

UNIVERSITAS PENDIDIKAN GANESHA
PASCASARJANA
PROGRAM STUDI PENELITIAN DAN EVALUASI PENDIDIKAN

**TARAF KESUKARAN, DAYA BEDA, DAN
EFEKTIVITAS PENGECOH**

Taraf Kesukaran Butir

Taraf kesukaran butir yang dinyatakan dengan indeks kesukaran butir didefinisikan sebagai proporsi peserta tes yang menjawab butir tersebut dengan benar. Jadi rumus yang digunakan untuk menghitung indeks kesukaran butir adalah:

$$I = \frac{B}{N}$$

I = Indeks kesukaran Butir

B = Banyaknya siswa yang menjawab butir tersebut dengan benar

N = Jumlah siswa yang mengikuti tes

Misalnya untuk butir tes nomor 1, siswa menjawab benar 19 orang dari 30 orang peserta tes. Dengan demikian dapat dihitung:

$$I = \frac{19}{30} = 0,63.$$

Indeks kesukaran butir akan bervariasi dari 0 sampai dengan 1. Butir tes dengan indeks kesukaran 0,2 lebih sukar daripada butir tes yang memiliki indeks kesukaran 0,7 karena lebih sedikit peserta tes yang menjawab benar butir tes tersebut. Butir tes dengan indeks kesukaran 0,0 merupakan butir tes yang amat sukar karena tidak satupun peserta tes menjawab benar butir tersebut. Sebaliknya, butir tes dengan indeks

UNIVERSITAS PENDIDIKAN GANESHA
PASCASARJANA
PROGRAM STUDI PENELITIAN DAN EVALUASI PENDIDIKAN

kesukaran 1,0 merupakan butir tes yang amat mudah karena semua peserta tes menjawab benar butir tersebut.

Indeks kesukaran butir bermanfaat untuk memilih mana butir tes yang akan dipilih atau dipertahankan dan mana butir tes yang akan digugurkan. Pertanyaannya, berapa indeks kesukaran butir tes yang dianggap ideal. Umumnya, butir yang dipilih adalah butir tes dengan indeks kesukaran 0,5. Akan tetapi, kondisi di lapangan menunjukkan bahwa sangat sulit untuk mendapatkan butir-butir tes dengan indeks kesukaran 0,5. Oleh karena itu, dibuat kriteria untuk indeks kesukaran butir. Salah satu kriteria yang sering digunakan adalah sebagai berikut.

Butir dengan P 0,00 sampai 0,30 tergolong sukar

Butir dengan P 0,31 sampai 0,70 tergolong sedang

Butir dengan P 0,71 sampai 1,00 tergolong mudah

Daya Beda Butir

Secara operasional, daya beda butir didefinisikan sebagai efektivitas butir untuk membedakan peserta tes yang memperoleh skor tinggi dengan peserta tes yang memperoleh skor rendah. Ada beberapa indikator daya beda butir, baik secara tidak langsung maupun secara langsung. Secara tidak langsung, koefisien korelasi biserial (γ_{pbi}) antara skor butir dengan skor total juga menyatakan daya beda butir. Akan tetapi, dalam kesempatan ini daya beda butir dikaji menggunakan pengukuran langsung, yakni dengan menggunakan perbandingan kinerja peserta tes pada kelompok atas (upper region) dan kelompok bawah (lower region) pada skor total.

Secara umum, kelompok atas dan kelompok bawah ditentukan berdasarkan skor total dengan rentangan 10 sampai 33 persen. Apabila skor total berdistribusi normal, maka 27 persen skor tertinggi dan 27

UNIVERSITAS PENDIDIKAN GANESHA
PASCASARJANA
PROGRAM STUDI PENELITIAN DAN EVALUASI PENDIDIKAN

persen sekor terendah masing-masing akan menjadi kelompok atas dan kelompok bawah. Apabila distribusi sekor total lebih datar (landai) daripada kurve normal, maka persentase yang diambil lebih besar, yakni mendekati 33 persen. Allen & Yen (1979) memberi batasan aplikasi kelompok atas dan kelompok bawah antara 25 dan 33 persen.

Indeks daya beda butir dinyatakan dengan ***d*** dan dihitung dengan menggunakan rumus berikut ini.

$$d = \frac{U - L}{N}$$

Yang mana:

U = banyak peserta tes pada kelompok atas yang menjawab butir dengan benar

L = banyak peserta tes pada kelompok bawah yang menjawab butir dengan benar

N = banyak peserta tes.

Pertanyaan yang muncul sekarang adalah seberapa indeks daya beda butir yang dianggap masuk akal. Ebel antara lain memberi batasan terhadap indeks daya beda butir seperti berikut.

Jika $d \geq 0,40$, maka butir tergolong sangat baik.

Jika $0,30 \leq d < 0,40$, maka butir tergolong cukup tetapi perlu perbaikan.

Jika $0,20 \leq d < 0,30$, maka butir tergolong kurang dan harus direvisi.

Jika $d < 0,20$, maka butir tergolong jelek, sehingga harus digugurkan.

UNIVERSITAS PENDIDIKAN GANESHA
PASCASARJANA
PROGRAM STUDI PENELITIAN DAN EVALUASI PENDIDIKAN

Sebagai contoh, berikut ini disajikan sebuah ilustrasi indeks daya beda butir. Pada ilustrasi tersebut instrumen hasil belajar yang terdiri dari 5 butir tes objektif diujicobakan terhadap 100 peserta tes. Sekor total yang diperoleh diranking dan selanjutnya ditentukan 27% sekor tertinggi sebagai kelompok atas dan 27% sekor terendah sebagai kelompok bawah. Dengan demikian diperoleh 27 orang peserta tes masuk kelompok atas dan 27 orang peserta tes masuk kelompok bawah. Selanjutnya dikaji peserta tes pada kelompok atas yang menjawab butir dengan benar (U) dan peserta tes pada kelompok bawah yang menjawab butir dengan benar (L). Ternyata diperoleh hasil seperti tercantum pada tabel di bawah ini.

No. Butir	U	L	$d = (U-L)/N$	Keterangan
1	16	9	0,26	Baik, daya beda tinggi
2	15	4	0,41	Sangat baik, daya beda sangat tinggi
3	21	4	0,63	Hampir sempurna, namun jarang diperoleh
4	27	0	1,00	Sempurna, namun tidak pernah diperoleh
5	15	15	0,00	Jelek harus direvisi
6	4	20	-0,59	Sangat jelek harus digugurkan
7	0	27	-1,00	Paling jelek harus digugurkan

Ferguson memperkenalkan formula untuk menghitung daya beda tes yang lebih umum, sehingga dapat diterapkan pula pada jenis

UNIVERSITAS PENDIDIKAN GANESHA
PASCASARJANA
PROGRAM STUDI PENELITIAN DAN EVALUASI PENDIDIKAN

instrumen non dikotomi, seperti tes esai. Formula daya beda Ferguson menyatakan rasio antara banyak perbedaan yang terjadi secara empiris saat uji coba tes dengan perbedaan maksimal yang dapat terjadi pada tes. Formula Ferguson untuk indeks daya beda butir adalah sebagai berikut.

$$d = \frac{N^2 - \sum f_i^2}{N^2 - \frac{N^2}{n+1}} = \frac{(n+1)(N^2 - \sum f_i^2)}{nN^2}$$

Yang mana:

d = Indeks daya beda

N = banyak peserta tes

f_i = frekuensi pada tiap-tiap sekor

n = banyak butir.

Sebagai contoh berikut ini disajikan contoh penerapan formula indeks daya beda Ferguson pada uji coba sebuah tes uraian yang terdiri dari 10 butir yang diujicobakan terhadap 15 orang responden yang datanya sudah dianalisis pada perhitungan indeks kesukaran di atas. . Skor maksimum masing-masing butir adalah 10 dan skor minimum masing-masing butir adalah 0. Hasil uji coba tampak pada tabel kerja di bawah ini. Pada tabel kerja di bawah ini disajikan sekor hasil uji coba untuk butir nomor 1 beserta frekuensi responden yang memperoleh masing-masing sekor (f_i).

Skor Butir No.1	f_i	f_i^2
--------------------	-------	---------

UNIVERSITAS PENDIDIKAN GANESHA
PASCASARJANA
PROGRAM STUDI PENELITIAN DAN EVALUASI PENDIDIKAN

9	1	1
8	1	1
7	4	16
6	4	16
5	3	9
4	3	9
Total	16	52

Berdasarkan tabel kerja di atas dapat dihitung indeks daya beda butir dengan formula Ferguson seperti berikut ini.

$$d = \frac{(n+1)(N^2 - \sum f_i^2)}{nN^2}$$

$$d = \frac{(10+1)(15^2 - 52)}{10 \times 15^2} = 0,85.$$

Efektivitas Pengecoh (*Distractor*)

Analisis efektifitas pengecoh (*distractor*) atau analisis pola jawaban dilakukan dengan menghitung peserta tes yang memilih tiap alternatif jawaban pada masing-masing Butir. Kriteria pengecoh yang baik adalah apabila pengecoh tersebut dipilih oleh paling sedikit 5% dari peserta tes. Sebagai ilustrasi, berikut ini disajikan hasil uji coba instrumen hasil belajar yang terdiri dari 7 butir tes objektif yang terdiri dari 4 pilihan diujicobakan terhadap 100 orang peserta tes. Mengingat banyak peserta tes 100 orang, pengecoh dikatakan baik jika dipilih oleh minimal 5% dari 100 orang, atau minimal dipilih oleh 5 orang peserta tes.

UNIVERSITAS PENDIDIKAN GANESHA
PASCASARJANA
PROGRAM STUDI PENELITIAN DAN EVALUASI PENDIDIKAN

Data tentang banyaknya peserta tes yang memilih tiap-tiap pilihan pada masing-masing butir, beserta keterangan baik-tidaknya pengecoh adalah seperti tercantum pada tabel berikut.

Nomor Butir	Frekuensi Pilihan Jawaban				Kunci Jawaban	Keterangan
	a	b	c	d		
1	70	10	11	9	a	Semua pengecoh baik
2	14	56	20	10	b	Semua pengecoh baik
3	4	2	11	83	d	Pengecoh a dan b kurang baik
4	6	14	69	6	c	Semua pengecoh baik
5	3	90	3	4	b	Semua pengecoh kurang baik baik
6	14	11	58	17	c	Semua pengecoh baik
7	6	1	0	93	d	Pengecoh b dan c kurang baik