

KRITERIA ALAT EVALUASI PEMBELAJARAN BAHASA

(CRITERIA OF LANGUAGE LEARNING EVALUATION TOOL)

Siti Fatinah
Balai Bahasa Provinsi Sulawesi Tengah
Jalan Untad I, Bumi Roviga, Tondo, Palu 94118
E-mail: fatinahgari@yahoo.com
Telepon 085241308828

Abstract

Every learning activity, including language learning, needs to be evaluated in order to measure the success rate of the learners. The learning success is measured by either test or non-test. Evaluation tool should have good validity and reliability. In other words, language test should meet the criteria of validity and credibility. In order to meet the criteria of validity, the language tests should be estimated by using validity estimation formulas, such as Pearson product-moment correlation coefficient or Spearman rank-correlation different. Moreover, In order to determine the reliability level of language test, there are six reliability estimation formula used. There are (1) re-test reliability, (2) the two-sided reliability, (3) Kuder-Richardson formula reliability 20, (4) Kuder-Richardson formula reliability 21, (5) Alpha Cronbach reliability, and (6) parallel-formed reliability.

Key words: learning, evaluation tool, validity, reliability

Abstrak

Setiap kegiatan pembelajaran, termasuk pembelajaran bahasa, perlu dievaluasi untuk menentukan tingkat keberhasilan para pemelajar. Keberhasilan pemelajar diukur melalui alat evaluasi, baik tes maupun non-tes. Alat evaluasi sebagai alat ukur hasil belajar, termasuk hasil belajar bahasa seharusnya memiliki kriteria tes yang baik, yakni validitas dan reliabilitas. Dengan kata lain, tes bahasa harus memiliki kriteria kevalidan dan kereliabelan. Agar dikategorikan memenuhi kriteria kevalidan, tes bahasa itu perlu diestimasi dengan menggunakan beberapa rumus estimasi validitas, antara lain koefisien korelasi *Pearson product-moment* (*Pearson product-moment correlation coefficient*) atau koefisien korelasi *Spearman rank-difference* (*Spearman rank-different correlation*). Selain diuji validitasnya, tes bahasa juga perlu diukur tingkat reliabilitasnya. Untuk menentukan tingkat reliabilitas tes bahasa, ada enam rumus estimasi reliabilitas yang digunakan, yaitu (1) reliabilitas ulang uji, (2) reliabilitas belah dua, (3) reliabilitas rumus Kuder-Richardson 20, (4) reliabilitas rumus Kuder-Richardson 21, (5) reliabilitas Alpha Cronbach, dan (6) reliabilitas bentuk paralel.

Kata kunci: pembelajaran, alat evaluasi, validitas, reliabilitas

1. Pendahuluan

Dalam pelaksanaan pembelajaran bahasa, sebagaimana halnya dalam pelaksanaan pembelajaran bidang studi yang lain, evaluasi merupakan bagian yang tidak dapat dipisahkan dari penyelenggaraan pembelajaran secara keseluruhan. Pada hakikatnya kedudukan evaluasi dalam desain penyelenggaraan pembelajaran merupakan bagian akhir dari rangkaian tiga komponen pokok penyelenggaraan pembelajaran, yakni tujuan pembelajaran, kegiatan pembelajaran, dan evaluasi hasil kegiatan pembelajaran. Djiwandono (2011: 1) menyatakan bahwa sebagai suatu pembelajaran, pembelajaran bahasa diselenggarakan untuk mencapai berbagai tujuan pembelajaran yang telah diidentifikasi dan dirumuskan berdasarkan kajian yang mendalam terhadap kebutuhan yang perlu dipenuhi. Tujuan-tujuan pembelajaran itu diupayakan pencapaiannya melalui serangkaian kegiatan pembelajaran yang dirancang secara matang dan seksama serta diselenggarakan secara sungguh-sungguh agar tujuan pembelajaran yang telah dirumuskan dapat tercapai secara maksimal. Pencapaian tujuan-tujuan pembelajaran itu dilakukan melalui pelaksanaan evaluasi terhadap hasil pembelajaran yang telah diselenggarakan selama kurun waktu tertentu yang telah direncanakan.

Evaluasi dalam penyelenggaraan pembelajaran bahasa ditujukan pada tingkat kemampuan bahasa sebagai pencerminan hasil pembelajaran. Secara tidak langsung hasil evaluasi itu dimaksudkan sebagai (1) umpan balik bagi seluruh aspek penyelenggaraan pembelajaran, yang meliputi kurikulum yang dijadikan landasan pembelajaran, bahan ajar yang digunakan, latihan dan tugas-tugas yang diberikan kepada pemelajar; (2) umpan balik bagi pengajar terhadap cara mengajar, pilihan dan kesesuaian bahan ajar, tes yang digunakan, dan

lain-lain (Djiwandono, 2008: 163). Semua bentuk umpan balik tersebut pada akhirnya merupakan umpan balik bagi tujuan pembelajaran yang telah diidentifikasi dan dirumuskan sebagai dasar bagi seluruh penyelenggaraan pembelajaran. Dengan kata lain, dalam kaitannya dengan penyelenggaraan pembelajaran, peranan evaluasi tidak terbatas semata-mata pada memenuhi kebutuhan untuk meningkatkan tingkat pencapaian dan keberhasilan pemelajar, meskipun hal itu pada umumnya merupakan fokus utama. Evaluasi diharapkan mampu memberikan umpan balik bagi penyelenggaraan pembelajaran secara keseluruhan.

Dalam evaluasi, termasuk evaluasi pembelajaran bahasa diperlukan suatu alat yang harus dipakai agar kegiatan evaluasi berlangsung sesuai dengan rencana. Alat itu dinamakan alat evaluasi. Nurgiyantoro (2010: 89) menyatakan bahwa secara garis besar, alat evaluasi terdiri atas dua, yaitu tes dan non-tes. Tes dan non-tes itu digunakan untuk memperoleh informasi atau data penilaian tentang subjek belajar yang dinilai secara berhasil guna jika dipakai secara tepat. Artinya, evaluator harus mampu menentukan kapan menggunakan alat tes dan kapan menggunakan alat non-tes, termasuk memilih bentuk tes yang mana dan bentuk non-tes yang mana. Pemilihan secara tepat terhadap kedua alat evaluasi tersebut tidak terlepas dari tujuan evaluasi dan jenis informasi yang diharapkan.

Tes sebagai salah satu alat pengukur hasil belajar peserta didik diharapkan mampu memberikan informasi yang dapat dipertanggungjawabkan kebenarannya. Artinya, alat tes yang digunakan harus dapat memberikan informasi tentang peserta didik sesuai dengan keadaan yang mendekati realita yang sesungguhnya. Hal itu penting karena informasi itu akan digunakan untuk mempertimbangkan, lalu memutuskan berbagai kebijakan, baik yang berkenaan dengan pe-

serta didik maupun kegiatan pembelajaran secara umum. Agar dapat memberikan informasi yang dapat dipertanggungjawabkan dan tepat, alat tes yang digunakan harus bermutu dan dapat dipertanggungjawabkan, serta memiliki kriteria yang dipersyaratkan, terutama validitas dan reliabilitas.

2. Pembahasan

Alat evaluasi yang digunakan dalam kegiatan pembelajaran harus memenuhi kriteria tes yang baik, termasuk tes bahasa. Sebuah tes dikategorikan sebagai tes yang baik jika memenuhi kriteria validitas dan reliabilitas. Begitu pula halnya dengan tes bahasa, harus memenuhi kedua kriteria tersebut.

2.1 Validitas

Gronlud (dalam Djiwandono, 2008: 164) mengemukakan bahwa meskipun validitas lebih tepat diartikan sebagai kesesuaian interpretasi hasil tes daripada tes sebagai alat evaluasi, secara lebih praktis dan sederhana validitas itu dikaitkan dengan kesesuaian tes sebagai alat ukur dengan sasaran pokok yang perlu diukur. Tes bahasa yang valid sebagai alat ukur kemampuan bahasa memusatkan pengukurannya pada kemampuan bahasa peserta tes, bukan kemampuan lain, misalnya pengetahuan tentang sejarah perkembangan bahasa. Tes kemampuan menulis yang valid sebagai alat ukur kemampuan menulis memusatkan pengukurannya pada kemampuan mengungkapkan gagasan secara tertulis, bukan kemampuan memahami isi bacaan atau kemampuan menyimak isi bacaan.

Berdasarkan paparan tersebut dapat dinyatakan bahwa validitas atau kesahihan adalah relevansi, kecocokan, atau kesesuaian alat evaluasi dengan fungsi evaluasi dan sasaran evaluasi. Sebuah alat evaluasi dapat dikatakan valid apabila dapat mengukur hasil belajar secara tepat.

2.1.1 Hakikat Validitas

Secara konvensional dikenal adanya beberapa jenis validitas, sedangkan secara konseptual, validitas pada hakikatnya bersifat tunggal (*unitary concept*), tidak berjenis-jenis. Yang beragam adalah cara pembuktian keberadaan validitas yang dapat dilakukan melalui salah satu dari tiga cara pokok pengumpulan bukti kesesuaian tes dengan sasaran tes. Gronlud, dalam Djiwandono (2008: 165) mengemukakan bahwa validitas dapat dikenali dan dibuktikan keberadaannya melalui kajian terhadap kesesuaiannya dengan isi (validitas isi), kesesuaiannya dengan kriteria (validitas kriteria), dan kesesuaiannya dengan konstruk (validitas konstruk). Sejalan dengan itu, Nitko (dalam Djiwandono, 2008: 165) menyatakan bahwa pembuktian validitas isi dapat dilakukan dengan menyusun kisi-kisi yang membandingkan tugas yang dituntut dalam mengerjakan tes dan isi butir-butir tes atau dengan isi kurikulum (validitas kurikulum). Validitas kriteria dibuktikan dengan membandingkan (mengorelasikan) skor tes yang ditelaah dengan skor tes yang serupa yang diketahui ciri-cirinya yang baik, yang diselenggarakan pada waktu yang hampir bersamaan (validitas sama waktu atau *concurrent validity*). Validitas konstruk, dalam arti konstruksi berupa paparan teoretis untuk menjelaskan tingkah laku (*theoretical construction that are used to explain behavior*) dibuktikan secara analitis-argumentatif atau secara empirik (Gronlud dalam Djiwandono, 2008: 165).

Menurut Gronlud (dalam Nurgiyantoro, 2010:151 – 152), jika kita berpikir tentang validitas dalam kaitannya dengan tes, hal-hal yang perlu dipertimbangkan adalah sebagai berikut.

- 1) Validitas menunjuk pada kelayakan interpretasi yang dibuat berdasarkan skor hasil tes yang berkaitan dengan penggunaan tertentu, bukan terhadap instru-

men atau tes tersebut. Penggunaan istilah yang lazim digunakan “validitas alat tes” sebenarnya yang lebih tepat adalah validitas interpretasi terhadap hasil tes.

- 2) Validitas adalah masalah kadar (*matter of degree*) sehingga perlu dihindari pemikiran bahwa sebuah hasil tes itu valid atau tidak valid.
- 3) Validitas berkaitan dengan penggunaan khusus karena tidak ada satu tes pun yang valid untuk semua tujuan. Oleh karena itu, penilaian terhadap validitas tes harus berkaitan dengan tujuan penggunaan hasil tes itu.

2.1.1.2 Pembuktian Validitas

Validitas sebuah tes tidak terwujud begitu saja, tetapi perlu diupayakan, kemudian dijelaskan dan dibuktikan. Siapa pun yang menyusun dan mengembangkan tes perlu memberi penjelasan dan pembuktian bahwa tesnya memang valid, terlebih jika hasil dan informasinya digunakan sebagai bahan pertimbangan sebuah keputusan yang penting, seperti kenaikan kelas, kelulusan, pemberian beasiswa, dan sebagainya. Untuk keperluan seperti itu, tes yang digunakan tidak seharusnya sekadar dianggap valid tanpa disertai penjelasan tentang aspek dan tingkat validitasnya, serta bukti-bukti yang dapat diverifikasi. Sesuai dengan data yang diperoleh, penjelasan itu dapat disajikan secara empirik dengan angka-angka statistik berupa hasil perhitungan dan penerapan rumus tertentu, atau melalui penalaran dan penjelasan kualitatif yang lugas dan logis.

Pembuktian validitas secara empirik dilakukan dengan menyajikan bukti empirik berupa hasil perhitungan korelasi antara dua deret skor yang diperoleh kelompok peserta tes yang sama dari dua tes yang berbeda, yaitu tes yang divalidasi dan tes

sejenis yang tingkat validasinya telah dibuktikan. Tingkat korelasi dari kedua tes tersebut diperoleh melalui penerapan rumus korelasi tertentu. Berdasarkan jenis data yang ada, tingkat korelasi dua deret skor semacam itu dapat dilakukan untuk menentukan koefisien korelasi *Pearson product-moment* (*Pearson product-moment correlation coefficient*) atau koefisien korelasi *Spearman rank-difference* (*Spearman rank-difference correlation*).

Rumus koefisien korelasi *product-moment* adalah sebagai berikut.

$$r_{1.2} = \frac{N\sum X_1X_2 - (\sum X_1)(\sum X_2)}{\sqrt{(N\sum X_1^2 - (\sum X_1)^2)(N\sum X_2^2 - (\sum X_2)^2)}}$$

Keterangan:

$r_{1.2}$ = koefisien korelasi yang dicari

N = jumlah siswa

X1 = skor hasil tes pertama (membaca pemahaman)

X2 = skor hasil tes kedua (criteria: bersastra)

Sumber: Nurgiyantoro (2010: 161)

Penerapan rumus tersebut menghasilkan angka yang menunjukkan tingkat korelasi antara dua deret skor yang dikorelasikan. Sebagai angka korelasi, angka tertinggi adalah 1,00 yang mengindikasikan tingkat korelasi paling tinggi atau sempurna, yang dalam praktiknya lebih bersifat teoretis. Artinya, hampir tidak pernah ditemukan dalam kenyataan yang sebenarnya. Angka korelasi hasil perhitungan seperti itu hampir selalu kurang dari angka 1,00; yang terentang antara 0,99; 0,78; 0,12; 0,04; dan seterusnya (Djiwandono, 2008: 167). Angka-angka kuantitatif itu diberi label kualitatif yang menunjukkan tinggi rendahnya tingkat korelasi.

URUTAN TINGKAT KORELASI

Koefisien	Kadar Korelasi
0,80--1,00	Tinggi
0,60--0,79	Lumayan Tinggi
0,40--0,59	Sedang
0,20--0,39	Rendah
0,00--0,19	Dapat Diabaikan

Sumber: Best, John W. 1977 (dalam Djiwandono, 2008: 168)

2.2 Reliabilitas

Tes sebagai alat ukur yang hasilnya dapat digunakan untuk membuat berbagai keputusan penting, selain memenuhi kriteria validitas, juga harus memiliki kriteria reliabilitas. Validitas dan reliabilitas adalah dua hal yang banyak dibicarakan secara bersamaan berkaitan dengan hasil pengukuran. Dalam banyak hal, reliabilitas akan memengaruhi atau menentukan validitas meskipun tidak semua yang reliabel pasti valid. Reliabilitas merujuk pada konsistensi pengukuran dan bukan pada ketepatan pengukuran.

2.2.1 Hakikat Reliabilitas

Sebuah tes dikatakan reliabel apabila hasil-hasil tes tersebut benar-benar dapat dipercaya karena bersifat ajeg dan tidak berubah secara mencolok. Artinya, jika kepada para siswa diberikan tes yang sama pada waktu yang berlainan, setiap siswa akan tetap berada pada urutan (ranking) yang sama di dalam kelompoknya. Tes sebagai alat ukur kemampuan; tes bahasa sebagai alat ukur kemampuan bahasa seharusnya reliabilitas. Dengan tes bahasa yang reliabel tingkat penguasaan bahasa peserta tes yang dinyatakan baik seharusnya memang baik dan akan tetap dinyatakan baik seandainya dites ulang pada peserta tes yang sama dalam waktu yang berbeda. Hal itulah yang merupakan hakikat reliabilitas. Alat tes tersebut dapat mengukur secara konsisten, ajeg, ti-

dak berubah-ubah. Tuckman (dalam Nurgiyantoro, 2010: 165) menyatakan bahwa istilah reliabilitas tes menunjuk pada pengertian apakah suatu tes dapat mengukur secara konsisten sesuatu yang akan diukur dari waktu ke waktu.

Grondlund (dalam Nurgiyantoro, 2010: 165) mengemukakan bahwa reliabilitas menunjuk pada pengertian konsistensi pengukuran, yaitu seberapa konsisten skor tes atau hasil evaluasi dari satu pengukuran ke pengukuran yang lain. Tinggi rendahnya pengukuran akan memengaruhi validitas. Tinggi rendahnya reliabilitas akan memungkinkan tercapainya validitas. Selanjutnya, Grondlund menegaskan beberapa hal berikut.

- 1) Reliabilitas menunjuk pada hasil yang diperoleh yang diuji dengan suatu instrumen tes dan bukan alat tes tersebut. Dengan kata lain, akan lebih tepat dikatakan reliabilitas skor tes atau pengukuran daripada reliabilitas alat tes.
- 2) Reliabilitas berkaitan dengan hal-hal yang khusus dan tidak dapat digeneralisasikan untuk semua situasi. Ia akan berbeda, misalnya jika waktu pengukuran berbeda atau sampel soal yang berbeda. Konsistensi skor tes hanya berlaku untuk satu kesempatan atau situasi dan bukan untuk berbagai kesempatan atau situasi.
- 3) Reliabilitas itu diperlukan, tetapi belum menjamin capaian validitas. Meskipun demikian, untuk meraih validitas harus memenuhi reliabilitas. Jadi, reliabilitas dapat dianggap sebagai prasyarat untuk mencapai validitas.
- 4) Reliabilitas berkaitan dengan perhitungan statistik sehingga membutuhkan bukti skor tes.

Pengertian konsistensi dalam reliabilitas tes berhubungan dengan hal-hal (a) tes dapat memberikan hasil yang relatif tetap

terhadap sesuatu yang diukur; (b) jawaban peserta didik terhadap butir-butir tes relatif tetap; dan (c) hasil tes diperiksa oleh siapa pun juga akan menghasilkan skor yang kurang lebih sama (dalam Nurgiyantoro, 2010: 166).

Allison (1999: 85) menyatakan bahwa reliabilitas tes menyangkut akurasi dan keterpercayaan hasil-hasilnya. Jika memori siswa pada saat ujian dapat dihapus, kita dapat melakukan tes atau ujian ulang dan hasilnya hampir sama. Hasil tes reliabilitas yang akurat akan mencerminkan pemahaman tiap-tiap siswa yang sedang diuji. Siswa yang jawabannya hanya menebak-nebak akan menyebabkan rendahnya reliabilitas tes.

Pengukuran yang dilakukan tidak hanya mencerminkan berapa banyak peserta didik telah berhasil dalam kegiatan belajar, tetapi juga bagaimana keakuratan alat tes yang digunakan. Keakuratan tes akan memengaruhi skor peserta didik. Skor yang diperoleh peserta didik tidak akan secara sempurna mencerminkan kemampuan hasil belajarnya. Karena tidak ada satu pun tes yang sempurna, kita perlu mengusahakan konsistensi hasil pengukuran dan mengetahui berapa tinggi tingkat reliabilitas itu. Informasi tentang tinggi rendahnya tingkat reliabilitas tes tersebut berguna untuk menentukan langkah selanjutnya, yakni perlu atau tidaknya tes itu diperbaiki.

2.2.2 Cara Estimasi Reliabilitas

Untuk mengetahui bahwa tes yang digunakan dapat mengukur secara konsisten sesuatu yang diukur digunakan prosedur untuk mengestimasi reliabilitas tes. Tuckman (dalam Nurgiyantoro, 2010: 167) mengemukakan bahwa prosedur atau teknik yang dapat digunakan untuk mengestimasi reliabilitas pengukuran ada empat, yaitu (1) teknik belah dua (*split-half*), (2) pengukuran dengan rumus Kuder Richardson 20 dan 21, (3) teknik

butir paralel (*parallel-item*), dan (4) teknik bentuk paralel (*alternate-forms*).

Fernandes (dalam Nurgiyantoro, 2010: 167) membedakan jenis reliabilitas tes ke dalam tiga golongan, yaitu (1) jenis konsistensi internal (*internal consistency*), yang terdiri atas empat macam, yaitu (a) koefisien Alpha, (b) Kuder-Rechardson 20, (c) Kuder-Rechardson 21, dan (d) belah-dua Spearman Brown; (2) stabilitas (*stability*), yaitu berupa teknik uji ulang; dan (3) ekuivalensi (*equivalence*), yakni berupa teknik bentuk paralel.

Berdasarkan pendapat kedua pakar tersebut dapat disimpulkan bahwa prosedur yang dapat digunakan untuk mengestimasi reliabilitas tes ada enam, yaitu (1) reliabilitas ulang uji, (2) reliabilitas belah dua, (3) reliabilitas rumus Kuder-Richardson 20, (4) reliabilitas rumus Kuder-Richardson 21, (5) reliabilitas Alpha Cronbach, dan (6) reliabilitas bentuk paralel.

2.2.2.1 Reliabilitas Ulang Uji

Teknik tes ulang uji adalah teknik memperkirakan tingkat reliabilitas tes dengan melakukan kegiatan pengukuran sebanyak dua kali terhadap tes yang sama kepada peserta didik yang sama pula (Nurgiyantoro, 2010:167). Hasil tes pertama dan kedua dikorelasikan. Jika koefisien korelasi yang diperoleh cukup tinggi, hasil pengukuran tes yang diujicobakan itu dinyatakan memiliki reliabilitas tinggi.

2.2.2.2 Reliabilitas Belah Dua

Pengujian reliabilitas tes dengan teknik belah dua (*split half*) dilakukan dengan memisahkan skor hasil ke dalam dua kelompok, yaitu kelompok ganjil dan kelompok genap atau kelompok awal dan kelompok akhir (Nurgiyantoro, 2010: 168). Caranya adalah dengan menghitung jumlah skor setiap butir soal yang bernomor ganjil dan soal yang bernomor genap. Jumlah kedua skor tersebut dikorelasikan untuk mendapatkan

koefisien korelasi antara keduanya. Koefisien korelasi dari hasil perhitungan tersebut hanya menunjukkan reliabilitas tes untuk seperdua soal. Untuk mendapatkan koefisien korelasi reliabilitas keseluruhan tes digunakan rumus Spearman-Brown. Rumus Spearman-Brown adalah sebagai berikut (dalam Nurgiyantoro, 2010: 169).

$$\text{reliabilitas seluruh tes} = \frac{1 + 2 \times \text{reliabilitas separuh tes}}{1 + \text{reliabilitas separuh tes}}$$

$$\text{atau} \quad r = \frac{2 \times r}{1 + r}$$

Untuk menentukan tingkat reliabilitas hasil perhitungan dapat digunakan tabel nilai-nilai korelasi karena berbagai perhitungan yang menggunakan koefisien r dengan angka paling besar 1,00 (pada umumnya di bawah 1,00) pada prinsipnya memakai logika yang sama. Taraf signifikansi nilai r yang diperoleh dipengaruhi oleh besarnya N (jumlah subjek).

2.2.2.3 Reliabilitas Rumus Kuder-Richardson 20 dan 21

Pengujian reliabilitas tes dengan menggunakan rumus Kuder-Richardson (K-R) 20 dan 21 dilakukan dengan membandingkan skor butir-butir tes. Jika butir-butir tes itu menunjukkan tingkat kesesuaian yang tinggi (*degree of agreement*), kita dapat menyimpulkan bahwa hasil pengukuran tes tersebut konsisten (Nurgiyantoro, 2010: 169).

Rumus K-R 20 adalah sebagai berikut.

$$r = \frac{n}{n-1} \left(1 - \frac{\sum pq}{s^2} \right)$$

Keterangan:

r = koefisien reliabilitas tes

n = jumlah butir soal

p = proporsi jawaban betul

q = proporsi jawaban salah ($q = 1 - p$)

s = simpangan baku, s^2 ; varian

Rumus K-R 21 adalah sebagai berikut

$$r = \frac{n}{n-1} \left(1 - \frac{\bar{X}(n-\bar{X})}{ns^2} \right)$$

Keterangan:

\bar{X} = rata-rata hitung (mean)

Sumber: Nurgiyantoro (2010: 170)

2.2.2.4 Reliabilitas Alpha Cronbach

Naga 1992 (dalam Nurgiyantoro, 2010: 171) mengemukakan bahwa koefisien reliabilitas Alpha Cronbach diterapkan pada tes yang mempunyai skor berskala dan dikotomis sekaligus. Artinya, prosedur uji reliabilitas ini diterapkan pada hasil pengukuran yang berjenjang, misalnya: 1-4, 1-5, 1-6, atau bergantung pada tujuan penyusunannya.

Rumus koefisien reliabilitas Alpha Cronbach adalah sebagai berikut.

$$r = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_t^2} \right)$$

Keterangan:

k = jumlah butir soal

$\sum s_i^2$ = jumlah varian butir-butir

s_t^2 = varian total (untuk seluruh butir tes)

Sumber: Fernandes dalam Nurgiyantoro (2010: 171)

2.2.2.5 Reliabilitas Bentuk Paralel

Nurgiyantoro (2010: 172) menyatakan bahwa pengujian reliabilitas hasil pengukuran tes dengan teknik butir paralel dilakukan pada dua perangkat tes yang bersifat paralel. Kedua perangkat tes itu dimaksudkan untuk mengukur tujuan atau kompetensi yang sama, dengan jumlah butir, susunan, dan tingkat kesulitan yang kurang lebih sama. Jadi, dua perangkat tes dibuat berdasarkan spesifikasi yang sama.

Untuk menguji reliabilitas hasil pengukuran tes, kedua perangkat tes tersebut diujicobakan kepada subjek yang sama, kemudian hasilnya dikorelasikan. Tinggi rendahnya koefisien korelasi akan mencerminkan

kan reliabilitas hasil pengukuran kedua perangkat tes tersebut.

3. Simpulan

Hasil evaluasi dalam kegiatan pembelajaran bahasa dimaksudkan sebagai (1) umpan balik bagi seluruh aspek penyelenggaraan pembelajaran yang meliputi kurikulum yang dijadikan landasan pembelajaran, bahan ajar yang digunakan, latihan, dan tugas-tugas yang diberikan kepada pemelajar; (2) umpan balik bagi pengajar terhadap cara mengajar, pilihan dan kesesuaian bahan ajar, tes yang digunakan, dan lain-lain. Untuk mengetahui hasil evaluasi diperlukan suatu alat evaluasi, yakni tes dan non-tes. Tes sebagai salah satu alat evaluasi hasil belajar para siswa diharapkan mampu memberikan informasi yang dapat dipertanggungjawabkan kebenarannya. Untuk itu, tes yang digunakan harus bermutu dan dapat dipertanggungjawabkan, serta memiliki ciri-ciri yang dipersyaratkan, terutama validitas dan reliabilitas.

Validitas atau kesahihan adalah relevansi, kecocokan, atau kesesuaian alat evaluasi dengan fungsi evaluasi dan sasaran evaluasi. Sebuah alat evaluasi dapat dikatakan valid apabila dapat mengukur hasil belajar secara tepat. Begitu pula dengan tes yang reliabel. Sebuah tes dikatakan reliabel atau terpercaya apabila hasil-hasil tes tersebut benar-benar dapat dipercaya karena bersifat ajeg dan tidak berubah secara mencolok. Artinya, jika para siswa diberikan tes yang sama pada waktu yang berlainan, setiap siswa akan tetap berada pada urutan (ranking) yang sama di dalam kelompoknya.

Daftar Pustaka

- Allison, Desmond. 1999. *Language Testing and Evaluation an Introductory Course*. Singapore: Singapore University Press.
- Djiwandono, M. Soenardi. 2008. *Tes Bahasa*. Jakarta: Indeks.
- . 2011. *Tes Bahasa: Pegangan bagi Pengajar Bahasa*. Edisi 2. Jakarta: Indeks.
- Nurgiyantoro, Burhan. 2010. *Penilaian Pembelajaran Bahasa: Berbasis Kompetensi*. Edisi Pertama. Yogyakarta: BPFE.