

ITEM AND TEST ANALYSIS (ITEMAN)*)

Oleh :
Badrun Kartowagiran)**

**PASCASARJANA
UNIVERSITAS NEGERI YOGYAKARTA
2011**



**) Makalah disampaikan pada pelatihan Asesmen Pembelajaran bagi Dosen Muda
UPI, tanggal 3 – 7 Oktober 2011 di Pascasarjana UNY*

***) Dosen Fakultas Teknik dan Pascasarjana UNY*

ITEM AND TEST ANALYSIS (ITEMAN)

Oleh: Badrun Kartowagiran

PENDAHULUAN

Untuk mendapatkan instrumen berkualitas tinggi, selain dilakukan analisis secara teori (telaah butir) maka perlu juga dilakukan analisis butir secara empirik. Secara garis besar, analisis butir secara empirik ini dapat dibedakan menjadi dua, yaitu: dengan pendekatan teori tes klasik dan teori respon butir (*Item Response Theory =IRT*). Pada kesempatan ini hanya akan dibicarakan analisis butir empirik model klasik yang menggunakan program Item and Test Analysis (ITEMAN).

Ada dua kegiatan yang dapat dilakukan dengan program ITEMAN, yaitu: (1) analisis instrumen yang jawabannya dikotomi, misal soal benar-salah, atau soal pilihan ganda, dan (2) instrumen dengan jawaban ganda (*multipoint*), misal skala yang menggunakan lima jawaban seperti tipe Likert. Untuk instrumen yang jawabannya dikotomi, kunci jawaban dapat dinyatakan dengan huruf mulai dari A sampai dengan I atau dengan angka mulai dari 0 sampai dengan 9. Sedangkan untuk instrumen dengan jawaban ganda, kunci jawaban dapat dinyatakan dengan tanda “ + “ dan “ - „. Bila satu butir diberi kunci “ + “ , skor butir itu sama dengan angka jawaban, sedangkan butir yang diberi kunci “ - „ , maka skor butir itu sama dengan komplemen dari jumlah alternatif (misal jawabannya 1 maka skor butir itu = 5 bila alternatif jawabannya 5).

INSTRUMEN DENGAN JAWABAN DIKOTOMI

Dalam kesempatan ini dipilih soal objektif bentuk pilihan ganda dengan 4 alternatif jawaban. Data dituliskan dalam bentuk alfabetik (A, B, C, D). Langkah pertama yang dilakukan adalah memasukkan data ke file dalam bentuk ASCII atau DOS Text yang dapat diketik dalam program Notepad.

Contoh:

020 o N 10
DCABBCADAACBACDACBDC
444444444444444444444444
YYYYYYYYYYYYYYYYYYYY
0042001 BADCCDABBDABBCDABDCD
0042002 BCADDACBCDABCADBADCC
dst
0042040 ABDCABDACCDAACBC

Keterangan :

Baris pertama

Kolom 1 – 3 : jumlah butir soal, maksimum 250 butir (dalam hal ini 020)
4 : kosong/spasi
5 : untuk jawaban kosong (omit) (dalam contoh o)
6 : kosong/spasi
7 : untuk butir soal yang belum dikerjakan (dalam contoh N)
8 : kosong/spasi
9 – 10 : jumlah identitas data siswa (maks 80) (dalam contoh 10)

Baris kedua : Kunci jawaban

Baris ketiga : Jumlah pilihan jawaban

Baris ke empat : kode Y = Yes, butir soal dianalisis, N = butir soal tidak dianalisis

Baris ke lima dst : jawaban responden. Bila data sudah masuk semua kemudian disimpan dalam sub direktori tersendiri, yaitu satu sub direktori dengan program ITEMAN. Dalam kesempatan ini data ditulis dengan “notepad” kemudian disimpan dengan menggunakan tipe *All files* dalam folder, yaitu jadi satu folder dengan program ITEMAN. Misalkan data yang baru selesai dimasukkan itu diberi nama CONTOH-1.DAT. Data yang

Menjalankan program

Apabila program ITEMAN dan data sudah dalam 1 folder maka cara menjalankannya adalah sebagai berikut.

1. Klik dua kali program ITEMAN, maka di layar akan tampak

ITEM & TEST ANALYSIS PROGRAM

>>>***** <<<<

Enter the name of the input file:

- 2. Ketik nama file yang akan dianalisis, misalnya CONTOH-1.DAT Enter

Enter the name of the outout file:

Komputer meminta diisikan nama output file (file hasil analisis)

5. Ketik nama file output (hasil) yang dikehendaki, misal CONTOH-1.OUT
lalu tekan ENTER, maka muncul:

Do you want the score written to a file?:

Y = bila dikehendaki hasil analisis direkam

N = bila hasil analisis tidak direkam.

Bila diketik Y maka akan muncul:

Enter the name of the score file:

Komputer minta diisikan nama file untuk skor peserta tes.

6. Ketikkan nama file untuk hasil skor, misal CONTOH-1.SCR lalu tekan
ENTER

Dalam waktu beberapa detik, akan muncul tampilan:

ITEM ANALYSIS IS COMPLETE

Ini menunjukkan bahwa proses analisis telah selesai. Hasil dapat diprint,
namun sebelumnya ditata dulu agar hasilnya tidak terpotong-potong. Contoh
hasil analisis dapat dilihat pada halaman berikut.

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 CONTOH-1.DAT

Page 1

Seq. No.	Item Statistics			Alternative Statistics					Point Biser. Key
	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	
1	0-1	0.175	0.625	0.424	A	0.150	-1.000	-0.971	
					B	0.500	0.344	0.275	
					C	0.175	0.186	0.126	
					D	0.175	0.625	0.424 *	
					Other	0.000	-9.000	-9.000	
5	0-5	0.175	0.186	0.126	A	0.325	0.248	0.191 ?	
					B	0.500	-0.344	-0.275	
					C	0.175	0.186	0.126 *	
					D	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

CHECK THE KEY
 C was specified, A works better

dst

There were 46 examinees in the data file.

Scale Statistics

Scale:	0
N of Items	20
N of Examinees	40
Mean	5.575
Variance	2.394
Std. Dev.	1.547
Skew	-1.699
Kurtosis	1.403
Minimum	2.000
Maximum	7.000
Median	6.000
Alpha	0.490
SEM	1.889
Mean P	0.279
Mean Item-Tot.	0.198
Mean Biserial	0.272

Keterangan Statistik Butir Soal:

1. Seq. No. adalah nomor urut butir soal
2. Scala-Item adalah nomor urut butir soal dalam tes/instrumen
3. Prop-Correct adalah proporsi peserta tes yang menjawab benar butir soal
4. Biser adalah indeks daya beda butir soal dengan menggunakan koefisien korelasi biserial. Nilai positif menunjukkan bahwa peserta tes yang menjawab benar butir soal, mempunyai skor yang relatif tinggi dalam tes tersebut. Untuk statistik pilihan jawaban (alternatif) korelasi biserial negatif sangat tidak dikehendaki untuk kunci jawaban.
5. Point biserial indeks daya beda butir soal dengan menggunakan koefisien korelasi point-biserial. Keterangan selanjutnya sama dengan yang ada pada Biser.

Keterangan Statistik Tes:

1. N of Item adalah jumlah butir soal
2. N of Examinees adalah jumlah peserta tes
3. Mean adalah skor rerata peserta tes
4. Variance adalah varian dari distribusi skor peserta tes yang memberikan gambaran tentang sebaran skor peserta tes.
5. Std.dev. adalah standar deviasi dari distribusi skor peserta tes.
6. Skew adalah kemiringan distribusi skor peserta tes. Juling negatif menunjukkan bahwa sebagian besar skor berada di bagian atas (skor tinggi) dari distribusi skor, dan sebaliknya
7. Kurtosis adalah puncak distribusi skor yang menggambarkan kelandaian distribusi skor peserta tes dibanding dengan distribusi normal. Nilai positif menunjukkan distribusi lebih lancip, dan nilai negatif menunjukkan distribusi yang lebih landai (merata). Kurtosis untuk distribusi normal adalah nol.
8. Alpha adalah koefisien reliabilitas alpha untuk tes tersebut.
9. SEM (*standard error of measurement*) adalah kesalahan baku pengukuran untuk setiap tes.
10. Mean P adalah rata-rata tingkat kesukaran semua butir soal dalam tes secara klasikal dihitung dengan cara mencari rata-rata proporsi peserta tes yang menjawab benar untuk semua butir dalam soal tes tersebut.
11. Mean Item-Tot adalah nilai rata-rata indeks daya beda dari semua butir dalam tes yang diperoleh dengan menghitung nilai rata-rata point biserial dari semua butir dalam tes/skala.
12. Mean Biserial adalah nilai rata-rata indeks daya beda dari semua butir dalam tes yang diperoleh dengan menghitung nilai rata-rata biserial dari semua butir dalam tes/skala.

INSTRUMEN DENGAN JAWABAN GANDA (MULTIPOINT)

Dalam kesempatan ini dipilih instrumen dengan 4 jawaban. Jawaban dituliskan dalam bentuk angka (1, 2, 3, dan 4). Langkah pertama yang dilakukan adalah memasukkan data ke file dalam bentuk ASCII atau DOS Text.

Contoh :

```
020 o N 10
++--++-----++--++++--+
4444444444444444444444444444
YYYYYYYYYYYYYYYYYYYYYYYY
0042001    12344312314322341123
0042002    43434323134134321341
dst
0042040    43243213213431231432
```

Keterangan:

Baris pertama

Seperti pada soal pilihan ganda di atas

Baris kedua : Kunci jawaban dengan tanda “+” dan “-”.

Baris ketiga : Skor maksimum yang disediakan untuk setiap butir

Baris ke empat : kode Y = Yes, butir soal dianalisis, N = butir soal tidak dianalisis

Baris ke lima dst : jawaban responden. Untuk butir yang memiliki kunci positif (+), skor butir sama dengan jawaban responden, misal jawaban responden 2 maka skor butir itu juga 2. Untuk butir yang memiliki jawaban negatif (-) maka skor butir itu komplemen dari jawaban responden, misal jawaban responden 1 maka skor butir itu sama dengan 4.

Menjalankan Program

Cara menjalankan program sama dengan cara menjalankan program sewaktu menganalisis untuk soal pilihan ganda di atas. Bedanya adalah sebelum program menganalisis data maka akan muncul :

```
-----
These missing-data options are available for scale with multipoint items:
-----
```

untuk itu dapat diketikkan angka 1, 2, atau 3, dengan penjelasan sbb:

1. ITEMAN menghitung statistik setiap butir dengan memperhatikan jumlah peserta tes yang menjawab butir tersebut.
2. ITEMAN menghapus peserta tes yang tidak ada datanya dari analisis untuk setiap instrumen secara terpisah.
3. ITEMAN menghapus peserta tes yang tidak ada datanya dari semua analisis. (Pilihan ini digunakan apabila instrumen itu terdiri dari sub-sub instrumen).

Contoh dari hasil analisis adalah sebagai berikut.

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 contoh-2.dat

Page 1

Seq. No.	Scale -Item	Item Statistics			N per Item	Alternative Statistics		
		Item Mean	Item Var.	Item-Scale Correlation		Alter-native	Proportion Endorsing	Key
1	0-1	2.900	1.490	0.417	40	1 2 3 4 Other	0.200 0.200 0.100 0.500 0.000	+
6	0-6	2.600	0.840	-0.088	40	1 2 3 4 Other	0.100 0.400 0.300 0.200 0.000	+
CHECK THE KEY + was specified, - works better								

There were 40 examinees in the data file.

Scale Statistics

Scale:	0

N of Items	20
N of Examinees	40
Mean	2.470
Variance	0.044
Std. Dev.	0.210
Skew	0.164
Kurtosis	0.529
Minimum	2.050
Maximum	2.900
Median	2.450
Alpha	0.748
SEM	0.225
Mean P	N/A
Mean Item-Tot.	0.214
Mean Biserial	N/A

Keterangan untuk Statistik Butir :

1. Penjelasan Seq.No. dan Scale-Item sama dengan penjelasan pada soal pilihan ganda di atas.
2. Item-Mean. Ini merupakan rata-rata jawaban terhadap butir itu dan besarnya dipengaruhi pilihan *missing data* yang telah ditentukan, apakah 1, 2, atau 3. Untuk pilihan (1) atau *itemwise* rata-rata didasarkan pada jumlah peserta tes tiap-tiap butir. Pilihan (2) atau *scalewise* rata-rata didasarkan pada jumlah peserta yang lengkap datanya dan diambil sama untuk seluruh instrumen. Pilihan (3) *casewise* rata-rata didasarkan pada jumlah peserta tes yang memiliki data lengkap untuk semua instrumen. Apabila suatu instrumen tidak terdiri dari sub-sub instrumen maka pilihan (2) sama dengan pilihan (3).
3. Item Var adalah Varians skor butir
4. Item-Scale Correlation adalah korelasi Pearson antara skor butir dengan rata-rata skor
5. N per Item adalah jumlah kasus (jumlah peserta) yang besarnya tergantung pada *missing data* yang dipilih.

Keterangan untuk Statistik Instrumen:

1. N of Items. Jumlah butir yang dimasukkan dalam analisis
2. N of Examinees. Jumlah peserta tes yang dimasukkan pada analisis dan ini sangat tergantung pada tipe *missing data* yang dipilih.
3. Mean. Rerata skor, yaitu skor jawaban betul dibagi dengan jumlah peserta tes yang besarnya tergantung pada tipe *missing data* yang dipilih.
4. Variance, Std. Dev., Skew, Kurtosis, mirip dengan penjelasan pada soal pilihan ganda.
5. Median adalah skor yang membagi 50% kelompok atas dan kelompok bawah.
6. Alpha dan SEM penjelasannya mirip dengan penjelasan pada soal pilihan ganda.
7. Mean P. Rerata tingkat kesukaran, tidak bisa dihitung.
8. Mean Biserial. Rerata Biserial, tidak bisa dihitung.

PENUTUP

Dalam melakukan analisis empirik suatu instrumen menggunakan program ITEMAN, hal penting yang harus diperhatikan adalah jenis instrumen. Apakah instrumen itu memiliki butir dengan jawaban dikotomi (*dichotomously scored items*) atau instrumen yang memiliki butir dengan jawaban ganda (*multipoint items*).

Untuk instrumen yang memiliki butir dengan jawaban dikotomi tidak perlu menentukan tipe *missing data*, sedangkan pada instrumen yang memiliki butir dengan jawaban ganda (*multipoint items*) harus menentukan tipe *missing data*. Oleh karenanya dalam melakukan analisis empirik pada instrumen yang memiliki butir dengan jawaban ganda (*multipoint items*) harus lebih hati-hati.

Untuk mendapatkan instrumen berkualitas tinggi memang tidak mudah. Meskipun demikian harus selalu diupayakan, lebih-lebih bila instrumen itu akan digunakan secara meluas atau secara nasional. Dengan instrumen yang berkualitas tinggi akan diperoleh data yang akurat dan pada gilirannya akan dapat diambil keputusan yang tepat.

DAFTAR PUSTAKA

- Tim. 1999. *Manual ITEMAN*. Jakarta: PUSISJIAN Balitbang Depdiknas.
- N. N. 1986. *User' manual for ITEMAN, RASCAL and ASCAL*. N.C. ASCAR

Lampiran

RAMBU-RAMBU PENERIMAAN BUTIR MENURUT TEORI

KLASIK

1. Ebel (1972) yang mengatakan bahwa alat ukur yang memiliki koefisien reliabilitas 0,8 sudah baik Nunnally (1978) berpendapat bahwa soal uraian yang memiliki koefisien reliabilitas 0,6 – 0,7 dan untuk soal pilihan ganda yang memiliki koefisien reliabilitas 0,75 – 0,90 sudah dapat dikatakan baik. Feldt dan Brehmman (1989) mengatakan bahwa suatu instrumen yang memiliki koefisien reliabilitas $r_{xy} \geq 0,7$ sudah dikatakan reliabel.
2. Menurut Dawson (1972) butir soal yang memiliki tingkat kesukaran 0,25 – 0,75 dikatakan baik.
3. Menurut Ebel (1972) butir yang memiliki daya pembeda 0,41 atau lebih besar dikatakan baik, sedangkan Fernandes mengatakan butir yang memiliki daya pembeda lebih besar dari 0,2 dikatakan baik.
4. Fernandes (1984) yang menjelaskan distraktor dikatakan baik apabila paling tidak dipilih oleh 2 % dari seluruh peserta. Nitko (1996) mengatakan distraktor dikatakan berfungsi manakala paling tidak dipilih **oleh seorang peserta** tes dari kelompok rendah.