



Panduan Analisis Butir Soal

**Departemen Pendidikan Nasional
Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah
Direktorat Pembinaan Sekolah Menengah Atas
Tahun 2008**

DAFTAR ISI

	<i>Halaman</i>
DAFTAR ISI	i
I. PENDAHULUAN	1
A. Pengertian	1
B. Manfaat Soal yang Telah Ditelaah	1
II. ANALISIS BUTIR SOAL SECARA KUALITATIF	3
A. Pengertian	3
B. Teknik Analisis Secara Kualitatif	3
C. Prosedur Analisis Secara Kualitatif	3
III. ANALISIS BUTIR SOAL SECARA KUANTITATIF	9
A. Pengertian.....	9
B. Analisis Butir Soal	9
IV. ANALISIS BUTIR SOAL DENGAN KALKULATOR	21
A. Pengertian	21
B. Pembersihan Data	21
C. Fungsi SD	21
D. Fungsi LR	22
E. Contoh Merandom Data	22
F. Contoh Uji Validitas Butir Soal Pilihan Ganda	23
V. ANALISIS BUTIR SOAL DENGAN KOMPUTER	25
A. Pengertian	25
B. Iteaman	25
C. Excel	30
D. SPSS (Statistical Program for Social Science)	30
VI. DAFTAR PUSTAKA	39

A. Pengertian

Kegiatan menganalisis butir soal merupakan suatu kegiatan yang harus dilakukan guru untuk meningkatkan mutu soal yang telah ditulis. Kegiatan ini merupakan proses pengumpulan, peringkasan, dan penggunaan informasi dari jawaban siswa untuk membuat keputusan tentang setiap penilaian (Nitko, 1996: 308). Tujuan penelaahan adalah untuk mengkaji dan menelaah setiap butir soal agar diperoleh soal yang bermutu sebelum soal digunakan. Di samping itu, tujuan analisis butir soal juga untuk membantu meningkatkan tes melalui revisi atau membuang soal yang tidak efektif, serta untuk mengetahui informasi diagnostik pada siswa apakah mereka sudah/belum memahami materi yang telah diajarkan (Aiken, 1994: 63). Soal yang bermutu adalah soal yang dapat memberikan informasi setepat-tepatnya sesuai dengan tujuannya di antaranya dapat menentukan peserta didik mana yang sudah atau belum menguasai materi yang diajarkan guru.

Dalam melaksanakan analisis butir soal, para penulis soal dapat menganalisis secara kualitatif, dalam kaitan dengan isi dan bentuknya, dan kuantitatif dalam kaitan dengan ciri-ciri statistiknya (Anastasi dan Urbina, 1997: 172) atau prosedur peningkatan secara judgment dan prosedur peningkatan secara empirik (Popham, 1995: 195). Analisis kualitatif mencakup pertimbangan validitas isi dan konstruk, sedangkan analisis kuantitatif mencakup pengukuran kesulitan butir soal dan diskriminasi soal yang termasuk validitas soal dan reliabilitasnya.

Jadi, ada dua cara yang dapat digunakan dalam penelaahan butir soal yaitu penelaahan soal secara kualitatif dan kuantitatif. Kedua teknik ini masing-masing memiliki keunggulan dan kelemahan. Oleh karena itu teknik terbaik adalah menggunakan keduanya (penggabungan). Kedua cara ini diuraikan secara rinci dalam buku ini.

B. Manfaat Soal yang Telah Ditelaah

Tujuan utama analisis butir soal dalam sebuah tes yang dibuat guru adalah untuk mengidentifikasi kekurangan-kekurangan dalam tes atau dalam pembelajaran (Anastasi dan Urbina, 1997:184). Berdasarkan tujuan ini, maka kegiatan analisis butir soal memiliki banyak manfaat, di antaranya adalah: (1) dapat membantu para pengguna tes dalam evaluasi atas tes yang digunakan, (2) sangat relevan bagi penyusunan tes informal dan lokal seperti tes yang disiapkan guru untuk siswa di kelas, (3) mendukung penulisan butir soal yang efektif, (4) secara materi dapat memperbaiki tes di kelas, (5) meningkatkan validitas soal dan reliabilitas (Anastasi and Urbina, 1997:172). Di samping itu, manfaat lainnya adalah: (1) menentukan apakah suatu fungsi butir soal sesuai dengan yang diharapkan, (2) memberi masukan kepada siswa tentang kemampuan dan sebagai dasar untuk bahan diskusi di kelas, (3) memberi masukan kepada guru tentang kesulitan siswa, (4) memberi masukan pada aspek tertentu untuk pengembangan kurikulum, (5) merevisi materi yang dinilai atau diukur, (6) meningkatkan keterampilan penulisan soal (Nitko, 1996: 308-309).

Linn dan Gronlund (1995: 315) juga menambahkan tentang pelaksanaan kegiatan analisis butir soal yang hiasanya didesain untuk menjawab pertanyaan-pertanyaan berikut ini.

(1) Apakah fungsi soal sudah tepat? (2) Apakah soal ini memiliki tingkat kesukaran yang tepat? (3) Apakah soal bebas dari hal-hal yang tidak relevan? (4) Apakah pilihan jawabannya efektif? Lebih lanjut Linn dan Gronlund (1995: 316-318) menyatakan bahwa kegunaan analisis butir soal bukan hanya terbatas untuk meningkatkan butir soal, tetapi ada beberapa hal, yaitu bahwa

data analisis butir soal bermanfaat sebagai dasar: (1) diskusi kelas efisien tentang hasil tes, (2) untuk kerja remedial, (3) untuk peningkatan secara umum pembelajaran di kelas, dan (3) untuk peningkatan keterampilan pada konstruksi tes.

Berbagai uraian di atas menunjukkan bahwa analisis butir soal adalah: (1) untuk menentukan soal-soal yang cacat atau tidak berfungsi penggunaannya; (2) untuk meningkatkan butir soal melalui tiga komponen analisis yaitu tingkat kesukaran, daya pembeda, dan pengecoh soal, serta meningkatkan pembelajaran melalui ambiguitas soal dan keterampilan tertentu yang menyebabkan peserta didik sulit. Di samping itu, butir soal yang telah dianalisis dapat memberikan informasi kepada peserta didik dan guru seperti contoh berikut ini.

DATA KEMAMPUAN PESERTA DIDIK

NAMA SISWA	NOMOR SOAL*										SKOR TOTAL#	KETERANGAN
	5	10	2	6	9	2	7	3	8	4		
A	1	1	1	1	1	1	0	1	0	0	7	Normal
B	1	1	1	1	1	0	1	0	0	0	6	Normal
C	0	0	0	1	0	1	1	0	1	1	5	Mengantuk dll.
D	1	0	1	0	0	0	0	0	1	1	4	Menebak
E	1	1	1	0	0	0	0	0	0	0	3	Lamban, berat
JUMLAH	4	3	4	3	2	2	2	1	2	2		

Keterangan:

1 = soal yang dijawab benar

0 = soal yang dijawab salah

* Soal disusun dari soal yang paling mudah sampai dengan soal yang paling sukar

Disusun dari skor yang paling tinggi sampai dengan skor paling rendah

Dari data di atas seperti soal nomor 3, 8, dan 4 (hanya dapat dijawab benar oleh 1, 2, dan 2 peserta didik) dapat memberikan informasi kepada guru atau pengawas tentang materi soal itu yang telah diajarkan kepada peserta didik. Mereka dapat memperbaiki diri berdasarkan informasi/data di atas. Informasi itu misalnya berupa 10 pertanyaan introspeksi diri atau penilaian diri seperti berikut ini.

PENILAIAN DIRI

NO	ASPEK YANG DITANYAKAN	YA	TIDAK
1.	Apakah guru membuat persiapan mengajar khususnya materi yang bersangkutan?		
2.	Apakah guru menguasai materi yang bersangkutan?		
3.	Apakah guru telah mengajarkan secara maksimal materi yang sesuai dengan tuntutan kompetensi yang harus dikuasai peserta didik?		
4.	Apakah perilaku yang diukur pada materi yang ditanyakan dalam soal itu sudah tepat (harus dikuasai siswa)?		
5.	Apakah materi yang ditanyakan merupakan materi urgensi, kontinyuitas, relevansi, dan keterpakaian dalam kehidupan sehari-hari tinggi?		
6.	Apakah guru memiliki kreativitas dalam memelajarkan materi yang bersangkutan?		
7.	Apakah guru mampu membangkitkan minat dan kegiatan belajar peserta didik khususnya dalam membelajarkan materi yang bersangkutan?		
8.	Apakah guru telah menyusun kisi-kisi dengan tepat sebelum menulis		

9.	soal? Apakah guru menulis soal berdasarkan indikator dalam kisi-kisi dan kaidah penulisan soal serta menyusun pedoman penskoran atau pedoman pengamatannya?		
10.	Apakah soal nomor 3, 8, dan 4 valid yaitu memiliki daya beda tinggi, tidak salah kunci jawaban, pengecohnya berfungsi, atau memang materinya belum diajarkan?		

Keterangan: Secara jujur berilah tanda (V) pada kolom Ya dan Tidak.

II. ANALISIS BUTIR SOAL SECARA KUALITATIF

A. Pengertian

Pada prinsipnya analisis butir soal secara kualitatif dilaksanakan berdasarkan kaidah penulisan soal (tes tertulis, perbuatan, dan sikap). Penelaahan ini biasanya dilakukan sebelum soal digunakan/diujikan.

Aspek yang diperhatikan di dalam penelaahan secara kualitatif ini adalah setiap soal ditelaah dari segi materi, konstruksi, bahasa/budaya, dan kunci jawaban/pedoman penskorannya. Dalam melakukan penelaahan setiap butir soal, penelaah perlu mempersiapkan bahan-bahan penunjang seperti: (1) kisi-kisi tes, (2) kurikulum yang digunakan, (3) buku sumber, dan (4) kamus bahasa Indonesia.

B. Teknik Analisis Secara Kualitatif

Ada beberapa teknik yang dapat digunakan untuk menganalisis butir soal secara kualitatif, diantaranya adalah teknik moderator dan teknik panel.

Teknik moderator merupakan teknik berdiskusi yang di dalamnya terdapat satu orang sebagai penengah. Berdasarkan teknik ini, setiap butir soal didiskusikan secara bersama-sama dengan beberapa ahli seperti guru yang mengajarkan materi, ahli materi, penyusun/pengembang kurikulum, ahli penilaian, ahli bahasa, berlatar belakang psikologi. Teknik ini sangat baik karena setiap butir soal dilihat secara bersama-sama berdasarkan kaidah penulisan. Di samping itu, para penelaah dipersilakan mengomentari/memperbaiki berdasarkan ilmu yang dimilikinya. Setiap komentar/masukan dari peserta diskusi dicatat oleh notulis. Setiap butir soal dapat dituntaskan secara bersama-sama, perbaikannya seperti apa. Namun, kelemahan teknik ini adalah memerlukan waktu lama untuk mendiskusikan setiap satu butir soal.

Teknik panel merupakan suatu teknik menelaah butir soal yang setiap butir soalnya ditelaah berdasarkan kaidah penulisan butir soal, yaitu ditelaah dari segi materi, konstruksi, bahasa/budaya, kebenaran kunci jawaban/pedoman penskorannya yang dilakukan oleh beberapa penelaah. Caranya adalah beberapa penelaah diberikan: butir-butir soal yang akan ditelaah, format penelaahan, dan pedoman penilaian/ penelaahannya. Pada tahap awal para penelaah diberikan pengarahan, kemudian tahap berikutnya para penelaah berkerja sendiri-sendiri di tempat yang tidak sama. Para penelaah dipersilakan memperbaiki langsung pada teks soal dan memberikan komentarnya serta memberikan nilai pada setiap butir soalnya yang kriterianya adalah: baik, diperbaiki, atau diganti.

Secara ideal penelaah butir soal di samping memiliki latar belakang materi yang diujikan, beberapa penelaah yang diminta untuk menelaah butir soal memiliki keterampilan, seperti guru yang mengajarkan materi itu, ahli materi, ahli pengembang kurikulum, ahli penilaian, psikolog, ahli bahasa, ahli kebijakan pendidikan, atau lainnya.

C. Prosedur Analisis Secara Kualitatif

Dalam menganalisis butir soal secara kualitatif, penggunaan format penelaahan soal akan sangat membantu dan mempermudah prosedur pelaksanaannya. Format penelaahan soal digunakan sebagai dasar untuk menganalisis setiap butir soal. Format penelaahan soal yang dimaksud

adalah format penelaahan butir soal: uraian, pilihan ganda, tes perbuatan dan instrumen non-tes.

Agar penelaah dapat dengan mudah menggunakan format penelaahan soal, maka para penelaah perlu memperhatikan petunjuk pengisian formatnya. Petunjuknya adalah seperti berikut ini.

1. Analisislah setiap butir soal berdasarkan semua kriteria yang tertera di dalam format!
2. Berilah tanda cek (V) pada kolom "Ya" bila soal yang ditelaah sudah sesuai dengan kriteria!
3. Berilah tanda cek (V) pada kolom "Tidak" bila soal yang ditelaah tidak sesuai dengan kriteria, kemudian tuliskan alasan pada ruang catatan atau pada teks soal dan perbaikannya.

a. Format Penelaahan Butir Soal Bentuk Uraian

FORMAT PENELAAHAN BUTIR SOAL BENTUK URAIAN

Mata Pelajaran :

Kelas/semester :

Penelaah :

No.	Aspek yang ditelaah	Nomor Soal									
		1	2	3	4	5	6	7	8	9	...
A.	Materi										
1	Soal sesuai dengan indikator (menuntut tes tertulis untuk bentuk Uraian)										
2	Batasan pertanyaan dan jawaban yang diharapkan sudah sesuai										
3	Materi yang ditanyakan sesuai dengan kompetensi (urgensi, relevansi, kontinuitas, keterpakaian sehari-hari tinggi)										
4	Isi materi yang ditanyakan sesuai dengan jenjang jenis sekolah atau tingkat kelas										
B	Konstruksi										
5	Menggunakan kata tanya atau perintah yang menuntut jawaban uraian										
6	Ada petunjuk yang jelas tentang cara mengerjakan soal										
7	Ada pedoman penskorannya										
8	Tabel, gambar, grafik, peta, atau yang sejenisnya disajikan dengan jelas dan terbaca										

No.	Aspek yang ditelaah	Nomor Soal									
		1	2	3	4	5	6	7	8	9	...
C.	Bahasa/Budaya										
9	Rumusan kalimat coal										
10	komunikatif										
	Butir soal menggunakan bahasa										
11	Indonesia yang baku										
	Tidak menggunakan										
	kata/ungkapan yang										
12	menimbulkan penafsiran ganda										
	atau salah pengertian										
13	Tidak menggunakan bahasa										
	yang berlaku setempat/tabu										
	Rumusan soal tidak										
	mengandung										

Keterangan: Berilah tanda (V) bila tidak sesuai dengan aspek yang ditelaah!

b. Format Penelaahan Soal Bentuk Pilihan Ganda

FORMAT PENELAAHAN SOAL BENTUK PILIHAN GANDA

Mata Pelajaran :

Kelas/semester :

Penelaah :

No.	Aspek yang ditelaah	Nomor Soal					
		1	2	3	4	5	...
A.	Materi						
1	Soal sesuai dengan indikator (menuntut tes tertulis untuk bentuk pilihan ganda						
2.	Materi yang ditanyakan sesuai dengan kompetensi (urgensi, relevansi, kontinuitas, keterpakaian sehari-hari tinggi)						
3.	Pilihan jawaban homogen dan logis						
4.	Hanya ada satu kunci jawaban						
B.	Konstruksi						
5.	Pokok soal dirumuskan dengan singkat, jelas, dan tegas						
6.	Rumusan pokok soal dan pilihan jawaban merupakan pernyataan yang diperlukan saja						
7.	Pokok soal tidak memberi petunjuk kunci jawaban						
8	Pokok soal bebas dan pernyataan yang						

No.	Aspek yang ditelaah	Nomor Soal					
		1	2	3	4	5	...
9.	bersifat negatif ganda Pilihan jawaban homogen dan logis ditinjau dari segi materi						
10.	Gambar, grafik, tabel, diagram, atau sejenisnya jelas dan berfungsi						
11.	Panjang pilihan jawaban relatif sama						
12.	Pilihan jawaban tidak menggunakan pernyataan "semua jawaban di atas salah/benar" dan sejenisnya						
13.	Pilihan jawaban yang berbentuk angka/waktu disusun berdasarkan urutan besar kecilnya angka atau kronologisnya						
14.	Butir soal tidak bergantung pada jawaban soal sebelumnya						
C.	Bahasa/Budaya						
15.	Menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia						
16.	Menggunakan bahasa yang komunikatif						
17.	Tidak menggunakan bahasa yang berlaku setempat/tabu						
18.	Pilihan jawaban tidak mengulang kata/kelompok kata yang sama, kecuali merupakan satu kesatuan pengertian						

Keterangan: Berilah tanda (V) bila tidak sesuai dengan aspek yang ditelaah!

c. **Format Penelaahan untuk Instrumen Perbuatan**

FORMAT PENELAAHAN SOAL TES PERBUATAN

Mata Pelajaran :

Kelas/semester :

Penelaah :

No.	Aspek yang ditelaah	Nomor Soal			
		1	2	3	...
A.	Materi				
1.	Soal sudah sesuai dengan indikator (menuntut tes perbuatan: kinerja, hasil karya, atau penugasan)				
2.	Pertanyaan dan jawaban yang diharapkan sudah sesuai				
3.	Materi sesuai dengan tuntutan kompetensi (urgensi, relevansi, kontinuitas, keterpakaian sehari-hari tinggi)				

No.	Aspek yang ditelaah	Nomor Soal			
		1	2	3	...
	Isi materi yang ditanyakan sesuai dengan jenjang jenis sekolah taua tingkat kelas				
	B. Konstruksi				
	5. Menggunakan kata tanya atau perintah yang menuntut jawaban perbuatan/praktik				
	6. Ada petunjuk yang jelas tentang cara				
	7. menjejakan soal				
	8. Ada pedoman penskorannya				
	Tabel, peta, gambar, grafik, atau sejenisnya disajikan dengan jelas dan terbaca				
	C. Bahasa/Budaya				
	9. Rumusan soal komunikatif				
	10. Butir soal menggunakan bahasa Indonesia yang				
	11. baku				
	Tidak menggunakan kata /ungkapan yang				
	12. menimbulkan penafsiran ganda atau salah				
	13. pengertian				
	Tidak menggunakan bahasa yang berlaku setempat/tabu				
	Rumusan soal tidak mengandung kata/ungkapan yang dapat menyinggung perasaan siswa				

Keterangan: Berilah tanda (V) bila tidak sesuai dengan aspek yang ditelaah!

d. Format Penelaahan untuk Instrumen Non-Tes

FORMAT PENELAAHAN SOAL NON-TES

Nama Tes :
 Kelas/semester :
 Penelaah :

No.	Aspek yang ditelaah	Nomor Soal			
		1	2	3	...
A.	Materi				
1.	Pernyataan/soal sudah sesuai dengan rumusan indikator dalam kisi-kisi.				
2.	Aspek yang diukur pada setiap pernyataan sudah sesuai dengan tuntutan dalam kisi-kisi (misal untuk tes sikap: aspek koginisi, afeksi, atau konasinya dan pernyataan positif atau negatifnya).				
B.	Konstruksi				
3.	Pernyataan dirumuskan dengan singkat (tidak melebihi 20 kata) dan jelas.				
4.	Kalimatnya bebas dari pernyataan yang tidak relevan objek yang dipersoalkan atau kalimatnya merupakan pernyataan yang diperlukan saja.				
5.	Kalimatnya bebas dari pernyataan yang bersifat negatif ganda.				
6.	Kalimatnya bebas dari pernyataan yang mengacu pada masa lalu.				
7.	Kalimatnya bebas dari pernyataan faktual atau dapat diinterpretasikan sebagai fakta.				
8.	Kalimatnya bebas dari pernyataan dapat diinterpretasikan lebih d Kalimatnya bebas dari pernyataan yang mungkin disetujui atau dikosongkan oleh hampir semua responden.				
9.	Setiap pernyataan hanya berisi satu gagasan secara lengkap.				
10.	Kalimatnya bebas dari pernyataan yang tidak pasti pasti seperti semua, selalu, kadang-kadang, tidak satupun, tidak pernah.				
11.	Jangan banyak menggunakan kata hanya, sekedar, semata-mata.				
12.	Gunakan seperlunya.				
C.	Bahasa/Budaya				
13.	Bahsa soa harus komunikatif dan sesuai dengan jenjang pendidikan siswa atau responden.				
14.					

No.	Aspek yang ditelaah	Nomor Soal			
		1	2	3	...
15.	Soal harus menggunakan bahasa Indonesia baku. Soal tidak menggunakan bahasa yang berlaku setempat/tabu.				

Keterangan: Berilah tanda (V) bila tidak sesuai dengan aspek yang ditelaah!

III. ANALISIS BUTIR SOAL SECARA KUANTITATIF

A. Pengertian

Penelaahan soal secara kuantitatif maksudnya adalah penelaahan butir soal didasarkan pada data empirik dari butir soal yang bersangkutan. Data empirik ini diperoleh dari soal yang telah diujikan.

B. Analisis Butir Soal

Ada dua pendekatan dalam analisis secara kuantitatif, yaitu pendekatan secara klasik dan modern.

1. Klasik

Analisis butir soal secara klasik adalah proses penelaahan butir soal melalui informasi dari jawaban peserta didik guna meningkatkan mutu butir soal yang bersangkutan dengan menggunakan teori tes klasik.

Kelebihan analisis butir soal secara klasik adalah murah, dapat dilaksanakan sehari-hari dengan cepat menggunakan komputer, murah, sederhana, familier dan dapat menggunakan data dari beberapa peserta didik atau sampel kecil (Millman dan Greene, 1993: 358).

Adapun proses analisisnya sudah banyak dilaksanakan para guru di sekolah seperti beberapa contoh di bawah ini.

- a. Langkah pertama yang dilakukan adalah menabulasi jawaban yang telah dibuat pada setiap butir soal yang meliputi berapa peserta didik yang: (1) menjawab benar pada setiap soal, (2) menjawab salah (option pengecoh), (3) tidak menjawab soal. Berdasarkan tabulasi ini, dapat diketahui tingkat kesukaran setiap butir soal, daya pembeda soal, alternatif jawaban yang dipilih peserta didik.
- b. Misalnya analisis untuk 32 siswa, maka langkah (1) urutkan skor siswa dari yang tertinggi sampai yang terendah. (2) Pilih 10 lembar jawaban pada kelompok atas dan 10 lembar jawaban pada kelompok bawah. (3) Ambil kelompok tengah (12 lembar jawaban) dan tidak disertakan dalam analisis. (4) Untuk masing-masing soal, susun jumlah siswa kelompok atas dan bawah pada setiap pilihan jawaban. (5) Hitung tingkat kesukaran pada setiap butir soal. (6) Hitung daya pembeda soal. (7) Analisis efektivitas pengecoh pada setiap soal (Linn dan Gronlund, 1995: 318-319).

Aspek yang perlu diperhatikan dalam analisis butir soal secara klasik adalah setiap butir soal ditelaah dari segi: tingkat kesukaran butir, daya pembeda butir, dan penyebaran pilihan jawaban (untuk soal bentuk obyektif) atau frekuensi jawaban pada setiap pilihan jawaban.

a. Tingkat Kesukaran (TK)

Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Indeks tingkat kesukaran ini pada umumnya dinyatakan dalam bentuk proporsi yang besarnya berkisar 0,00 - 1,00 (Aiken (1994: 66). Semakin besar indeks tingkat kesukaran

yang diperoleh dari hasil hitungan, berarti semakin mudah soal itu. Suatu soal memiliki TK= 0,00 artinya bahwa tidak ada siswa yang menjawab benar dan bila memiliki TK= 1,00 artinya bahwa siswa menjawab benar. Perhitungan indeks tingkat kesukaran ini dilakukan untuk setiap nomor soal. Pada prinsipnya, skor rata-rata yang diperoleh peserta didik pada butir soal yang bersangkutan dinamakan tingkat kesukaran butir soal itu. Rumus ini dipergunakan untuk soal obyektif. Rumusnya adalah seperti berikut ini (Nitko, 1996: 310).

$$\text{Tingkat Kesukaran (TK)} = \frac{\text{Jumlah siswa yang menjawab benar butir soal}}{\text{Jumlah siswa yang mengikutites}}$$

Fungsi tingkat kesukaran butir soal biasanya dikaitkan dengan tujuan tes. Misalnya untuk keperluan ujian semester digunakan butir soal yang memiliki tingkat kesukaran sedang, untuk keperluan seleksi digunakan butir soal yang memiliki tingkat kesukaran tinggi/sukar, dan untuk keperluan diagnostik biasanya digunakan butir soal yang memiliki tingkat kesukaran rendah/mudah.

Untuk mengetahui tingkat kesukaran soal bentuk uraian digunakan rumus berikut ini.

$$\text{Mean} = \frac{\text{Jumlah skor siswa peserta tes pada suatu soal}}{\text{Jumlah peserta didik yang mengikutites}}$$

$$\text{Tingkat Kesulitan} = \frac{\text{Mean}}{\text{Skor maksimum yang ditetapkan}}$$

Hasil perhitungan dengan menggunakan rumus di atas menggambarkan tingkat kesukaran soal itu. Klasifikasi tingkat kesukaran soal dapat dicontohkan seperti berikut ini.

0,00 - 0,30 soal tergolong sukar

0,31 - 0,70 soal tergolong sedang

0,71 - 1,00 soal tergolong mudah

Tingkat kesukaran butir soal dapat mempengaruhi bentuk distribusi total skor tes. Untuk tes yang sangat sukar (TK= < 0,25) distribusinya berbentuk positif skewed, sedangkan tes yang mudah dengan TK= >0,80 distribusinya berbentuk negatif skewed.

Tingkat kesukaran butir soal memiliki 2 kegunaan, yaitu kegunaan bagi guru dan kegunaan bagi pengujian dan pengajaran (Nitko, 1996: 310-313). Kegunaannya bagi guru adalah: (1) sebagai pengenalan konsep terhadap pembelajaran ulang dan memberi masukan kepada siswa tentang hasil belajar mereka, (2) memperoleh informasi tentang penekanan kurikulum atau mencurigai terhadap butir soal yang bias. Adapun kegunaannya bagi pengujian dan pengajaran adalah: (a) pengenalan konsep yang diperlukan untuk diajarkan ulang, (b) tanda-tanda terhadap kelebihan dan kelemahan pada kurikulum sekolah, (c) memberi masukan kepada siswa, (d) tanda-tanda kemungkinan adanya butir soal yang bias, (e) merakit tes yang memiliki ketepatan data soal.

Di samping kedua kegunaan di atas, dalam konstruksi tes, tingkat kesukaran butir soal sangat penting karena tingkat kesukaran butir dapat: (1) mempengaruhi karakteristik distribusi skor (mempengaruhi bentuk dan penyebaran skor tes atau jumlah soal dan korelasi antarsoal), (2) berhubungan dengan reliabilitas. Menurut koefisien alfa clan KR-20, semakin tinggi korelasi antarsoal, semakin tinggi reliabilitas (Nunnally, 1981: 270-271).

Tingkat kesukaran butir soal juga dapat digunakan untuk memprediksi alat ukur itu sendiri (soal) dan kemampuan peserta didik dalam memahami materi yang diajarkan guru. Misalnya satu butir soal termasuk kategori mudah, maka prediksi terhadap informasi ini adalah seperti berikut.

- 1) Pengecoh butir soal itu tidak berfungsi.
- 2) Sebagian besar siswa menjawab benar butir soal itu; artinya bahwa sebagian besar siswa telah memahami materi yang ditanyakan.

Bila suatu butir soal termasuk kategori sukar, maka prediksi terhadap informasi ini adalah seperti berikut.

- 1) Butir soal itu "mungkin" salah kunci jawaban.
- 2) Butir soal itu mempunyai 2 atau lebih jawaban yang benar.
- 3) Materi yang ditanyakan belum diajarkan atau belum tuntas pembelajarannya, sehingga kompetensi minimum yang harus dikuasai siswa belum tercapai.
- 4) Materi yang diukur tidak cocok ditanyakan dengan menggunakan bentuk soal yang diberikan (misalnya meringkas cerita atau mengarang ditanyakan dalam bentuk pilihan ganda).
- 5) Pernyataan atau kalimat soal terlalu kompleks dan panjang.

Namun, analisis secara klasik ini memang memiliki keterbatasan, yaitu bahwa tingkat kesukaran sangat sulit untuk mengestimasi secara tepat karena estimasi tingkat kesukaran dibiaskan oleh sampel (Haladyna, 1994: 145). Jika sampel berkemampuan tinggi, maka soal akan sangat mudah ($TK = >0,90$). Jika sampel berkemampuan rendah, maka soal akan sangat sulit ($TK = < 0,40$). Oleh karena itu memang merupakan kelebihan analisis secara IRT, karena IRT dapat mengestimasi tingkat kesukaran soal tanpa menentukan siapa peserta tesnya (invariance). Dalam IRT, komposisi sampel dapat mengestimasi parameter dan tingkat kesukaran soal tanpa bias.

b. Daya Pembeda (DP)

Daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan antara warga belajar/siswa yang telah menguasai materi yang ditanyakan dan warga belajar/siswa yang tidak/kurang/belum menguasai materi yang ditanyakan. Manfaat daya pembeda butir soal adalah seperti berikut ini.

- 1) Untuk meningkatkan mutu setiap butir soal melalui data empiriknya. Berdasarkan indeks daya pembeda, setiap butir soal dapat diketahui apakah butir soal itu baik, direvisi, atau ditolak.

2) Untuk mengetahui seberapa jauh setiap butir soal dapat mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan guru. Apabila suatu butir soal tidak dapat membedakan kedua kemampuan siswa itu, maka butir soal itu dapat dicurigai "kemungkinannya" seperti berikut ini.

- Kunci jawaban butir soal itu tidak tepat.
- Butir soal itu memiliki 2 atau lebih kunci jawaban yang benar
- Kompetensi yang diukur tidak jelas
- Pengecoh tidak berfungsi
- Materi yang ditanyakan terlalu sulit, sehingga banyak siswa yang menebak
- Sebagian besar siswa yang memahami materi yang ditanyakan berpikir ada yang salah informasi dalam butir soalnya

Indeks daya pembeda setiap butir soal biasanya juga dinyatakan dalam bentuk proporsi. Semakin tinggi indeks daya pembeda soal berarti semakin mampu soal yang bersangkutan membedakan warga belajar/siswa yang telah memahami materi dengan warga belajar/peserta didik yang belum memahami materi. Indeks daya pembeda berkisar antara -1,00 sampai dengan +1,00. Semakin tinggi daya pembeda suatu soal, maka semakin kuat/baik soal itu. Jika daya pembeda negatif (<0) berarti lebih banyak kelompok bawah (warga belajar/peserta didik yang tidak memahami materi) menjawab benar soal dibanding dengan kelompok atas (warga belajar/peserta didik yang memahami materi yang diajarkan guru).

Untuk mengetahui daya pembeda soal bentuk pilihan ganda adalah dengan menggunakan rumus berikut ini.

$$DP = \frac{BA - BB}{\frac{1}{2}N} \text{ atau } DP = \frac{2(BA - BB)}{N}$$

DP = daya pembeda soal,

BA = jumlah jawaban benar pada kelompok atas,

BB = jumlah jawaban benar pada kelompok bawah, N=jumlah siswa yang mengerjakan tes.

Di samping rumus di atas, untuk mengetahui daya pembeda soal bentuk pilihan ganda dapat dipergunakan rumus korelasi point biserial (r_{pbis}) dan korelasi biserial (r_{bis}) (Miliman and (ireene, 1993: 359-360) dan (Glass and Stanley, 1970: 169-170) seperti berikut.

$$r_{pbis} = \frac{\bar{X}_b - \bar{X}_s}{SD} \sqrt{pq} \quad \text{dan} \quad r_{bis} = \frac{\bar{Y}_b - \bar{Y}_s}{SD} \cdot \frac{nb.ns}{un\sqrt{n^2 - n}}$$

\bar{X}_b , \bar{Y}_b adalah rata-rata skor warga belajar/siswa yang menjawab benar

\bar{X}_s , \bar{Y}_s adalah rata-rata skor warga belajar siswa yang menjawab salah

SDt adalah simpangan baku skor total

nb dan n, adalah jumlah siswa yang menjawab benar dan jumlah siswa yang menjawab salah, serta $nb + n_s = n$.
 p adalah proporsi jawaban benar terhadap semua jawaban siswa
 q adalah $1 - p$
 U adalah ordinat kurva normal.

Untuk mengetahui daya pembeda soal bentuk uraian adalah dengan menggunakan rumus berikut ini.

$$DP = \frac{\text{Mean kelompok atas} - \text{Mean kelompok bawah}}{\text{Skor maksimum soal}}$$

Hasil perhitungan dengan menggunakan rumus di atas dapat menggambarkan tingkat kemampuan soal dalam membedakan antar peserta didik yang sudah memahami materi yang diujikan dengan peserta didik yang belum/tidak memahami materi yang diujikan. Adapun klasifikasinya adalah seperti berikut ini (Crocker dan Algina, 1986: 315).

0,40 - 1,00 soal diterima baik
 0,30 - 0,39 soal diterima tetapi perlu diperbaiki
 0,20 - 0,29 soal diperbaiki
 0,19 - 0,00 soal tidak dipakai/dibuang

r_{pbis} merupakan korelasi product moment antara skor dikotomis dan pengukuran kriterium, sedangkan r_{bis} merupakan korelasi product moment antara variabel latent distribusi normal berdasarkan dikotomi benar-salah dan pengukuran kriterium. Oleh karena itu, untuk perhitungan pada data yang sama $r_{pbis} = 0$, sedangkan r_{bis} paling sedikit 25% lebih besar daripada r_{pbis} . Kedua korelasi ini masing-masing memiliki kelebihan (Millman and Greene, 1993: 360) walaupun para guru/pengambil kebijakan banyak yang suka menggunakan r_{pbis} .

Kelebihan korelasi point biserial: (1) memberikan refleksi kontribusi soal secara sesungguhnya terhadap fungsi tes. Maksudnya ini mengukur bagaimana baiknya soal berkorelasi dengan criterion (tidak bagaimana baiknya beberapa/secara abstrak); (2) sederhana dan langsung berhubungan dengan statistik tes, (3) tidak pernah mempunyai value 1,00 karena hanya variabel-variabel dengan distribusi bentuk yang sama yang dapat berkorelasi secara tepat, dan variabel kontinyu (kriterium) dan skor dikotomis tidak mempunyai bentuk yang sama.

Adapun kelebihan korelasi biserial adalah: (1) cenderung lebih stabil dari sampel ke sampel, (2) penilaian lebih akurat tentang bagaimana soal dapat diharapkan untuk membedakan pada beberapa perbedaan point di skala abilitas, (3) value r_{bis} yang sederhana lebih langsung berhubungan dengan indikator diskriminasi ICC.

Contoh menghitung korelasi point biserial (r_{pbis}).

DAFTAR SKOR SISWA SOAL NOMOR 5

Nomor siswa yang menjawab benar	Jumlah skor keseluruhan	Nomor siswa yang menjawab salah	Jumlah skor keseluruhan
1	19	14	17
2	18	15	16
3	18	16	15
4	16	17	14
5	16	18	14
6	16	19	12
7	15	20	12
8	13	21	12
9	13	22	12
10	13	23	12
	12	24	11
	12	25	11
	11	26	10
		27	9
		28	8
		29	8
		30	7
Jumlah	192		200

Jumlah siswa yang menjawab benar = 13

Jumlah siswa yang menjawab salah = 17

Jumlah siswa keseluruhan = 30

Rata-rata siswa yang menjawab benar = $192:13 = 14,7692$

Rata-rata siswa yang menjawab salah = $200:17 = 11,7647$

Rata-rata skor siswa keseluruhan = $(192+200) : 30 = 13,0667$

Simpangan baku skor total = 3,0954

Jumlah skor keseluruhan = 392

$$r_{pbis} = \frac{\bar{xb} - \bar{xs}}{SD} \sqrt{pq}$$

$$r_{pbis} = \frac{14,7692 - 11,7647}{3,0954} \sqrt{\frac{13}{30} \cdot \frac{17}{30}}$$

$$r_{pbis} = \frac{3,0045}{3,0954} \sqrt{(0,433333)(0,566666)}$$

$$= (0,9706338) (0,4955355)$$

$$= 0,4809835$$

= 0,48 (Artinya butir soal nomor 5 diterima/baik)

Di samping menggunakan kriteria di atas, untuk menentukan diterima/tidaknya (signifikansi) suatu butir dapat ditentukan dengan menggunakan tabel Z bila $n \geq 30$ dengan menggunakan rumus $Z = r \sqrt{N-1}$ atau tabel t bila $n < 30$ dengan rumus $t = r \sqrt{(N-1)/(1-r^2)}$ (Bruning dan Kintz, 1987: 179-180). Contoh untuk data di atas digunakan tabel Z.

$$Z = r\sqrt{N-1}$$

$$Z = 0,48\sqrt{30-1}$$

$$Z = 2,58$$

Dalam tabel Z dapat diketahui untuk $\alpha = 0,05$ dengan 2 sisi (2 tailed), Z kritiknya adalah $\pm 1,96$ dan $Z=2,58$ probabilitasnya ("area di atas Z" atau "bidang tersempit") = 0,0049. Caranya adalah lihat Tabel Z pada lampiran buku ini.

c. Penyebaran (distribusi) jawaban

Penyebaran pilihan jawaban dijadikan dasar dalam penelaahan soal. Hal ini dimaksudkan untuk mengetahui berfungsi/tidaknya jawaban yang tersedia. Suatu pilihan jawaban (pengecoh) dapat dikatakan berfungsi apabila pengecoh:

- 1) paling tidak dipilih oleh 5 % peserta tes/siswa,
- 2) lebih banyak dipilih oleh kelompok siswa yang belum paham materi.

d. Reliabilitas Skor Tes

Tujuan utama menghitung reliabilitas skor tes adalah untuk mengetahui tingkat ketepatan (precision) dan keajegan (consistency) skor tes. Indeks reliabilitas berkisar antara 0 - 1. Semakin tinggi koefisien reliabilitas suatu tes (mendekati 1), makin tinggi pula keajegan/ketepatannya.

Tes yang memiliki konsistensi reliabilitas tinggi adalah akurat, reproducibel, dan generalized terhadap kesempatan testing dan instrumen tes lainnya. Secara rinci faktor yang mempengaruhi reliabilitas skor tes di antaranya:

- 1) Semakin banyak jumlah butir soal, semakin ajek suatu tes.
- 2) Semakin lama waktu tes, semakin ajek.
- 3) Semakin sempit range kesukaran butir soal, semakin besar keajegan.
- 4) Soal-soal yang saling berhubungan akan mengurangi keajegan.
- 5) Semakin objektif pemberian skor, semakin besar keajegan.
- 6) Ketidaktepatan pemberian skor.
- 7) Menjawab besar soal dengan cara menebak.
- 8) Semakin homogen materi semakin besar keajegan.
- 9) Pengalaman peserta ujian.
- 10) Salah penafsiran terhadap butir soal.
- 11) Menjawab soal dengan buru-buru/cepat.
- 12) Kesiapan mental peserta ujian.
- 13) Adanya gangguan dalam pelaksanaan tes.

- 14) Jarak antara tes pertama dengan tes kedua.
- 15) Mencontek dalam mengerjakan tes.
- 16) Posisi individu dalam belajar.
- 17) Kondisi fisik peserta ujian.

Ada 3 cara yang dapat dilakukan untuk menentukan reliabilitas skor tes, yaitu :

- 1) Keajegan pengukuran ulang: kesesuaian antara hasil pengukuran pertama dan kedua dari sesuatu alat ukur terhadap kelompok yang sama.
- 2) Keajegan pengukuran setara: kesesuaian hasil pengukuran dan 2 atau lebih alat ukur berdasarkan kompetensi kisi-kisi yang lama.
- 3) Keajegan belah dua: kesesuaian antara hasil pengukuran belahan pertama dan belahan kedua dari alat ukur yang sama.

Penggunaan rumus untuk mengetahui koefisien ketiga jenis reliabilitas di atas dijelaskan secara rinci berikut ini.

e. Reliabilitas Instrumen Tes (soal bentuk pilihan ganda)

Untuk mengetahui koefisien reliabilitas tes soal bentuk pilihan ganda digunakan rumus Kuder Richadson 20 (KR-20) seperti berikut ini.

$$KR - 20 = \frac{k}{k-1} \left[1 - \frac{\sum p(1-p)}{(SD)^2} \right]$$

Keterangan:

k : Jumlah butir soal
 $(SD)^2$: Varian

Contoh menghitung KR-20:

Siswa	Soal				Σ Skor	\bar{X}	$x - \bar{X}$	$(x - \bar{X})^2$
	1	2	3	4				
A	1	0	0	0	1	2	-1	1
B	1	1	0	0	2	2	0	0
C	0	0	1	1	2	2	0	0
D	0	0	0	0	0	2	-2	4
E	1	1	0	1	3	2	-1	1
F	1	1	1	1	4	2	-2	4
p	0,67	0,50	0,33	0,50	12			10

(1-p) : 0,33 0,50 0,67 0,50
p(1-p) : 0,22 0,25 0,22 0,25
Σp(1-p) : 0,22 + 0,25 + 0,22 + 0,25 = 0,944
Jumlah siswa = 6 orang
Jumlah skor = 12
Variance = $\frac{\sum (x - \bar{x})^2}{N}$

$$=10:6$$

$$=1,67$$

Standar Deviasi = $\sqrt{1,67}$

$$=1,29$$

$$KR-20 = \frac{k}{k-1} \left[1 - \frac{\sum p(1-p)}{(SD)^2} \right]$$

$$KR-20 = \frac{4}{4-1} \left[1 - \frac{0,944}{1,67} \right]$$

= 0,58 (Artinya bahwa tingkat keajegan tes ini rendah. Hal ini disebabkan butir soal yang dianalisis hanya 4 butir soal)

Di samping KR-20 di atas, ada teknik lain untuk menghitung reliabilitas tes, yaitu yang dikembangkan oleh Spearman-Brown. Caranya adalah dengan mengelompokkan nomor butir yang ganjil dan genap. Perhatikan contoh berikut ini.

Nama Peserta Didik	Soal				Σ Skor
	1	2	3	4	
Nur Chasanah	1	0	0	0	1
Salim Alkhasan	1	1	0	0	2
Abdul Latif	0	0	1	1	2
Choeroddin	0	0	0	0	0
Moh Chanif	1	1	0	1	3
Rofi'ah	1	1	1	1	4
p	0,67	0,50	0,33	0,50	12

Nama Peserta Didik	Butir Ganjil (1+3)	Butir Genap (2+4)	Skor Z untuk		Z _{ganjil} × Z _{genap}
			Ganjil	Genap	
Nur Chasanah	1	0	0	-1,22	0
Salim Alkhasan	1	1	0	0	0
Abdul Latif	1	1	0	0	0
Choeroddin	0	0	-1,72	-1,22	2,10
Moh Chanif	1	2	0	+1,22	0
Rofi'ah	2	2	+1,72	+1,22	2,10
p	0,67	0,50	0,33	0,50	12

n= 6 Mean = 1,0 1,0 Jumlah= 4,2

SD = 0,58 0,82

$$r_{12} = \frac{\sum Z_{ganjil} \times Z_{genap}}{n}$$

$$r_{12} = \frac{4,2}{6}$$

$$= 0,70$$

$$\text{reliabilitas Spearman Brown} = \frac{2r_{1.2}}{1+r_{1.2}}$$

$$= \frac{2 \cdot (0,70)}{1+0,70}$$

$$= 0,82$$

(Artinya bahwa tingkat keajegan/konsistensi tes ini adalah tinggi, sehingga skor tes ini dapat dipercaya penggunaannya.)

2. Modern

Analisis butir soal secara modern yaitu penelaahan butir soal dengan menggunakan Item Response Theory (IRT) atau teori jawaban butir soal. Teori ini merupakan suatu teori yang menggunakan fungsi matematika untuk menghubungkan antara peluang menjawab benar suatu soal dengan kemampuan siswa. Nama lain IRT adalah latent trait theory (LTT), atau characteristics curve theory (ICC).

Asal mula IRT adalah kombinasi suatu versi hukum phi-gamma dengan suatu analisis faktor butir soal (item factor analysis) kemudian bernama Teori Trait Latent (Latent Trait Theory), kemudian sekarang secara umum dikenal menjadi teori jawaban butir soal (Item Response Theory) (McDonald, 1999: 8).

Dalam subbab ini akan disajikan kelebihan analisis secara IRT dan kalibrasi butir soal dan pengukuran kemampuan orang.

1. Kelebihan Analisis IRT

Untuk mengetahui kelebihan analisis IRT, maka para guru perlu mengetahui keterbatasan analisis secara klasik. Keterbatasan model pengukuran secara klasik bila dibandingkan dengan teori jawaban butir soal adalah seperti berikut (Hambleton, Swaminathan, dan Rogers, 1991: 2-5). (1) Tingkat kemampuan dalam teori klasik adalah "true score". Jika tes sulit artinya tingkat kemampuan peserta didik mudah. Jika tes mudah artinya tingkat kemampuan peserta didik tinggi. (2) Tingkat kesukaran soal didefinisikan sebagai proporsi peserta didik dalam grup yang menjawab benar soal. Mudah/sulitnya butir soal tergantung pada kemampuan peserta didik yang dites dan kemampuan tes yang diberikan. (3) Daya pembeda, reliabilitas, dan validitas soal/tes didefinisikan berdasarkan grup peserta didik. Adapun kelebihan IRT adalah bahwa: (1) IRT tidak berdasarkan grup dependent, (2) skor siswa dideskripsikan bukan test dependent, (3) model ini menekankan pada tingkat butir soal bukan tes, (4) IRT tidak memerlukan paralel tes untuk

menentukan reliabilitas tes, (5) IRT suatu model yang memerlukan suatu pengukuran ketepatan untuk setiap skor tingkat kemampuan. Kelemahan teori tes klasik di atas diperkuat Hambleton dan Swaminathan (1985: 1-3) yaitu: (1) tingkat kesukaran dan daya pembeda tergantung pada sampel; (2) penggunaan metode dan teknik untuk desain dan analisis tes dengan memperbandingkan kemampuan siswa pada pernbagian kelompok atas, tengah, bawah. Meningkatnya validitas skor tes diperoleh dari tingkat kesukaran tes dihubungkan dengan tingkat kemampuan setiap siswa; (3) konsep reliabilitas tes didefinisikan dari istilah tes paralel; (4) tidak ada dasar teori untuk menentukan bagaimana siswa memperoleh tes yang sesuai dengan kemampuan siswa; (5) Standar error of measurement (SEM) hanya berlaku untuk seluruh peserta didik.

Selanjutnya Hambleton dan Swaminathan (1985: 13) menyatakan bahwa tujuan utama IRT adalah memberikan kesamaan antara statistik soal dan estimasi kemampuan. Ada tiga keuntungan IRT adalah: (1) asumsi banyak soal yang diukur pada trait yang sama, perkiraan tingkat kemampuan peserta didik adalah independen; (2) asumsi pada populasi tingkat kesukaran, daya pembeda merupakan independen sampel yang menggambarkan untuk tujuan kalibrasi soal; (3) statistik yang digunakan untuk menghitung tingkat kemampuan siswa diperkirakan dapat terlaksana, (Hableton dan Swaminathan, 1985: 11). Jadi IRT merupakan hubungan antara probabilitas jawaban suatu butir soal yang benar dan kemampuan siswa atau tingkatan/level prestasi siswa. Namun kelemahan bekerja dengan model IRT adalah bekerja melalui suatu proses yang sulit karena kelebihan IRT adalah: (1) tanpa varian pada parameter butir soal, (2) tanpa varian pada parameter abilitas, (3) adanya ketepatan pada pengukuran lokal, (Bejar, 1983: 3-4).

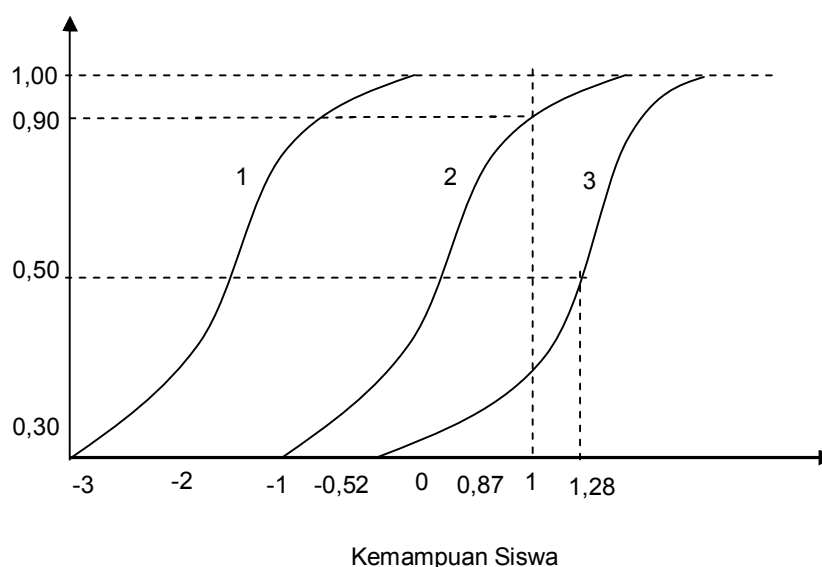
Ada empat macam model 1RT (Hambleton, 1993: 154-157; Hambleton dan Swaminathan, 1985: 34-50). (1) Model satu parameter (Model Rasch), yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran coal. (2) Model dua parameter, yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran dan daya pembeda soal. (3) Model tiga parameter, yaitu untuk menganalisis data yang menitikberatkan pada parameter tingkat kesukaran soal, daya pembeda soal, dan menebak (guessing). (4) Model empat parameter, yaitu untuk menganalisis data yang menitikberatkan pada parameter tingkat kesukaran soal, daya beda soal, menebak, dan penyebab lain.

Hambleton dan Swaminathan (1985: 48) menjelaskan bahwa siswa yang memiliki kemampuan tinggi tidak selalu menjawab soal dengan betul. Kadang-kadang mereka sembrono (mengerjakan dengan serampangan), memiliki informasi yang berlebihan, sehingga mereka menjawab salah pada suatu soal. Untuk mengatasi masalah ini diperlukan model 4 parameter.

Dari keempat model itu tidak sama penekanannya dan sudah barang tentu tiap-tiap model itu memiliki kelebihan dan kekurangan. Kelebihan dan kekurangan itu dapat diklasifikasikan sesuai dengan jumlah parameter yang ditentukan pada masing-masing model dan tujuan menggunakan model yang bersangkutan.

Adapun contoh kurva ciri soal model satu parameter atau Rasch terlihat seperti pada grafik di bawah ini.

Peluang menjawab benar



2. Kalibrasi Butir Soal dan Pengukuran Kemampuan Orang.

Kalibrasi butir soal dan pengukuran kemampuan orang merupakan proses estimasi parameter pada model respon butir. Model persamaan dasar Rasch adalah model probabilistik yang mencakup hasil dari suatu interaksi butir soal-orang. Proses mengestimasi kemampuan orang dinamakan pengukuran, sedangkan proses mengestimasi parameter tingkat kesukaran butir soal dinamakan kalibrasi. Jadi kalibrasi soal merupakan proses penyamaan skala soal yang didasarkan pada tingkat kesukaran butir soal dan tingkat kemampuan siswa. Adapun ciri suatu skala adalah mempunyai titik awal, biasanya 0, dan mempunyai satuan ukuran atau unit pengukuran.

Prosedur estimasi dapat dilakukan dengan tangan atau komputer. Ada beberapa langkah yang dapat dilakukan dalam mengkalibrasi butir dan mengukur kemampuan orang dengan tangan (Wright and Linacre, 1992: 32-45) seperti berikut ini.

- Menyusun jawaban peserta didik untuk setiap butir soal ke dalam tabel.
Dalam menyusun jawaban peserta didik untuk setiap butir ke dalam tabel perlu disediakan kolom: (1) siswa, (2) butir soal, (3) skor siswa, dan (4) skor butir soal. Data berbentuk angka 1 untuk jawaban benar dan 0 untuk jawaban salah.
- Mengedit data
Berdasarkan model Rasch, butir soal yang dijawab siswa betul semua atau salah semua dan siswa yang dapat menjawab dengan betul semua atau salah semua, soal atau siswa yang bersangkutan tidak dianalisis atau dikeluarkan dari tabel. Pada langkah kedua ini perlu disediakan tambahan kolom: (1) proporsi skor siswa dan (2) proporsi skor butir soal. Proporsi skor peserta didik adalah skor siswa : jumlah butir soal; sedangkan proporsi skor soal adalah skor soal : jumlah siswa.
- Menghitung distribusi skor soal
Berdasarkan skor soal yang sudah diedit, maka skor soal diklasifikasikan menjadi

beberapa kelompok berdasarkan skor yang sama. Untuk memudahkan penghitungan Distribusi skor butir soal, maka perlu disusun beberapa kolom di dalam tabel, seperti kolom: (1) kelompok skor soal (i) yaitu kelompok skor yang didasarkan pada skor soal yang sama, kolom ini berhubungan langsung dengan kolom 2 dan kolom 3; (2) nomor butir soal, (3) skor soal (Si), (4) frekuensi soal (Fi) yaitu jumlah soal yang memiliki skorsoal sama; (5) proporsi benar (Pi) yaitu Si : jumlah peserta tes; (6) proporsi salah (1-Pi), (7) logit (log odds unit)-proporsi salah (Xi) yaitu $\ln [(1 - Pi)/Pi]$, (8) hasil kali frekuensi soal dengan logit proporsi salah (FiXi), (9) kuadrat logit proporsi salah (FiXi)², (10) hasil kali frekuensi soal dengan kuadrat logit proporsi salah (FiXi²), (11) inisial kalibrasi butir soal yaitu $d_i = X_i - \text{nilai rata-rata skor soal}$, dan (12) hasil kali antara frekuensi soal dengan kuadrat nilai rata-rata skor coal (FIX ?).

- d. Menghitung distribusi skor peserta didik.
Untuk memudahkan di dalam menghitung distribusi skor peserta didik perlu disusun beberapa kolom yaitu kolom: (1) kemungkinan skor peserta didik (r) yang disusun secara berurutan dimulai dan skor terendah sampai tertinggi; (2) skor peserta didik, yaitu berupa toli skor peserta didik; (3) frekuensi peserta didik (nr) yang memperoleh skor; (4) proporsi benar (Pi) yaitu skor peserta didik dibagi jumlah soal, (5) logit proporsi benar (Yr) yaitu $\ln [Pr/(1-Pr)]$; (6) perkalian antara frekuensi siswa dengan logit proporsi benar (nrYr); (7) logic proporsi benar yang dikuadratkan (Yr kuadrat); (8) hasil perkalian antara frekuensi peserta didik dengan logic proporsi benar yang dikuadratkan (nrYr kuadrat); (9) inisial pengukuran kemampuan peserta didik (br Yr); (10) perkalian antara frekuensi peserta didik dengan nilai rata-rata skor peserta didik (nrYr kuadrat).
- e. Menghitung faktor ekspansi kemampuan peserta didik (x) dan kesukaran butir soal (Y). Dalam menghitung faktor ekspansi diperlukan variasi distribusi kelompok skor soal (U) dan variance distribusi kelompok skor siswa (V). Faktor ekspansi kemampuan peserta didik terhadap keluasan tes adalah $X = [(1 - U/2,89) / (1 - UV/8,35)]^{1/2}$ Faktor ekspansi kemampuan peserta didik terhadap penyebaran sampel adalah $X = [(1 + U/2,89) / (1 - UV/8,35)]^{1/2}$
- f. Menghitung tingkat kesukaran dan kesalahan standar butir soal
Dalam menghitung tingkat kesukaran dan kesalahan standar soal perlu disusun beberapa kolom di dalam tabel, yaitu kolom: (1) kelompok skor soal (1); (2) nomor soal; (3) inisial kalibrasi soal (d); (4) faktor ekspansi kesukaran soal terhadap penyebaran sampel (Y); (5) tingkat kesukaran soal atau $Yd_i = d_i$; (6) skor soal (S); (7) kesalahan standar kalibrasi soal yang dikoreksi $[SE(d_i)]$ atau $SE = [N/S_i (N-S_i)]^{1/2}$
- g. Menghitung tingkat kemampuan dan kesalahan standar siswa
Dalam menghitung tingkat kemampuan dan kesalahan standar siswa disusun beberapa kolom, yaitu kolom: (1) kemungkinan skor siswa (r); (2) initial pengukuran kemampuan siswa (br); (3) faktor ekspansi kemampuan siswa terhadap keluasan tes (X); (4) tingkat kemampuan siswa (br) atau (Xbr); (5) kesalahan standar pengukuran kemampuan siswa yang dikoreksi $[SE (br)]$ yaitu $X [L/r (L-r)]^{1/2}$; (6) peserta tes.
- h. Menghitung probabilitas atau peluang menjawab benar setiap butir soal $[P(0)]$.
Untuk menghitung peluang menjawab benar setiap butir pada model Rasch atau

model satu parameter digunakan rumus berikut ini.

$$P_i(O) = \frac{e^{X^o - b_i}}{1 + e^{D(O - b_i)}} \quad \text{atau} \quad P_i(O) = \frac{1}{1 + e^{D(E) - b_i}}$$

Estimasi data yang lebih teliti dan akurat hasilnya adalah menggunakan komputer seperti menggunakan program Bigsteps. Dalam program Bigsteps, estimasi data digunakan metode Approximation Maximum Likelihood (PROX) dan Unconditional Maximum Likelihood (UCON). Untuk menghasilkan hasil yang akurat, estimasi data dengan komputer dapat melakukan iterasi maksimum untuk metode PROX, misal bisa sampai 20

kali kemudian dilanjutkan dengan metode UCON sampai dengan 50 kali tergantung banyaknya data. Perbedaan hasil kalibrasi pada setiap iterasi semakin lama semakin kecil dan akan berhenti bila prosesnya sudah terpenuhi (converge) atau lebih kecil dari 0,01.

Kriteria data sesuai dengan model Rasch adalah apabila hasil korelasi point biserial tidak negatif dan outfitnya < 2 baik outfit butir soal maupun outfit orang. Hal ini menunjukkan bahwa data adalah fit dengan model. Maksudnya bahwa data soal sesuai dengan model Rasch atau valid yang memiliki mean= 0 dan SD=1. Metode pengujian fit tergantung pada jumlah butir soal dalam tes: (a) tes sangat pendek (10 atau beberapa butir), (b) tes pendek (11-20 butir), atau (c) tes panjang (>20 butir).

Outfit orang maksudnya statistik orang menunjukkan bagaimana perilaku yang tidak diharapkan pada butir soal yang mempunyai tingkat kesukaran jauh dengan kemampuan orang yang bersangkutan. Adapun Outfit butir maksudnya statistik butir soal menunjukkan bagaimana perilaku yang tidak diharapkan dari orang yang mempunyai kemampuan lebih dengan tingkat kesukaran butir yang bersangkutan.

Dalam pelaksanaannya, analisis secara IRT tidak serumit seperti penjelasan di atas. Pelaksanaannya sangat mudah dipahami oleh para guru karena dalam analisis digunakan program komputer, seperti program RASCAL, PASCAL, BIGSTEPS, atau QUEST. Untuk mengenal lebih jauh program-program ini, bacalah pada bab berikut.

IV. ANALISIS BUTIR SOAL DENGAN KALKULATOR

A. Pengertian

Analisis butir soal dengan kalkulator maksudnya adalah penelaahan butir soal secara kuantitatif yang penghitungannya menggunakan bantuan kalkulator. Kalkulator yang digunakan di dalam menganalisis data adalah kalkulator scientific atau kalkulator statistik, misalnya seperti CASIO fx - 3600P. Setiap kalkulator, khususnya kalkulator statistik, cara pengoperasiannya tergantung pada versinya masing-masing. Setiap versi memiliki ciri khusus dalam pengoperasiannya. Oleh karena itu, apabila para guru membeli kalkulator statistik pada versi terbaru, bacalah buku manualnya. Karena semua petunjuk pengoperasionalnya ada di dalamnya.

Sebagai pengenalan awal dalam buku ini, kalkulator yang digunakan untuk memberi penjelasan adalah menggunakan kalkulator "lama" yaitu CASIO fx-3600P. Adapun penggunaannya ada 4 aspek yang perlu diperhatikan, yaitu: (1) pembersihan data, (2) fungsi SD, (3) fungsi LR, (4) teknik merandom data.

B. Pembersihan Data

Sebelum kalkulator digunakan untuk menganalisis data sebaiknya data yang berada di dalam kalkulator perlu dibersihkan terlebih dahulu. Maksudnya agar hasil analisisnya tidak tercemari dengan data-data atau angka yang sudah digunakan di dalam kalkulator.

Cara pembersihannya adalah tekan tombol ON, INV, AC. Apabila masih belum bersih, tekanlah tombol MR, M+. Apabila masih belum bersih, tekanlah tombol MODE, . , INV, AC.

C. Fungsi SD

Fungsi SD merupakan perhitungan yang berhubungan dengan standard deviasi. Sebelum memulai memasukkan data, munculkanlah kata SD pada layar kalkulator. Caranya adalah dengan menekan tombol MODE, 3. Setelah di layar kalkulator muncul SD, maka langkah selanjutnya adalah memulai memasukkan data.

Caranya adalah memasukkan hanya skor siswa (55, 54, 51, 55, 53; tidak perlu memasukkan "nomor/nama siswa") seperti berikut.

No.	Siswa	Skor X	Tekan tombol
1.	A	55	RUN
2.	B	5	RUN
3.	C	51	RUN
4.	D	55	RUN
	E	53	RUN (Tampak di layar kalkulator 53)

Hasilnya adalah seperti berikut ini.

Menghitung	Tekan tombol	Tampak di layar kalkulator
- SD sampel	INV, 3	1.673320
- SD populasi	INV, 2	1.496662
- Mean	INV, 1	53.6
- Jumlah data	K OUT, 3	5.
- Jumlah skor	K OUT, 2	268
- Jumlah kuadrat skor	K OUT, 1	14376

D. Fungsi LR

Fungsi LR merupakan perhitungan yang berhubungan dengan Linier Regression. Sebelum memulai memasukkan data, munculkanlah kata LR pada layar kalkulator. Caranya adalah dengan menekan tombol **MODE**, 2. Setelah di layar kalkulator muncul LR, maka langkah selanjutnya adalah memulai memasukkan data. Caranya adalah memasukkan hanya skor siswa (tidak perlu memasukkan "nomor/nama siswa") seperti berikut.

No. Siswa	Skor X	Tekan tombol	Skor Y	Tekan tombol	
1.	A	55	[(...	75	RUN
2.	B	52	[(...	60	RUN
3.	C	54	[(...	66	RUN
4.	D	53	[(...	80	RUN
5.	E	53	[(...	85	RUN
6.	F	54	[(...	70	RUN

(Tampak di layar kalkulator 70.)

Hasilnya adalah seperti berikut ini.

Menghitung	Tekan tombol	Tampak di layar kalkulator
- Mean X	INV, 1	53.5
- SD sampel X	INV, 3	1.0488088
- SD populasi X	INV, 2	0.9574271
- Mean Y	INV, 4	72.66666
- SD sampel Y	INV, 6	9.201449
- SD populasi Y	INV, 5	8.399735
- Korelasi XY	INV, 9	0.165793
- A Constant in regression	INV, 7	-5.1515
- B Regression coefficients	INV, 8	1.4545
- Y	K OUT, 6	23334
- XY	K OUT, 1	17179
- $S X^{-1}$	K OUT, 2	321
- ZX	K OUT, 3	6
- Tn	K OUT, 4	32106
- \sqrt{Y}	K OUT, 5	436

E. Contoh Merandom data

Untuk merandom data, tekan tomhol INV dan tanda titik. Tampak di layar misalnya angka 0,425. Bila yang dirandom menggunakan satu digit, maka angka yang digunakan adalah satu angka setelah koma, yaitu angka 4. Bila dua digit yang digunakan adalah dua angka setelah koma, yaitu 42. Bila tiga digit angka yang digunakan adalah tiga angka setelah koma, yaitu 425.

Contoh misalnya merandom kunci jawaban butir soal untuk pilihan ganda. Kunci A= 1, B=2, C=3, D=4. Angka yang digunakan adalah hanya satu digit. Jadi berdasarkan hasil random dari kalkulator di atas, maka soal nomor 1 kunci jawabannya adalah D (karena angka 4= D). Kemudian ditekan tombol INV dan tanda titik lagi. Tampak di layar misalnya angka 0,184; maka kunci jawaban soal nomor 2 adalah A (karena angka 1= A). Ditekan tombol INV dan tanda titik lagi. Tampak di layar misalnya angka 0, 865. Angka ini tidak kita pergunakan karena batas angka yang dicari hanya sampai nomor 4, sedangkan yang muncul adalah nomor 8. Ditekan tombol INV dan tanda titik lagi dan seterusnya sampai selesai jumlah butir soalnya. Selamat mencoba!

F. Contoh Uji Validitas Butir Soal Bentuk Pilihan Ganda

Karena di dalam program kalkulator tidak tersedia uji validitas butir (korelasi point biserial) yaitu korelasi antara data nominal dan data kontinyu, maka kita perlu menghitungnya dengan menggunakan rumus seperti berikut ini.

$$r_{pbis} = \frac{\bar{X}_b - \bar{X}_s}{SD} \sqrt{pq}$$

Keterangan:

Xb: adalah rata-rata skor kemampuan peserta didik yang menjawab benar

Xs: adalah rata-rata skor kemampuan peserta didik yang menjawab salah

SD: adalah simpangan baku skor total

p : adalah proporsi jawaban benar terhadap semua jawaban siswa

q adalah 1-p

Caranya adalah ketiklah jawaban peserta didik/responden dengan menggunakan angka 1 (jawaban benar) dan 0 (jawaban salah).

V. ANALISIS BUTIR SOAL

A. Pengertian

Analisis butir soal dengan komputer maksudnya adalah penelaahan butir soal secara kuantitatif yang penghitungannya menggunakan bantuan program komputer. Analisis data dengan menggunakan program komputer adalah sangat tepat. Karena tingkat keakuratan hitungan dengan menggunakan program komputer lebih tinggi bila dibandingkan dengan diolah secara manual atau menggunakan kalkulator/ tangan. Program komputer yang digunakan untuk menganalisis data modelnya bermacam-macam tergantung tujuan dan maksud analisis yang diperlukan.

Program yang sudah dikenal secara umum adalah EXCEL, SPSS (Statistical Program for Social Science), atau program khusus seperti ITEMAN (analisis secara kiasik), RASCAL, ASCAL, BILOG (analisis secara item respon teori atau IRT), FACETS (analisis model Rasch untuk data kualitatif). Untuk memahami program-program komputer di atas, bacalah manual programnya secara saksama, kemudian praktikkan dengan menggunakan program komputer sebagai latihannya. Berikut ini akan disajikan contoh program analisis data dengan menggunakan komputer, seperti program ITEMAN, RASCAL, ASCAL, BIGSTEP, QUEST. Selamat berlatih!

B. ITEMAN

ITEMAN merupakan program komputer yang digunakan untuk menganalisis butir soal secara klasik. Program ini termasuk satu paket program dalam MicroCATⁿ yang dikembangkan oleh Assessment Systems Corporation mulai tahun 1982 dan mengalami revisi pada tahun 1984, 1986, 1988, dan 1993; mulai dari versi 2.00 sampai dengan versi 3.50. Alamatnya adalah Assessment Systems Corporation, 2233 University Avenue, Suite 400, St Paul, Minesota 55114, United States of America.

Program ini dapat digunakan untuk: (1) menganalisis data file (format ASCII) jawaban butir soal yang dihasilkan melalui manual entry data atau dari mesin scanner; (2) menskor dan menganalisis data soal pilihan ganda dan skala Likert untuk 30.000 siswa dan 250 butir soal; (3) menganalisis sebuah tes yang terdiri dari 10 skala (subtes) dan memberikan informasi tentang validitas setiap butir (daya pembeda, tingkat kesukaran, proporsi jawaban pada setiap option), reliabilitas (KR-20/Alpha), standar error of measurement, mean, variance, standar deviasi, skew, kurtosis untuk jumlah skor pada jawaban benar, skor minimum dan maksimum, skor median, dan frekuensi distribusi skor, Saat ini telah tersedia ITEMAN tiner Windows 95, 98, NT, 2000, ME, dan XP dengan harga \$299. Sebelum menggunakan program Iteman, bacalah manualnya/buku petunjuk pengoperasionalnya secara seksama. Sebagai contoh, tahap awal adalah membuat "file data" (control tile) yang berisi 5 komponen utama.

1. Baris pertama adalah baris pengontrol yang mendeskripsikan data.
2. Baris kedua adalah daftar kunci jawaban setiap butir soal.
3. Baris ketiga adalah daftar jumlah option untuk setiap butir soal.
4. Baris keempat adalah daftar butir soal yang hendak dianalisis (jika butir yang akan dianalisis diberi tanda Y (yes), jika tidak diikuti dalam analisis diberi tanda N (no)).
5. Baris kelima dan seterusnya adalah data siswa dan pilihan jawaban siswa.

Setiap pilihan jawaban siswa (untuk soal bentuk pilihan ganda) diketik dengan menggunakan huruf, misal ABCD atau angka 1234 untuk 4 pilihan jawaban atau ABCDE atau 12345 untuk 5 pilihan jawaban.

Cara menggunakan program ini, pertama data diketik di DOS atau Windows. Cara termudah adalah menggunakan program Windows yaitu dengan mengetik data di tempat Notepad. Caranya adalah klik Start-Programs-Accessories-Notepad.

Contoh pengetikan data untuk soal bentuk pilihan ganda

```
30 o n 6 [Jumlah soal, kode omit, kode tidak
dijawab, jmlh karakter]
43142442113424141324213411334 [Kunci jawaban dapat ditulis
dengan angka atau huruf]
444444444444444444444444444444 [Jumlah pilihan]
YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY [Soal yang
dianalisis, bila tidak dianalisis ditulis NJ]
Dita 123123244113424143324213211334 (Jawaban
siswa, dapat ditulis Fauria
423142243413424141124213111233 dengan angka atau
huruf) Fara
423142242113424141324213411334
Nafis 143142242433434141324413431334
Raufan 243142242413434141411213211134
Dina 423342224113423141421213044331
```

Contoh pengetikan data untuk skala Likert.

```
30 x Y 10 [Jumlah soal, kods omit, kode
tidak dijawab, jmlh karakter]
+++++++ --- +++++ ----+++++-- [Positif/negative pernyataan]
7777777777777777777777777777 [Jumlah pilihan]
77 [Kode skala]
Nurul 2112141232423434231112312437 [Jawaban siswa, dapat
Imam 3122142142424434232245623325 dengan angka atau
Ali 2242123313324431243254624371
Kiki 22421112X4324433232265566641
Chanan 32421424234244344322653546X3
```

Contoh lain pengetikan data untuk soal bentuk pilihan ganda

```

25 0 N 24
ABDCEBCEDAABEDCCBDBAEDCAB Kuncine
55555555555555555555555555555555 Pilihane
YYYYYYYYYYYYYYYYYYYYYYYYYYYYY
IWAN SUYAWAN          ABDCEBCEDAABEDCEADBAEEECB
TIKA HATIKAH          ACCEEBCDBAABEECBDBAEEAAB
YENNY SUKHRAINI       ABDDDBCEDAABCACCBDDBCDCAB
WIJI PURWANTA         ACBCEBCEDDCEEDCCAADAEDBBB
HENNY LISTIANA        ABDCECBDDAABDEACBDBBBECAB
UJANG HERMAWAN        CDDCEBCEDCDCEDCBCCADDCAE
NIKEN IRIANTI         CDDCEBACDAABEBBCBDBAADAAB
MIMIK RIATIN          ABDDDBCEDAABCACCBDDBCDCAB
NUR WAHYU RISDIANTO   ABDECDCEDAABBCDCBDBAAACAB
RURI SUSIYANTI        AEDEEBCEDBBDEDCCBDCBDCAB
RYSA DWI INDAH YATI   ABCDEBCEDAABCACCBDBDEBCAB
ANDRIKO               ACDCEBCECBCBEDCADABAEBBCB
JOKO SLAMET           AAAABBBCCCDDEEAABBCCDDEEA
LUKMAN NURHUDA        ACDBEBCECDBBEDCCBBAAEDCBB
OTAH PIANTO           DBBCEBAECAABDCBCBDBAEAEAB
AKHMAD SYAMSURIZAL    ADDCEBCEDCBCDDCCBDBEEDCAB
DENY TRI SETIAWAN     ABCDABCEDABCBDCCBDEAEDCAB
DEWI SETYOWATI        ACCBEBDCDBABEDBCEBDBDCBCAC
ISMAIL SHOLEH         ABDECDCEDAABBCDCBDBAAACAB
JEMI INTARYO          ACCEEBCDBAABEECBDBAEEAAB

```

Langkah kedua data yang telah diketik disimpan, misal disimpan pada file: **Tes1.txt**. Selanjutnya untuk menggunakan program Iteman yaitu dengan mengklik icon **Iteman**. Kemudian isilah pertanyaan-pertanyaan yang muncul di layar computer seperti berikut.

```

Enter the name of the input file: Tes1.txt <enter>
Enter the name of the output file: haltes1.txt <enter>
Do you want the scores written to a file? (Y/N) Y <enter>
Enter the name of the score file: scrtes1.txt <enter>
**ITEMAN ANALYSIS IS COMPLETE**

```

Langkah ketiga adalah membaca hasil, yaitu dengan mengklik icon **hs1tes1**. Hasilnya adalah seperti pada contoh berikut.

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

Item analysis for data from file tes1.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 CHECK THE KEY A was specified, C works better	0.850	-0.018	-0.012	A	0.850	-0.018	-0.012	*
					B	0.000	-9.000	-9.000	
					C	0.100	0.047	0.028	?
					D	0.050	-0.040	-0.019	
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
2	0-2	0.450	0.534	0.425	A	0.050	-1.000	-0.856	
					B	0.450	0.534	0.425	*
					C	0.300	-0.262	-0.199	
					D	0.150	0.231	0.151	
					E	0.050	0.121	0.057	
					Other	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
3	0-3	0.600	0.515	0.406	A	0.050	-1.000	-0.856	
					B	0.100	-0.142	-0.083	
					C	0.250	0.039	0.029	
					D	0.600	0.515	0.406	*
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
4	0-4 CHECK THE KEY C was specified, D works better	0.400	0.172	0.135	A	0.050	-1.000	-0.856	
					B	0.200	-0.059	-0.041	
					C	0.400	0.172	0.135	*
					D	0.200	0.474	0.332	?
					E	0.150	0.018	0.012	
					Other	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
5	0-5 CHECK THE KEY E was specified, D works better	0.700	0.215	0.163	A	0.050	0.281	0.133	
					B	0.050	-1.000	-0.856	
					C	0.100	0.142	0.083	
					D	0.100	0.331	0.194	?
					E	0.700	0.215	0.163	*
					Other	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
6	0-6 CHECK THE KEY B was specified, D works better	0.850	-0.089	-0.058	A	0.000	-9.000	-9.000	
					B	0.850	-0.089	-0.058	*
					C	0.050	-0.040	-0.019	
					D	0.100	0.142	0.083	?
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				Key
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	
25	0-25	0.850	1.000	0.685	A	0.050	-1.000	-0.856	
					B	0.850	1.000	0.685	*
					C	0.050	-0.523	-0.247	
					D	0.000	-9.000	-9.000	
					E	0.050	-0.040	-0.019	
					Other	0.000	-9.000	-9.000	

Keterangan :

Prop. Correct= tingkat kesukaran butir:,
 Biser dan Point Biser.= korelasi Biserial dan Korelasi Point
 Biserial,
 Alt.= alternative/pilihan jawaban,
 Prop. Endorsing= proporsi Jawaban pada setiap option

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 tes1.txt

Page 6

There were 20 examinees in the data file.

Scale Statistics

Scale:	0

N of Items	25
N of Examinees	20
Mean	16.250
Variance	9.087
Std. Dev.	3.015
Skew	-2.463
Kurtosis	6.976
Minimum	5.000
Maximum	20.000
Median	17.000
Alpha	0.437
SEM	2.261
Mean P	0.650
Mean Item-Tot.	0.266
Mean Biserial	0.352

Hasil scor butir soal pilihan ganda dari ITEMAN versi 3.00

24	1	Scores for examinees from file tes1.txt
IWAN SUYAWAN		20.00
TIKA HATIKAH		16.00
YENNY SUKHRAINI		18.00
WIJI PURWANTA		15.00
HENNY LISTIANA		16.00
UJANG HERMAWAN		16.00
NIKEN IRIANTI		17.00
MIMIK RIATIN		18.00
NUR WAHYU RISDIANTO		17.00
RURI SUSIYANTI		17.00
RYSA DWI INDAH YATI		19.00
ANDRIKO		15.00
JOKO SLAMET		5.00
LUKMAN NURHUDA		17.00
OTAH PIANTO		16.00
AKHMAD SYAMSURIZAL		19.00
DENY TRI SETIAWAN		18.00
DEWI SETYOWATI		13.00
ISMAIL SHOLEH		17.00
JEMI INTARYO		16.00

Hasil korelasi point-biserial (r_{pbi}) dan korelasi biserial (r_{pbis}) berasal dari perhitungan rumus berikut.

$$r_{pbi} = \frac{\bar{Y}_p - \bar{Y}_t}{St} \sqrt{\frac{p}{1-p}} \text{ atau } r_{bis} = \frac{\bar{Y}_p - \bar{Y}_t}{St} \sqrt{\frac{p(1-p)}{U}}$$

- \bar{Y}_p = mean skor pada kriterion siswa yang menjawab benar soal.
 \bar{Y}_t dan St = mean dan standard deviasi kriterion seluruh siswa.
 p = proporsi siswa yang menjawab benar soal.
 U = ordinat kurva normal.

Korelasi point-biserial (r_{pbi}) tidak sama dengan 0, korelasi biserial (r_{bis}) paling sedikit 25% lebih besar daripada r_{pbi} untuk perhitungan pada data yang sama. Korelasi point-biserial (r_{pbi}) merupakan korelasi product moment antara skor dikotomis dan pengukuran kriterion; sedangkan korelasi biserial (r_{bis}) merupakan korelasi product moment antara variabel latent distribusi normal berdasarkan dikotomi benar-salah dan pengukuran kriterion.

Menurut Millman dan Greene (1989) dalam *Educational Measurement*, kedua korelasi ini memiliki kelebihan masing-masing. Kelebihan korelasi point biserial adalah: (1) memberikan refleksi kontribusi soal secara sesungguhnya terhadap fungsi tes. Maksudnya ini mengukur bagaimana baiknya soal berkorelasi dengan kriterion (tidak bagaimana baiknya beberapasecara abstrak); (2) sederhana dan langsung berhubungan dengan statistik tes; (3) tidak pernah mempunyai value 1,00 karena hanya variabel-variabel dengan distribusi bentuk yang sama yang dapat berkorelasi secara sempurna, dan variabel kontinyu (kriterion) dan skor dikotomis tidak mempunyai bentuk yang sama. Kelebihan korelasi biserial adalah: (1) cenderung lebih stabil dari sampel ke sampel, (2) penilaian lebih akurat tentang bagaimana soal dapat diharapkan untuk membedakan pada beberapa perbedaan point di skala abilitas, (3) value r_{bis} yang sederhana lebih langsung berhubungan dengan indikator diskriminasi kurva karakteristik butir (Item Characteristic Curve atau ICC). Kebanyakan para ahli pendidikan, khususnya di Indonesia, banyak yang menggunakan korelasi point biserial daripada korelasi biserial.

Kriteria baik tidaknya butir soal menurut Ebel dan Frisbie (1991) dalam *Essentials of Educational Measurement* halaman 232 adalah bila korelasi point biserial: >0.40 =butir soal sangat baik; $0.30 - 0.39$ =soal baik, tetapi perlu perbaikan; $0.20 - 0.29$ =soal dengan beberapa catatan, biasanya diperlukan perbaikan; < 0.19 =soal jelek, dibuang, atau diperbaiki melalui revisi. Adapun tingkat kesukaran butir soal memiliki skala 0 - 1. Semakin mendekati 1 soal tergolong mudah dan mendekati 0 soal tergolong sukar.

C. EXCEL

Excel merupakan sebuah program pengolah data yang biasa dinamakan "spreadsheet". Karena program ini dapat digunakan untuk mengolah data yang berupa angka ataupun lainnya. Ada dua cara mengolah data dengan Excel, yaitu (1) melalui program bantu khusus perhitungan statistik dan (2) melalui fungsi statistik yang terdapat di dalam Excel.

Oleh karena itu tidak semua program Statistik ada di program Excel, seperti halnya Uji Validitas butir soal baik soal pilihan ganda maupun bentuk uraian, uji reliabilitas baik bentuk pilihan ganda, uraian maupun reliabilitas non-tes, dalam hal ini harus disain secara manual. Karena di dalam program ini tidak tersedia program tersebut.

D. SPSS (Statistical Program for Social Science)

SPSS merupakan sebuah program pengolah data yang sudah sangat dikenal di dalam dunia pendidikan. Penggunaannya sangat mudah untuk dipahami para guru di sekolah. Semua data diketik di dalam format SPSS yang sudah disediakan. Setelah selesai, kemudian tinggal memilih statistik yang akan digunakan pada menu STATISTIC/ANALYZE. Misalnya uji validitas butir atau reliabilitas tes, diklik pada menu ANALYZE kemudian pilih CORELATE, pilih BIVARIAT, untuk uji reliabilitas pilih RELIABILITY. Di samping itu, program ini dapat digunakan untuk analisis data kuantitatif secara umum, misalnya untuk uji normalitas, homogenitas, dan linearitas data.

Agar mudah pengoperasiannya dalam menggunakan program ini, sebaiknya para guru membaca terlebih dahulu manual/buku pedoman pengoperasiannya secara saksama. Berikut ini disajikan salah satu contoh penggunaan program SPSS yang digunakan untuk menguji uji normalitas, homogenitas, dan linearitas data, serta uji kesesuaian antara butir soal dan kisi-kisinya (analisis faktor). Program SPSS selama ini sudah diproduksi beberapa versi, diantaranya versi 11, 12, maupun versi 13. Untuk lebih jelasnya perhatikan contoh pengetikan data dan analisisnya berikut ini.

Motivasi Belajar (X)	Prestasibelajar (Y)	Jenis Kelamin
60	65	1
61	68	2
75	85	1
70	76	2
60	65	1
80	89	2
70	74	1
60	62	2
79	81	1
69	75	2

Setelah program SPSS dibuka, data di atas di masukkan ke dalam format SPSS. Caranya sangat mudah yaitu seperti berikut.

1. Klik "Variable View" (letaknya di sebelah kiri bawah).
2. Ketik X pada kolom "Name".
3. Klik pada kolom "Label" kemudian ketik Motivasi Belajar.
4. Ketik Y pada kolom "Name" (di bawah X).
5. Klik pada kolom "Label" kemudian ketik Prestasi Belajar.
6. Ketik JK pada kolom "Name" (di bawah Y)
7. Klik pada kolom "Label" kemudian ketik Jenis Kelamin.
8. Klik pada kolom "Scale" kemudian klik pada "Nominal".

68.00	1	10.0	10.0	40.0
74.00	1	20.0	20.0	50.0
75.00	1	10.0	10.0	60.0
76.00	1	10.0	10.0	70.0
81.00	1	10.0	10.0	80.0
85.00	1	10.0	10.0	90.0
89.00	1	10.0	10.0	100.0

Total 10 100.0 100.0

Jenis Kelamin

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1.00	5	50.0	50.0	50.0
2.00	5	50.0	50.0	100.0

Total 10 100.0 100.0

b. Cara kedua

Analyze

Descriptive statistics

Descriptives

- Semua variable dimasukkan ke dalam kotak "Variables"
- Klik: "Options"
- Klik: - mean, sum
 - std deviation, variance, range, minimum, maximum, S.E. mean
 - kurtosis, skewness
 - Ascending means
- Klik: "Continue"
- Klik: "OK"

Descriptive Statistic

	N Statistic	Range Statistic	Minimum Statistic	Maximum Statistic	Sum Statistic
Motivasi belajar	10	20.00	60.0	80.00	684.00
Prestasi belajar	10	27.00	62.0	89.00	740.00
Jenis Kelamain	10	1.00	1.00	2.00	15.00
Valid N (listwise)					

Descriptive Statistic

	Mean		Std. Statistic	Variance Statistic
	Statistic	Std. error		
Motivasi belajar	68.4000	2.4998	7.90499	62.489
Prestasi belajar	74.0000	2.8713	9.07989	82.444
Jenis Kelamain	1.5000	.1667	.52705	.278
Valid N (listwise)	10			

Descriptive Statistic

	Skewness		Kurtosis	
	Statistic	Std. error	Statistic	Std. error
Motivasi belajar	.243	.687	-1.512	1.334
Prestasi belajar	.307	.687	-1.037	1.334
Jenis Kelamain	.000	.687	-2.571	1.334
Valid N (listwise)	10	.587		

2. Uji Persyaratan Analisis

a. Contoh Uji Normalitas

Analyze
 Descriptive statistics
 Explore

- Variabel X dan Y dimasukkan ke dalam kotak "Dependent List:"
- Klik kotak "Plot" kemudian klik pada "Normality plots with tests".
- Klik "Continue"
- Klik "OK"

Rumusan hipotesis.

H0 : sample berasal dari populasi berdistribusi normal.

H1 : sample tidak berasal dari populasi berdistribusi normal.

Kaidah penetapan:

- Jika signifikan > 0,05, sampel berasal dari populasi berdistribusi normal.
- Jika signifikan < 0,05, sampel tidak berasal dari populasi berdistribusi normal.

b. Uji Homogenitas

Analyze
 Descriptive statistics
 Explore

- Variabel X dan Y dimasukkan ke dalam kotak "Dependent List."
- Variabel jenis kelamin dimasukkan ke dalam kotak ":Factor List:"
- Klik kotak "Plot" kemudian klik pada "Normality plots with tests" dan "Untransformed"
- Klik "Continue"
- Klik "OK"

Rumusan hipotesis:

H0: variansi pada setiap kelompok sama (homogen).

H1 : variansi pada setiap kelompok tidak sama (tidak homogen)..

Kaidah penetapan:

- Jika signifikan $> 0,05$, variansi setiap sampel sama (homogen).

c. Contoh Uji Linearitas

Analyze

Compare Means

Means

- Variabel X dimasukkan ke dalam kotak "Dependent List:"
- Variabel Y dimasukkan ke dalam kotak "Independent List:"
- Klik kotak "Option" kemudian klik pada "Anova table and eta" dan "Test for linearity"
- Klik "Continue"
- Klik "OK"

Rumusan hipotesis:

H0: Linearitas tidak dipenuhi.

H1: Linearitas dipenuhi.

Kaidah penetapan:

- Jika signifikan $> 0,05$, linearitas tidak dipenuhi.
- Jika signifikan $< 0,05$, linearitas dipenuhi.
- Jika signifikan $< 0,05$, variansi setiap sampel tidak sama (tidak homogen).

3. Contoh Uji Perbedaan dengan t-tes

Analyze

Compare Means

Independent-Sample T Test

- Variabel Y dimasukkan ke kotak "Test Variables"
- Variabel jenis kelamin dimasukkan ke kotak "Grouping variable"
- Klik "Define Groups" kemudian ketik 1 pada Group 1 dan ketik 2 pada Group 2. -Klik "Continue"
- Klik "OK"

Rumusan hipotesis:

H0 :tidak terdapat perbedaan antara variable X dan variable Y...

H1 :terdapat perbedaan antara variable X dan variable Y ...

Kaidah penetapan:

- Jika signifikan $> 0,05$, H0 diterima.
- Jika signifikan $< 0,05$, H0 ditolak.

4. Contoh Uji Perbedaan/Pengaruh dengan ANOVA

Analyze

Compare means

One-way ANOVA

- Variabel Y (pada eksperimen dan control) dimasukkan ke dalam "Dependent List:"
- Variabel jenis kelamin dimasukkan ke dalam "Factor:"
- Klik "Options" kemudian klik "Homogeneity of variance test".
- Klik "Continue"
- Klik "OK"

Rumusan hipotesis:

H0: tidak terdapat perbedaan/pengaruh antara variable X dan variable Y

H1: terdapat perbedaan/pengaruh antara variable X dan variable Y

Kaidah penetapan:

- Jika signifikan $> 0,05$, H0 diterima.
- Jika signifikan $< 0,05$, H0 ditolak.

5. Contoh Uji Hubungan dengan Korelasi

Analyze

Correlate

Bivariate

- Variabel X dan Y dimasukkan ke dalam kotak "Variables"
- Klik "Pearson" "Two-Tailed"
- Klik "Options" kemudian klik "means and standard deviations"
- Klik "Continue"
- Klik "OK"

Rumusan hipotesis:

H0 : tidak terdapat hubungan antara variable X dan variable Y.

H1 : terdapat hubungan antara variable X dan variable Y.

Kaidah penetapan:

- Jika signifikan $> 0,05$, H0 diterima.
- Jika signifikan $< 0,05$, H0 ditolak.

6. Contoh Uji Hubungan dengan Regresi Linear

Analyze

Regression

Linear

- Variabel Y dimasukkan ke kotak "Dependent"
- Variabel X dimasukkan ke kotak "Independents"
- Klik "Statistics" kemudian klik "estimates", "model fit", dan klik "Continue".
- Klik "OK"

Rumusan hipotesis:

H0 : tidak terdapat hubungan antara variable X dan variable Y.

H1 : terdapat hubungan antara variable X dan variable Y.

Kaidah penetapan:

- Jika signifikan $> 0,05$, H0 diterima.
- Jika signifikan $< 0,05$, H0 ditolak.

7. Uji Kesesuaian antara Butir Soal dan Kisi-kisinya (Uji Konstruktif dengan Analisis Faktor)

a. Analisis Faktor Eksploratori

Kegiatan memvalidasi konstruk dilaksanakan setelah tes digunakan/diuji coba. Analisis faktor terdiri dari dua yaitu analisis faktor eksploratori dan konfirmatori. Analisis faktor konfirmatori menekankan pada estimasi parameter dan tes hipotesis, sedangkan analisis faktor eksploratori menekankan pada beberapa faktor yang menjelaskan hubungan antar-indikator dan estimasi muatan faktor.

Untuk menguji validitas kesesuaian antara butir soal dan kisi-kisi konstruksinya digunakan analisis faktor. Konsep validitas berhubungan dengan: (1) ketepatan, (2) kebermaknaan, dan (3) kegunaan suatu skor tes (Gable, 1986: 71). Macam-macam validitas adalah validitas: (1) konten yang meliputi: definisi konsep dan definisi operasional; (2) konstruk, dan (3) kriterion-related (Gable, 1986: 72-77). Terdapat empat teknik untuk menganalisis konstruk, yaitu dengan: (1) korelasi antarvariabel, (2) analisis multitrait multimethod, (3) analisis faktor, dan (4) prosedur known-groups (Gable, 1986: 77).

Analisis faktor dikembangkan oleh Charles Spearman tahun 1904 di USA (Harman, 1976: 3). Analisis faktor adalah suatu nama generik yang diberikan pada suatu kelas metode statistik multivariat yang tujuan utamanya adalah Untuk mendefinisikan struktur dalam matriks data (Hair et. al, 1998: 90). Tujuan utama analisis faktor adalah untuk menguji secara empirik hubungan antar butir soal dan untuk menentukan kelompok soal yang saling menentukan sebagai suatu faktor/konstruk yang diukur melalui instrumen (Gable, 1986: 85). Jadi tujuan utamanya dapat disimpulkan menjadi 3, yaitu untuk menentukan: (1) faktor umum yang diperlukan terhadap jumlah patern korelasi antar semua pasangan tes dalam satu set tes; (2) faktor umum sesungguhnya (asli) yang menghitung untuk tes interkorelasi; (3) proporsi varian untuk suatu variabel observasi yang dihubungkan dengan varian faktor umum (Crocker and Algina, 1986: 305-306) atau sebagai pengenalan struktur melalui peringkasan data atau reduksi/pengurangan data (Hair et al., 1998: 95).

Adapun manfaat analisis faktor adalah: (1) memberitahu kita tes-tes dan ukuran-ukuran yang saling dapat serasi atau sama tujuannya dan sejauhmana kesamaannya, (2) membantu menemukan dan mengidentifikasi kebutuhan-kebutuhan atau sifat-sifat fundamental yang melandasi tes dan pengukuran (Kerlinger, 1993: 1000).

Langkah atau prosedur penggunaan analisis factor eksploratori selalu memproses melalui 4 tahap, yaitu: (1) perhitungan korelasi matriks untuk semua variabel, (2) ekstraksi faktor untuk menentukan jumlah faktor, (3) rotasi, untuk membuat faktor lebih bermakna, dan (4) perhitungan skor setiap faktor untuk setiap case.

Cara pengoperasional dalam program SPSS adalah seperti berikut.

Pilih menu STATISTIC atau ANALYZE
DATA REDUCTION
FACTOR

Pada boks dialog variabel yang akan dianalisis dimasukkan ke kotak VARIABLES. Klik pada kotak DESCRIPTIVE (misal: klik "initial solution" pada kolom statistics dan "KMO and Bartlett's test of sphericity" pada kolom correlation Matrix), EXTRACTION, ROTATION, SCORES, atau OPTION. Hasil print outnya terdiri dari beberapa tabel dan sebuah grafik "scree plot".

Berikut ini dijelaskan beberapa hasil print out analisis faktor eksploratori dan penafsirannya.

(1) Statistik Deskriptif

Dalam tabel statistik deskriptif berisi informasi yang bersifat deskriptif seperti mean dan standard deviasi setiap variabel. Jika besarnya mean variabel sangat dekat/ekstrem pada skala jawaban dan standar deviasinya rendah, maka korelasi antarvariabel akan rendah dan berakibat rendah pula pada hasil analisis faktor (Gabel, 1986:91).

(2) Bartlett test of sphericity

Tes ini digunakan untuk mengetes hipotesis yang korelasi matriknya merupakan suatu matriks identitas, yaitu semua diagonal adalah 1 dan semua yang tidak diagonal (off-diagonal) adalah 0. Hasil tes menunjukkan bahwa sample data berasal dari suatu populasi normal multivariat atau tidak. Jadi bila nilai tes statistik dari sphericity luas/tinggi dan level signifikannya kecil, maka dapat dikatakan bahwa matriks korelasi populasi adalah signifikan (Norusis, 1993:50).

(3) Pengukuran Sampling Kaiser Meyer Olkin (KMO)

KMO merupakan suatu indeks perbandingan besarnya koefisien korelasi observed dan besarnya koefisien korelasi parsial. Jika jumlah kuadrat korelasi parsial pada semua pasangan variabel adalah kecil bila dibandingkan dengan jumlah kuadrat koefisien korelasinya, maka besar KMO mendekati 1. Jika besar KMO kecil atau rendah maka hasil analisis faktornya adalah tidak baik.

Kaiser (1974) dalam Norusis (1993: 52) mengklasifikasi tentang besarnya KMO adalah bila besarnya 0,90 bagus sekali (marvelous), 0,80 bermanfaat (meritorious), 0,70 sedang/cukup (middling), 0,60 sedikit cukup (mediocre), 0,50 gawat/menyedihkan (miserable), dan di bawah 0,50 tidak dapat diterima (unacceptable).

(4) Matriks Korelasi antarbutir

Korelasi antarbutir menunjukkan adanya beberapa butir yang saling berhubungan secara wajar. Jika korelasi antarvariabel adalah kecil, maka variabel-variabel itu berhubungan dengan faktor-faktor secara umum (share common factors) (Norusis, 1993:50).

(5) Matriks Korelasi Anti-image

Matrik ini berisi korelasi anti-image, maksudnya adalah koefisien korelasi parsial yang negatif. Jika proporsi untuk koefisien yang banyak adalah tinggi, maka kita dipersilakan untuk mempertimbangkan kembali tepat atau tidak menggunakan analisis faktor.

(6) Ekstraksi Faktor

Ekstraksi merupakan hubungan antara faktor-faktor dan variabel individu. Tujuan utama ekstraksi faktor adalah untuk menentukan jumlah faktor. Beberapa jumlah faktor yang diperlukan untuk merepresen data. Hal ini sangat membantu dalam menguji persentase total varian (eigenvalues) untuk masing-masing faktor. Total varian merupakan jumlah varian masing-masing variabel. Di samping itu, untuk menentukan jumlah faktor dapat dilihat pada "scree test" atau "scree plot" Dari tes atau plot itu dapat diketahui jumlah faktor yang ditunjukkan dengan beberapa garis yang panjang dan curam serta diikuti dengan jumlah garis yang pendek-pendek.

(7) Residuals

Keterangan residu terdapat di bawah matrik koefisien korelasi estimate. Jika residu lebih besar dari 0,05 adalah residunya luas. Artinya model tidak fit dengan data dan data perlu diperbaiki (Norusis, 1993:59).

(8) Rotasi

Rotasi analisis faktor adalah membantu lebih mudah untuk menginterpretasikan data. Tujuan rotasi adalah untuk menentukan suatu struktur sederhana. Artinya di setiap faktor tidak dikehendaki adanya nilai nol pada faktor loading untuk setiap variabel. Rotasi tidak berpengaruh pada fitnya faktor. Rotasi mendistribusikan

kembali penjelasan varian untuk faktor individu.

Adapun metode rotasi dapat digunakan sesuai dengan tujuan, yaitu orthogonal seperti: varimax, equamax, quartimax, atau oblique seperti direct oblimin.

Thurstone dalam Kerlinger (1993: 1019-1020) memberikan panduan dalam melakukan rotasi, yaitu menetapkan 5 prinsip atau struktur sederhana yang berlaku untuk rotasi yang tegak lurus (ortogonal atau sudut 90 derajat) maupun yang tidak/miring (jika sudut yang dibentuk oleh dua sumbu merupakan sudut lancip/ tumpul). Prinsip-prinsip struktur sederhana yang dimaksud adalah: (1) setiap larik dari matriks faktor harus setidaknya-tidaknya memiliki satu muatan yang mendekati nol; (2) untuk setiap kolom pada matriks faktor harus terdapat setidaknya-tidaknya variabel bermuatan nol atau mendekati nol yang sama banyaknya dengan banyaknya faktor; (3) untuk setiap pasangan faktor (kolom) harus terdapat sejumlah variabel yang mempunyai muatan pada satu faktor (kolom) tetapi tidak bermuatan pada faktor lainnya; (4) kalau ada empat faktor atau lebih, sebagian besar dari variabel-variabel itu harus memiliki muatan yang dapat diabaikan (mendekati nol) pada sebarang pasangan faktor; (5) untuk setiap pasangan faktor (kolom) pada matriks faktor itu harus ada hanya kolom sekaligus. Kriteria ini menghendaki sebanyak mungkin variabel "murni" yakni setiap variabel memuat sedikit mungkin faktor dan nol yang sebanyak mungkin dalam matriks faktor yang dirotasi (Kerlinger, 1933: 1021).

- Aiken, Lewis R. (1994). Psychological Testing and Assessment, (Eight Edition), Boston: Allyn and Bacon.
- Anastasi, Anne and Urbina, Susana. (1997). Psicohological Testing. (Seventh Edition). New Jersey: Prentice-Hall, Inc.
- Assessment Systems Corporation. (1984). User's Manual for the MiicroCat Testing System, USA.
- Atkinson, John W. (1978). Personality Motivation and Achievemcnt. Sashington. Hemisphere Publishing Corporation.
- Bejar, Isaac I. (1983). Introduction to Item Response Theory and Their Assumptions. Hambleton, Ronald K. (Editor). Applications of Item Response Theory. Canada: Educational Research Institute of British Columbia.
- Bruning, James L. and Kintz, B. L. (1987). Computational Handbook of Statistics. Third Edition. Illinois: Scott, Foresman and Company.
- Crocker, L. & Algina, J. (1986). Introduction to Classical and Modern Test, Theory. New York: Holt, Rinehart and Winston, Inc.
- Ebel, Robert L, and Frisbie, David A. *Essentials of Educatiornul Measurement*. New Jersey: Prentice Hall, 1991.
- Gable, Robert K. (1986). Instrument Development in the Affective Domain Boston: Kluwer-Nijhoff Publishing.
- Glass, Gene V. and Stanley, Julian C. (1970). Statistical Methods in Education and Psychology. New Jersey: Prentice Hall, Inc.
- Hair, J. F.; Anderson, R. E., Tatham, R. L., and Black, W. C. (1998). Multivariate Data, Analysis. New Jersey. Prentice-Hall International, Inc.
- Haladyna, Thomas M. (1994). Developing and Validating Multiple-Choice Test Items. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Hambleton, Ronald K (1993). Principles and Selected Applications of Item Response Theory. In Linn, Robert L. (Editor). Educational Measurement. Third Edition. Phoenix: American Council on Education, Series on Higher Education Oryx Press.