

## ANALISIS BUTIR SOAL

### A. PENDAHULUAN

Tes adalah suatu pernyataan, tugas atau seperangkat tugas yang direncanakan untuk memperoleh informasi tentang trait atau atribut pendidikan dan psikologi. Setiap butir pertanyaan atau tugas tersebut mempunyai jawaban atau ketentuan yang dianggap benar. Tes dapat diklasifikasikan menurut bentuk, tipe dan ragamnya (Asmawi Zainul, dkk :1997).

Pengukuran adalah pemberian angka kepada suatu atribut atau karakteristik tertentu yang dimiliki oleh orang, hal atau obyek tertentu menurut aturan atau formulasi yang jelas. Karakteristik dari pengukuran adalah penggunaan angka atau skala tertentu dan menggunakan aturan atau formula tertentu (Asmawi Zainul, dkk :1997).

Penilaian adalah suatu proses untuk mengambil keputusan dengan menggunakan informasi yang diperoleh melalui pengukuran hasil belajar, baik yang menggunakan instrumen tes atau non tes. Dengan kata lain, penilaian adalah pemberian nilai terhadap kualitas sesuatu.

Keterkaitan antara tes, pengukuran dan penilaian adalah penilaian hasil belajar baru dapat dilakukan dengan baik dan benar bila menggunakan informasi yang diperoleh melalui pengukuran hasil belajar yang menggunakan tes sebagai alat ukurnya. Kegunaan tes, pengukuran dan penilaian dalam pendidikan antara lain adalah untuk seleksi, penempatan, diagnosa, remedial, umpan balik, memotivasi dan membimbing, perbaikan kurikulum, program pendidikan serta pengembangan ilmu.

Perencanaan dalam pengujian sangat penting karena tes baru akan berarti bila terdiri dari butir-butir soal yang menguji tujuan yang penting dan mewakili ranah pengetahuan, kemampuan dan keterampilan secara representatif. Ada enam hal yang perlu dipertimbangkan dalam perencanaan tes yaitu: *pengambilan sampel dan pemilihan butir soal, tipe tes yang akan digunakan, aspek yang akan diuji, format butir soal, jumlah butir soal dan distribusi tingkat kesukaran butir soal* (Asmawi Zainul, dkk :1997).

Kelemahan butir soal tidak terletak pada bentuk atau tipe butir soal, tetapi lebih banyak ditentukan oleh butir soal yang dikonstruksi dengan baik atau tidak baik. Butir soal obyektif akan sama baiknya dengan butir soal uraian untuk mengukur keberhasilan belajar yang dikonstruksi secara baik. Bahkan dalam beberapa hal butir soal uraian jauh lebih besar resikonya daripada butir soal obyektif. Hal ini disebabkan mutu butir soal uraian tidak hanya terletak pada kemampuan siswa untuk menjawab soal tersebut, tetapi lebih banyak ditentukan oleh kemampuan dan obyektifitas pembuat soal dalam memberikan skor pada hasil tes tersebut.

Butir soal obyektif dapat dianalisa secara lebih akurat dan bertanggung jawab sehingga dapat diketahui kelemahannya secara tepat. Butir soal tes obyektif dapat digunakan berulang-ulang, asalkan tidak dalam perangkat tes yang sama. Oleh karena itu ada manfaat atau kegunaan analisis butir soal, kemudian direvisi sehingga butir soal yang kurang baik konstruksinya dapat diperbaiki. Akhirnya akan diperoleh butir soal yang telah teruji dan secara akurat mengukur hasil belajar yang ingin diukur.

Ada beberapa alasan mengapa diperlukan analisis butir soal. Menurut (Asmawi Zainul, dkk :1997) alasan tersebut antara lain :

- a. Untuk mengetahui kekuatan dan kelemahan butir tes, sehingga dapat dilakukan seleksi dan revisi butir soal.
- b. Untuk menyediakan informasi tentang spesifikasi butir soal secara lengkap, sehingga akan lebih memudahkan bagi pembuat soal dalam menyusun perangkat soal yang akan memenuhi kebutuhan ujian dalam bidang dan tingkat tertentu.
- c. Untuk segera dapat mengetahui masalah yang terkandung dalam butir soal, seperti: kemenduaan butir soal, kesalahan meletakkan kunci jawaban, soal yang terlalu sukar dan terlalu mudah, atau soal yang mempunyai daya beda rendah. Masalah ini bila diketahui dengan segera akan memungkinkan bagi pembuat soal untuk mengambil keputusan apakah butir soal yang bermasalah itu akan digugurkan atau direvisi guna menentukan nilai peserta didik.
- d. Untuk dijadikan alat guna menilai butir soal yang akan disimpan dalam kumpulan soal.

- e. Untuk memperoleh informasi tentang butir soal sehingga memungkinkan untuk menyusun beberapa perangkat soal yang paralel. Penyusunan perangkat seperti ini sangat bermanfaat bila akan melakukan ujian ulang atau mengukur kemampuan beberapa kelompok peserta tes dalam waktu yang berbeda.

## **B. PEMBAHASAN**

### **1. Deskripsi Analisis Butir Soal**

Penilaian terhadap butir soal pada dasarnya merupakan analisis butir soal, dan selama ini pada umumnya para ahli pengukuran mengatakan bahwa analisis butir soal maksudnya adalah penilaian terhadap soal. Telah diketahui bersama bahwa penyusunan tes sangat mempengaruhi kualitas butir soal.

Pendekatan untuk menganalisis butir soal yang berkembang saat ini terdiri dari dua pendekatan yaitu pendekatan klasik dan pendekatan modern. Kedua pendekatan ini masing-masing memiliki kelebihan dan kekurangan. Namun keduanya masih sering digunakan dalam analisis butir soal. Analisis butir soal dengan pendekatan klasik diantaranya dapat dilakukan menggunakan **Program Iteman**.

Dengan melihat beberapa persyaratan yang harus dipenuhi, penyusunan tes dituntut untuk mengikuti pedoman penyusunan tes dan melakukan ujicoba. Kemudian berdasarkan hasil ujicoba, respon peserta dianalisis menggunakan Program Iteman untuk mendapatkan karakteristik butir soal.

Data hasil analisis dengan Program Iteman dianalisis kembali menggunakan instrumen penilaian butir soal yang memenuhi syarat sebagai alat ukur yang baik. Suryabrata (1999) menyatakan bahwa analisis butir soal mencakup telaah soal atau analisis kualitatif dan analisis terhadap data empirik hasil ujicoba atau analisis kuantitatif.

Analisis butir soal secara kualitatif menekankan penilaian dari ketiga segi yaitu materi, konstruksi, dan bahasa. Namun demikian dalam pembahasan ini dikhususkan untuk menjelaskan analisis butir soal secara kuantitatif.

Analisis ini dilakukan berdasarkan data yang diperoleh secara empiris melalui ujicoba dari suatu perangkat tes. Analisis kuantitatif sering disebut dengan analisis item yang menghasilkan karakteristik atau parameter butir dan tes, yaitu: tingkat kesukaran, daya beda dan distribusi jawaban dan kunci setiap butir, serta reliabilitas dan kesalahan pengukuran (SEM) dalam tes.

Telah disinggung di depan bahwa analisis soal antara lain bertujuan untuk mengadakan identifikasi soal-soal yang baik, kurang lebih atau sedang dan soal yang tidak baik. Dengan analisis soal dapat diperoleh informasi tentang kekurangan sebuah soal tes dan “petunjuk” untuk mengadakan perbaikan.

Dalam tes dan pengukuran, dikenal beberapa karakteristik butir soal. Untuk tes hasil belajar pada umumnya dipertimbangkan tiga karakteristik butir soal, yaitu : tingkat kesukaran, daya beda dan distribusi jawaban atau berfungsi tidaknya pilihan jawaban (distraktor). Ketiga karakteristik butir soal ini secara bersama-sama akan menentukan mutu butir soal. Bila salah satu dari ketiga karakteristik ini tidak memenuhi persyaratan maka mutu butir soal akan turun.

## a. Karakteristik Butir Soal

### 1). Tingkat Kesukaran (*Difficulty level*)

Menurut Asmawi Zainul, dkk (1997) tingkat kesukaran butir soal adalah proporsi peserta tes menjawab benar terhadap butir soal tersebut. Tingkat kesukaran butir soal biasanya dilambangkan dengan  $p$ . Makin besar nilai  $p$  yang berarti makin besar proporsi yang menjawab benar terhadap butir soal tersebut, makin rendah tingkat kesukaran butir soal itu. Hal ini mengandung arti bahwa soal itu makin mudah, demikian pula sebaliknya.

Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang mahasiswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan mahasiswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya (Suharsimi Arikunto : 2001).

Tingkat kesukaran butir soal tidaklah menunjukkan bahwa butir soal itu baik atau tidak. Tingkat kesukaran butir hanya menunjukkan bahwa butir soal itu sukar atau mudah untuk kelompok peserta tes tertentu. Butir soal hasil belajar yang terlalu sukar atau terlalu mudah tidak banyak memberi informasi tentang butir soal atau peserta tes (Asmawi Zainul, dkk : 1997).

Pada analisis butir soal secara klasikal, seperti yang dijelaskan oleh Depdikbud (1997) tingkat kesukaran dapat diperoleh dengan beberapa cara antara lain : a). skala kesukaran linier; b). skala bivariat; c). indeks davis; d). proporsi menjawab benar.

Cara yang paling umum digunakan adalah proporsi menjawab benar atau *proportion correct*, yaitu jumlah peserta tes yang menjawab benar pada soal yang dianalisis dibandingkan dengan peserta tes seluruhnya. Dalam analisis item ini digunakan *proportion correct* ( $p$ ), untuk menilai tingkat kesukaran butir soal, yang dapat dilihat berdasarkan hasil analisis iteman pada lampiran 6.

Besarnya tingkat kesukaran berkisar antara 0,00 sampai 1,00. Untuk sederhananya, tingkat kesukaran butir dan perangkat soal dapat dibagi menjadi tiga kelompok, yaitu mudah, sedang dan sukar. Sebagai patokan menurut (Asmawi Zainul, dkk : 1997) dapat digunakan tabel sebagai berikut:

**Tabel 1. Klasifikasi Tingkat Kesukaran Butir Soal**

Tingkat Kesukaran	Nilai p
Sukar	0,00 – 0,25
Sedang	0,26 – 0,75
Mudah	0,76 – 1,00

Untuk menyusun suatu naskah ujian sebaiknya digunakan butir soal yang mempunyai tingkat kesukaran berimbang, yaitu : soal berkategori sukar sebanyak 25%, kategori sedang 50% dan kategori mudah 25%.

Dalam penggunaan butir soal dengan komposisi seperti di atas, maka dapat diterapkan penilaian berdasar acuan norma atau acuan patokan. Bila komposisi butir soal dalam suatu naskah ujian tidak berimbang, maka penggunaan penilaian acuan norma tidaklah tepat, karena informasi kemampuan yang dihasilkan tidaklah akan berdistribusi normal.

Walaupun demikian ada yang berpendapat bahwa soal-soal yang dianggap baik adalah soal-soal yang sedang, yaitu soal-soal yang mempunyai indeks kesukaran berkisar antara 0,26 – 0,75. Berbagai kriteria tersebut mempunyai kecenderungan bahwa butir soal yang memiliki indeks kesukaran kurang dari 0,25 dan lebih dari 0,75 sebaiknya dihindari atau tidak digunakan, karena butir soal yang demikian terlalu sukar atau terlalu mudah, sehingga kurang mencerminkan alat ukur yang baik.

Namun demikian menurut Suharsimi Arikunto (2001) soal-soal yang terlalu mudah atau terlalu sukar tidak berarti *tidak boleh digunakan*. Hal ini tergantung dari tujuan penggunaannya. Jika dari peserta tes banyak, padahal yang dikehendaki lulus hanya sedikit maka diambil peserta yang terbaik, untuk itu diambilkan butir soal tes yang sukar. Demikian sebaliknya jika kekurangan peserta tes, maka dipilihkan soal-soal yang mudah. Selain itu, soal-soal yang sukar akan menambah motivasi belajar bagi siswa-siswa yang pandai, sedangkan soal-soal yang mudah akan membangkitkan semangat kepada siswa yang lemah.

## 2). Daya beda

Daya beda butir soal ialah indeks yang menunjukkan tingkat kemampuan butir soal membedakan kelompok yang berprestasi tinggi (kelompok atas) dari kelompok yang berprestasi rendah (kelompok bawah) diantara para peserta tes (Asmawi Zainul, dkk : 1997). Suryabrata (1999) menyatakan tujuan pokok mencari daya beda adalah untuk menentukan apakah butir soal tersebut memiliki kemampuan membedakan kelompok dalam aspek yang diukur, sesuai dengan perbedaan yang ada pada kelompok itu.

Daya beda butir soal yang sering digunakan dalam tes hasil belajar adalah dengan menggunakan indeks korelasi antara skor butir dengan skor totalnya. Daya beda dengan cara ini sering disebut validitas internal, karena nilai korelasi diperoleh dari dalam tes itu sendiri. Daya beda dapat dilihat dari besarnya koefisien korelasi biserial maupun koefisien korelasi point biserial.

Dalam analisis ini digunakan nilai koefisien korelasi biserial untuk menentukan daya beda butir soal. Koefisien korelasi biserial menunjukkan hubungan antara dua skor, yaitu skor butir soal dan skor keseluruhan dari peserta tes yang sama.

Koefisien daya beda berkisar antara  $-1,00$  sampai dengan  $+1,00$ . Daya beda  $+1,00$  berarti bahwa semua anggota kelompok atas menjawab benar terhadap butir soal itu, sedangkan kelompok bawah seluruhnya menjawab salah terhadap butir soal itu. Sebaliknya daya beda  $-1,00$  berarti bahwa semua anggota kelompok atas menjawab salah butir soal itu, sedangkan kelompok bawah seluruhnya menjawab benar terhadap soal itu.

Daya beda yang dianggap masih memadai untuk sebutir soal ialah apabila sama atau lebih besar dari  $+0,30$ . Bila lebih kecil dari itu, maka butir soal tersebut dianggap kurang mampu membedakan peserta tes yang mempersiapkan diri dalam menghadapi tes dari peserta yang tidak mempersiapkan diri. Bahkan bila daya beda itu menjadi negatif, maka butir soal itu sama sekali tidak dapat dipakai sebagai alat ukur prestasi belajar. Oleh karena itu butir soal tersebut harus dikeluarkan dari perangkat soal. Makin tinggi daya beda suatu butir soal, maka makin baik butir soal tersebut, dan sebaliknya makin rendah daya bedanya, maka butir soal itu dianggap tidak baik (Asmawi Zainul, dkk : 1997).

Menurut Dali S Naga (1992) kriteria besarnya koefisien daya beda diklasifikasikan menjadi empat kategori. Secara lebih rinci dijelaskan dalam tabel di bawah ini.

**Tabel 2. Klasifikasi Daya Beda Butir Soal**

Kategori Daya Beda	Koefisien Korelasi
Baik	0,40 – 1,00
Sedang (tidak perlu revisi)	0,30 – 0,39
Perlu direvisi	0,20 – 0,29
Tidak baik	-1,00 – 0,19

### 3). Distribusi Jawaban (Distraktor)

Dilihat dari konstruksi butir soal terdiri dari dua bagian, yaitu pokok soal dan alternatif jawaban. Alternatif jawaban jawaban juga terdiri dari dua bagian, yaitu kunci jawaban dan pengecoh. Pengecoh dikatakan berfungsi apabila semakin rendah tingkat kemampuan peserta tes semakin banyak memilih pengecoh, atau makin tinggi tingkat kemampuan peserta tes akan semakin sedikit memilih pengecoh.

Hal demikian dapat ditunjukkan dengan adanya korelasi yang tinggi, rendah atau negatif pada hasil analisis. Apabila proporsi peserta tes yang menjawab dengan salah atau memilih pengecoh kurang dari 0,025 maka pengecoh tersebut harus direvisi. Dan untuk pengecoh yang ditolak apabila tidak ada yang memilih atau proporsinya 0,00 (Depdikbud : 1997).

Proporsi alternatif jawaban masing-masing butir soal dapat dilihat pada kolom *proportion endorsing* pada hasil analisis iteman.

Selain memperhatikan fungsi daya tarik untuk dipilih oleh peserta tes, pengecoh soal juga perlu memperhatikan daya beda (koefisien korelasi) yang ditunjukkan oleh masing-masing alternatif jawaban. Setiap pengecoh diharapkan memiliki daya beda negatif, artinya suatu pengecoh diharapkan lebih sedikit dipilih oleh kelompok tinggi dibandingkan dengan kelompok bawah. Atau daya beda pengecoh tidak lebih besar dari daya beda kunci jawaban setiap butir soal.



Menurut Depdikbud (1997) untuk menilai pengecoh (distraktor) dari masing-masing butir soal dapat dikategorikan sebagai berikut:

**Tabel 3. Klasifikasi Distraktor Butir Soal**

Kategori Distraktor	Nilai Proportion Endorsing
Baik	$\geq 0,025$
Revisi	$< 0,025$
Tidak Baik / Tolak	0,000

**b. Kriteria Kualitas Butir Soal**

Berdasarkan uraian di atas, menurut pandangan teori tes klasik secara empiris mutu butir soal ditentukan oleh statistik butir soal yang meliputi : tingkat kesukaran, daya beda dan efektifitas distraktor. Menurut statistik butir, kualitas butir soal secara keseluruhan dapat dikategorikan sebagai berikut :

**Tabel 4. Klasifikasi Kualitas Butir Soal**

Kategori	Kriteria Penilaian
Baik	Apabila (1). Tingkat kesukaran $0,25 \leq p \leq 0,75$ , (2). Korelasi biserial butir soal $\geq 0,40$ dan (3). Korelasi biserial alternatif jawaban (distraktor) bernilai negatif.
Revisi	Apabila (1). Tingkat kesukaran $p < 0,25$ atau $p > 0,75$ tetapi korelasi biserial butir $\geq 0,40$ dan korelasi biserial distraktor bernilai negatif, (2). Tingkat kesukaran $0,25 \leq p \leq 0,75$ dan korelasi biserial butir soal $\geq 0,40$ tetapi ada korelasi biserial pada distraktor yang bernilai positif, (3). Tingkat kesukaran $0,25 \leq p \leq 0,75$ dan korelasi biserial butir soal antara 0,20 sampai 0,30 tetapi korelasi distraktor bernilai negatif selain kunci atau tidak ada yang lebih besar nilainya dari kunci jawaban.
Tidak baik	Apabila (1). Tingkat kesukaran $p < 0,25$ atau $p > 0,75$ dan ada korelasi biserial pada distraktor bernilai positif, (2). Korelasi biserial butir soal $< 0,20$ , (3). Korelasi biserial butir soal $< 0,30$ dan korelasi biserial distraktor bernilai positif.

### c. Reliabilitas

Reliabilitas adalah suatu hal yang sangat penting pada alat pengukuran standar. Reliabilitas dihubungkan dengan pengertian adanya ketepatan tes dalam pengukurannya. Reliabilitas adalah kestabilan skor yang diperoleh peserta tes yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda atau dari suatu pengukuran ke pengukuran lainnya. Dengan kata lain reliabilitas merupakan tingkat konsistensi atau kemantapan hasil terhadap hasil dua pengukuran hal yang sama. Dapat juga diartikan sebagai tingkat kepercayaan dari suatu alat ukur (Depdikbud : 1997).

Hasil pengukuran diharapkan akan sama apabila pengukuran itu diulangi. Dengan perangkat tes yang reliabel, apabila tes itu diberikan dua kali pada peserta yang sama tetapi dalam selang waktu yang berbeda sepanjang tidak ada perubahan dalam kemampuan maka skor yang diperoleh akan konstan. Kriteria untuk menentukan tinggi rendahnya reliabilitas sebuah perangkat tes, menurut (Suharsimi Arikunto : 2001) dilihat pada rentangan koefisien korelasi sebagai berikut :

**Tabel 10. Klasifikasi Tingkat Reliabilitas Tes**

Kategori Reliabilitas Tes	Nilai Koefisien Korelasi
Sangat Tinggi	0,800 – 1,000
Tinggi	0,600 – 0,799
Cukup	0,400 – 0,599
Rendah	0,200 – 0,399
Sangat Rendah	0,000 – 0,199

### C. PENUTUP

Perencanaan dalam pengujian sangat penting karena tes baru akan berarti bila terdiri dari butir-butir soal yang menguji tujuan yang penting dan mewakili ranah pengetahuan, kemampuan dan keterampilan secara representatif. Ada enam hal yang perlu dipertimbangkan dalam perencanaan tes yaitu: pengambilan sampel dan pemilihan butir soal, tipe tes yang akan digunakan, aspek yang akan diuji, format butir soal, jumlah butir soal dan distribusi tingkat kesukaran butir soal

Jadi jelaslah bahwa dibutuhkan adanya alat yang dapat dipercaya untuk mengukur apakah alat ukur (butir soal) yang digunakan memang dapat dijadikan dasar untuk menentukan keputusan yang bijaksana. Inilah peran yang harus dimainkan oleh analisis butir soal, yaitu mengukur butir soal yang akan digunakan. Butir soal yang ternyata terlalu lemah, akan sukar dipertanggungjawabkan untuk dijadikan sebagai dasar penentuan keputusan, terutama keputusan yang sifatnya individual.

#### **D. DAFTAR PUSTAKA**

- Asmawi Zainul dan Noehi Nasoetion. 1997. *Penilaian Hasil Belajar*. Pusat Antar Universitas, Direktorat Jenderal Pendidikan Tinggi: Departemen Pendidikan Dan kebudayaan.
- Dali, S Naga. 1992. *Pengantar Teori Sekor Pada Pengukuran Pendidikan*. Gunadarma: Jakarta.
- Departemen Pendidikan Dan Kebudayaan. 1997. *Manual Item And Test Analysis (Iteman)*. Badan Penelitian dan Pengembangan Pendidikan dan Kebudayaan: Pusat Penelitian dan Pengembangan Sistem Pengujian.
- Suharsimi Arikunto. 2001. *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara: Jakarta.
- Suryabrata, S. 1999. *Pengembangan Alat Ukur Psikologis*. Direktorat jenderal Pendidikan Tinggi: Departemen Pendidikan dan Kebudayaan.



### 3. Statistik Tes atau Skala

Beberapa informasi penting yang dapat diperoleh dari print out statistik tes / skala adalah sebagai berikut :

a. Mean

Mean adalah skor rata-rata atau nilai tengah peserta tes. Berdasarkan hasil analisis menunjukkan bahwa nilai rata-rata dari 34 peserta tes adalah 22,176. Hal ini berarti perolehan skor masing-masing peserta secara klasikal mendekati nilai tersebut.

b. Variance

Variance merupakan varian dari distribusi skor peserta tes yang memberikan gambaran tentang sebaran skor peserta tes. Berdasarkan hasil analisis iteman diperoleh nilai variance sebesar 20,851. Hal ini mengandung pengertian bahwa sebaran skor peserta tes berada pada kisaran angka 20,851.

c. Standar Deviasi

Standar deviasi merupakan distribusi skor peserta tes. Skor deviasi merupakan perbedaan selisih antara skor peserta tes dengan rata-rata skor kelompok. Dan standar deviasi merupakan akar dari varian. Hasil analisis menunjukkan besarnya nilai standar deviasi adalah 4,566.

d. Skew

Skew menunjukkan kemiringan distribusi skor peserta tes 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 berarti skor berdistribusi secara simetris di sekitar skor

rata-rata atau nilai meannya. Dari hasil analisis diketahui besarnya nilai skew adalah 0,514. Hal ini menggambarkan bahwa sebagian besar skor berada pada bagian bawah atau skor rendah dari distribusi skor peserta tes.

e. Kurtosis

Kurtosis merupakan puncak distribusi skor yang menggambarkan kelandaian distribusi skor dibandingkan dengan distribusi normal. Nilai positif menunjukkan distribusi yang lebih lancip atau memuncak dan nilai negatif menunjukkan distribusi yang lebih landai atau merata. Dari hasil analisis menunjukkan nilai kurtosis sebesar  $-0,893$ . Hal ini berarti puncak distribusi skor mempunyai kecenderungan lebih landai atau merata karena bernilai negatif.

f. Minimum

Minimum adalah skor terendah peserta tes. Hasil analisis menunjukkan nilai terendah yang diperoleh peserta tes adalah 16.

g. Maximum

Maximum adalah skor tertinggi peserta tes. Hasil analisis menunjukkan nilai tertinggi yang diperoleh peserta tes adalah 31.

h. Median

Median adalah skor tengah, dimana 50% skor berada pada atau lebih rendah dari skor tersebut. Hasil analisis menunjukkan nilai median skor adalah 22.

i. Reliabilitas

Reliabilitas adalah suatu hal yang sangat penting pada alat pengukuran standar. Reliabilitas dihubungkan dengan pengertian adanya ketepatan tes dalam pengukurannya. Reliabilitas adalah kestabilan skor yang diperoleh peserta tes yang sama ketika diuji ulang dengan tes yang sama pada situasi

yang berbeda atau dari suatu pengukuran ke pengukuran lainnya. Dengan kata lain reliabilitas merupakan tingkat konsistensi atau kemantapan hasil terhadap hasil dua pengukuran hal yang sama. Dapat juga diartikan sebagai tingkat kepercayaan dari suatu alat ukur (Depdikbud : 1997).

Hasil pengukuran diharapkan akan sama apabila pengukuran itu diulangi. Dengan perangkat tes yang reliabel, apabila tes itu diberikan dua kali pada peserta yang sama tetapi dalam selang waktu yang berbeda sepanjang tidak ada perubahan dalam kemampuan maka skor yang diperoleh akan konstan. Kriteria untuk menentukan tinggi rendahnya reliabilitas sebuah perangkat tes, menurut (Suharsimi Arikunto : 2001) dilihat pada rentangan koefesien korelasi sebagai berikut :

**Tabel 10. Klasifikasi Tingkat Reliabilitas Tes**

Kategori Reliabilitas Tes	Nilai Koefesien Korelasi
Sangat Tinggi	0,800 – 1,000
Tinggi	0,600 – 0,799
Cukup	0,400 – 0,599
Rendah	0,200 – 0,399
Sangat Rendah	0,000 – 0,199

Berdasarkan analisis data, diperoleh nilai koefesien reliabilitas alpha sebesar 0,694 yang merupakan indeks homogenitas tes.

Hal ini menunjukkan bahwa reliabilitas soal tes Gizi dan Diit berada dalam kategori tinggi dan merupakan perangkat tes yang berkualitas baik.

j. Standart Error of Measurement (SEM)

Unsur lain yang berkaitan langsung dengan reliabilitas tes adalah standar kesalahan pengukuran. Kesalahan pengukuran dalam tes disebabkan oleh kesalahan pengambilan sampel peserta tes (*sampling error*), dan kesalahan pelaksanaan tes itu sendiri (*measurement error*). Berdasarkan hasil analisis

diperoleh nilai SEM sebesar 2,526. Hal ini menunjukkan bahwa nilai estimasi terhadap deviasi standar kesalahan pengukuran dalam skor tes adalah 2,526.

k. Mean P

Mean P adalah rata-rata tingkat kesukaran semua butir soal dalam tes secara klasikal. Perhitungan dilakukan dengan cara mencari rata-rata proporsi peserta tes yang menjawab benar untuk semua butir soal dalam tes. Besarnya nilai mean P yang ditunjukkan oleh hasil analisis adalah 0,554. Berdasarkan kriteria tingkat kesukaran nilai tersebut berada dalam kategori sedang, artinya perangkat soal Gizi dan Diit mempunyai tingkat kesukaran sedang.

l. Mean Biserial

Mean biserial adalah nilai rata-rata indeks daya pembeda dari semua soal dalam tes. Nilai ini diperoleh dengan cara menghitung nilai rata-rata koefisien korelasi biserial dari semua soal dalam tes. Besarnya nilai mean biserial berdasar hasil analisis adalah 0,394.

Hal ini menunjukkan bahwa indeks daya beda dari semua soal tes dalam kategori sedang atau tidak perlu direvisi. Jadi dapat dikatakan bahwa perangkat tes Gizi dan Diit secara klasikal dapat membedakan kelompok atas dan kelompok bawah, namun demikian agar diperoleh daya beda yang lebih tinggi sebaiknya diadakan perbaikan terhadap butir soal yang mempunyai indeks daya beda rendah.



### **C. FILE SKOR**

Hasil analisis pada file skor menampilkan skor yang diperoleh seluruh peserta tes. Dari hasil tersebut diperoleh skor tertinggi adalah 31,00 dan skor terendah 16,00, secara lengkap dapat dilihat pada lampiran 4.

Berdasarkan klasifikasi atau kriteria di atas dapat dikategorikan perangkat tes Gizi dan Diet untuk mengetahui kualitas butir soal secara keseluruhan. Hasil analisis selengkapnya disajikan dalam tabel di bawah ini :

**Tabel 9. Hasil Analisis Kualitas Butir Soal**

Kategori	Nomor Butir Soal	Jumlah	Prosentase (%)
Baik	1,4,8,9,10,12,19,24,28,31	10	25
Revisi	2,5,13,16,17,20,21,22,23,26,27,29, 30,32,33,34,35,37,39,40	20	50
Tidak Baik	3,6,7,11,14,15,18,25,36,38	10	25
Jumlah		40	100

Berdasarkan tabel di atas, dapat diketahui bahwa kualitas butir soal termasuk kategori baik sebanyak 10 butir (25%), kategori revisi 20 butir (50%) dan kategori tidak baik berjumlah 10 butir (25%). Dengan demikian butir soal tersebut sebagian

besar harus direvisi terutama dalam pembuatan distraktor. Hal ini dapat dilihat dalam print out hasil analisis iteman, yang menunjukkan nilai proportion endorsing masih rendah bahkan tidak sedikit yang bernilai 0,00 sehingga menyebabkan butir soal tersebut harus direvisi kembali. Butir soal yang diterima atau termasuk dalam kategori baik adalah butir soal yang mempunyai tingkat kesukaran sedang.

### **3. Statistik Tes atau Skala**

Beberapa informasi penting yang dapat diperoleh dari print out statistik tes / skala adalah sebagai berikut :

a. Mean

Mean adalah skor rata-rata atau nilai tengah peserta tes. Berdasarkan hasil analisis menunjukkan bahwa nilai rata-rata dari 34 peserta tes adalah 22,176. Hal ini berarti perolehan skor masing-masing peserta secara klasikal mendekati nilai tersebut.

b. Variance

Variance merupakan varian dari distribusi skor peserta tes yang memberikan gambaran tentang sebaran skor peserta tes. Berdasarkan hasil analisis iteman diperoleh nilai variance sebesar 20,851. Hal ini mengandung pengertian bahwa sebaran skor peserta tes berada pada kisaran angka 20,851.

c. Standar Deviasi

Standar deviasi merupakan distribusi skor peserta tes. Skor deviasi merupakan perbedaan selisih antara skor peserta tes dengan rata-rata skor kelompok. Dan standar deviasi merupakan akar dari varian. Hasil analisis menunjukkan besarnya nilai standar deviasi adalah 4,566.

d. Skew

Skew menunjukkan kemiringan distribusi skor peserta tes 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 berarti skor berdistribusi secara simetris di sekitar skor rata-rata atau nilai meannya. Dari hasil analisis diketahui besarnya nilai skew adalah 0,514. Hal ini menggambarkan bahwa sebagian besar skor berada pada bagian bawah atau skor rendah dari distribusi skor peserta tes.

e. Kurtosis

Kurtosis merupakan puncak distribusi skor yang menggambarkan kelandaian distribusi skor dibandingkan dengan distribusi normal. Nilai positif menunjukkan distribusi yang lebih lancip atau memuncak dan nilai negatif menunjukkan distribusi yang lebih landai atau merata. Dari hasil analisis menunjukkan nilai kurtosis sebesar  $-0,893$ . Hal ini berarti puncak distribusi skor mempunyai kecenderungan lebih landai atau merata karena bernilai negatif.

f. Minimum

Minimum adalah skor terendah peserta tes. Hasil analisis menunjukkan nilai terendah yang diperoleh peserta tes adalah 16.

g. Maximum

Maximum adalah skor tertinggi peserta tes. Hasil analisis menunjukkan nilai tertinggi yang diperoleh peserta tes adalah 31.

h. Median

Median adalah skor tengah, dimana 50% skor berada pada atau lebih rendah dari skor tersebut. Hasil analisis menunjukkan nilai median skor adalah 22.

i. Reliabilitas

Reliabilitas adalah suatu hal yang sangat penting pada alat pengukuran standar. Reliabilitas dihubungkan dengan pengertian adanya ketepatan tes dalam pengukurannya. Reliabilitas adalah kestabilan skor yang diperoleh peserta tes yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda atau dari suatu pengukuran ke pengukuran lainnya. Dengan kata lain reliabilitas merupakan tingkat konsistensi atau kemantapan hasil terhadap hasil dua pengukuran hal yang sama. Dapat juga diartikan sebagai tingkat kepercayaan dari suatu alat ukur (Depdikbud : 1997).

Hasil pengukuran diharapkan akan sama apabila pengukuran itu diulangi. Dengan perangkat tes yang reliabel, apabila tes itu diberikan dua kali pada peserta yang sama tetapi dalam selang waktu yang berbeda sepanjang tidak ada perubahan dalam kemampuan maka skor yang diperoleh akan konstan. Kriteria untuk menentukan tinggi rendahnya reliabilitas sebuah perangkat tes, menurut (Suharsimi Arikunto : 2001) dilihat pada rentangan koefesien korelasi sebagai berikut :

**Tabel 10. Klasifikasi Tingkat Reliabilitas Tes**

Kategori Reliabilitas Tes	Nilai Koefesien Korelasi
Sangat Tinggi	0,800 – 1,000
Tinggi	0,600 – 0,799
Cukup	0,400 – 0,599
Rendah	0,200 – 0,399
Sangat Rendah	0,000 – 0,199

Berdasarkan analisis data, diperoleh nilai koefesien reliabilitas alpha sebesar 0,694 yang merupakan indeks homogenitas tes.

Hal ini menunjukkan bahwa reliabilitas soal tes Gizi dan Diit berada dalam kategori tinggi dan merupakan perangkat tes yang berkualitas baik.

j. Standart Error of Measurement (SEM)

Unsur lain yang berkaitan langsung dengan reliabilitas tes adalah standar kesalahan pengukuran. Kesalahan pengukuran dalam tes disebabkan oleh kesalahan pengambilan sampel peserta tes (*sampling error*), dan kesalahan pelaksanaan tes itu sendiri (*measurement error*). Berdasarkan hasil analisis diperoleh nilai SEM sebesar 2,526. Hal ini menunjukkan bahwa nilai estimasi terhadap deviasi standar kesalahan pengukuran dalam skor tes adalah 2,526.

k. Mean P

Mean P adalah rata-rata tingkat kesukaran semua butir soal dalam tes secara klasikal. Perhitungan dilakukan dengan cara mencari rata-rata proporsi peserta tes yang menjawab benar untuk semua butir soal dalam tes. Besarnya nilai mean P yang ditunjukkan oleh hasil analisis adalah 0,554. Berdasarkan kriteria tingkat kesukaran nilai tersebut berada dalam kategori sedang, artinya perangkat soal Gizi dan Diit mempunyai tingkat kesukaran sedang.

#### I. Mean Biserial

Mean biserial adalah nilai rata-rata indeks daya pembeda dari semua soal dalam tes. Nilai ini diperoleh dengan cara menghitung nilai rata-rata koefisien korelasi biserial dari semua soal dalam tes. Besarnya nilai mean biserial berdasar hasil analisis adalah 0,394.

Hal ini menunjukkan bahwa indeks daya beda dari semua soal tes dalam kategori sedang atau tidak perlu direvisi. Jadi dapat dikatakan bahwa perangkat tes Gizi dan Diit secara klasikal dapat membedakan kelompok atas dan kelompok bawah, namun demikian agar diperoleh daya beda yang lebih tinggi sebaiknya diadakan perbaikan terhadap butir soal yang mempunyai indeks daya beda rendah.

#### C. FILE SKOR

Hasil analisis pada file skor menampilkan skor yang diperoleh seluruh peserta tes. Dari hasil tersebut diperoleh skor tertinggi adalah 31,00 dan skor terendah 16,00, secara lengkap dapat dilihat pada lampiran 4.

### **BAB III**

### **PENUTUP**

Berdasarkan hasil analisis butir soal dengan Program Iteman beserta pembahasan perangkat tes mata kuliah Gizi dan Diit untuk mahasiswa Tata Boga Semester 4, Jurusan PKK, Fakultas Teknik, UNY tahun akademik 2002/2003 dapat disimpulkan bahwa kualitas butir soal masih perlu ditingkatkan karena terdapat 10 butir (25%) soal yang ditolak atau tidak baik, 20 butir (50%) harus di revisi dan 10 butir (25%) yang sudah baik dan tidak perlu direvisi. Namun demikian perangkat tes tersebut mempunyai indeks reliabilitas yang tinggi yaitu sebesar 0,694, sehingga tingkat konsistensi skor tes tidak perlu diragukan apabila soal tersebut diujikan lagi.

Untuk meningkatkan kualitas atau mutu soal tes Gizi dan Diit diperlukan perbaikan-perbaikan sesuai dengan petunjuk yang diberikan oleh hasil analisis butir dengan Program Iteman. Perbaikan tersebut tidak bersifat menyeluruh, akan tetapi pada butir-butir yang menunjukkan nilai kelemahan-kelemahan. Dengan demikian soal tersebut masih dapat digunakan lagi untuk kepentingan yang sama setelah dilakukan revisi agar kualitas tes meningkat.

## DAFTAR PUSTAKA

- Asmawi Zainul dan Noehi Nasoetion. 1997. *Penilaian Hasil Belajar*. Pusat Antar Universitas, Direktorat Jenderal Pendidikan Tinggi: Departemen Pendidikan Dan kebudayaan.
- Dali, S Naga. 1992. *Pengantar Teori Sekor Pada Pengukuran Pendidikan*. Gunadarma: Jakarta.
- Departemen Pendidikan Dan Kebudayaan. 1997. *Manual Item And Test Analysis (Iteman)*. Badan Penelitian dan Pengembangan Pendidikan dan Kebudayaan: Pusat Penelitian dan Pengembangan Sistem Pengujian.
- Suharsimi Arikunto. 2001. *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara: Jakarta.
- Suryabrata, S. 1999. *Pengembangan Alat Ukur Psikologis*. Direktorat jenderal Pendidikan Tinggi: Departemen Pendidikan dan Kebudayaan.



**Tabel Ringkasan Hasil Analisis Butir Soal Gizi dan Diit**

No Butir Soal	Tingkat Kesukaran	Daya Beda	Distraktor	Keterangan
01	Sedang	Baik	Tolak	Baik
02	Sukar	Baik	Baik	Revisi
03	Mudah	Tidak baik	Tolak	Tidak baik
04	Sedang	Baik	Tolak	Baik
05	Sukar	Baik	Baik	Revisi
06	Mudah	Sedang	Tolak	Tidak baik
07	Sedang	Tidak baik	Tolak	Tidak baik
08	Sedang	Baik	Baik	Baik
09	Sedang	Baik	Tolak	Baik
10	Sedang	Baik	Baik	Baik
11	Sukar	Baik	Baik	Tidak baik
12	Sedang	Baik	Tolak	Baik
13	Sedang	Sedang	Tolak	Revisi
14	Mudah	Sedang	Baik	Tidak baik
15	Mudah	Tidak baik	Tolak	Tidak baik
16	Sukar	Baik	Tolak	Revisi
17	Sedang	Sedang	Tolak	Revisi
18	Sedang	Tidak baik	Baik	Tidak baik
19	Mudah	Baik	Baik	Baik
20	Mudah	Revisi	Tolak	Revisi
21	Sedang	Sedang	Baik	Revisi
22	Mudah	Sedang	Baik	Revisi
23	Mudah	Baik	Tolak	Revisi
24	Sedang	Baik	Baik	Baik
25	Sukar	Tidak baik	Baik	Tidak baik
26	Sedang	Sedang	Baik	Revisi
27	Mudah	Baik	Tolak	Revisi
28	Mudah	Baik	Baik	Baik
29	Sedang	Revisi	Baik	Revisi
30	Mudah	Sedang	Tolak	Revisi
31	Mudah	Baik	Tolak	Baik
32	Mudah	Baik	Tolak	Revisi
33	Sukar	Revisi	Tolak	Revisi
34	Sedang	Baik	Baik	Revisi
35	Sedang	Revisi	Baik	Revisi
36	Sukar	Tidak baik	Tolak	Tidak baik
37	Sedang	Baik	Baik	Revisi
38	Sedang	Tidak baik	Tolak	Tidak baik
39	Sedang	Sedang	Baik	Revisi

40	Sedang	Baik	Baik	Revisi
----	--------	------	------	--------