Validitas dan Reliabilitas Alat Ukur

Syamsurizal

Jurusan Biologi FMIPA, Universitas Negeri Padang, West Sumatera, Indonesia

e-mail: syam_unp@fmipa.unp.ac.id

Abstrak, Instrumen penelitian pendidikan ranah pengetahuan dapat diukur dengan instrumen berbentuk tes tertulis. Syarat tes tertulis adalah: memiliki validitas yang baik, reliabilitas yang tinggi, tingkat kesukaran yang baik, daya beda yang dapat membedakan siswa berkemampuan tinggi atau rendah serta option yang berkualitas.

1. Validitas

Dalam bidang pendidikan kata validitas atau kesahihan digunakan sekurang-kurangnya dalam tiga konteks, yaitu: a) validitas penelitian (*research validity*), b) validitas soal (*item validity*), dan c) validitas alat ukur (*test validity*).

Validitas penelitian mempersoalkan derajat kesesuaian hasil penelitian dengan keadaan yang sebenarnya; hasil penelitian mencerminkan keadaan yang sebenarnya. Validitas penelitian mengandung dua hal, yaitu a) *validitas internal*, dan b) *validitas eksternal*. Validitas internal penelitian mempersoalkan kesesuaian antara data hasil penelitian dengan keadaan yang sebenarnya. Untuk memperoleh validitas internal penelitian yang memadai peneliti melakukannya melalui penggunaan instrumen pengambil data yang memenuhi persyaratan ilmiah tertentu. Validitas eksternal penelitian mempersoalkan derajat kesesuaian antara generalisasi hasil penelitian dengan keadaan yang sebenarnya; senyampang generalisasi hasil penelitian sesuai dengan keadaan sebenarnya. Untuk menjamin validitas eksternal hasil penelitian, peneliti melakukannya melalui penyusunan rancangan sampling yang cermat.

2. Validitas Soal

Validitas soal (*item validity*) merupakan istilah yang mudah menimbulkan kerancuan, diduga karena namanya yang mudah menimbulkan salah tafsir (*misleading*). Kita hendaklah mengingat benar bahwa validitas soal bukan validitas tes. Validitas soal merupakan derajat kesesuaian antara sesuatu soal dengan perangkat soal-soal lain (Syarif & Syamsurizal, 2019). . Ukuran validitas soal adalah korelasi antara skor pada soal itu dengan skor pada perangkat soal (*item total*

correlation) yang sering kali dihitung dengan korelasi biserial. Isi validitas soal adalah daya pembeda soal (*item discriminating power*) bukan validitas tes. Dengan demikian orang tidak dapat mengklaim bahwa karena telah memiliki kumpulan validitas soal berarti telah memiliki validitas tes. Informasi yang dimilikinya hanyalah bahwa kumpulan atau perangkat soal itu bersama-sama mengukur sesuatu, namun apakah sesuatu itu merupakan hal yang dimaksud oleh tes itu, informasinya belum ada.

3. Validitas Tes

Pengertian validitas tes atau validitas alat ukur yang sudah umum adalah “sejauh mana tes itu mengukur apa yang dimaksudkan untuk diukur”. Dengan demikian, validitas tes pada dasarnya menunjuk kepada derajat fungsi mengukurnya suatu tes, atau derajat kecermatan ukurnya sesuatu tes. Untuk mengkaji validitas alat ukur, yaitu sejauh mana alat ukur itu mengukur apa yang dimaksudkan untuk diukur, secara konvensional orang memandangnya dari tiga arah, yaitu dari arah: a) isi yang diukur, 2) rekaan teoritis (*construct*) atribut yang diukur, dan 3) kriteria alat ukur. Oleh karena itu dalam sumber studi tentang validitas tes, orang biasa membedakan validitas tes menjadi tiga macam, secara skematis disajikan dalam Gambar 1.

Gambar 1. Skema Validitas Tes

Keempat bentuk validitas tersebut sangat tepat sesuai dengan tujuannya. Sebagai contoh, untuk penilaian pendidikan misalnya, validitas isi merupakan hal yang paling berguna. Untuk kepentingan pribadi misalnya, validitas prediktif lebih bermanfaat. Landy (1987) memandang keempat bentuk ini sebagai kumpulan peranko (*stamp collecting*). Ia berpandangan bahwa keempat validitas ini harus dipandang dari skor ketimbang hanya membendakan bentuk. Ketimbang membedakan bentuk, para ahli dewasa ini lebih sepakat untuk melihat validitas dalam kerangka memahami skor dan implikasinya. Messick (1989) misalnya mengemukakan bahwa validitas senantiasa berkaitan dengan penemuan sains untuk memaknai skor.

1. Validitas isi (*content validity*)

Validitas isi (*content validity*) sering pula dinamakan validitas kurikulum yang mengandung arti bahwa suatu alat ukur dipandang valid apabila sesuai dengan isi kurikulum yang hendak diukur (Syarif & Syamsurizal, 2019). Validitas isi tes

menunjuk kepada kedalaman tes, yang merupakan seperangkat soal-soal, dilihat dari isinya memang mengukur sesuatu yang dimaksudkan untuk diukur. Ukuran kedalaman ini ditentukan berdasar derajat representatif isi tes itu bagi isi hal yang diukur. Salah satu cara yang digunakan untuk menentukan validitas adalah dengan mengkaji isi tes itu. Sebuah tes misalnya terdiri atas 25 soal penjumlahan dan pengurangan sangat baik digunakan untuk mengukur kemampuan matematika dibandingkan dengan tes yang terdiri atas 10 soal tentang olah raga tetapi tidak ada hal-hal yang berkaitan dengan penjumlahan dan pengurangan. Validitas isi ditentukan dengan melihat apakah soal-soal yang digunakan telah menunjukkan sampel atribut yang diukur. Dengan demikian menurut Guion (1977), validitas isi sangat bergantung kepada dua hal yaitu tes itu sendiri dan proses yang mempengaruhi dalam merespon tes. Sebagai contoh misalnya tes tertulis yang dipersiapkan untuk pekerjaan mungkin tidak menyajikan pengukuran yang valid untuk kemampuan pegawai melakukan pekerjaan, sekalipun mungkin saja tes itu sudah merupakan alat yang valid untuk mengukur pengetahuan tentang apa yang harus dikerjakan. Salah satu cara untuk memperoleh validitas isi adalah dengan melihat soal-soal yang membentuk tes itu. Jika keseluruhan soal tampak mengukur apa yang seharusnya tes itu digunakan, tidak diragukan lagi bahwa validitas isi sudah terpenuhi.

Dalam dunia pendidikan, sebuah tes dikatakan memiliki isi apabila mengukur sesuai dengan domain dan tujuan khusus tertentu Yang sama dengan isi pelajaran Yang telah diberikan di kelas. Soal matematika dikatakan valid apabila hanya mengukur kemampuan matematika, bukannya mengukur kemampuan bahasa. Ketika kita mengatakan akan mengukur kemampuan X peserta tes, kita harus mengukur atribut atau karakteristik khusus Yang berkaitan dengan X peserta tes yang akan diukur (Guion, 1977). Sebagai contoh, sebuah tes dirancang untuk mengukur kemampuan bermain bola basket dalam mata pelajaran Penjaskes misalnya, tentunya hal yang diukur haruslah antara lain berkaitan dengan kemampuan berlari, membawa bola, menembak bola, dan mendribble bola. Hal-hal yang diukur mungkin sangat luas seperti untuk kemampuan membaca misalnya, atau bahkan sangat sempit seperti untuk kemampuan penjumlahan.

Sebagian ahli tes berpendapat bahwa tidak ada satupun pendekatan statistik yang dapat digunakan untuk menentukan validitas isi suatu tes. Menurut Guion (1977), validitas isi hanya dapat ditentukan berdasarkan judgment para ahli. Prosedur yang dapat digunakan antara lain:

- (1) mendefinisikan domain yang hendak diukur
- (2) menentukan domain yang akan diukur oleh masing-masing soal

(3) membandingkan masing-masing soal dengan domain yang sudah ditetapkan.

Sekalipun prosedur ini tampak sederhana, tetapi dalam praktek terkadang sulit dilakukan. Kesulitan utama dalam prosedur ini adalah *mendefinisikan domain yang hendak diukur*. Dalam buku tentang pedoman penulisan tes tertulis, domain ini sama halnya dengan kisi-kisi. Sebagai contoh misalnya dalam menentukan soal fisika yang berkaitan dengan *problem solving* atau *reasoning*; beberapa ahli mungkin masih berdebat apakah suatu soal benar-benar telah masuk dalam kategori problem solving atau reasoning. Hal yang paling penting adalah, adanya kesepakatan antara beberapa penulis tentang kemampuan yang diukur oleh suatu soal.

2. Validitas Konstruk (*construct validity*)

Konstruk (*construct*) adalah sesuatu yang berkaitan dengan fenomena dan objek yang abstrak, tetapi gejalanya dapat diamati dan diukur. Gravitasi, massa, kemampuan matematika, kemampuan bahasa Inggris, kebahagiaan, dan kesedihan antara lain termasuk konstruk. Gravitasi misalnya dapat dijadikan sebagai contoh bagaimana memahami konstruk. Ketika buah apel jatuh ke tanah, konstruk tentang gravitasi dapat digunakan untuk menjelaskan dan memperkirakan perilaku (jatuhnya buah apel misalnya) yang diamati. Namun demikian, kita tidak dapat melihat yang dimaksud dengan konstruk gravitasi itu sendiri. Hal yang dapat kita lihat hanyalah apel itu jatuh. Kita dapat mengukur gravitasi dan mengembangkan teori tentang gravitasi.

Validitas konstruk mengandung arti bahwa suatu alat ukur (dikatakan valid apabila telah cocok dengan konstruksi teoritik di mana tes itu dibuat. Sebuah tes dikatakan memiliki validitas konstruksi apabila *soal-soalnya mengukur setiap aspek berpikir seperti yang diuraikan dalam standar kompetensi, kompetensi dasar, maupun indikator yang terdapat dalam kurikulum*. Soal yang dapat dikembangkan dari kisi-kisi seperti tampak ada Tabel 2.2 haruslah berupa soal yang sesuai dengan kemampuan membandingkan piramida ekologi, mengatasi masalah lingkungan dengan menggunakan konsep rantai makanan, menjelaskan aliran energi serta membuat bagan daur biogeokimia (Carbon, Nitrogen, Sulfur, dan Pospor.

Konstruksi yang dimaksud pada validitas ini bukanlah merupakan konstruksi seperti bangunan atau susunan, tetapi berupa rekaan psikologis yang berkaitan dengan aspek-aspek ingatan, pemahaman, aplikasi, analisis, sintesis, dan evaluasi.

Format Analisis Konstruksi Soal Pilihan Ganda

No. Soal	1	2	3	4	5	6	7	...	40	Jumlah
A. KONTEN ILMU										
1) Butir soal sesuai indikator										
2) Hanya ada satu kunci atau jawaban yang benar										
3) Isi materi sesuai dengan tujuan pengukuran.										
4) Isi materi sesuai dengan jenjang, jenis sekolah dan tingkatan kelas										
5) Pilihan benar-benar berfungsi, jika pilihan merupakan hasil perhitungan, maka pengecoh dapat diperoleh karena salah rumus/salah hitung										
B. KONSTRUK										
6) Pokok soal (stem) dirumuskan dengan jelas										
7) Rumusan soal dan pilihan dirumuskan dengan tegas										
8) Pokok soal tidak memberi petunjuk/mengarah kepada pilihan jawaban yang benar										
9) Pokok soal tidak mengandung pernyataan negatif ganda										
10) Bila terpaksa menggunakan kata, negatif, harus digarisbawahi atau dicetak lain										
11) Pilihan jawaban homogen										
12) Hindari adanya alternative: "jawaban seluruh jawaban di atas benar" atau "tak satu jawaban di atas yang benar" dan yang sejenisnya										
13) Panjang alternatif/pilihan jawaban relatif sama, jangan ada yang sangat panjang dan ada yang sangat pendek										
14) Pilihan jawaban dalam bentuk angka/waktu diurutkan										
15) Wacana, gambar, atau grafik benar-benar berfungsi										
16) Antar butir tidak bergantung satu sama lain.										

C. ASPEK BAHASA										
17) Rumusan kalimat soal komunikatif										
18) Kalimat menggunakan bahasa yang baik dan benar sesuai dengan jenis bahasanya										
19) Rumusan kalimat tidak menimbulkan penafsiran ganda atau salah pengertian										
20) Menggunakan bahasa/kata yang umum (bukan bahasa lokal)										
21) Rumusan soal tidak mengandung kata-kata yang menyinggung perasaan mahasiswa										
22) Rumusan soal tidak mengandung SARAP										
Jumlah										

3. Validitas berdasar kriteria (*Criterion-related*)

Dalam validitas berdasar kriteria, validitas alat ukur itu dilihat dari besarnya hasil pengukuran dengan alat yang dipersoalkan itu sama atau mirip dengan hasil pengukuran alat lain yang dijadikan kriteria. Kriteria itu umumnya adalah hasil pengukuran atribut yang sama dengan alat lain yang diakui merupakan alat ukur yang baik (memenuhi persyaratan akademik dan profesional tertentu). Misalnya orang yang menyusun tes intelegensi mungkin menggunakan hasil pengukuran dengan tes Wechsler atau tes Stanford-Binet sebagai kriteria.

Validitas berdasar kriteria ini dibedakan menjadi dua macam, berdasar atas kapan kriteria itu dapat dimanfaatkan. Jika kriteria itu sekarang atau dalam waktu dekat dapat dimanfaatkan disebut validitas sama saat (*concurrent validity*), dan jika kriteria itu baru beberapa waktu kemudian dapat dimanfaatkan disebut validitas prediktif (*predictive validity*).

Validitas prediksi (*predictive validity*)

Predictive validity menunjukkan kepada hubungan antara tes skor yang diperoleh peserta tes dengan keadaan yang akan terjadi diwaktu yang akan datang. Sebuah tes dikatakan memiliki validitas prediksi apabila mempunyai kemampuan untuk mempredik-sikan apa yang akan terjadi di masa yang akan datang.

Contoh sederhana misalnya apa yang terjadi pada penerimaan peserta tes berdasarkan hasil tes seleksi setelah mereka lulus SMA. Peserta tes yang memiliki nilai yang bagus di tes seleksi tersebut lalu diterima di perguruan tinggi, diperkirakan akan berhasil ketika mereka belajar di perguruan tinggi. Apabila hal itu terjadi, maka tes masuk perguruan tinggi tersebut dikatakan memiliki validitas prediksi bagus. Sebaliknya, apabila hasil di perguruan tinggi kurang baik, maka tes seleksi dimaksud tidak memiliki validitas yang bagus.

4. Validitas konkruen (*Concurrent validity*)

Concurrent validity atau *validitas ada sekarang* menunjuk pada hubungan antara tes skor dengan yang dicapai dengan keadaan sekarang. Validitas ini dikenal sebagai validitas empiris. Sebuah tes dikatakan memiliki *concurrent validity* apabila hasilnya sesuai dengan pengalaman.

RELIABILITAS

Reliabilitas alat ukur menunjukkan sejauh mana hasil pengukuran dengan alat tersebut dapat dipercaya. Hal ini ditunjukkan oleh taraf keajegan (konsistensi) skor yang diperoleh oleh para subjek yang diukur dengan alat yang sama, atau diukur dengan alat yang setara pada kondisi yang berbeda (Fiatma et al., 2017).

Reliabilitas menunjuk pada satu pengertian bahwa sesuatu instrumen cukup dapat dipercaya untuk digunakan sebagai alat pengumpul data karena instrumen tersebut sudah baik. Reliabilitas menunjuk pada tingkat keterandalan sesuatu. Reliabel artinya dapat dipercaya, jadi dapat diandalkan (Arieska et al., 2018; Syafti et al., 2019; Syamsurizal, 2019).

Pengertian umum menyatakan bahwa instrumen penelitian harus reliabel. Dengan pengertian ini sebenarnya kita dapat dipercaya adalah datanya, bukan semata-mata instrumennya (Nurhawa et al., 2019). Ungkapan yang menyatakan bahwa instrumen harus reliabel sebenarnya mengandung arti bahwa instrumen tersebut cukup baik sehingga mampu mengungkap data yang dapat dipercaya. Apabila pengertian ini sudah tertangkap maka akan tidak begitu menjumpai kesulitan dalam menentukan cara menguji reliabilitas instrumen.

Secara garis besar ada dua macam reliabilitas, yaitu reliabilitas eksternal dan reliabilitas internal. Jika ukuran atau kriteriumnya berada di luar instrumen maka dari hasil pengujian ini diperoleh reliabilitas eksternal. Sebaliknya jika perhitungan dilakukan berdasarkan data dari instrumen tersebut saja, akan menghasilkan reliabilitas internal.

1. *Reliabilitas eksternal*

Ada dua cara untuk menguji reliabilitas eksternal sesuatu instrumen yaitu dengan *teknik paralel* dan *teknik ulang*. Apabila peneliti ingin menggunakan teknik pertama yaitu teknik paralel, peneliti harus menyusun dua set instrumen. Kedua instrumen tersebut sama-sama diujicobakan kepada sekelompok responden saja (responden mengerjakan dua kali) selanjutnya hasil dari dua kali tes uji coba tersebut dikorelasikan, dengan teknik korelasi *product-moment* atau korelasi Pearson. Dari data dua kali uji coba dari dua instrumen yang satu dipandang sebagai nilai X, yang satu lainnya Y. Tinggi rendahnya indeks korelasi inilah menunjukkan tinggi rendahnya reliabilitas instrumen. Oleh karena itu dalam menggunakan teknik ini peneliti mempunyai *dua* instrumen dan melakukan dua kali tes, maka disebut teknik *doubel test doubel trial*.

Teknik reliabilitas eksternal kedua adalah dengan teknik ulang. Dengan menggunakan teknik ini peneliti hanya menyusun satu perangkat instrumen. Instrumen tersebut diujicobakan kepada sekelompok responden, hasilnya dicatat. Pada kali lain instrumen tersebut diberikan kepada sekelompok yang semula untuk dikerjakan lagi, dan hasil kedua juga dicatat. Selanjutnya kedua hasil tersebut dikorelasikan. Dengan teknik ini peneliti hanya menggunakan satu tes tetapi dilaksanakan *dua kali* uji coba. Teknik ini juga disebut sebagai *teknik single test double trial*.

2. Reliabilitas internal

Reliabilitas internal diperoleh dengan cara menganalisis data dari satu kali hasil pengujian. Ada bermacam-macam cara untuk mengetahui reliabilitas internal. Pemilihan sesuatu teknik didasarkan atas bentuk instrumen maupun selera peneliti. Kadang-kadang penggunaan teknik yang berbeda menghasilkan indeks reliabilitas yang berbeda pula. Hal ini wajar karena terkadang dipengaruhi oleh sifat atau karakteristik datanya sehingga dalam penghitungan diperoleh angka yang berbeda sebagai akibat pembulatan angka. Namun demikian untuk beberapa teknik diperlukan persyaratan-persyaratan tertentu sehingga peneliti tidak begitu saja memilih teknik-teknik tersebut.

Berbagai teknik mencari reliabilitas, misalnya saja adalah dengan rumus 1) Spearman-Brown; 2) Flanagan; 3) Rulon; 4) K-R 20; 5) K-R 21; 6) Hoyt; 7) Alpha.

Mengukur Validitas dan Reliabilitas

Salah satu cara untuk menentukan validitas alat ukur adalah dengan menggunakan korelasi *product moment* dengan simpangan yang dikemukakan oleh Pearson seperti berikut:

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

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

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

x^2 = kuadrat dari x

y^2 = kuadrat dari y

Sebagai contoh koefisien korelasi dapat diperoleh, misalnya dengan: a) Teknik split-half (teknik belah dua), yaitu mengkorelasikan skor setengah pertama dari suatu tes dengan setengah kedua. Untuk memperoleh skor setengah itu dapat diperoleh dengan mengkalkulasi skor nomor ganjil dengan nomor genap. Korelasi dari kedua skor tersebut akan menunjukkan homogenitas antar butir soal yang digunakan dalam perangkat tes itu secara keseluruhan; b) Teknik Kuder-Richardson, dengan menggunakan rumus yang dikembangkan oleh Kuder dan Richardson, yaitu rumus yang keduapuluh satu (KR-21).

Rumus tersebut adalah:

$$r_{KR-21} = \frac{k}{k-1} \left\{ 1 - \frac{(k-X)}{k S^2} \right\}$$

dimana k = jumlah butir pertanyaan, X skor rerata; s^2 = varian (kuadrat dari standar deviasi)

Rumus Alpha digunakan untuk mencari reliabilitas instrumen yang skornya bukan 1 dan 0, misalnya angket atau soal bentuk uraian.

Rumus Alpha:

$$r_{11} = \left\{ \frac{k}{k-1} \right\} \left\{ 1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right\}$$

dimana:

r_{11} = reliabilitas instrumen

k = banyaknya butir pertanyaan atau banyaknya soal

$\sum \sigma_b^2$ = jumlah varians butir

σ_t^2 = varians total

DAFTAR PUSTAKA

- Arikunto, S. 1998. *Prosedur Penelitian: Suatu Pendekatan Praktek*. Edisi Revisi IV. Jakarta: Penebit Rineka Cipta.
- Departemen Pendidikan Nasional. 2003. *Kurikulum 2004 SMA: Pedoman Khusus Pengembangan Silabus dan Penilaian Mata Pelajaran Biologi*. Jakarta: Direktorat Jenderal Pendidikan Dasar dan Menengah Direktorat Pendidikan Menengah Umum.
- Guion, R.M. 1977. Content Validity: The source of my discontent. *Applied Psychological Measurement*. 1 (1—10).
- Surapranata, S. 2005. *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes*. Cetakan kedua. Bandung: Penerbit PT Remaja Rosdakarya.
- Arieska, M., Syamsurizal, S., & Sumarmin, R. (2018). Guiding Development Based Approach Practicum Vertebrates Taxonomy Scientific Study Program for Students of Biology Education. *IOP Conference Series: Materials Science and Engineering*, 335(1), 12096.
- Friatma, A., Syamsurizal, S., & Helendra, H. (2017). Analyzed Quality Question of Final Exam Whole Semester on Biology Course Class XI IPA SMA Negeri District South of Solok Selatan Academic Year 2015/2016. *Bioeducation Journal*, 1(2), 50–67.

- Nurhawa, W. O., Yogica, R., Hartanto, I., & Syamsurizal, S. (2019). The Effect of Talking Chips Model Containing Science Literacy on Students' Learning Competencies In the Material of Climate Change and Its Impact on Ecosystems at Junior High School 20 Padang. *Atrium Pendidikan Biologi*, 4(4), 17–25.
- Syafti, N. K., Darussyamsu, R., Selaras, G. H., & Syamsurizal, S. (2019). Analyzed Quality Question Type Multiple Choices Questions (MCQ) of Final Exam Even Semester on Biology Course Class XI SMA Negeri Se-Kecamatan IV Jurai Academic Year 2017/2018. *Atrium Pendidikan Biologi*, 4(1), 200–210.
- Syamsurizal, S. (2019). Pengembangan Asesmen Berbasis Kelas untuk Peningkatan Kompetensi Guru IPA SMP dan MTs Kecamatan Koto Tengah. *Pelita Eksakta*, 1(02).
- Syarif, E. A., & Syamsurizal, S. (2019). Analyzed Quality of Senior High School Biology Olympiad Questions at West Sumatera, Riau, Jambi, and Bengkulu in 2018. *Bioeducation Journal*, 3(2), 142–150.