

MENGANALISIS BUTIR SOAL

Revoltje O.W. Kaunang

Fakultas Ilmu Sosial Universitas Negeri Gorontalo

Abstrak: Salah satu tujuan dari analisis butir soal adalah untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek. Dengan analisis butir soal dapat diperoleh informasi tentang kejelekan dan kebaikan sebuah soal dan petunjuk untuk mengadakan perbaikan. Ada tiga masalah yang berhubungan dengan analisis butir soal, yaitu taraf kesukaran, daya pembeda, dan pola jawaban soal. a) Taraf kesukaran, adalah suatu bilangan yang menunjukkan sukar dan mudahnya suatu soal; b) Daya pembeda adalah kemampuan sesuatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh (berkemampuan rendah); c) Pola jawaban soal adalah distribusi testee dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda.

Kata-kata kunci: taraf kesukaran, daya pembeda, pola jawaban soal.

Tidak ada usaha guru yang lebih baik selain usaha untuk selalu meningkatkan mutu tes yang disusunnya. Guru yang sudah berpengalaman, mengajar dan menyusun soal-soal tes, juga masih sukar menyadari bahwa tesnya masih belum sempurna. Oleh karena itu, cara yang paling baik adalah secara jujur melihat hasil yang diperoleh siswa.

Secara teori, siswa dalam satu kelas merupakan populasi atau kelompok yang keadaannya heterogen. Dengan demikian, maka apabila dikenai sebuah tes akan tercermin hasilnya dalam suatu kurva normal. Sebagian besar siswa berada di daerah sedang, sebagian kecil berada di ekor kiri, dan sebagian kecil yang lain berada di ekor kanan kurva.

Apabila keadaan setelah hasil tes dianalisis tidak seperti yang diharapkan dalam kurva normal, maka tentu ada “apa-apa” dengan soal tesnya. Apabila hampir seluruh siswa memperoleh skor jelek, berarti bahwa tes yang disusun mungkin terlalu sukar. Sebaliknya jika seluruh siswa memperoleh skor baik, dapat diartikan bahwa tesnya terlalu mudah. Tentu saja interpretasi terhadap soal tes akan lain seandainya tes itu sudah disusun sebaik-baiknya sehingga memenuhi persyaratan sebagai tes.

Dengan demikian maka apabila kita memperoleh keterangan tentang hasil tes, akan membantu kita dalam mengadakan penilaian secara objektif terhadap tes yang kita susun. Arikunto (2005: 205) mengemukakan ada 4 cara untuk menganalisis butir tes, yakni:

- a. Cara pertama meneliti secara jujur soal-soal yang sudah di susun, kadang-kadang dapat diperoleh jawaban tentang ketidakjelasan perintah dan bahasa, taraf kesukaran, dan lain-lain keadaan soal tersebut. Pertanyaan-pertanyaan tersebut antara lain: 1) Apakah banyaknya soal untuk tiap topic sudah seimbang?; 2) Apakah semua soal menanyakan bahan yang telah diajarkan?; 3) Apakah semua soal menanyakan bahan yang telah diajarkan?; 4) Apakah soal itu tidak sukar untuk dimengerti?; 5) Apakah soal itu dapat dikerjakan oleh sebagian besar siswa?
- b. Cara kedua adalah mengadakan analisis soal. Analisis butir soal adalah suatu prosedur yang sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes yang kita susun.

Faedah mengadakan analisis butir soal: 1) Membantu guru dalam mengidentifikasi butir-butir soal yang jelek; 2) Memperoleh informasi yang akan dapat digunakan untuk menyempurnakan soal-soal untuk kepentingan lebih lanjut; Memperoleh gambaran secara selintas tentang keadaan yang kita susun. Analisis butir soal terutama dapat dilakuakn untuk tes objektif. Hal ini tidak berarti bahwa tes uraian tidak dapat dianalisis, akan tetapi memang dalam menganalisis butir tes uraian, belum ada pedoman secara standar.

- c. Cara ketiga adalah mengadakan checking validitas. Validitas yang paling penting dari tes buatan guru adalah validitas kurikuler (*content validity*). Untuk mengadakan checking validitas kurikuler, kita harus merumuskan tujuan setiap bagian pelajaran secara khusus dan jelas sehingga setiap soal dapat kita jodohkan denganq setiap tujuan khusus tersebut.

Tes yang tidak mempunyai validitas kurikuler atau walaupun mempunyai tetapi kecil maka dapat juga terjadi salah satu atau beberapa tujuan khusus yang tidak dicantumkan, berarti bahwa validitas kurikulernya semakin kecil.

Dalam hal ini Terry D. Ten Brink (1990) mengemukakan pendapatnya demikian: 1) Untuk tes yang dirancang akan menggunakan *norm-referenced* tidak harus menuliskan setiap tujuan khusus, tetapi cukup

dengan tujuan-tujuan yang esensial saja; 2) Untuk tes yang dirancang akan menggunakan *criterion referenced*, maka setiap tujuan khusus harus dicantumkan dalam tabel spesifikasi

- d. Cara keempat adalah dengan mengadakan checking reliabilitas.

Salah satu indikator untuk tes yang mempunyai reliabilitas yang tinggi adalah bahwa kebanyakan dari soal-soal tes itu mempunyai daya pembeda yang tinggi.

Salah satu tujuan dari analisis butir soal adalah untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek. Dengan analisis butir soal dapat diperoleh informasi tentang kejelekan sebuah soal dan petunjuk untuk mengadakan perbaikan.

Kapan sebuah soal dikatakan baik? Untuk memberikan jawaban terhadap pertanyaan ini, perlu diterangkan tiga masalah yang berhubungan dengan analisis butir soal, yaitu taraf kesukaran, daya pembeda, dan pola jawaban soal.

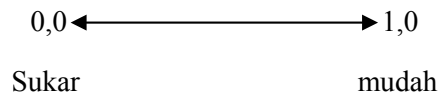
a. Taraf Kesukaran

Nitko (1983) (*dalam*, Arikunto: 2005) mengemukakan Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena diluar jangkauannya.

Seorang siswa akan menjadi hafal akan kebiasaan guru-gurunya dalam hal pembuatan soal ini. Misalnya guru A dalam memberikan ulangan soalnya mudah-mudah, sebaliknya guru B kalau memberikan ulangan soalnya sukar-sukar. Dengan pengetahuannya tentang kebiasaan ini, maka siswa akan belajar giat jika menghadapi ulangan dari guru B dan sebaliknya jika akan mendapat ulangan dari guru A, tidak mau belajar giat atau bahkan mungkin tidak mau belajar sama sekali.

Bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran (*difficulty index*). Besarnya indeks kesukaran antara 0,00 sampai dengan 1,0. Indeks kesukaran ini menunjukkan taraf kesukaran soal.

Soal dengan indeks kesukaran 0,0 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1,0 menunjukkan bahwa soalnya terlalu mudah.



Di dalam istilah evaluasi, indeks kesukaran ini diberi symbol P (p besar), singkatan dari kata “proporsi”. Dengan demikian maka soal dengan $P=0,70$ lebih mudah jika dibandingkan dengan $P = 0,20$. Sebaliknya soal dengan $P = 30$ lebih sukar dari pada soal dengan $P = 0,80$.

Melihat besarnya bilangan indeks ini maka lebih cocok jika bukan disebut sebagai indeks kesukaran tetapi indeks kemudahan atau indeks fasilitas, karena semakin mudah soal itu, semakin besar pula bilangan indeksnya. Akan tetapi telah disepakati bahwa walaupun semakin tinggi indeksnya menunjukkan soal yang semakin mudah, tetapi tetap disebut indeks kesukaran.

Rumus mencari P adalah:

$$P = \frac{B}{JS}$$

Di mana:

P = indeks kesukaran

B = banyaknya siswa yang menjawab soal itu dengan betul

JS = jumlah seluruh siswa peserta tes

Contoh penggunaan

Misalnya jumlah siswa peserta tes dalam suatu kelas ada 20 orang. Dari 20 orang siswa tersebut ada 10 orang yang dapat mengerjakan soal nomor 2 dengan betul. Maka indeks kesukarannya adalah:

$$P = \frac{10}{20} = 0,5$$

Menurut ketentuan yang sering diikuti, indeks kesukaran sering diklasifikasikan sebagai berikut:

Soal dengan P 1,00 sampai 0,30 adalah soal sukar

Soal dengan P 0,30 sampai 0,70 adalah soal sedang

Soal dengan P 0,70 sampai 1,00 adalah soal mudah

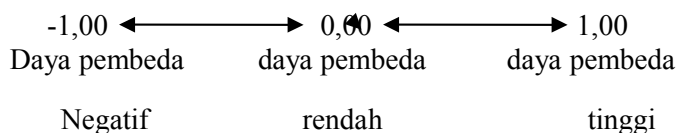
Walaupun demikian ada yang berpendapat bahwa soal-soal yang dianggap baik, yaitu soal-soal sedang, adalah soal-soal yang mempunyai indeks kesukaran 0,30 sampai dengan 0,70.

Perlu diketahui bahwa soal-soal yang terlalu mudah atau terlalu sukar, lalu tidak berarti tidak boleh digunakan. Hal ini tergantung dari penggunaannya. Jika dari pengikut yang banyak, kita menghendaki yang lulus hanya sedikit, kita ambil siswa yang paling top. Untuk ini maka lebih baik diambilkan butir-butir tes yang sukar. Sebaliknya jika kekurangan pengikut ujian, kita pilihkan soal-soal yang mudah. Selain itu, soal yang sukar akan menambah gairah belajar siswa yang paidai, sedangkan soal-soal yang terlalu mudah, akan membangkitkan semangat kepada siswa yang lemah.

b. Daya pembeda

Daya pembeda soal, adalah kemampuan sesuatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh (berkemampuan rendah). Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminasi, disingkat D (d besar). Seperti

halnya indeks kesukaran, indeks diskriminasi (daya pembeda) ini berkisar antara 0,00 sampai 1,00. Hanya bedanya, indeks kesukaran tidak mengenal tanda negative (-), tetapi pada indeks diskriminasi ada tanda negatif. Tanda negatif pada indeks diskriminatif digunakan jika sesuatu soal “terbalik” menunjukkan kualitas testee. Yaitu anak pandai disebut bodoh dan anak bodoh disebut pandai. Dengan demikian ada tiga titik pada daya pembeda yaitu:



Bagi suatu soal yang dapat dijawab benar oleh siswa pandai maupun siswa bodoh, maka soal itu tidak mempunyai daya pembeda. Demikian pula jika semua siswa baik pandai maupun bodoh tidak dapat menjawab dengan benar. Soal tersebut tidak baik juga karena tidak mempunyai daya pembeda. Soal yang baik adalah soal yang dapat dijawab benar oleh siswa-siswa yang pandai saja.

Seluruh pengikut tes dikelompokkan menjadi 2 kelompok, yaitu kelompok pandai atau kelompok atas (*upper group*) dan kelompok bodoh atau kelompok bawah (*lower group*). Jika seluruh kelompok atas dapat menjawab soal tersebut dengan benar, sedang seluruh kelompok bawah menjawab salah, maka soal tersebut mempunyai D paling besar, yaitu 1,00. Sebaliknya jika semua kelompok atas menjawab salah, tetapi semua kelompok bawah menjawab betul, maka nilai D-nya -1,00. Tetapi jika siswa kelompok atas dan siswa kelompok bawah sama-sama menjawab benar atau sama-sama menjawab salah, maka soal tersebut mempunyai nilai D 0,00. Karena tidak mempunyai daya pembeda sama sekali.

Cara menentukan daya pembeda (nilai D)

Untuk ini perlu dibedakan antara kelompok kecil (kurang dari 100) dan kelompok besar (100 Orang ke atas).

a) Untuk kelompok kecil

Seluruh kelompok testee dibagi dua sama besar, 50% kelompok atas dan 50% kelompok bawah.

Contoh:

Siswa	Skor	
A	9	}
B	8	
C	7	
D	7	kelompok atas (JA)
E	6	
<hr/>		
F	5	}
G	5	
H	4	
I	4	
J	3	kelompok bawah (JB)

Seluruh pengikut tes, dideretkan mulai dari skor teratas sampai terbawah, lalu dibagi 2.

b) Untuk kelompok besar

Mengingat biaya dan waktu untuk menganalisis, maka untuk kelompok besar biasanya hanya diambil kedua kutubnya saja, yaitu 27% skor teratas sebagai kelompok atas (JA) dan 27% skor terbawah sebagai kelompok bawah (JB).

JA = Jumlah kelompok atas

JB = Jumlah kelompok bawah

Contoh:

9 }
9 }
8 }
8 }
8 } 27% sebagai JA
.
.
.
-
.
.

.	
-	
.	}
.	
.	
.	
.	
2	27% sebagai JB
1	
1	
1	
0	

Rumus mencari D

Rumus untuk menentukan indeks diskriminasi adalah:

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

Dimana:

J = jumlah peserta

JA = banyaknya peserta kelompok atas

JB = banyaknya peserta kelompok bawah

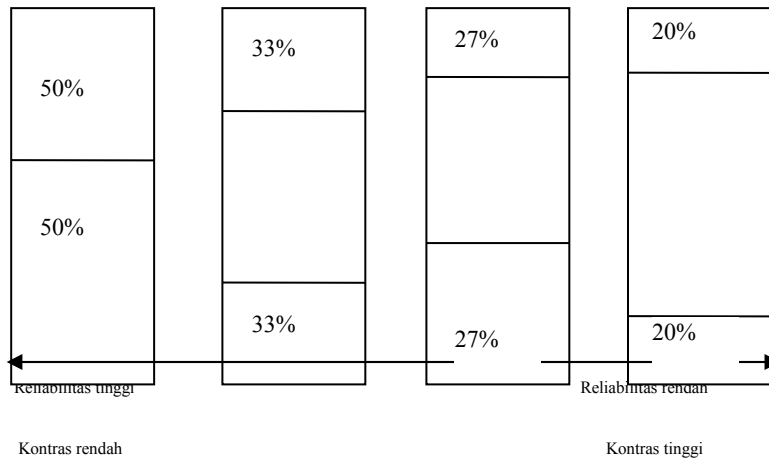
BA = banyaknya peserta kelompok atas yang menjawab soal dengan benar

BB = banyaknya peserta kelompok bawah yang menjawab soal dengan benar

$PA = \frac{BA}{JA}$ = Proporsi peserta kelompok atas yang menjawab benar

$PB = \frac{BB}{JB}$ = Proporsi peserta kelompok bawah yang menjawab benar

Dali S Naga, (1992:51-54) mengemukakan pembagian peserta ke dalam bagian tinggi dan rendah untuk menemukan kelompok tinggi dan kelompok rendah yakni:



Tampak pada gambar bahwa di sebelah kiri terdapat pembagian menjadi dua bagian sama besar 50%, dan disebelah kanan berturut-turut terdapat pembagian menjadi tiga bagian dengan bagian paling atas dan paling bawah sebesar 30%, 27%, dan 20%.

Makin ke kanan, ukuran kelompok tinggi dan kelompok rendah makin kecil, dari 50% menurun sampai 20%. Ini berarti bahwa perbedaan di antara kelompok tinggi dan kelompok rendah, makin kekanan makin besar atau dengan kata lain, makin kontras atau makin jelas.

Sekalipun demikian, kekontrasan ini dibayar dengan makin rendahnya reliabilitas. Makin kekanan makin sedikit jumlah peserta yang diikuti atau makin banyak peserta yang disisihkan. Dengan makin sedikitnya jumlah peserta yang diikutsertakan maka makin rendah pula tingkat

kepercayaan kita akan informasi itu atau dengan kata lain, makin rendah reliabilitasnya.

Mencari ukuran terbaik yakni ukuran kelompok tinggi dan kelompok rendah yang cukup kontras tetapi juga cukup reliable. Ini berarti bahwa ukuran itu tidak boleh terlalu kekiri dan tidak juga terlalu ke kanan yakni 33% dan 27%.

Klasifikasi daya pembeda:

D : 0,00 – 0,20 : jelek (poor)

D : 0,20 – 0,40 : cukup (satisfactory)

D : 0,40 – 0,70 : baik (good)

D : 0,70 – 1,00 : baik sekali (excellent)

D : negative, semuanya tidak baik, jadi semua butir soal yang mempunyai nilai D negative sebaiknya dibuang saja.

c. Pola Jawaban Soal

Yang dimaksud pola jawaban soal adalah distribusi testee dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Pola jawaban soal diperoleh dengan menghitung banyaknya testee yang memilih pilihan jawaban a, b, c, atau d atau yang tidak memilih pilihan manapun (blangko). Dalam istilah evaluasi disebut omit, disingkat O.

Dari pola jawaban soal dapat ditentukan apakah pengecoh (distractor) berfungsi sebagai pengecoh dengan baik atau tidak. Pengecoh yang tidak dipilih sama sekali oleh testee berarti bahwa pengecoh itu jelek, terlalu menyolok menyesatkan. Sebaliknya sebuah distraktor (pengecoh) dapat dikatakan berfungsi dengan baik apabila distractor tersebut mempunyai daya tarik yang besar bagi pengikut-pengikut tes yang kurang memahami konsep atau kurang menguasai bahan.

Dengan melihat pola jawaban soal, dapat diketahui:

1. Taraf kesukaran soal
2. Daya pembeda soal
3. Baik dan tidaknya distraktor.

Suatu distraktor dapat diperlakukan dengan tiga cara:

- a. Diterima, karena sudah baik
- b. Ditolak, karena tidak baik
- c. Ditulis kembali, karena kurang baik.

Kekurangannya mungkin hanya terletak pada rumusan kalimatnya sehingga hanya perlu ditulis kembali, dengan perubahan seperlunya. Suatu distraktor dapat dikatakan berfungsi baik jika paling sedikit dipilih oleh 5% pengikut tes.

Simpulan

Analisis butir soal merupakan suatu usaha guru untuk meningkatkan kualitas pembelajaran di kelas. Apabila hampir seluruh siswa memperoleh skor jelek, berarti bahwa tes yang disusun mungkin terlalu sukar. Sebaliknya jika seluruh siswa memperoleh skor baik, dapat diartikan bahwa tesnya terlalu mudah. Tentu saja interpretasi terhadap soal tes akan lain seandainya tes itu sudah disusun sebaik-baiknya sehingga memenuhi persyaratan sebagai tes. Dengan demikian maka apabila kita memperoleh keterangan tentang hasil tes, akan membantu kita dalam mengadakan penilaian secara objektif terhadap tes yang kita susun.

DAFTAR PUSTAKA

Arikunto, S. 2005. *Dasar-Dasar Evaluasi Pendidikan*, Jakarta. Bumi Aksara

- Hopkins Carles D, & Antes Richard L. 1990. *Classroom Measurement and Evaluastion*. Minneapolis. Was typeset by Stanton Publication Services, Inc.
- Naga, S.D. 1992. *Pengantar Teori Skor Pada Pengukuran Pendidikan*. Jakarta. Gunadarma
- Nitko, A. 1983. *Educational Test and Measurement an Introduction*. New York. Harcurt Brace Jovanovics, Inc.
- Ten, Brink Terry D. 1974. *Evaluation, A Practical Guide for Teacher*, Mc.Graw Hill Book Company.