

MANUAL ITEM AND TEST ANALIYSIS (ITEMAN)

PEDOMAN PENGGUNAAN ITEMAN

Oleh: Kana Hidayati, M.Pd.

A. GAMBARAN UMUM

ITEM AND TEST ANALYSIS (ITEMAN) merupakan perangkat lunak (*software*) yang dibuat melalui bahasa pemrograman komputer dan dibuat khusus untuk analisis butir soal dan tes. Hasil analisis meliputi: tingkat kesukaran butir soal, daya pembeda soal, statistik sebaran jawaban, kehandalan/reliabilitas tes, kesalahan pengukuran (*standar error*), dan distribusi skor serta skor setiap peserta tes.

Untuk menginstal program ITEMAN cukup dengan mengcopy file ITEMAN.EXE. Untuk lebih memudahkan dalam melakukan analisis sebaiknya file DATA yang akan dianalisis, file ITEMAN.EXE dan file HASIL analisis ditempatkan dalam satu sub direktori.

FORMAT FILE DATA

Langkah pertama dilakukan dengan memasukkan (entry) jawaban siswa (respons) ke dalam suatu file data dalam bentuk text yang dapat dilakukan melalui klik *start-all programs-accessories-notepad*. Data tersebut dapat berupa *alphabetic* (A,B,C,D,.....) atau *numeric* (1,2,3,4,.....). Sebagai contoh kita akan menganalisis butir soal dari suatu tes yang terdiri atas 20 butir soal. Data respons siswa dapat diketik ke dalam file data (misal CONTOH.TXT) sebagai berikut:

```
020 0 n 04
ACCDABEABDABECCBDDEE
55555555555555555555
YYYYYYYYYYYYYYYYYYYY
001 ACCDABEABDABECCBDDEE
002 ACCAABEABDABECCBDDEE
003 ACCBABEABDABECCBDAEE
004 ACCDABEABACBECCBDDAE
005 ACCDAABABDABBCCBDDEA
006 ACCCABECBAABECCBDBBB
007 AACCAABEABCCAEDCCDCCA
008 ACCEABEABECBECCBDDDC
009 ABCBABECADBAEBCCDAEC
010 BCCBABEBCCDBECCBDDAD
011 ACCABBCDBADBECCBADED
012 ACACACEBABCBCCCBDDDB
013 CDCCABBAABEEEACDDBEE
014 ECCABBDACBEDECCDBCCC
015 DCCACBCADCDEECCEBDDE
016 DCBEDBEACEDDECAACEEE
017 EEDEDDEABDECDDAEADAA
018 ECCBEBBEDCBAECCACBEB
019 BCEACBDDEBBDECB AEDEB
020 BAABEBAEEEECEEECEDEE
```

Agar file data dapat dianalisis dengan ITEMAN diperlukan 4 baris perintah sebagai baris kontrol yang diketik mulai dari baris pertama kolom pertama. Empat baris pertama sebagai baris kontrol adalah sebagai berikut:

***) Baris pertama**

Kolom	Keterangan	Dalam contoh
1-3	jumlah butir soal	020
4	kosong/spasi	-
5	untuk jawaban omit (kosong)	0
6	kosong/spasi	-
7	untuk soal yang belum dikerjakan	n
8	kosong/spasi	-
9-10	jumlah identitas data siswa	04

Keterangan:

Kolom 1-3 : Diisi dengan angka yang menunjukkan jumlah butir soal yang dianalisis. Kolom 1 untuk ratusan, kolom 2 untuk puluhan, dan kolom 3 untuk satuan. Artinya untuk menuliskan jumlah butir soal sebanyak 20 angka 2 dimulai pada kolom kedua dan angka 0 pada kolom ketiga. Atau dapat ditulis 020.

Kolom 5 : Diisikan angka '9' atau '0' untuk butir soal yang tidak dijawab.

Kolom 7 : Diisikan 'N' atau 'n' untuk soal yang belum sempat dikerjakan.

Kolom 9-10: Diisi dengan panjang karakter untuk identitas siswa. Sebagai contoh untuk identitas sebanyak 4 karakter dapat ditulis '04' atau '4'.

***) Baris kedua:** Berisi kunci jawaban (option yang benar). Penulisan kunci jawaban harus sesuai dengan kode untuk data baik *alphabetic* maupun *numeric*.

***) Baris ketiga:** Berisi jumlah pilihan jawaban (option). Diisi dengan angka, misalnya 2,3,4,... dan seterusnya sesuai dengan jumlah option.

***) Baris keempat:** Berisi kode/skala tes: 'Y' = butir soal yang dianalisis dan 'N' =butir soal yang tidak dianalisis. Dapat pula diisikan bilangan 0,1,2,... dan seterusnya untuk tes berskala.

***) Baris kelima dan seterusnya:** Berisi jawaban siswa yang diketik dengan *alphabetic* maupun *numeric*, dengan ketentuan setiap satu baris menunjukkan jawaban dari satu responden.

B. MENJALANKAN PROGRAM

Aktifkan Personal Computer yang kita gunakan, selanjutnya lakukan langkah-langkah sebagai berikut:

- Klik *Start* lalu *All Programs* pilih *Accessories* dan klik *Notepad* kemudian buka file baru atau klik *file* lalu *open*.
- Masukkan data yang akan dianalisis dengan ketentuan yang digunakan sebagaimana dikemukakan di atas kemudian disimpan, misal CONTOH.TXT
- Yakinkan bahwa file ITEMAN.EXE sudah diinstal (dicopy).
- Double Klik file ITEMAN.EXE maka akan tampil di layar:

```

ITEM & TEST ANALISIS PROGRAM
>>>*****<<<
Enter the name of the input file:_

```

Komputer meminta diisikan nama input file data.

- Ketikkan nama file data yang akan dianalisis, misalnya CONTOH-1.TXT lalu tekan tombol ENTER. Selanjutnya akan tampil di layar:

Enter the name of the output file: _

Komputer meminta diisikan nama output file (file hasil analisis).

- f. Ketikkan nama file output (hasil) yang kita kehendaki, misalnya HASIL.TXT lalu tekan tombol ENTER, maka akan muncul tampilan:

Do you want the scores written to a file? (Y/N)

Komputer meminta jawaban Y atau N

- g. Ketikkan Y (=yes) bila kita inginkan file hasil skor, atau N(=No) bila kita tidak menghendakinya. Bila kita ketik Y maka akan muncul:

Enter the name of the score file: _

Komputer meminta disikan nama file untuk skor peserta tes.

- h. Ketikkan nama file untuk hasil skor, misalnya SKOR.TXT lalu tekan tombol ENTER dalam waktu beberapa detik akan muncul tampilan:

ITEM ANALYSIS IS COMPLETE.

Ini menunjukkan bahwa proses analisis telah selesai. File hasil analisis (HASIL.TXT) dan hasil skor (SKOR.TXT) berada dalam sub direktori yang sama.

C. HASIL ANALISIS

Hasil dari analisis ITEMAN ini berupa dua file yaitu file statistik dan file skor. Keduanya berupa file ASCII yang dapat dilihat dengan menggunakan program pengolah kata (*word procesor*).

1. File Statistik

Statistik hasil analisis ITEMAN dapat dibedakan ke dalam 2 bagian yaitu: (a) Statistik butir soal dan (b) statistik tes/skala. Berikut menunjukkan hasil analisis statistik butir soal dan hasil analisis statistik tes/skala:

```
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

      Item Statistics                Alternative Statistics
-----
Seq.  Scale  Prop.    Point    Prop.    Point
No.  -Item  Correct  Biser.  Biser.  Alt.  Endorsing  Biser.  Biser.  Key
-----
   1  0-1    0.550    0.942    0.749    A    0.550    0.942    0.749  *
      0.150    -0.518    -0.338
      0.050    -0.251    -0.119
      0.100    -0.356    -0.208
      0.150    -0.702    -0.459
      Other  0.000    -9.000    -9.000

.
DAN SETERUSNYA.....

There were 20 examinees in the data file.
```

Scale Statistics

Scale:	0
N of Items	20
N of Examinees	20
Mean	11.400
Variance	21.540
Std. Dev.	4.641
Skew	0.381
Kurtosis	-1.109
Minimum	5.000
Maximum	20.000
Median	10.000
Alpha	0.847
SEM	1.817
Mean P	0.570
Mean Item-Tot.	0.500
Mean Biserial	0.653

Keterangan:

a. Statistik butir soal.

Untuk tes yang terdiri dari butir-butir soal yang bersifat dikotomi misalnya pilihan ganda, statistik berikut adalah output dari setiap butir soal yang dianalisis.

1. **Seq. No.** adalah nomor urut butir soal dalam file data.
2. **Scala item** adalah nomor urut butir soal dalam tes.
3. **Prop. Correct** adalah Proporsi siswa yang menjawab benar butir tes. Nilai ekstrem mendekati nol atau satu menunjukkan bahwa butir soal tersebut terlalu sukar atau terlalu mudah untuk peserta tes. Indeks ini disebut juga indeks tingkat kesukaran soal secara klasikal.
4. **Biser** adalah indeks daya pembeda soal dengan menggunakan koefisien korelasi biserial. Nilai positif menunjukkan bahwa peserta tes yang menjawab benar butir soal mempunyai skor relatif tinggi dalam tes tersebut. Sebaliknya nilai negatif menunjukkan bahwa peserta tes yang menjawab benar butir soal memperoleh skor yang relatif rendah dalam tes. Untuk statistik pilihan jawaban (alternatif), korelasi biserial negatif sangat tidak dikehendaki untuk kunci jawaban dan sangat dikehendaki untuk pilihan jawaban yang lain.
5. **Point biserial** adalah juga indeks daya pembeda soal dan pilihan jawaban (alternatif) dengan menggunakan koefisien korelasi point biserial, penafsirannya sama dengan statistik biserial.

Catatan:

- Nilai -9.000 menunjukkan bahwa statistik butir soal atas pilihan jawaban tidak dapat dihitung. Hal ini seringkali terjadi apabila tidak ada peserta tes yang menjawab butir soal/pilihan jawaban tersebut. Statistik pilihan jawaban/alternatif memberikan informasi yang sama dengan statistik butir soal. Perbedaannya adalah bahwa statistik pilihan jawaban dihitung secara terpisah. Untuk setiap pilihan jawaban dan didasarkan pada dipilih tidaknya alternatif tersebut, bukan pada benarnya jawaban. Tanda bintang (*) yang muncul disebelah kanan hasil analisis menunjukkan kunci jawaban.

b. Statistik tes.

Program ini memberikan hasil analisis statistik untuk tes sebagai berikut:

1. **N of Items:** Jumlah butir soal dalam tes yang dianalisis. Untuk tes yang terdiri atas butir-butir soal dikotomi hal ini merupakan jumlah total butir soal dalam tes.
2. **N of Examines:** Jumlah peserta tes yang digunakan dalam analisis.
3. **Mean:** skor/rata-rata peserta tes.
4. **Variance:** Varian dari distribusi skor peserta tes yang memberikan gambaran tentang sebaran skor peserta tes.

5. **Std. Deviasi:** Deviasi standar dari distribusi skor peserta tes. Deviasi standar merupakan akar dari variance.
6. **Skew. :** Kemiringan distribusi skor peserta yang memberikan gambaran tentang bentuk distribusi skor peserta tes. Kemiringan negatif menunjukkan bahwa sebagian besar skor berada pada bagian atas/skor tinggi dari distribusi skor. Sebaliknya kemiringan positif menunjukkan bahwa sebagian besar skor berada pada bagian bawah/skor rendah dari distribusi skor. Kemiringan nol menunjukkan bahwa skor berdistribusi secara simetris di sekitar rata-rata/mean.
7. **Kurtosis:** Puncak distribusi skor yang menggambarkan kelandaian distribusi skor dibanding dengan distribusi normal. Nilai positif menunjukkan distribusi yang lebih lancip/memuncak dan nilai negatif menunjukkan distribusi yang lebih landai/merata. Kurtosis untuk distribusi normal adalah nol.
8. **Minimum:** Skor terendah peserta dalam tes.
9. **Maximum:** Skor tertinggi peserta dalam tes.
10. **Median:** Skor tengah dimana 50% berada pada/lebih rendah dari skor tersebut.
11. **Alpha:** Koefisien reliabilitas alpha untuk tes yang merupakan indeks homogenitas tes. Koefisien alpha bergerak dari 0.0 sampai 1.0. Koefisien alpha hanya cocok digunakan pada tes yang bukan mengukur kecepatan dan hanya mengukur satu dimensi (single trait). Reliabilitas adalah kestabilan skor yang diperoleh orang yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda. Jadi reliabilitas dapat dikatakan sebagai tingkat konsistensi hasil dua pengukuran terhadap hal yang sama. Dengan perangkat soal yang reliable apabila tes diberikan dua kali pada orang yang sama tetapi dalam selang waktu yang berbeda sepanjang tidak ada perubahan kemampuan maka skor yang diperoleh akan konstan.
12. **SEM:** kesalahan pengukuran standar untuk setiap tes. SEM merupakan estimasi dari deviasi standar kesalahan pengukuran dalam skor tes. Kesalahan pengukuran membantu pemakai tes dalam memahami kesalahan yang bersifat random yang mempengaruhi skor peserta dalam pelaksanaan tes.
13. **Mean p:** 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 soal dalam tes.
14. **Mean item tot:** Nilai rata-rata indeks daya pembeda dari semua soal dalam tes yang diperoleh dengan menghitung nilai rata-rata point biserial dari semua soal dalam tes.
15. **Mean biserial:** Nilai rata-rata indeks daya pembeda yang diperoleh dengan menghitung nilai rata-rata korelasi biserial dari semua butir soal dalam tes.
16. **Scale intercorelation:** Indeks korelasi antara skor-skor peserta tes yang diperoleh dari setiap subtes.

2. File Skor.

Program ITEMAN juga memberikan hasil skor untuk setiap peserta tes yang menunjukkan jumlah benar dari seluruh jawaban. Dari contoh di atas, dapat dilihat skor peserta tes pada file SKOR.TXT sebagai berikut:

```
4 1 Scores for examinees from file CONTOH.TXT
001 20.00
002 19.00
003 18.00
004 17.00
005 16.00
006 14.00
007 10.00
008 15.00
009 10.00
010 12.00
011 12.00
```

012 9.00
 013 9.00
 014 7.00
 015 9.00
 016 8.00
 017 5.00
 018 7.00
 019 6.00
 020 5.00

Baris pertama dari output menunjukkan jumlah karakter untuk identitas peserta tes (dalam contoh diatas: 4). Jumlah skala (dalam contoh diatas 1) dan nama file input. Kemudian hasil skala diberikan secara berurutan sesuai dengan urutan peserta tes dalam file data.

Catatan:

Program ITEMAN mampu menganalisis maksimal 250 butir soal dalam satu file dengan kapasitas 3.000 responden.

D. INTERPRETASI HASIL ANALISIS (File Statistik)

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.TXT

Page 1

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
1	0-1	0.550	0.942	0.749	A	0.550	0.942	0.749	*
					B	0.150	-0.518	-0.338	
					C	0.050	-0.251	-0.119	
					D	0.100	-0.356	-0.208	
					E	0.150	-0.702	-0.459	
					Other	0.000	-9.000	-9.000	
2	0-2	0.750	0.610	0.448	A	0.100	-0.479	-0.280	
					B	0.050	-0.146	-0.069	
					C	0.750	0.610	0.448	*
					D	0.050	-0.251	-0.119	
					E	0.050	-0.669	-0.316	
					Other	0.000	-9.000	-9.000	
3	0-3	0.750	0.814	0.597	A	0.100	-0.540	-0.316	
					B	0.050	-0.355	-0.168	
					C	0.750	0.814	0.597	*
					D	0.050	-0.669	-0.316	
					E	0.050	-0.564	-0.267	
					Other	0.000	-9.000	-9.000	
4	0-4	0.150	0.869	0.567	A	0.250	-0.136	-0.100	
					B	0.250	-0.170	-0.124	
					C	0.200	-0.139	-0.097	
					D	0.150	0.869	0.567	*
					E	0.150	-0.286	-0.187	
					Other	0.000	-9.000	-9.000	
5	0-5	0.600	0.898	0.708	A	0.600	0.898	0.708	*
					B	0.100	-0.233	-0.136	
					C	0.100	-0.479	-0.280	
					D	0.100	-0.602	-0.352	
					E	0.100	-0.663	-0.388	
					Other	0.000	-9.000	-9.000	
6	0-6	0.850	0.194	0.127	A	0.050	0.481	0.227	?
					B	0.850	0.194	0.127	*
					C	0.050	-0.251	-0.119	
					D	0.050	-0.669	-0.316	
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

CHECK THE KEY
 B was specified, A works better

7	0-7	0.600	0.563	0.444	A	0.050	-0.669	-0.316	*		
					B	0.150	-0.102	-0.066			
					C	0.100	-0.110	-0.065			
					D	0.100	-0.602	-0.352			
					E	0.600	0.563	0.444			
					Other	0.000	-9.000	-9.000			
8	0-8	0.600	0.452	0.356	A	0.600	0.452	0.356	*		
					B	0.100	-0.110	-0.065			
					C	0.100	0.074	0.043			
					D	0.100	-0.295	-0.172			
					E	0.100	-0.663	-0.388			
					Other	0.000	-9.000	-9.000			
9	0-9	0.500	0.864	0.689	A	0.150	-0.286	-0.187	*		
					B	0.500	0.864	0.689			
					C	0.150	-0.333	-0.217			
					D	0.100	-0.417	-0.244			
					E	0.100	-0.724	-0.424			
					Other	0.000	-9.000	-9.000			
10	0-10	0.300	0.607	0.461	A	0.150	0.407	0.266	*		
					B	0.200	-0.562	-0.393			
					C	0.200	-0.292	-0.205			
					D	0.300	0.607	0.461			
					E	0.150	-0.286	-0.187			
					Other	0.000	-9.000	-9.000			
11	0-11	0.250	1.000	0.746	A	0.250	1.000	0.746	*		
					B	0.200	-0.677	-0.474			
					C	0.200	0.208	0.145			
					D	0.200	-0.177	-0.124			
					E	0.150	-0.610	-0.398			
					Other	0.000	-9.000	-9.000			
12	0-12	0.500	1.000	0.819	A	0.150	-0.333	-0.217	*		
					B	0.500	1.000	0.819			
					C	0.100	-0.786	-0.460			
					D	0.150	-0.610	-0.398			
					E	0.100	-0.295	-0.172			
					Other	0.000	-9.000	-9.000			
13	0-13	0.850	0.194	0.127	A	0.000	-9.000	-9.000	?		
					B	0.050	0.481	0.227			
					CHECK THE KEY		C	0.050		-0.251	-0.119
					E was specified, B works better		D	0.050		-0.669	-0.316
						E	0.850	0.194		0.127	*
						Other	0.000	-9.000		-9.000	
14	0-14	0.750	0.610	0.448	A	0.050	-0.251	-0.119	*		
					B	0.050	-0.146	-0.069			
					C	0.750	0.610	0.448			
					D	0.100	-0.479	-0.280			
					E	0.050	-0.669	-0.316			
					Other	0.000	-9.000	-9.000			
15	0-15	0.750	0.814	0.597	A	0.100	-0.602	-0.352	*		
					B	0.050	-0.564	-0.267			
					C	0.750	0.814	0.597			
					D	0.050	-0.251	-0.119			
					E	0.050	-0.669	-0.316			
					Other	0.000	-9.000	-9.000			
16	0-16	0.500	1.000	0.819	A	0.150	-0.610	-0.398	*		
					B	0.500	1.000	0.819			
					C	0.150	-0.425	-0.278			
					D	0.100	-0.417	-0.244			
					E	0.100	-0.540	-0.316			
					Other	0.000	-9.000	-9.000			
17	0-17	0.600	0.898	0.708	A	0.100	-0.356	-0.208	*		
					B	0.100	-0.417	-0.244			
					C	0.100	-0.479	-0.280			
					D	0.600	0.898	0.708			
					E	0.100	-0.724	-0.424			
					Other	0.000	-9.000	-9.000			

18	0-18	0.600	0.229	0.180	A	0.100	0.319	0.187	?
					B	0.150	-0.194	-0.127	
					C	0.100	-0.356	-0.208	
					D	0.600	0.229	0.180	*
					E	0.050	-0.355	-0.168	
					Other	0.000	-9.000	-9.000	
19	0-19	0.550	0.125	0.100	A	0.150	-0.009	-0.006	
					B	0.100	0.012	0.007	
					C	0.100	-0.356	-0.208	
					D	0.100	0.074	0.043	
					E	0.550	0.125	0.100	*
					Other	0.000	-9.000	-9.000	
20	0-20	0.400	0.385	0.303	A	0.150	-0.148	-0.097	
					B	0.150	-0.333	-0.217	
					C	0.150	-0.102	-0.066	
					D	0.150	-0.055	-0.036	
					E	0.400	0.385	0.303	*
					Other	0.000	-9.000	-9.000	

There were 20 examinees in the data file.

Scale Statistics

```

-----
Scale:          0
-----
N of Items      20
N of Examinees 20
Mean           11.400
Variance       21.540
Std. Dev.      4.641
Skew           0.381
Kurtosis       -1.109
Minimum        5.000
Maximum       20.000
Median         10.000
Alpha          0.847
SEM            1.817
Mean P         0.570
Mean Item-Tot. 0.500
Mean Biserial  0.653

```

Secara umum kriteria pemilihan soal pilihan ganda:

Kriteria	Koefisien	Keputusan
Tingkat Kesukaran	0,30 s.d. 0,70 (sedang)	Diterima
	0,10 s.d. 0,29 atau 0,70 s.d. 0,90 (sukar atau mudah)	Direvisi
	< 0,10 atau > 0,90 (sangat sukar atau sangat mudah)	Ditolak
Daya Pembeda	> 0,3	Diterima
	0,10 s.d 0,29	Direvisi
	< 0,10	Ditolak
Proporsi Jawaban	> 0,05	Berfungsi baik

Catatan:

- ~~~~~
- Ditinjau dari tujuan pelaksanaan tes, perlu diperhatikan bahwa soal yang terlalu mudah atau terlalu sukar mungkin memang kurang memberikan informasi yang berguna bagi peserta tes pada umumnya, di antaranya kemungkinan karena belum berfungsinya pengecoh dengan baik, namun demikian pada soal yang terlalu mudah atau terlalu sukar ini, apabila pengecohnya berfungsi dengan baik yakni *Prop Endorsing* memenuhi ($>0,05$) serta daya

pembedanya negatif maka soal tersebut masih memenuhi untuk diterima, terlebih kalau memang soal tersebut akan digunakan untuk seleksi yang sifatnya sangat menentukan.

- Suatu soal kadang-kadang dikategorikan ke dalam keadaan ekstrem sukar yaitu apabila p mendekati nol atau ekstrem mudah yaitu apabila p mendekati 1.
- Soal yang ekstrem mudah atau soal yang ekstrem sukar tidak memberikan informasi yang berguna bagi sebagian besar peserta tes. Oleh sebab itu soal seperti ini kemungkinan distribusi jawaban pada alternatif jawaban ada yang tidak memenuhi syarat. Sekalipun soal terlalu sukar atau terlalu mudah apabila setiap pengecoh (distribusi jawaban) pada soal tersebut menunjukkan kondisi distribusi jawaban merata atau distribusi jawabannya logis dan daya pembedanya negatif (kecuali kunci) maka soal-soal tersebut masih memenuhi syarat untuk diterima. Sekalipun soal terlalu sukar 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 dalam bank soal.
- Apabila soal ekstrem mudah atau ekstrem sukar serta daya pembeda dan statistik pengecohnya belum memenuhi kriteria maka soal tersebut perlu direvisi dan diujicoba lagi.
- Daya pembeda 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 peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah. Korelasi point biserial maupun korelasi biserial adalah korelasi product moment yang diterapkan pada data, variabel yang dikorelasikan masing-masing bersifat berbeda satu sama lain. Variabel soal bersifat dikotomi sedangkan variabel skor total bersifat kontinu. Variabel soal dikotomi karena skor pada soal hanya ada nol dan satu. Variabel skor total bersifat kontinu yang diperoleh dari jumlah jawaban yang benar.

Butir Soal Nomor 1

Seq. No.	Scale -Item	Prop. Correct	Point Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
1	0-1	0.550	0.942	0.749	A	0.550	0.942	0.749	*
					B	0.150	-0.518	-0.338	
					C	0.050	-0.251	-0.119	
					D	0.100	-0.356	-0.208	
					E	0.150	-0.702	-0.459	
					Other	0.000	-9.000	-9.000	

Tingkat kesukaran soal ini adalah 0,550 (sedang), cukup bagus, yakni sebanyak 55,0% peserta tes dapat menjawab soal dengan benar. Daya pembedanya, $r_{bis}=0,942$ dan $r_{pbis}=0,749$ menunjukkan keduanya bertanda positif. Hal ini berarti bahwa peserta tes yang pintar (skor totalnya tinggi) cenderung menjawab benar soal ini dan peserta tes yang kurang pintar cenderung menjawab salah soal ini. Karena alternatif jawaban A merupakan kunci, maka tanda positif ini menunjukkan bahwa kunci jawaban sudah berfungsi sebagaimana mestinya. Ditinjau dari distribusi jawaban yaitu persentase peserta tes merespons alternatif jawaban, semua pengecoh tampak telah berfungsi dengan baik. Dapat dilihat pada kolom *Prop Endorsing*, tampak bahwa sebanyak 15% peserta tes merespons alternatif jawaban B, 5% merespons alternatif jawaban C, 10% merespons alternatif jawaban D, dan 15% merespons alternatif jawaban D. Ditinjau dari daya pembeda masing-masing pengecoh juga dapat dikatakan berfungsi baik karena r_{bis} atau r_{pbis} untuk alternatif jawaban A, B, C, dan D seluruhnya bernilai negatif artinya peserta tes yang pintar cenderung tidak memilih alternatif jawaban itu dan siswa yang kurang pintar cenderung memilih alternatif jawaban tersebut. Dengan kata

lain, peserta tes yang skornya rendah lebih memilih pengecoh sebagai jawaban yang benar. **Kesimpulannya** butir soal nomor 1 ini sudah cukup baik dan dapat digunakan. Sama halnya dengan butir soal nomor 1 ini, butir soal lain yang sudah dapat dikatakan cukup baik dan dapat digunakan adalah butir soal nomor 2, 3, 4, 5, 7, 9, 12, 14, 15, 16, 17, dan 20.

Butir Soal Nomor 6

6	0-6	0.850	0.194	0.127	A	0.050	0.481	0.227	?
					B	0.850	0.194	0.127	*
					C	0.050	-0.251	-0.119	
					D	0.050	-0.669	-0.316	
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

CHECK THE KEY
B was specified, A works better

Tingkat kesukaran soal ini adalah 0,850 (mudah), sebanyak 85,0% peserta tes dapat menjawab soal tersebut dengan benar. Daya pembedanya, $r_{bis}=0,194$ dan $r_{pbis}=0,127$ menunjukkan keduanya bertanda positif. Hal ini berarti bahwa peserta tes yang pintar cenderung menjawab benar soal ini dan peserta tes yang kurang pintar cenderung menjawab salah soal ini. Karena alternatif jawaban B merupakan kunci, maka tanda positif ini menunjukkan bahwa kunci jawaban sudah berfungsi sebagaimana mestinya. Namun demikian, baik ditinjau dari besarnya nilai tingkat kesukaran maupun daya pembeda, soal ini termasuk dalam kategori **perlu direvisi**. Hal ini didukung pula oleh belum berfungsinya seluruh pengecoh dengan baik. yang tampak dari distribusi jawaban yaitu sebanyak 5% peserta tes merespons alternatif jawaban A, 5% merespons alternatif jawaban C, 5% merespons alternatif jawaban D, dan 0% merespons alternatif jawaban E. Tampak bahwa alternatif jawaban E belum berfungsi sebagai pengecoh yang baik. Selain itu, ditinjau dari daya pembedanya, alternatif jawaban B dapat dikatakan belum berfungsi baik karena r_{bis} atau r_{pbis} untuk alternatif jawaban A ternyata juga bernilai positif dan lebih besar nilainya daripada alternatif jawaban B sebagai kuncinya. Hal ini menunjukkan bahwa ada kecenderungan peserta tes yang pintar memilih jawaban A ini dan siswa yang kurang pintar cenderung tidak memilih alternatif jawaban A, walaupun besarnya hanya 5% dari peserta tes. Dengan kata lain, ada peserta tes yang skornya tinggi lebih memilih alternatif jawaban A sebagai jawaban yang benar. **Kesimpulannya** butir soal nomor 6 ini perlu ditinjau lagi dan sekiranya akan digunakan maka perlu direvisi. Oleh karena itu dari hasil analisis tampak adanya peringatan `CHECK THE KEY B was specified, A works better` yang menunjukkan bahwa kunci jawaban B kurang tepat dan alternatif jawaban A tampak berfungsi lebih baik. Oleh karena itu yang perlu dilakukan guna merevisi soal ini antara lain adalah periksa kembali kunci jawaban, apabila kunci jawaban ternyata salah lakukan kembali analisis soal, dan apabila ternyata kunci jawaban sudah benar maka kemungkinan kesalahan terletak pada kesalahan penguasaan konsep oleh peserta didik. Sama halnya dengan butir soal nomor 6 ini, butir soal lain yang perlu direvisi adalah butir soal nomor 13 dan butir soal nomor 18.

Butir Soal Nomor 8

8	0-8	0.600	0.452	0.356	A	0.600	0.452	0.356	*
					B	0.100	-0.110	-0.065	
					C	0.100	0.074	0.043	
					D	0.100	-0.295	-0.172	
					E	0.100	-0.663	-0.388	
					Other	0.000	-9.000	-9.000	

Pada soal nomor 8, tingkat kesukarannya adalah 0,600 (sedang), sebanyak 60,0% peserta tes dapat menjawab soal tersebut dengan benar. Daya pembedanya, $r_{bis}=0,452$ dan $r_{pbis}=0,356$ menunjukkan keduanya bertanda positif. Hal ini berarti bahwa peserta tes yang pintar cenderung menjawab benar soal ini dan peserta tes yang kurang pintar cenderung menjawab salah. Karena alternatif jawaban A merupakan kunci, maka tanda positif ini menunjukkan bahwa kunci jawaban sudah berfungsi sebagaimana mestinya. Dengan demikian soal ini termasuk kategori soal yang cukup baik dan dapat digunakan *walaupun* ditinjau dari pengecohnya tampak bahwa alternatif jawaban C, daya pembedanya juga bernilai positif dengan sebanyak 10% peserta tes merespons alternatif jawaban C ini. Pada soal ini tidak ada peringatan untuk mengecek kembali, karena besarnya nilai daya pembeda pada kunci

jawaban (A) tetap lebih besar dari alternative jawaban C, sehingga butir soal ini tetap bisa diterima dan baik untuk digunakan.

Sama halnya dengan butir soal no 8 ini, butir soal lain yang sudah baik dan tetap dapat digunakan walaupun memiliki pengecoh yang dapat dikatakan cukup berfungsi adalah butir soal nomor 10.

Butir Soal Nomor 11

11	0-11	0.250	1.000	0.746	A	0.250	1.000	0.746	*
					B	0.200	-0.677	-0.474	
					C	0.200	0.208	0.145	
					D	0.200	-0.177	-0.124	
					E	0.150	-0.610	-0.398	
					Other	0.000	-9.000	-9.000	

Pada soal nomor 11 ini, tingkat kesukarannya adalah 0,250 (sukar), sebanyak 25,0% peserta tes yang dapat menjawab soal tersebut dengan benar. Daya pembedanya, $r_{bis}=1,000$ dan $r_{pbis}=0,746$ menunjukkan keduanya bertanda positif. Hal ini berarti bahwa soal ini cukup dapat membedakan peserta tes yang pintar dan kurang pintar. Ditinjau dari distribusi jawaban yaitu persentase peserta tes merespons alternatif jawaban, semua pengecoh tampak telah berfungsi dengan baik. Dapat dilihat pada kolom *Prop Endorsing*, tampak bahwa sebanyak 20% peserta tes merespons alternatif jawaban B, 20% merespons alternatif jawaban C, 20% merespons alternatif jawaban D, dan 15% merespons alternatif jawaban E. Dengan demikian soal ini termasuk kategori soal yang cukup baik dan dapat digunakan **walaupun** ditinjau dari pengecohnya tampak bahwa alternatif jawaban C, daya pembedanya juga bernilai positif. Sama seperti soal nomor 8, karena besarnya daya pembeda pada kunci jawaban tetap lebih besar daripada daya pembeda pada alternatif jawaban C, maka butir soal ini tetap bisa diterima dan baik untuk digunakan.

Butir Soal Nomor 19

19	0-19	0.550	0.125	0.100	A	0.150	-0.009	-0.006	
					B	0.100	0.012	0.007	
					C	0.100	-0.356	-0.208	
					D	0.100	0.074	0.043	
					E	0.550	0.125	0.100	*
					Other	0.000	-9.000	-9.000	

Pada soal nomor 19, tingkat kesukarannya adalah 0,550 (sedang), sebanyak 55,0% peserta tes dapat menjawab soal tersebut dengan benar. Daya pembedanya, $r_{bis}=0,125$ dan $r_{pbis}=0,100$ menunjukkan keduanya bertanda positif. Hal ini berarti bahwa soal ini cukup dapat membedakan peserta tes yang pintar dan kurang pintar. Ditinjau dari distribusi jawaban yaitu persentase peserta tes merespons alternatif jawaban, semua pengecoh telah berfungsi dengan baik. Dapat dilihat pada kolom *Prop Endorsing*, tampak bahwa sebanyak 15% peserta tes merespons alternatif jawaban A, 10% merespons alternatif jawaban B, 10% merespons alternatif jawaban C, dan 10% merespons alternatif jawaban D. Soal ini termasuk kategori soal yang cukup baik dan dapat digunakan **walaupun** ditinjau dari daya pembedanya termasuk dalam kategori perlu direvisi. Hal ini tampak dengan tidak adanya peringatan dari hasil analisis. Hal ini dikarenakan pengecoh telah cukup berfungsi dengan baik ditinjau dari distribusi jawaban peserta tes walaupun tampak bahwa pada alternatif jawaban B, daya pembedanya juga bernilai positif. Sama seperti soal nomor 8, karena besarnya daya pembeda pada kunci jawaban tetap lebih besar daripada alternatif jawaban B, maka butir soal ini tetap bisa diterima dan dapat digunakan.

Reliabilitas

Sebenarnya tidak terdapat suatu ukuran yang pasti mengenai berapa tinggi koefisien reliabilitas. Reliabilitas yang baik tergantung pada tujuan atau kegunaan tes. Menurut Remmers et al. (1960) menyatakan bahwa kebanyakan tes-tes di bidang pendidikan pada umumnya memiliki koefisien reliabilitas minimal 0,8 untuk populasi yang sesuai. Sedangkan menurut Nunally (1972) dan Kaplan & Saccuzo, koefisien reliabilitas 0,7 s.d. 0,8 sudah termasuk cukup tinggi untuk penelitian dasar.

Berdasarkan hasil analisis, diperoleh:

There were 20 examinees in the data file.

Scale Statistics

```
-----  
Scale:          0  
-----  
N of Items      20  
N of Examinees  20  
Mean            11.400  
Variance        21.540  
Std. Dev.       4.641  
Skew            0.381  
Kurtosis        -1.109  
Minimum         5.000  
Maximum         20.000  
Median          10.000  
Alpha          0.847  
SEM             1.817  
Mean P          0.570  
Mean Item-Tot.  0.500  
Mean Biserial   0.653
```

Tampak bahwa koefisien reliabilitas alpha sebesar 0,847. Hal ini berarti reliabilitas tes ini dapat dikatakan termasuk dalam kategori **baik** .

Referensi:

- Allen, Mary .J., dan Yen, Wendy, M. (1979). *Introduction to measurement theory*. Monterey: Cole Publishing Company.
- Azwar, Saifudin. (1997). *Reliabilitas dan validitas*. (edisi ke-3).Yogyakarta: Pustaka Pelajar.
- Bahrul Hayat, Ph.D., Drs. SumarnaS., M.S., Drs. Suprananto yang disunting oleh Jahja Umar, Ph.D. dan Herwindo Ariwibowo. (1997). Ph.D. PEDOMAN PENGGUNAAN ITEMAN [*Manual Item and Test Analysis* (ITEMAN) Jakarta: Pusat Penelitian dan Pengembangan Sistem Pengujian, Balitbang Dikbud. 1997
- Ebel, Robert L. (1972). *Essential of educational measurement and evaluating in education and psycology*. New York: Holt, Rine hart, and Winston. Inc.
- Gronlund, Norman E. (1981). *Measurement and evaluating in teaching*. New York: Macmillan Publishing Co., Inc.
- Mardapi, Djemari. (1998). Analisis butir soal dengan teori tes klasik dan teori respons butir. *Jurnal kependidikan*. Edisi khusus Dies-tahun XXVIII.
- Mehrens, W. A. & Lehman, I. J. (1973). *Measuement and evaluation in education and psycology*. New York: Holt, Rinehart, and Winston , Inc.