

A. Pengantar

B. Validitas dan Reliabilitas instrumen

Setiap instrumen yang akan digunakan diuji secara sistematis untuk menjamin validitas dan reliabilitasnya.

Instrumen tes sebagai alat ukur harus memenuhi beberapa syarat, di antaranya reliabilitas, validitas, dan kepraktisan (Harris, 1969; Davies, 1990) atau validitas, reliabilitas, skorabilitas, ekonomi, dan administrabilitas (Lado, 1961). Tiga syarat terakhir yang dikemukakan oleh Lado itu pada dasarnya sama dengan kepraktisan.

Validitas adalah kemampuan soal untuk mengetes keterampilan bahasa sesuai dengan yang diinginkan pengetes. Instrumen harus sah dalam konstruk dan isi, menghindari validitas yang semu, memiliki kemampuan kiris-balik, pragmatis, konkuren, dan prediktif. Validitas menunjukkan apakah suatu alat ukur benar-benar mengukur sesuatu yang harus diukur dengan hasil yang tepat. Bersama dengan reliabilitas, validitas menunjukkan kualitas alat ukur. Dapat juga dikatakan bahwa validitas adalah tingkat ketepatan suatu tes mengukur apa yang dimasukkan peneliti untuk diukur. Validitas dibedakan atas validitas isi (*content validity*), validitas kriteria (*criterion validity*), dan validitas konstruk (*construct validity*) (Wiersma, 1990: 183-194).

Dalam paradigma komunikatif, terdapat prinsip-prinsip sebagai berikut: ada interaksi antarpeserta, bentuk dan isi wacana tidak dapat diprediksi, purposif, wacana dalam lingkup yang relevan, bermakna, serta ada integrasi beberapa keterampilan bahasa, *direct testing method* yang dapat menunjukkan proses yang lebih realistik.

Reliabilitas adalah kemampuan untuk mempertahankan konsistensi mutu hasil tes. Ada tiga aspek yang harus diperhatikan, yaitu *inter-marker*, *intra-marker*, dan *parallel form reliability* (*future alternative form* sesuai dengan perkembangan teori dan praktik pengajaran bahasa). Reliabilitas juga adalah ketetapan (keajegan) skor tes. Reliabilitas dipengaruhi oleh banyak faktor, di antaranya ketepatan sampel. Reliabilitas dapat diuji dengan berbagai cara; salah satu di antaranya yang paling mudah adalah tes-ulang (*re-test*); cara yang lain adalah tes bentuk lain (*alternate form*) dan belah-dua (*split-half*) (Mueller, 1986: 58-59).

Kepraktisan instrumen tes maksudnya adalah kemungkinan instrumen tes itu dapat dilaksanakan. Kepraktisan menyangkut segi ekonomi, kemudahan administrasi, penyekoran, dan interpretasi. Dari segi ekonomi, apakah tes memerlukan banyak biaya dan waktu; dari segi administrasi dan penyekoran apakah tes dapat dilaksanakan mengingat tenaga dan perlengkapan yang diperlukan. Segi kemudahan interpretasi menyangkut apakah hasil tesnya, walaupun tes itu reliabel dan valid, masih relevan dan tidak kadaluwarsa. Dengan kata lain, tes harus *efisiensi*, kepraktisan, ekonomis, mudah dalam pengadministrasian, penilaian (skoring), dan penafsiran hasilnya.

Selain ketiga prinsip utama tersebut, dalam perancangan instrumen tes bahasa, beberapa hal di bawah ini juga mendapatkan perhatian (Weir, 1990: 36-39).

Pertama, butir soal yang disusun itu sebaiknya diambil dari teks yang spesifik dan orisinal (*specificity and authenticity*), yakni teks yang digunakan dalam kehidupan nyata di masyarakat. Teks demikian dapat diambil dari berbagai sumber yang ada di masyarakat: koran, majalah, dan buku bacaan lainnya.

Kedua, beberapa parameter yang mendapatkan perhatian adalah sebagai berikut.

- ✎ Instrumen tes memperhatikan aktivitas yang harus dilakukan oleh peserta tes dalam menyelesaikan soal.
- ✎ Instrumen tes memperhatikan *setting* (susunan keadaan) fisik dan psiko-sosial.
- ✎ Instrumen tes memperhatikan juga interaksi sosial dan peran peserta tes.
- ✎ Instrumen tes memperhatikan instrumen yang akan digunakan.
- ✎ Instrumen tes memperhatikan dialek yang digunakan masyarakat.
- ✎ Instrumen tes juga memperhatikan keterampilan pendukung lainnya dalam menyelesaikan soal .

Ketiga, karakteristik bentuk soal perlu juga memperhatikan beberapa hal sebagai berikut.

- ✎ konteks yang realistis;
- ✎ rentang informasi yang relevan;
- ✎ intersubjektivitas: siswa sebagai pembicara sekaligus pendengar;
- ✎ peluang untuk berkomunikasi secara mandiri dan kreatif;
- ✎ peluang untuk melakukan self-monitoring;
- ✎ peluang untuk memproses masukan yang sesuai dengan perkembangan siswa;
- ✎ peluang untuk menyelesaikan tugas sesuai dengan kendala waktu yang normal.

Keempat, dimensi soal juga memperhatikan:

- ✎ bentuk wacana: panjang-pendek, reseptif-produktif
- ✎ kompleksitas struktur (kalimat/wacana)

☞ rentangan fungsional: keragaman *illocutionary acts*

☞ rentangan referensial: keragaman makna/pengetahuan leksikal

Kelima, desain tes harus memungkinkan adanya pengembangan dalam kesahihan, mutu, dan efisiensi pengadministrasiannya.

Keenam, dalam pelaksanaan tes, langkah-langkah yang dipertimbangkan meliputi: persiapan, pencetakan dan distribusi soal (termasuk panduan pelaksanaan), dan pelaksana.

Ketujuh, pemantauan pasca pelaksanaan diperlukan untuk melihat kemungkinan perbaikan di masa yang akan datang, kemungkinan adanya kiris-balik (*washback*) tes tersebut terhadap proses belajar mengajar, dan dampak dari tes tersebut terhadap masyarakat secara keseluruhan.

C. Analisis dan revisi instrumen

Dari hasil uji coba yang telah dilaksanakan, langkah berikutnya adalah analisis soal. Penganalisisan biasanya difokuskan pada realibilitas, validitas, dan daya pembeda.

1. Reliabilitas

Pengukuran reliabilitas digunakan untuk mengetahui keajegan tes yang akan digunakan. Untuk keperluan pengukuran tersebut ada beberapa alternatif teknik pengukuran yang dapat dimanfaatkan, yaitu tes ulang, tes paralel, tes satu kali dan tes belah dua.

Dari alternatif yang dikemukakan di atas, penelitian ini teknik pelaksanaannya hanya satu kali tes. Harga koefisien reliabilitasnya dihitung dengan menggunakan rumus K - R (Kuder - Richardson). Rumusnya seperti berikut ini.

$$\frac{(K) (SD^2) - X (K - X)}{(SD^2)b(K-1)}$$

K = jumlah item dari tes

SD = standar deviasi

X = atau rata-rata

Sebagai tolok ukur menafsirkan harga koefisien yang diperoleh, disesuaikan dengan harga koefisien yang dikemukakan J.P. Guilford.

Kurang dari 0,20 = derajat keterpercayaan rendah

0,21 - 0,40 = derajat keterpercayaan sedang

- 0,41 - 0,70 = derajat keterpercayaan tinggi
- 0,71 - 1,00 = derajat keterpercayaan sangat tinggi

2. Validitas

Pengukuran validitas dihitung tiap-tiap butir soal, karena nilai validitas yang diperoleh masing-masing item soal akan berdaya dukung terhadap keandalan seluruh tes. Untuk menganalisisnya digunakan rumus pengukuran validitas biserial, yaitu:

$$r_{pb} = \frac{M_p - M_t \cdot p}{S_t \cdot q}$$

r_{pb} = koefisien korelasi biserial

M_t = rerata skor total

M_p = rerata subyek yang menjawab betul bagi item yang dicari validitasnya

p = proporsi siswa yang menjawab benar

(p) = banyaknya siswa yang menjawab benar
jumlah seluruh siswa

q = proporsi siswa yang menjawab salah ($q = 1 - p$)

Untuk menafsirkan harga nilai validitas hitung, nilai validitas hitung diinterpretasikan terhadap koefisien korelasi sebagai berikut ini (Arikunto, 1990: 71).

- antara 0,800 sampai dengan 1,00 : sangat tinggi
- antara 0,600 sampai dengan 0,800 : tinggi
- antara 0,400 sampai dengan 0,600 : cukup
- antara 0,200 sampai dengan 0,400 : rendah
- antara 0,000 sampai dengan 0,200 : sangat rendah

Dengan bercermin pada nilai koefisien korelasi di atas, nilai validitas uji coba tes dapat dilihat pada lampiran.

3. Daya Pembeda

Faktor lain yang diperhatikan atas kelaikan tes adalah daya pembeda. Daya pembeda digunakan untuk membedakan antara siswa yang pandai dan kurang pandai.

Perhitungan daya pembeda ditempuh langkah-langkah:

1. Mengurut hasil tes dari nilai yang terbesar sampai yang terkecil;
2. Mengambil validitas hitung dengan mengambil dua kutub yang berbeda, 27% kelompok atas dan 27% kelompok bawah;
3. Menghitung daya pembeda dengan rumus:

$$D = \frac{BA}{JA} - \frac{BB}{JB}$$

$$D = \frac{BA}{JA} - \frac{BB}{JB}$$

D = indeks diskriminasi

BA = banyaknya peserta kelompok atas yang menjawab benar

BB = banyaknya peserta kelompok bawah yang menjawab benar

JA = banyaknya peserta kelompok atas

JB = banyaknya peserta kelompok bawah

Sebagai tolok ukur daya pembeda, indeks diskriminasi daya pembeda dibedakan sebagai berikut ini (Arikunto, 1990:221)

0,00 - 0,20 = daya pembeda jelek

0,20 - 0,40 = daya pembeda cukup

0,40 - 0,70 = daya pembeda baik

0,70 - 1,00 = daya pembeda baik sekali

Soal yang dijadikan alat pengumpul data difokuskan kepada pertimbangan dua faktor, (1) nilai validitas item, dan (2) daya pembeda. Dua faktor tersebut dimanfaatkan sebagai pertimbangan memilih soal, karena masing-masing item mempunyai nilai koefisien sehingga memudahkan memilih item yang akan digunakan sebagai instrumen penelitian, sedangkan nilai reliabilitas yang hanya dinilai berdasarkan keseluruhan tes atau soal.

Item-item soal tersebut masih perlu diuji keabsahannya. Hal tersebut sesuai dengan pendapat Arikunto (1980:81) yang menyatakan bahwa sebuah tes mungkin reliabel tetapi tidak valid. Namun, sebuah tes yang valid biasanya reliabel. Pernyataan ini mengisyaratkan bahwa meskipun suatu tes tidak memiliki reliabilitas yang memadai bukan berarti harus dikesampingkan melainkan harus diseleksi item-item mana yang menyebabkan soal tidak reliabel. Apalagi soal-soal yang memiliki reliabel yang sangat tinggi akan dipergunakan sebagai pengumpul data.

Dari hasil perhitungan uji coba yang dilakukan, adakalanya soal yang valid tidak diimbangi oleh daya pembeda yang baik atau sebaliknya. Berdasarkan kenyataan ini, maka soal yang dipilih sebagai alat pengumpul data adalah yang sesuai dengan nilai reliabel dan kedua kriteria tersebut atau salah satu yang terpenuhi, baik daya beda atau validitas.

Tahap terakhir setelah melakukan seleksi soal adalah merevisi soal. Revisi soal dilakukan apabila terdapat soal yang kurang baik, tetapi soal tersebut harus dihadirkan pada alat pengumpul data. Soal yang bernilai validitas dan daya beda negatif tidak dipergunakan.