

WORKSHOP DIREKTOR DIKLAT

**“ANALISIS BUTIR SOAL TERHADAP INSTRUMEN
EVALUASI KEGIATAN DIKLAT”**



Oleh:
Mutaqi, MPd., MT

Kerjasama

**UDIKLAT PT PLN (PERSERO) SEMARANG
DAN
JURUSAN TEKNIK ELEKTRO
FT – UNY YOGYAKARTA
TAHUN 2007**

ANALISIS BUTIR SOAL TERHADAP INSTRUMEN EVALUASI KEGIATAN DIKLAT *)

Oleh: Mutaqin **)

Salah satu ciri pembakuan (standardisasi) dari tes prestasi adalah adanya langkah ujicoba dan analisis butir soal. Jawaban (respon) siswa merupakan data empirik yang dapat dijadikan sebagai alat untuk mengukur apakah suatu butir soal dapat berfungsi seperti yang diharapkan. Terdapat beberapa macam metode pengukuran untuk menganalisis butir soal. Metode umum yang dipakai adalah dengan pendekatan item respon theory (IRT). Salah satu program (soft ware) komputer yang digunakan adalah program : Item and Test Analysis (ITEMAN).

Analisis terhadap butir soal (soal obyektif), yang antara lain meliputi: tingkat kesukaran butir soal, daya pembeda soal, reliabilitas tes, kesalahan pengukuran, varians dan deviasi standar

1. Tingkat kesukaran Butir soal

Pada analisis butir soal secara klasikal, tingkat kesukaran (p) dapat diperoleh dengan beberapa cara, antara lain : dengan skala kesukaran linier, skala bivariat, atau proporsi menjawab benar. Cara yang paling mudah dan paling umum digunakan adalah skala rata-rata atau proporsi menjawab benar (*proportioan correct*) atau juga disebut tingkat kesukaran soal (p). Tingkat kesukaran soal adalah bahwa proporsi peserta tes yang menjawab benar pada soal yang dianalisis dibandingkan dengan jumlah peserta tes seluruhnya yang menjawab dan skor maksimum. Persamaannya adalah sebagai berikut :

$$p = \frac{\sum B}{S_m N}$$

Keterangan : p = proporsi menjawab benar pada butir soal tertentu

$\sum B$ = banyaknya peserta tes menjawab benar pada soal tertentu

N = jumlah peserta tes yang menjawab

S_m = skor maksimum peserta pada item tertentu

Tingkat kesukaran (p) sebenarnya merupakan nilai rata-rata dari kelompok peserta tes. Oleh karena itu tingkat kesukaran sebenarnya adalah rerata dari suatu distribusi skor kelompok dari suatu soal. Tingkat kesukaran ini dihitung atas dasar nilai rata-rata, atau sering disebut kesukaran rata-rata.

*) Disampaikan pada Workshop Direktur Diklat di UDIKLAT PT PLN (PERSERO) Semarang, 9 April 2007

***) Dosen Jurusan Pendidikan Teknik Elektro FT UNY

Tingkat kesukaran ini sebenarnya merupakan ukuran kemudahan soal. Makin tinggi indeks tingkat kesukaran (p) maka makin mudah soal tersebut, dan sebaliknya. Besarnya tingkat kesukaran (p) besarnya antara 0 s.d 1. Pengelompokan tingkat kesukaran dapat dilihat pada tabel berikut :

Proportion correct	Kategori Soal
$p > 0,70$	Mudah
$0.30 \leq p \leq 0,70$	Sedang
$p < 0,30$	sukar

No	Nama siswa	Nomor Item										Skro Total
		1	2	3	4	5	6	7	8	9	10	
1	Andi	1	0	1	0	1	1	0	0	0	0	5
2	Badu	1	1	0	0	0	1	0	1	0	0	4
3	Cici	1	1	1	0	0	1	1	1	0	0	6
4	Dono	1	1	1	0	0	1	1	1	1	0	7
5	Eko	1	0	0	0	1	1	0	0	0	0	3
..
36	Lulu	1	1	1	1	1	0	1	1	0	1	8
	$\sum x$	36	19	23	20	17	18	15	16	18	0	
	N	36	36	36	36	36	36	36	36	36	36	
	Tingkt Kesukaran (p)	1,00	0.528	0.639	0.555	0.472	0.500	0.417	0.444	0.500	0.00	

Suatu soal kadang-kadang dikategorikan kedalam kestrim sukar, yaitu apabila p mendekati nol, dan ekstrim mudah yaitu apabila p mendekati satu. Oleh karena itu apabila kita akan memasukkan soal analisis tingkat kesukaran ke dalam bank soal, ada beberapa hal yang perlu diperhatikan dalam mengevaluasi tingkat kesukaran soal, antara lain :

- 1) Soal yang ekstrim mudah atau soal ekstrim yang sukar tidak memberikan informasi yang berguna bagi sebagian besar peserta tes. Oleh sebab itu hal seperti ini kemungkinan distribusi jawaban alternatif jawaban ada yang tidak memenuhi syarat.
- 2) Sekalipun soal terlalu sulit atau terlalu mudah, apabila setiap pengecoh (distribusi jawaban) pada soal tersebut menunjukkan kondisi :
 - * distribusi jawaban merata atau disebut jawaban logis
 - * daya bedanya negatif (kecuali kunci), maka soal-soal tersebut masih memenuhi syarat untuk diterima.

- 3) Kalipun soal terlalu sulit atau terlalu mudah, namun apabila memiliki daya pembeda dan statistik pengecoh memenuhi kriteria , maka soal tersebut dapat dipilih dan diterima sebagai salah satu alternatif untuk disimpan sebagai bank soal.
- 4) Apabila soal ekstrim mudah atau ekstrim sukar, serta daya pembeda dan statistik pengecohnya belum memenuhi kriteria , maka soal tersebut perlu direvisi dan ujicoba lagi.

2. Daya pembeda soal.

Daya pembeda atau daya beda suatu soal berfungsi untuk menentukan dapat tidaknya suatu soal membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada pada kelompok itu. Tujuan dari pengujian daya pembeda adalah untuk melihat kemampuan butir soal dalam membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah. Ada beberapa cara yang digunakan untuk menentukan daya pembeda, antara lain dengan menggunakan indeks korelasi. Persamaan yang dipakai adalah dengan : Korelasi Point Biserial

Korelasi point biserial maupun biserial adalah korelasi product moment yang diterapkan pada data, variabel-variabel yang dikorelasikan sifatnya masing-masing berbeda satu sama lain. Variabel soal bersifat dikotomi, sedangkan variabel skor total atau sub skor total bersifat dikotonum. Variabel soal dinamakan dikotomi karena skor-skor yang terdapat pada soal hanya ada dua alternatif, misalnya satu dan nol. Sepertihalnya pada bentuk soal pilihan ganda, soal yang benar diberi nilai satu dan yang salah diberi nilai nol. Variabel skor total atau sub skor total peserta tes bersifat dikotonum atau non dikotomi , yang diperoleh dari jumlah jawaban yang benar. Korelasi point biserial digunakan persamaan :

$$D = \frac{JKa - JKb}{nKa} \qquad D = \frac{JKa - JKb}{nKb}$$

Keterangan :

D = Daya Pembeda

Jka = Jumlah peserta tes yang menjawab Benar Kelompok Atas

Jkb = Jumlah peserta tes yang menjawab Benar Kelompok Bawah

nKa = Jumlah peserta tes Kelompok Atas

nKb = Jumlah peserta tes Kelompok Bawah

b) Korelasi Biserial

Korelasi biserial dapat ditentukan dengan menggunakan persamaan :

$$r_{bis} = \frac{M_p - M_\tau}{S_\tau} \cdot \frac{p}{y}$$

Keterangan :

r_{bis} = koefisien korelasi biserial

M_p = mean skor tes dari peserta tes yang memiliki jawaban benar pada butir soal

M_τ = mean skor total

S_τ = deviasi standar

p = proporsi peserta tes yang jawabannya benar pada butir soal

y = ordinat p dalam distribusi normal

Nilai korelasi point biserial selalu lebih rendah dibandingkan dengan nilai korelasi biserial. Koefisien point biserial merupakan kombinasi hubungan antara soal dengan kriteria dan taraf kesukaran.

3. *Reliabilitas Tes*

Reliabilitas tes (*test reliability*) adalah suatu hal yang sangat penting pada alat pengukuran yang standar. Reliabilitas dihubungkan dengan pengertian adanya ketepatan suatu tes dalam pengukurannya. Reliabilitas adalah kestabilan skor yang diperoleh orang yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda atau dari suatu pengukuran ke pengukuran lainnya. Reliabilitas mengarah pada konsistensi atau kemantapan hasil dari dua pengukuran terhadap hal yang sama. Hasil pengukuran itu diharapkan akan sama apabila pengukuran itu diulangi. Dengan perangkat tes yang reliabel, apabila tes itu kita berikan dua kali pada orang yang sama tetapi dalam waktu selang yang berbeda, sepanjang tidak ada perubahan kemampuan, maka skor yang diperoleh akan konstan.

Reliabilitas dan validitas merupakan hal yang penting dalam suatu tes. Reliabilitas mendukung validitas. Suatu tes mungkin saja reliabel, tetapi belum tentu valid. Akan tetapi sebaliknya, tes yang valid sudah biasanya reliabel. Reliabilitas memiliki dua konsistensi. Pertama adalah konsistensi internal, yakni tingkat sejauhmana soal itu homogen baik dari tingkat segi kesukaran maupun bentuk soalnya. Kedua, konsistensi eksternal, yaitu tingkat sejauhmana skor yang dihasilkan tetap sama sepanjang kemampuan seseorang yang diukur

belum berubah. Apabila hasil skor tes pertama sama dengan hasil skor kedua, maka tes tersebut dikatakan memiliki reliabilitas yang tinggi, atau terdapat korelasi yang tinggi.

Konsistensi skor tes yang diperoleh disebut reliabilitas. Dengan kata lain reliabilitas tes adalah tingkat sejauhmana skor deviasi seseorang tetap konsisten dengan pengulangan tes yang sama .

Salah satu cara untuk menghitung indeks reliabilitas tes, digunakan dengan persamaan koefien alpha atau sering disebut Cronbach's Alpha. Koefisien alpha dihitung dengan rumus ;

$$\alpha = \frac{R}{R-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2}\right)$$

dimana R adalah jumlah butir soal dalam tes, σ_i^2 adalah varian butir soal dan σ_x^2 adalah varian tes soal (skor total).

Untuk soal yang bersifat dikotomi (seperti pilihan ganda) varian butir soal diperoleh dengan rumus : $\sigma_i^2 = p_i \cdot q_i$

Dimana p_i adalah tingkat kesukaran soal dan q_i adalah $(1-p_i)$.

4. Kesalahan Pengukuran

Kesalahan pengukuran (*standard error of measurement*) membantu pemakai tes dalam memahami kesalahan yang bersifat random (acak) yang mempengaruhi skor seseorang peserta tes dalam pelaksanaan tes. Kesalahan pengukuran dihitung dengan rumus :

$$\sigma_E = \sigma_x \sqrt{1 - p_x}$$

dimana σ_x adalah deviasi standar dari skor total dan p_x adalah koefisien reliabilitas .

5. Varian dan Deviasi Standar

Varian (variance) dan deviasi standar (standard deviation) merupakan ukuran variabilitas distribusi skor yang cukup stabil. Statistik varian dan deviasi standar didasarkan pada konsep yang sederhana tentang skor deviasi. Untuk setiap tes dengan skor mentah X, skor deviasi x didefinisikan sebagai berikut ;

$$X = X - \mu$$

Dimana μ adalah mean skor dari kelompok peserta tes. Skor deviasi merupakan perbedaan (selisih) antara skor seseorang dengan mean skor kelompok. Apabila skor

mentah lebih rendah dari dari mean, maka skor deviasi mempunyai tanda negatif. Sebaliknya apabila skor mentah lebih besar dari mean, maka skor deviasi mempunyai tanda positif.

Statistik deviasi untuk sekumpulan skor diperoleh dengan menghitung rata-rata skor deviasi kuadrat. Hal ini dinyatakan dalam rumus :

$$\sigma_x^2 = \frac{\sum (X - \mu)^2}{N}$$

statistik deviasi standar merupakan akar dari varian :

$$\sigma_x = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

Perlu diingat bahwa statistik varian dan deviasi didasarkan pada skor deviasi. Semakin besar nilai absolut skor deviasi, semakin besar pula varian dan deviasi standarnya.

5	0.085
Other	0.000

6	0-6	4.207	0.716	0.305	58	1	0.000	+
						2	0.034	
						3	0.169	
						4	0.339	
						5	0.441	
						Other	0.017	

MicroCAT (tm) Testing System
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

Item and Test Analysis Program -- ITEMAN (tm) Version 3.00

Item analysis for data from file LATIHCB.DAT

Page 2

Seq. No.	Scale -Item	Item Statistics				Alternative Statistics		
		Item Mean	Item Var.	Item-Scale Correlation	N per Item	Alter-native	Proportion Endorsing	Key
7	0-7	3.246	0.641	0.270	57	1	0.017	+
						2	0.119	
						3	0.492	
						4	0.288	
						5	0.051	
						Other	0.034	
8	0-8	3.780	0.511	0.561	59	1	0.000	+
						2	0.034	
						3	0.288	
						4	0.542	
						5	0.136	
						Other	0.000	
9	0-9	3.276	0.683	0.025	58	1	0.068	-
						2	0.288	
						3	0.492	
						4	0.119	
						5	0.017	
						Other	0.017	
10	0-10	2.898	0.498	0.356	59	1	0.017	-
						2	0.119	
						3	0.644	
						4	0.186	
						5	0.034	
						Other	0.000	

Missing-data option: Compute statistics on all available item responses

There were 50 examinees in the data file.

Scale Statistics

Scale:	0

N of Items	10
N of Examinees	50
Mean	3.244
Variance	0.059
Std. Dev.	0.244
Skew	-0.295
Kurtosis	0.343
Minimum	2.500
Maximum	3.800
Median	3.300
Alpha	-0.005
SEM	0.244
Mean P	N/A
Mean Item-Tot.	0.318
Mean Biserial	N/A