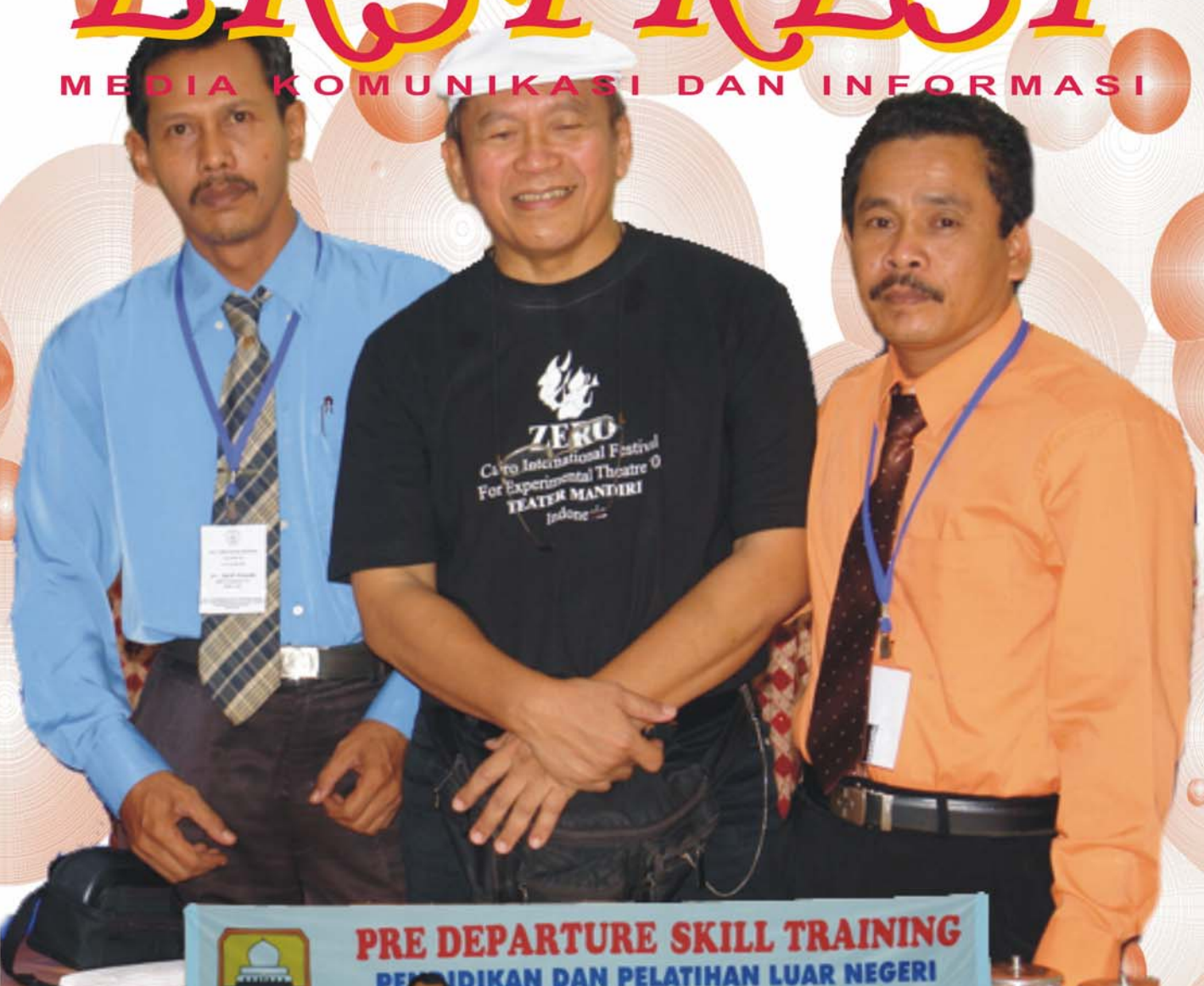


# EKSPRESI

MEDIA KOMUNIKASI DAN INFORMASI



**PRE DEPARTURE SKILL TRAINING**  
PENDIDIKAN DAN PELATIHAN LUAR NEGERI  
GURU BERPRESTASI PROPINSI JAMBI  
TAMBAH 2005



Departemen Pendidikan Nasional  
Direktorat Jenderal Peningkatan Mutu Pendidik dan Tenaga Kependidikan  
Pusat Pengembangan Penataran Guru Bahasa



# Mengapa dan Bagaimana Tes Bahasa yang Baik: Suatu Pengenalan

Oleh Widiatmoko



## Pengantar

Tes, sebagaimana diketahui, merupakan sekumpulan pertanyaan, pernyataan, atau tugas yang sistematis yang memerlukan respon dari peserta ujian (siswa)

dalam rangka keberukuran keterampilan, pengetahuan, intelegensi, kemampuan, atau minat yang dimilikinya (Widiatmoko, 2004: 5).

Istilah tersebut apabila dirujuk pada sejarah perkembangannya sesungguhnya mencakupi alat ukur tes yang memerlukan respon dikotomi (betul-salah) dan yang memerlukan respon berjenjang. Yang terakhir ini bertalian dengan skala. Dikenalilah skala Likert, semantik diferensial, skala Guttman (analisis skalogram), skala Coombs, skala Thurstone, dan sebagainya (Naga, 1992). Baik yang memerlukan respon dikotomi maupun kontinum tersebut, keduanya bertalian dengan atribut dari orang, objek, dan peristiwanya.

Dengan demikian, dikenal istilah hasil belajar siswa (atribut dari orang), sikap siswa (atribut dari orang), kedalaman ilmu (atribut dari objek), kecepatan penyampaian isi pidato (atribut dari peristiwa), dan sebagainya.

Seiring dengan pandangan strukturalis psikometris, tes berrespon dikotomi menjadi

dominan. Ini kemudian lebih dikenal sebagai tes pilihan berganda. Ditilik dari teori tes klasik, tes tersebut memiliki banyak kelemahan di dalam desain dan konstruksinya. Ditilik dari teori responsi butir, ia banyak kelebihanannya. Kelemahan-kelemahan itu antara lain berupa ketergantungan responden dari butir-butirnya, ketergantungan butir dari para responden, statistik butir berubah-ubah, dan sebagainya.

Sedangkan kelebihan-kelebihannya mencakupi: pembebasan ketergantungan responden dari butir-butirnya, pembebasan ketergantungan butir dari para respondennya, butir bersifat unidimensi, dan karakteristik butir selalu berada pada skala yang sama dengan karakteristik respondennya yang ditunjukkan dengan nilai statistik yang konstan (Hulin, Drasgow, & Parsons, 1983).

Dalam perkembangannya, baik teori tes klasik maupun teori responsi butir, keduanya secara simultan mempengaruhi disiplin ilmu pengetahuan lainnya. Satu di antara mereka adalah ilmu bahasa. Dengan demikian, dikenalilah istilah tes bahasa. Tes bahasa dalam uraian ini hanya bertalian dengan teori tes klasik.

## Mengapa Tes Bahasa Diberikan

Di dalam banyak literatur tes bahasa, ditemukan sejumlah macam tes bahasa.

*Une bone vision de la vie cree une bonne action, et une bonne action produit un bon sentiment (Dengan visi hidup yang baik dan kegiatan yang baik akan tercipta perasaan yang baik).*

Namun secara mendasar, sebelum tes diberikan kepada para peserta, dikenali secara baik alasan-alasan mengapa suatu tes diberikan. Heaton (1990) menyajikan beberapa alasan mengapa suatu tes diberikan, antara lain: untuk mencari kemajuan belajar, memotivasi belajar, menemukan kesulitan belajar, mengetahui prestasi belajar, mengetahui tingkat kemampuan, menyeleksi calon pembelajar, dan mengetahui kelancaran berbahasa.

Apabila tes diberikan dengan alasan untuk mencari kemajuan belajar siswa, tes tersebut adalah *progress test*. Jenis tes ini memiliki karakteristik, seperti kecenderungan 80%-90% terdapat nilai B (baik) atau A (baik sekali), mencakup materi yang tidak banyak dalam rentang waktu belajar yang pendek, diberikan dengan sedikit tidak formal. Jenis tes ini juga dapat dimaksudkan untuk mencari kesulitan belajar para siswa.

Ini bermakna bahwa tes ini selain sebagai alat untuk memotivasi belajar juga untuk mencari materi-materi yang masih lemah dikuasai. Harapan bahwa nilai sebagian besar siswa tinggi, ternyata yang ditemui sebaliknya, maka tes ini pun digunakan sebagai alat ukur untuk mendeteksi kesulitan belajar. Dengan demikian, siswa mampu melakukan perbaikan dalam masa *remedial teaching*.

Sedangkan jenis tes yang dimaksudkan untuk mengetahui prestasi belajar siswa, tes formatif atau tes sumatif atau lebih dikenal dengan tes tengah semester atau tes akhir semester biasa diberikan. Karakteristik antarkeduanya sedikit berbeda. Tes formatif mencakupi jumlah materi yang lebih sedikit dibandingkan dengan yang ada pada tes sumatif. Tes formatif diberikan dalam rentang waktu yang lebih pendek dari suatu program pembelajaran, sedangkan tes sumatif

diberikan dalam rentang waktu yang lebih lama dari suatu program pembelajaran.

Tes formatif diberikan selama proses pembelajaran, sedangkan tes sumatif diberikan pada akhir program. Meskipun terdapat perbedaan, kedua jenis tes tersebut juga memiliki kesamaan, yakni kedua jenis tersebut diberikan dalam keadaan yang lebih formal, memiliki kecenderungan rentangan nilai yang merata dari terendah hingga tertinggi dan berdistribusi normal.

Ada kalanya suatu tes diberikan sebelum proses pembelajaran. Jenis tes ini meliputi tes seleksi dan tes penempatan. Tes seleksi bertujuan untuk memilih calon pembelajar yang terbaik. Selain itu, tes ini juga memiliki karakteristik, seperti menggunakan acuan penilaian kriteria, mencakupi jumlah materi yang bersifat umum dan luas, membutuhkan waktu yang tidak sebentar untuk memenuhi jumlah pembelajar, dan lebih memberatkan pihak penyelenggara karena pada suatu saat tidak diperoleh calon pembelajar yang terbaik.

Sebaliknya, tes penempatan adalah tes yang bertujuan untuk memilih calon pembelajar sesuai dengan tingkatnya, misalnya tingkat mahir, tingkat semenjana, tingkat dasar, dan sebagainya. Biasanya yang sering menggunakan jenis tes ini adalah penyelenggara kursus bahasa. Namun, tidak menutup kemungkinan jenis tes ini dipadukan dengan jenis tes seleksi. Ini dapat dilakukan apabila penyelenggara tes memiliki calon pembelajar dalam jumlah yang sangat banyak, sedangkan yang dibutuhkan hanya separonya.

Dengan demikian, tes seleksi dilakukan untuk mencari separo calon pembelajar terbaik. Selain itu, dari separo jumlah peserta yang diterima tersebut, diklasifikasikan menurut nilai capaian yang mereka dapatkan, misalnya 20 pembelajar bernilai terbaik berada

*Aset manusia yang paling berharga dalam adalah sikap positif.*—**Bobbi DePorter**

di kelas A, 20 berikutnya di kelas B, dan seterusnya.

Ada juga tes yang diberikan tidak pada awal, proses, dan akhir program, melainkan berdiri sendiri. Jenis tes ini lazim ditemui, seperti *proficiency test*. Tes ini bertujuan untuk mengetahui kemahiran berbahasa seseorang. Biasanya seseorang yang melakukan tes ini pernah belajar bahasa tersebut dalam kurun waktu tertentu. Tes ini mencakupi *Test of English as a Foreign Language* (TOEFL), *Test of English for International Communication* (TOEIC), dan sebagainya. Tes pertama dimaksudkan untuk tujuan akademik, sedangkan tes kedua untuk tujuan vokasional.

### **Bagaimana Tes Bahasa yang Baik**

Agar tes bahasa dapat digunakan sebagai alat ukur, ia harus memenuhi persyaratan, yang meliputi reliabilitas dan validitas.

Reliabilitas suatu tes, menurut Nurgiyantoro (1995), berarti bahwa tes tersebut bersifat konsisten. Ia bermakna bahwa hasil ukur tes kurang lebih sama dengan sesuatu yang diukurinya meskipun tes tersebut diberikan di dalam kurun waktu atau oleh penguji yang berbeda. Naga (2002) mendefinisikan reliabilitas sebagai tingkat kecocokan antara hasil ukur dan keadaan sesungguhnya pada responden.

Di dalam literatur, disebutkan ada banyak jenis reliabilitas yang dapat digunakan. Ini mencakupi jenis konsistensi internal, stabilitas, dan ekuivalensi. Konsistensi internal terdiri atas pilah paruh Spearman Brown, koefisien Alpha Cronbach, Kuder-Richardson 20 (KR-20), Kuder-Richardson 21 (KR-21). Stabilitas terdiri atas uji-uji ulang. Sedangkan ekuivalensi terdiri atas uji-uji setara.

Pertama uji pilah paruh (*split half*) dilakukan dengan cara memisahkan skor hasil

tes ke dalam kelompok ganjil dan kelompok genap. Kemudian dilakukan perhitungan jumlah skor kelompok ganjil dan jumlah skor kelompok genap. Kedua jumlah skor tersebut dikorelasikan untuk mendapatkan koefisien korelasi ( $r$ ). Akhirnya koefisien korelasi seluruh tes dapat dihitung dengan rumus  $r = (2xr)/(1+r)$ , di mana  $r$  adalah reliabilitas. Dalam pengujiannya, dapat ditemukan berbagai macam koefisien.

Salah satu di antara mereka adalah koefisien Alpha Cronbach. Ini diterapkan pada tes yang memiliki skor berskala atau politomi. Ini bermakna bahwa skor tes itu memiliki sejumlah kemungkinan yang berjenjang, misalnya 1-5 atau yang lain bergantung pada maksud penyusunannya. Pertanyaan-pertanyaan yang menggunakan koefisien reliabilitas Alpha Cronbach ini biasanya bertalian dengan sikap, minat, motivasi, dan lain-lain yang jawabannya berskala. Rumus koefisien Alpha Cronbach adalah  $r = [K/(k-1)] \times [1 - (SSi^2)/(St^2)]$ ,  $K$  adalah jumlah butir,  $SSi^2$  adalah jumlah variansi butir,  $St^2$  adalah variansi total. Selain itu, dapat juga dihitung koefisien Kuder Richardson 20 dan 21 (KR-20 dan KR-21).

Pengujian reliabilitas dengan menggunakan KR-20 dan KR-21 dilakukan dengan membandingkan skor butir-butir tes. Apabila butir-butir tes menunjukkan tingginya tingkat kecocokan, disimpulkan bahwa tes tersebut akurat atau mengukur secara akurat. Penggunaan KR-20 menghasilkan koefisien lebih besar daripada yang dengan menggunakan KR-21. KR-20 juga lebih rumit. Tetapi jika untuk menguji tes yang bersifat heterogen dan mencakupi berbagai pokok bahasan, KR-20 lebih direkomendasikan.

Sebaliknya KR-21 sekalipun lebih sederhana dan mengukur secara lebih akurat dalam

Kikuwa ittoki no haji kikanu wa isshou ni haji (*Jangan malu untuk bertanya*).

memberikan penafsiran, ia tidak mampu menguji alat tes yang heterogen. Rumus KR-20 adalah  $r = [n/(n-1)] \times [1 - (Spq)/(S^2)]$ , n adalah jumlah butir, p adalah proporsi jawaban betul, q proporsi jawaban salah, dan  $S^2$  variansi. Sedangkan rumus KR-21 adalah  $r = [n/(n-1)] \times [1 - \{i(n-i)\}/\{nS^2\}]$ , i adalah rata-rata (*means*).

Kedua adalah uji-uji ulang. Teknik ini digunakan untuk memperkirakan tingkat stabilitas tes dengan melakukan kegiatan pengukuran dua kali pada tes yang sama kepada siswa yang sama. Hasil tes pertama dan kedua kemudian dikorelasikan. Apabila koefisien korelasi (*r*) tinggi, tes yang diujicobakan dinyatakan stabil atau reliabel.

Teknik uji-uji ulang memiliki beberapa kelemahan, antara lain: sulit untuk menghilangkan pengaruh jawaban tes yang pertama, adanya kemungkinan faktor-faktor yang mempengaruhi hasil tes kedua (misalnya berupa meningkatnya kemampuan siswa sebagai hasil belajar), sulit menciptakan dua kondisi penyelenggaraan dua kali tes, menuntut siswa mengalami dua kali tes yang dirasa kurang menguntungkan dan membebankan siswa.

Ketiga adalah uji-uji setara. Pengujian tingkat reliabilitas tes dengan uji-uji setara dilakukan terhadap dua perangkat tes yang setara. Kedua tes tersebut memiliki jumlah butir, susunan, tingkat kesulitan, dan tujuan pengukuran yang sama. Ini dilakukan dengan cara mengujicobakan kedua tes tersebut kepada subjek yang sama, kemudian hasilnya dikorelasikan. Pengujian ini hampir sama dengan uji-uji ulang. Tetapi, yang membedakan adalah bahwa uji-uji setara terdiri atas dua perangkat tes yang berbeda.

Selain syarat reliabilitas, suatu tes juga dituntut memenuhi syarat validitasnya. Naga (2002) mendefinisikan validitas sebagai ke-

cocokan antara alat ukur dan sasaran ukur. Harrison (1983) mendefinisikan validitas sebagai sejauh mana alat ukur mampu mengukur apa yang seharusnya diukur.

Secara umum, validitas terdiri atas validitas isi, validitas kriteria, dan validitas konstruk. Meskipun demikian, sebagian ahli membagi validitas menjadi validitas isi, validitas prediktif, validitas serentak, dan validitas konstruk.

Pertama, validitas isi menunjuk pada pengertian apakah alat ukur tes itu memiliki kesesuaian dengan tujuan dan deskripsi bahan ajar. Apabila butir-butir tes secara jelas dimaksudkan untuk mengukur tujuan-tujuan tertentu dan mewakili bahan yang diajarkan, dikatakan tes tersebut memiliki validitas isi.

Pemenuhan validitas isi tersebut biasanya dilihat dari ketersediaan kisi-kisi yang baik yang dipakai sebagai dasar penyusunan butir-butir tes di samping juga ketepatan masing-masing butir itu sendiri. Validitas isi pada dasarnya tidak memerlukan perhitungan statistik. Dengan demikian, sebenarnya validitas isi hampir sama dengan validitas wajah yang hanya memperhatikan kesesuaian elemen-elemen yang harus dipenuhi dalam kisi-kisi tes.

Kedua adalah validitas konstruk. Validitas ini bertalian dengan konstruk atau konsep bidang bahasa atau bidang lainnya yang akan diuji validitasnya. Ia menunjuk pada pengertian apakah tes yang disusun itu telah sesuai dengan konsep ilmu yang diteskan tersebut. Untuk menentukan tingkat validitas konstruk, penyusunan butir dilakukan dengan mendasarkan diri pada kisi-kisi. Dalam praktiknya, jenis validitas konstruk digunakan untuk mempertimbangkan tingkat validitas butir yang bertalian dengan masalah sikap, motivasi,

*Le monde que tu vis existe dans ton esprit, alors change ton monde en changeant ton esprit (Duniamu ada dalam benakmu maka ubahlah duniamu dengan mengubah cara berpikirmu).*

minat, dan lain-lain yang menggunakan skala bertingkat.

Ketiga adalah validitas kriteria. Validitas ini menunjuk pada pengertian seberapa jauh siswa yang sudah diajarkan bidang bahasa atau bidang lainnya menunjukkan kemampuan yang lebih tinggi daripada mereka yang belum diajarkan.

Naga (2002) menyebutkan validitas ini bertujuan untuk menentukan kecocokan antara hasil ukur berdasarkan pada sasaran ukur prediktor dan sasaran ukur kriteria. Hasil ukur atau skor prediktor merupakan hasil ukur yang diperoleh melalui penerapan alat ukur pada subjek yang validitasnya diperiksa. Sedangkan hasil ukur kriteria merupakan acuan untuk melihat kecocokannya dengan hasil ukur prediktor.

Sebagaimana diketahui, validitas ini terdiri atas validitas serentak dan validitas prediksi. Validitas serentak merupakan validitas di mana hasil ukur kriteria dan hasil ukur prediktor terjadi pada waktu yang sama. Sedangkan validitas prediksi merupakan validitas di mana hasil ukur kriteria terjadi kemudian setelah hasil ukur prediktor diketahui. Dengan demikian, secara sederhana dapat dikatakan bahwa agar tes bahasa memenuhi kriteria yang baik, seyogyanya ia memenuhi syarat valid dan reliabelnya.

### Penutup

Membicarakan tes tidak akan lepas dari sejarah perkembangannya. Diselusuri dari sejarahnya, suatu tes bertalian dengan jenis tes yang memiliki respon dikotomi dan yang memiliki respon kontinum (politomi/berjenjang). Tes bahasa yang berkembang saat ini tentu memiliki keterpengaruhannya dari itu semua. Tarik ulur pengaruh antara teori tes klasik dan teori responsi butir menjadikan tes bahasa menjadi menarik untuk selalu dikaji.

Kajian tes bahasa dalam pandangan teori responsi butir akan sangat banyak membutuhkan waktu dibandingkan dengan kajian tes bahasa dalam pandangan teori tes klasik. Namun demikian, tes bahasa dalam pandangan teori tes klasik pun masih sering diadun oleh para pakar pendidikan yang berkecimpung di wilayah ini. Secara praktis, kelemahan-kelemahan tes bahasa dalam pandangan ini dapat ditutupi oleh kriterium persyaratan yang mencakupi validitas dan reliabilitasnya. [ ]

### Pustaka Rujukan

- Harrison, Andrew. *A Language Testing Handbook*. London: Macmillan Press. 1983.
- Heaton, J.B. *Classroom Testing*. London: Longman Group. 1990.
- Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983). *Item response theory: Application to psychological measurement*. Homewood, Illinois: Dow Jones-Irwin.
- Naga, Dali S. *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma. 1992.
- Naga, Dali S. *Teori Tes dan Pengukuran*. Kertas kerja disajikan pada kuliah Program Doktor Penelitian dan Evaluasi Pendidikan, Universitas Negeri Jakarta. 2002.
- Nurdiyantoro, Burhan. *Penilaian dalam Pengajaran Bahasa dan Sastra*. Yogyakarta: BPFE. 1995.
- Widiatmoko. (2004). *Language Assessment: Bahan Ajar Diklat Tingkat Dasar Guru Bahasa Inggris Sekolah Menengah Atas*. Jakarta: PPPG Bahasa.

*Penulis adalah Widyaiswara Bahasa Inggris PPPG Bahasa.*

*Il faut aimer ce que l'on a (Kita harus mencintai apa yang kita miliki).*