wayan Nurkencana & Sunartana, *op.cit.*, hlm. 132

<sup>9</sup> *Ibid*, hlm. 134

Kualitas suatu tes hasil belajar sangat ditentukan oleh kualitas item-itemnya. Tes hasil belajar yang berisi item-item yang berkualitas tinggi walaupun dalam jumlah yang sedikit akan jauh lebih berguna dari pada tes hasil belajar yang berisi puluhan item berkualitas rendah. Item- item yang berkualitas rendah tidak saja menurunkan fungsi tes, tetapi juga akan memberikan hasil pengukuran yang menyesatkan. Salah satu cara untuk menentukan kualitas suatu tes hasil belajar adalah dengan melakukan analisis soal (*item analysis*). Analisis soal terutama dapat dilakukan untuk tes objektif. Hal ini tidak berarti bahwa tes uraian tidak dapat dianalisis, akan tetapi memang dalam menganalisis butir tes uraian, belum ada pedoman yang standar. Jadi, tes hasil belajar bentuk objektif lebih mudah dianalisis dari pada tes hasil belajar bentuk uraian, baik dari segi validitas, reliabilitas, tingkat kesukaran maupun daya pembedanya.<sup>10</sup>

#### 4. KESIMPULAN

Baik buruknya suatu tes atau suatu alat evaluasi dapat kita tinjau dari beberapa segi, yaitu: validitas, reliabilitas, tingkat kesukaran dan daya beda. Ketika mengkonstruksi atau memilih tes atau instrumen evaluasi, pertanyaan yang sering muncul dari pemikiran seorang guru adalah pada kondisi apakah interpretasi skor yang dihasilkan dari penggunaan instrumen tepat, bermanfaat, dan dapat digunakan pada keadaan sekolah yang ada.

Tanpa melihat jenis instrumen dan kemampuan guru dalam memilih jenis tes dalam melakukan evaluasi terhadap siswa, semua instrumen seperti yang disebut di atas agar memiliki kemampuan mengevaluasi perlu memenuhi karakteristik, yaitu valid, reliabel, dan dapat digunakan.

#### DAFTAR PUSTAKA

- Jurnal Geografi Media Infromasi Pengembangan Ilmu dan Profesi Kegeografian, Pradika Adi Wijayanto, Evaluasi Kualitas Instrumen Tes Dalam Pembelajaran Geografi Di MAN 2 Kota Batu, Jurnal Geografi Volume 13 No 2 (102 dari 224)
- Jurnal, Khaerudin, Kualitas Instrumen Tes Hasil Belajar, Jurnal Madaniyah, Volume 2 Edisi IX Agustus 2015
- Jurnal, Hidayatul Maulidah, Analisis Kualitas Instrumen Tes Fisika Kelas X Menggunakan Software Anates, Volume 2 Nomor 3, Desember 2022

---

<sup>10</sup> Jurnal, Khaerudin, Kualitas Instrumen Tes Hasil Belajar, Jurnal Madaniyah, Volume 2 Edisi IX Agustus 2015, hlm, 213

Sahrul Dahri, Pengembangan Instrumen Evaluasi Belajar Via Ponsel Pada Mata Kuliah Pendidikan Agama Islam, *Jurnal Lentera* Vol. 21 No. 1 (Maret, 2022), hlm, 128

Sawaluddin, Langkah-Langkah Dan Teknik Evaluasi Hasil Belajar Pendidikan Agama Islam, Vol. 6, No. 1, Januari – Juni 2020 (13-24)

Sukardi, *Evaluasi Pendidikan: Prinsip dan Operasionalnya*, Jakarta: PT. Bumi Aksara, 2010

Wayan Nurkancana & Sunartana, *Evaluasi Pendidikan*, Surabaya: Usana Offset Printing, 1986