

## ***ITEM AND TEST ANALYSIS (ITEMAN) 4.3***

### **A. Pengertian ITEMAN**

*Item And Test Analysis* (ITEMAN) merupakan program komputer yang digunakan untuk menganalisis butir soal dengan pendekatan teori tes klasik. Program ini termasuk satu paket program dalam MicroCAT yang dikembangkan oleh *Assessment Systems Corporation* mulai tahun 1982 dan mengalami revisi pada tahun 1984, 1986, 1988, dan 1993; mulai dari versi 2.00 versi 3.50, dan sampai versi 4.3 yang sekarang. 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 kuantitatif setiap butir (daya pembeda, tingkat kesukaran, efektivitas distraktor), reliabilitas (KR-20/*Alpha*, *split-half*, dan *Spearman-Brown (S-B)*), *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.

## **B. Keunggulan ITEMAN 4.3**

ITEMAN 4.3 menawarkan beberapa keunggulan substansial dibandingkan ITEMAN 3.0:

1. Keuntungan yang paling penting adalah penambahan grafis. Sekarang memungkinkan untuk menghasilkan plot kuantil item untuk setiap item. Selain itu, pengguna dapat mengontrol jumlah titik di plot.
2. Iteman 4.3 mampu menangani pretest (percobaan atau unscored) item-item yang tidak termasuk dalam nilai akhir tetapi untuk statistik yang masih diinginkan.
3. Lebih banyak statistik dihitung, termasuk *alpha* (KR-20) koefisien reliabilitas dengan setiap item yang dihapus, beberapa koefisien reliabilitas belah dua (baik dengan dan tanpa koreksi Spearman-Brown), kesalahan pengukuran baku bersyarat, dan subkelompok P (proporsi benar) statistik hingga tujuh kelompok yang dipesan.
4. Alih-alih file teks ASCII sederhana, hasilnya sekarang adalah file RTF yang disiapkan sebagai laporan formal, dan juga dalam format comma-separated value (CSV) yang dapat dimanipulasi

(diurutkan, disorot, dll.) dalam perangkat lunak spreadsheet. Ini juga menghasilkan file CSV skor ujian.

5. Skor skala dan subscores dapat ditambahkan ke output.
6. Skor dapat diklasifikasikan menjadi dua kelompok pada skor yang ditentukan, dan kedua kelompok dapat menggunakan label pengguna.
7. Item dapat dianalisis relatif terhadap skor eksternal daripada total skor pada tes.
8. Jumlah maksimum item yang dapat dianalisis telah ditingkatkan menjadi 10.000.
9. Jenis "batch" kemampuan, menggunakan "Multiple Runs File" telah ditambahkan untuk memungkinkan pengguna menjalankan beberapa set data tanpa harus menggunakan antarmuka pengguna grafis untuk setiap run. File Multiple Runs dapat dibuat di luar ITEMAN di editor teks atau secara interaktif di dalam ITEMAN.

### **C. Kelebihan dan Kekurangan ITEMAN 4.3**

1. Kelebihan Iteman
  - Menginput data dapat dibantu dengan program Microsoft excel.
  - Hasil analisis lebih mudah dibaca karena berupa tabel-tabel.

- Tampilan pada iteman lebih menarik daripada anates.
  - Tersedia banyak versi.
2. Kekurangan Iteman
- Sering terjadi kesalahan saat penulisan data respon siswa.
  - Sulit dioperasikan karena menggunakan bahasa Inggris.
  - Banyak pengguna kesulitan pada saat input format.

#### **D. Manfaat ITEMAN 4.3**

Manfaat dari ITEMAN 4.3 adalah:

1. Peserta dapat mengenal software aplikasi ITEMAN 4.3
2. Peserta dapat menjalankan software aplikasi ITEMAN 4.3
3. Peserta dapat menginterpretasikan output software aplikasi ITEMAN 4.3
4. Peserta dapat menganalisis butir soal yang baik sesuai dengan pendekatan teori tes klasik

# PENDEKATAN TEORI TES KLASIK DAN ASUMSINYA

## A. Teori Tes Klasik

Teori tes klasik (CTT) merupakan teori yang menggunakan model matematika yang sangat sederhana untuk menunjukkan hubungan antara skor amatan, skor sebenarnya, dan skor kesalahan, dengan Rumus  $X = T + E$  (Mardapi, 2012: 52). Model ini diikuti dengan sejumlah asumsi untuk menyederhanakan formula dalam mengestimasi indeks kehandalan dan kesahihan suatu instrumen. Menurut Allen & Yen (1979: 56) “*classical true-score theory is a simple, quite useful model that describes how errors of measurement can influence observed score*”. Allen & Yen menyatakan bahwa teori tes klasik adalah teori yang sederhana, cepat, dan berguna yang menggambarkan bagaimana kesalahan baku pengukuran bisa mempengaruhi skor pengamatan.

Selain memiliki rumus yang sederhana teori tes klasik juga memiliki asumsi yang mudah dipahami (Mardapi, 2012:53) yaitu 1) instrumen hanya mengukur satu dimensi. Instrumen atau tes yang dirancang menguji matematika, maka dimensi yang diukur juga hanya matematika; 2) tidak ada hubungan antara skor murni dan

skor kesalahan  $\rho_{TE} = 0$ ; 3) Tidak ada hubungan antara kesalahan pada pengukuran pertama dengan kesalahan pada pengukuran kedua, dan dapat ditulis  $\rho_{E1E2} = 0$ ; 4) Tidak ada hubungan antara skor murni pada pengukuran pertama dengan kesalahan pada pengukuran kedua, dan dapat ditulis dengan  $\rho_{T1E2} = 0$ ; 5) Tidak ada hubungan antara skor murni pada pengukuran kedua dengan kesalahan pada pengukuran pertama, dan dapat ditulis  $\rho_{T2E1} = 0$ ; dan 6) Rata kesalahan pengukuran pada populasi adalah 0 atau  $\mu_E = 0$ .

Asumsi-asumsi teori tes klasik sebagaimana disebutkan di atas, memungkinkan untuk dikembangkan dalam berbagai formula yang berguna dalam melakukan pengukuran psikologis. Validitas, reliabilitas, tingkat kesukaran, daya beda dan efektifitas distraktor, adalah formula penting dalam pemilihan butir soal.

## **B. Validitas**

Validitas suatu perangkat tes merupakan kemampuan suatu tes untuk mengukur apa yang seharusnya diukur (Mardapi, 2008:15). Tes dikatakan valid jika tes tersebut mengukur sesuai dengan tujuan pengukuran (Allen & Yen, 1979:95). Sebagai contoh tes yang digunakan untuk menentukan jenis pekerjaan, dikatakan valid jika skor yang diperoleh oleh peserta ujian berkorelasi dengan performansi pekerjaannya di masa datang.

Menurut Samuel Messick (Linn, 1989:13), validitas merupakan penilaian menyeluruh dimana bukti empiris dan logika teori mendukung pengambilan keputusan serta tindakan berdasarkan skor tes atau model-model penilaian yang lain, berdasarkan beberapa pendapat di atas, dapat disimpulkan bahwa tes dikatakan valid jika tes tersebut mampu mengukur apa yang hendak diukur.

Ada tiga tipe validitas, yaitu validitas kriteria, validitas isi, validitas konstruk. Validitas kriteria dibedakan menjadi dua, yaitu validitas prediktif dan validitas konkuren. Validitas kriteria (Retnawati, 2014:2) dimaksudkan untuk menjawab pertanyaan sejauh mana tes memprediksi kemampuan peserta di masa mendatang (*predictive validity*) atau mengestimasi kemampuan dengan alat ukur lain dengan tenggang waktu yang hampir bersamaan (*concurrent validity*). Besar kecilnya harga estimasi validitas prediktif suatu instrumen digambarkan dengan koefisien antar prediktor dengan kriteria. Validitas kriteria (Retnawati, 2014:16-17) diketahui dengan mengestimasi korelasi skor tes peserta dengan skor kriteria. Korelasi ini disebut koefisien validitas yang menyatakan derajat hubungan antara presiktor dengan kriteria. Hal ini bermanfaat untuk memprediksikan suatu skor kemampuan ke skor kriteria dalam rangka memprediksikan kemampuan atau

performen peserta tes. Prediksi ini dilakukan dengan persamaan regresi sederhana/tunggal atau regresi ganda.

Ada dua macam validitas isi, yakni validitas kenampakan dan validitas logika. Validitas isi berarti sejauh mana suatu perangkat tes mencerminkan keseluruhan *trait* yang hendak diukur. Suatu tes memiliki bukti validitas isi ditetapkan menurut analisis rasional terhadap isi tes, yang penilaiannya didasarkan atas pertimbangan subjektif individual (Mardapi, 2008:15). Validitas kenampakan didasarkan kepada pertanyaan apakah suatu butir dalam perangkat tes mengukur aspek yang relevan dengan domainnya. Validitas logika berkaitan dengan keseksamaan batasan pada domain yang hendak diukur, dan merupakan jawaban apakah keseluruhan butir merupakan sampel representatif dari keseluruhan butir yang mungkin dibuat. Validitas isi dapat dilakukan dengan kesepakatan ahli (*expert judgment*) kemudian untuk mengetahui kesepakatan ini, dapat digunakan indeks validitas butir yang diusulkan Aiken. Indeks validitas butir yang diusulkan Aiken ini dirumuskan sebagai berikut (Retnawati, 2014:3).

$$V = \frac{\sum s}{n(c-1)} \dots\dots\dots (1)$$

Keterangan rumus:

- $V$  : Indeks validitas butir
- $s$  : skor yang ditetapkan setiap rates dikurangi skor terendah dalam kategori yang dipakai
- $n$  : banyak rater

*c* : banyaknya kategori yang dapat dipilih rater

Validitas konstruk adalah validitas yang menunjukkan seberapa jauh suatu tes mengungkap trait yang hendak diukur (Allen & Yen, 1979:108). Prosedur validasi konstruk diawali dari suatu indentifikasi dan batasan mengenai variabel yang hendak diukur dan dinyatakan dalam bentuk konstruk logis berdasarkan teori mengenai variabel tersebut (Retnawati, 2014: 3). Dari teori ini ditarik suatu konstruksi praktis mengenai hasil pengukuran pada kondisi tertentu, dan konstruksi inilah yang akan diuji. Apabila hasilnya sesuai dengan harapan maka instrumen itu dianggap memiliki validitas konstruk yang baik. Validitas konstruk dapat dibuktikan dengan kebermaknaan skor hasil pengukuran. Pendekatan yang dipilih berupa pembuktian bahwa konstruk yang dihipotesiskan dan dikonfirmasi keberadaannya (Retnawati, 2014:7). Analisis yang digunakan dengan analisis faktor eksploratori maupun konfirmatori.

### **C. Reliabilitas**

Reliabilitas merupakan derajat keajegan di antara dua buah hasil pengukuran pada objek yang sama (Mehrens & Lehmann, 1973:102). Defisini ini dapat diilustrasikan dengan orang yang diukur tinggi badannya akan diperoleh hasil yang tidak berbeda walaupun menggunakan alat pengukur yang berbeda dan skala yang berbeda. Dalam dunia pendidikan, hasil belajar atau kemampuan

peserta didik dikatakan reliabel jika dilakukan pengukuran hasilnya akan sama dengan informasi sebelumnya, walaupun penguji berbeda, kotektornya yang berbeda atau butir soal yang berbeda tetapi memiliki karakteristik yang sama.

Allen & Yen (1979: 62), berpendapat bahwa tes dikatakan reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor sesungguhnya. Selanjutnya dikatakan bahwa reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes yang parallel. Senada dengan Crocker & Algina (1986: 127) menyatakan bahwa reliabilitas merupakan konsistensi peserta tes terhadap hasil pengukuran pada tes yang sama atau parallel. Jadi suatu tes itu reliabel jika hasil pengukuran mendekati keadaan peserta tes yang sesungguhnya.

Dalam pendidikan, pengukuran tidak dapat langsung dilakukan pada ciri atau karakter yang akan diukur. Ciri atau karakter tersebut bersifat abstrak sehingga sulit memperoleh alat ukur yang stabil untuk mengukur karakteristik peserta tes. Reliabilitas suatu alat ukur akan mengacu pada keajegan alat ukur yang mengandung makna kecermatan. Pengukuran yang tidak reliabel akan menghasilkan skor yang tidak dapat dipercaya karena perbedaan skor yang terjadi diantara individu lebih ditentukan oleh kesalahan

daripada perbedaan skor sesungguhnya. Semakin tinggi reliabilitas suatu tes maka kemungkinan kesalahan yang terjadi semakin kecil.

Estimasi reliabilitas tes dapat dilakukan dengan dua cara, pengukuran konsistensi eksternal dan pengukuran konsistensi internal (Retnawati, 2015:3). Pengukuran reliabilitas eksternal diperoleh dengan cara mengolah hasil pengetesan yang berbeda, baik dari instrumen yang berbeda maupun yang sama. Ada dua metode pengukuran konsistensi eksternal suatu instrumen yaitu metode tes ulang dan metode tes parallel. Estimasi reliabilitas dengan metode tersebut akan menghasilkan koefisien stabilitas yang dihitung dengan korelasi linier antara distribusi skor peserta pada pemberian tes pertama dengan skor peserta pada pemberian skor kedua.

Pengukuran konsistensi internal diperoleh dengan cara menganalisis data dari satu kali pengetesan. Secara umum ada tiga metode pengukuran konsistensi internal yaitu metode belah dua (rumus Spearman-Brown, rumus Flanagan, dan rumus Rulon), 2) metode Kuder-Richardson (KR-20 dan KR-21) dan metode Cronbach Alpha. Pengukuran konsistensi internal yang sering digunakan adalah metode Cronbach Alpha, karena bisa digunakan untuk menghitung reliabilitas suatu tes dengan bentuk pilihan ganda atau uraian. Rumus alpha (Retnawati, 2015:9) sebagai berikut.

$$r_i = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right) \dots\dots\dots (2)$$

Keterangan rumus:

- $r_i$  : Reliabilitas instrumen
- $k$  : Banyaknya butir pertanyaan atau banyaknya soal
- $\sum \sigma_b^2$  : Jumlah varian butir soal
- $\sigma_t^2$  : Varian total

Tabel 1. Kriteria Koefisien Reliabilitas

Koefisien Reliabilitas	Makna
$r \leq 0,20$	Sangat Rendah
$0,20 < r \leq 0,40$	Rendah
$0,40 < r \leq 0,60$	Sedang
$0,60 < r \leq 0,80$	Tinggi
$0,80 < r \leq 1,00$	Sangat Tinggi

Mengenai seberapa tinggi indeks reliabilitas menurut Ebel (1986: 71) alat ukur yang digunakan untuk menentukan mutu pendidikan seharusnya memiliki indeks reliabilitas 0,90. Menurut Mehrens & Lehmann (1973: 104) menyatakan alat ukur yang digunakan untuk membuat keputusan pada peserta secara perorangan harus memiliki indeks reliabilitas minimal sebesar 0,85. Pendapat lain (Allen & Yen, 1979:75; Saifuddin Azwar, 2012:98; Djemari Mardapi, 2008:122) mensyaratkan indeks reliabilitas minimal 0,70 sudah cukup baik.

#### D. Tingkat Kesukaran

Tingkat kesukaran merupakan salah satu parameter yang digunakan dalam teori tes klasik. Tingkat kesukaran dianalisis dengan tujuan untuk melihat apakah soal yang dibuat sudah sesuai dengan kemampuan peserta didik yang mengikuti tes. Proporsi peserta didik menjawab benar sedikit (kecil) maka dapat disimpulkan bahwa tingkat kesukaran itu tinggi, sebaliknya apabila proporsi menjawab benar banyak (besar) maka dapat disimpulkan bahwa tingkat kesukaran tersebut rendah. Reynolds, Livingston, & Wilson (2010:148) menyatakan bahwa “*item difficulty is defined as the percentage or proportion of tes takers who correctly answer the item*”. Pernyataan tersebut artinya tingkat kesukaran ditentukan sebagai persentasi atau proporsi dari pengambil tes yang menjawab benar.

Menurut Crocker & Algina (1986: 311) menyatakan bahwa “*the proportion of axaminees who answer the item correctly is usually denoted as  $p_i$  and is called the item difficulty*”. Maknanya bahwa proporsi dari peserta tes yang menjawab benar biasanya dinotasikan sebagai  $p_i$  adalah tingkat kesukaran soal. Berdasarkan pendapat di atas bahwa tingkat kesukaran merupakan proporsi peserta tes yang menjawab benar. Apabila proporsi menjawab benar kecil dapat disimpulkan bahwa tingkat kesukarannya tinggi,

sebaliknya apabila proporsi menjawab benar besar dapat disimpulkan bahwa tingkat kesukaran soal tersebut rendah.

Menurut Allen & Yen (1979: 121) tingkat kesukaran yang baik adalah 0,30 sampai 0,70. Senada dengan Djemari Mardapi (2012: 186) tingkat kesukarannya adalah  $0,30 \leq P \leq 0,70$  soal digunakan untuk tes. Butir dengan tingkat kesukaran dibawah 0,30 dianggap sukar sedangkan jika indeksny di atas 0,70 maka butir tersebut dianggap mudah. Namun demikian, besarnya tingkat kesukaran butir suatu perangkat tes dapat dirangcang dengan pertimbangan tujuan penyusunan perangkat tes tersebut.

Tabel 2. Kriteria Tingkat Kesukaran

Kriteria $P$	Keterangan
$P > 0,70$	Mudah
$0,30 \leq P \leq 0,70$	Sedang
$P < 0,30$	Sukar

## E. Daya Beda

Daya beda merupakan kemampuan item dalam membedakan antara siswa yang memiliki kemampuan tinggi dan siswa yang memiliki kemampuan rendah (Azwar, 2012:137). Senada dengan pendapat tersebut Reynolds, Livingston, & Willson (2010: 151) menyatakan bahwa “*item discrimination refers to how well an item can accurately discriminate between test takers who differ on the construct being measured*”. Pernyataan tersebut artinya daya pembeda mengacu kepada bagaimana sebuah item yang baik bisa

secara akurat membedakan antara pengikut tes yang berbeda pada rancangan yang diukur.

Daya beda (*index discrimination*) dapat dihitung dengan menggunakan rumus *Point Biserial Correlation*. Senada dengan pendapat tersebut Wright (2008:230) menyatakan bahwa “*one approach to determining the discrimination level of an item is to find the Point biserial correlation between individu student scores on the item and the student’s scores ont the whole test*”. Pengertian tersebut dapat diartikan bahwa satu pendekatan untuk menentukan level perbedaan dari sebuah item adalah menemukan korelasi poin biserial antara skor siswa pada item dan skor siswa pada tes seluruhnya. Rumus yang digunakan adalah sebagai berikut.

$$r_{pb} = \frac{\overline{X_1} - \overline{X_0} \sqrt{p(1-p)}}{S_x} \dots\dots\dots (3)$$

Keterangan rumus:

- $r_{pb}$  : Korelasi point biserial
- $\overline{X_1}$  : Nilai rata-rata peserta tes yang menjawab benar
- $\overline{X_0}$  : Nilai rata-rata peserta tes yang menjawab salah
- $p$  : Proporsi siswa menjawab item secara benar
- $S_x$  : Standar deviasi dari skor mentah

Rumus yang digunakan tersebut akan menghasilkan karakteristik dari daya beda suatu soal. Karakteristik dari soal tersebut akan mencerminkan apakah soal yang telah dirancang memiliki daya beda yang bisa diterima dan digunakan untuk tes.

Menurut Ebel (Azwar, 2012:140) menyarankan kriteria daya beda dalam empat kategori pada Tabel 3.

Tabel 3. Kriteria Daya Beda

Kriteria	Keterangan
0,40 – 1,00	Sangat Memuaskan
0,30 – 0,39	Memuaskan
0,20 – 0,29	Belum Memuaskan
Negatif – 0,19	Jelek

## F. Efektivitas Distraktor

Analisis akhir dari teori tes klasik adalah analisis distraktor. Pada soal pilihan ganda alternatif jawaban yang tidak benar bisa dijadikan pengecoh (distraktor) bagi peserta didik yang sebenarnya tidak mengetahui alternatif jawaban yang benar. Distraktor akan memberikan informasi kepada pendidik tentang peserta didik yang benar-benar menguasai materi pelajaran dan peserta didik yang kurang menguasai materi pelajaran. Reynolds, Livingston, & Willson (2010:157) menyatakan bahwa “*distraktor analysis allows you to examine how many examinees in the top and bottom groups selected each option on multiple-choice item*”. Pernyataan tersebut dapat diartikan analisis distraktor mengizinkan penyusun tes untuk memeriksa grup yang berkemampuan tinggi dan grup berkemampuan rendah memilih masing-masing pada pilihan ganda yang telah disediakan.

Pengecoh (distraktor) dikatakan berfungsi efektif, apabila dipilih minimal 5% dari seluruh peserta ujian (Mardapi, 2008:143).

### G. Kesalahan Pengukuran

Kesalahan pengukuran (*Standard Error Measurement, SEM*) dapat digunakan untuk memahami kesalahan yang bersifat acak atau random yang mempengaruhi skor peserta ujian dalam pelaksanaan tes. Kesalahan pengukuran yang disimbolkan dengan  $\sigma_E$ , dapat dihitung dengan rumus yang diturunkan dari rumus reliabilitas (Allen & Yen, 1979:73).

$$\sigma_E = \sigma_X \sqrt{1 - \rho_{xx'}} \dots\dots\dots (4)$$

Keterangan :

$\sigma_x$  : standar deviasi dari skor total

$\rho_{xx'}$  : indeks reliabilitas

# PROSEDUR PENGOPERASIAN

## ITEMAN 4.3

### A. Tahap Persiapan Penggunaan ITEMAN 4.3

Sebelum menggunakan ITEMAN 4.3, terlebih dahulu kita memerlukan beberapa komponen berikut;

1. Software ITEMAN 4.3
2. File Rekap Jawaban Siswa (file dalam bentuk txt notepad)

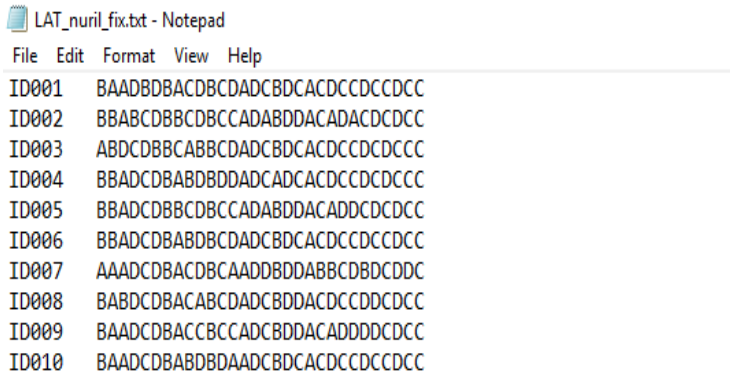
Tahap ini menurut saya merupakan langkah yang paling penting dan ribet, karena perlu ketelitian dan ketekunan. Sesuai dengan perkembangan dan perbaikan versi software sehingga menimbulkan adanya berbagai macam metode untuk membuat rekap jawaban siswa ini, diantaranya :

#### **Metode 1. Input Notepad**

Cara ini merupakan cara standar, dan mirip seperti cara pada ITEMAN 3.0

1. Buka file notepad baru
2. Isikan file notepad tersebut dengan nama siswa, NIM, atau apapun untuk menandakan atau membedakan "si subjek penjawab soal/examinee". dalam contoh ini saya menggunakan kode ID0001, ID0002, dst....

3. Di sebelah kanan nama/NIM, masukkan jawaban subjek ke dalam notepad, tanpa spasi.



```
LAT_nuril_fix.txt - Notepad
File Edit Format View Help
ID001 BAADBDBACDBCDADCBDACDCCDCCDCC
ID002 BBABCDBBCDBCCADABDDACADCDCCDCC
ID003 ABDCCBBACBBCCDADCBDACDCCDCCDCC
ID004 BBADCDBABDBDDADCADCACDCCDCCDCC
ID005 BBADCDBBCDBCCADABDDACADDCCDCCDCC
ID006 BBADCDBABDBCDADCBDACDCCDCCDCC
ID007 AAADCDBACDBCAADDBDDABBCBDCDDC
ID008 BABDCDBACBCDADCBDACDCCDCCDCC
ID009 BAADCDBACCBCADCBDACDCCDCCDCC
ID010 BAADCDBABDBDAADCBDACDCCDCCDCC
```

Gambar 3.1 Jawaban Siswa txt

## Metode 2. Input Excel

cara ini lebih sederhana dan cenderung banyak digunakan, serta minimal resiko terjadinya kekeliruan. Sehingga, walaupun agak ribet, tapi kecenderungan untuk salah data menjadi lebih kecil.

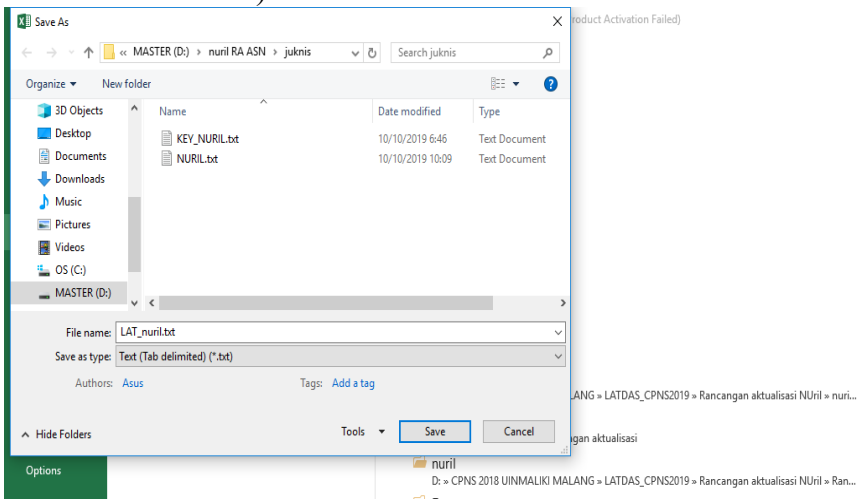
1. Buka excel baru.
2. Isi excel dengan data examinee, dan juga jawaban examinee.

Silahkan Entry hasil jawaban soal pilihan ganda siswa, pada microsoft excel. Model entry dibuat seperti pada tampilan gambar di bawah ini.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3	ID0001	B	A	A	D	B	D	B	A	C	D	B	C	D	A	D	C	B	D	C	A	C	D	C	C
4	ID0002	B	B	A	B	C	D	B	B	C	D	B	C	C	A	D	A	B	D	D	A	C	A	D	A
5	ID0003	A	B	D	C	D	B	B	C	A	B	B	C	D	A	D	C	B	D	C	A	C	D	C	C
6	ID0004	B	B	A	D	C	D	B	A	B	D	B	D	D	A	D	C	A	D	C	A	C	D	C	C
7	ID0005	B	B	A	D	C	D	B	B	C	D	B	C	C	A	D	A	B	D	D	A	C	A	D	D
8	ID0006	B	B	A	D	C	D	B	A	B	D	B	C	D	A	D	C	B	D	C	A	C	D	C	C
9	ID0007	A	A	A	D	C	D	B	A	C	D	B	C	A	A	D	D	B	D	D	A	B	B	C	D
10	ID0008	B	A	B	D	C	D	B	A	C	A	B	C	D	A	D	C	B	D	D	A	C	D	C	C
11	ID0009	B	A	A	D	C	D	B	A	C	C	B	C	C	A	D	C	B	D	D	A	C	A	D	D
12	ID0010	B	A	A	D	C	D	B	A	B	D	B	D	A	A	D	C	B	D	C	A	C	D	C	C
13	ID0011	B	A	D	C	D	B	B	B	D	B	C	B	A	A	D	C	B	D	C	A	C	D	C	C
14	ID0012	B	A	A	D	C	D	B	A	B	D	B	D	A	A	D	C	B	D	C	A	C	D	C	C

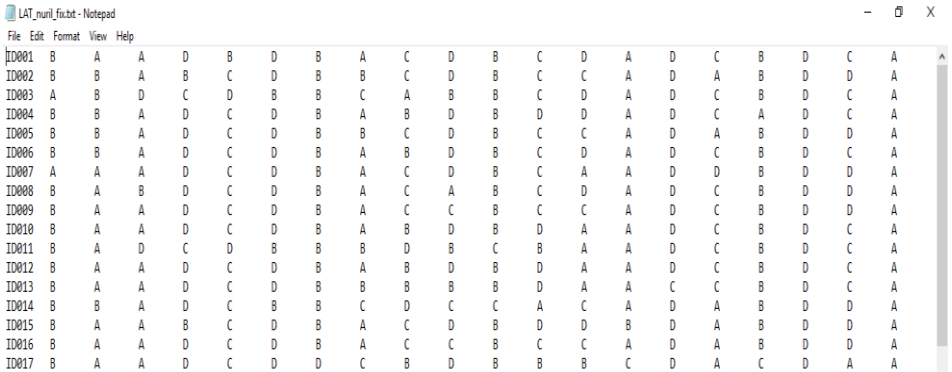
Gambar. 3. 2. Jawaban siswa excel

3. Lalu lakukan “Save As”, dengan memilih option “Text (Tab Delimited).txt”



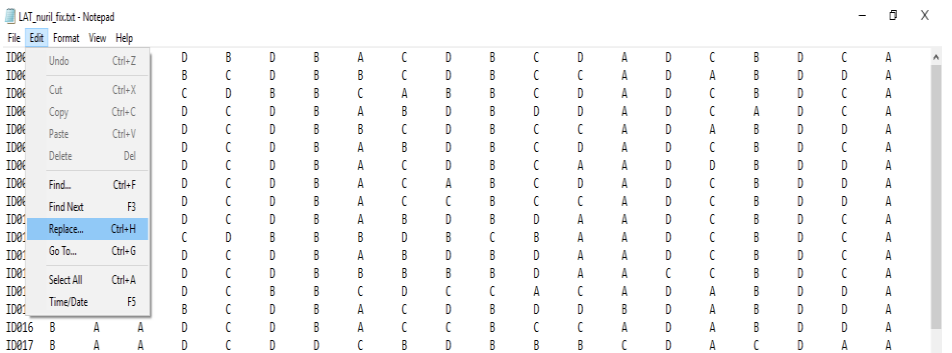
Gambar. 3. 3. Jawaban siswa excel di Save AS txt

4. Lalu dibuka file “Text (Tab Delimited).txt” akan muncul seperti tampilan berikut.

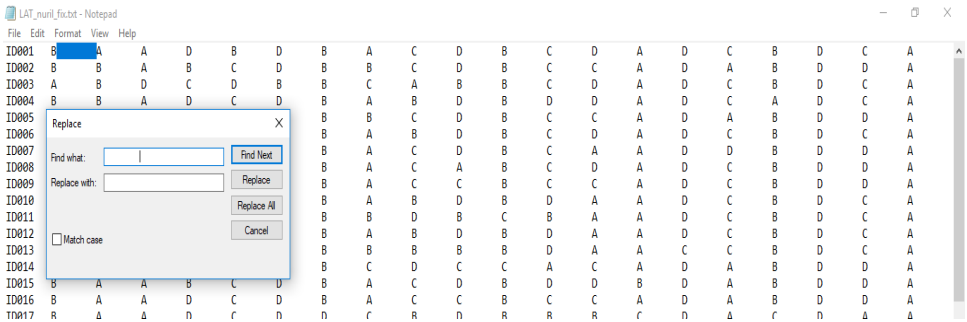


Gambar. 3. 4. Jawaban siswa txt

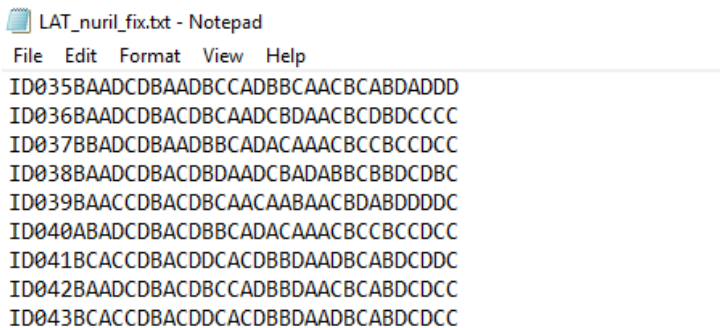
5. Lalu lakukan langkah berikut untuk membuat spasi.



Gambar. 3.5. Cara menghilangkan spasi dengan Repalce



Gambar. 3.6. Kotak Dialog Repalce



Gambar. 3.7. Hasil Menghilangkan Spasi dengan Repalce

### 3. File Kunci Jawaban Soal (file dalam bentuk txt notepad)

Jika pada ITEMAN 3.0 kunci jawaban langsung bersama dengan rekap jawaban examinee, maka di ITEMAN 4.3 ini kunci jawabannya bisa diletakkan pada file terpisah. Membuat kunci jawaban sama seperti membuat rekap: bisa menggunakan metode notepad, atau menggunakan excel. Untuk menghemat waktu, dalam panduan ini, saya langsung gunakan metode excel untuk membuat kunci jawaban.

#### 1. Buat file excel baru

2. Input data kunci, sesuai dengan aturan berikut.

Tabel 4. Format Kunci

ITEM ID	Kunci Jawaban	Jumlah Jawaban Alternatif	Domain penilaian	Status Inklusi	Tipe Item
Item01, Item02, ... dsb.	- Pilihan ganda (ABCD atau ABCDE)	Banyak opsi jawaban 4 (ABCD) atau 5 (ABCDE) pada pilihan ganda	Biasanya kita memilih satu perangkap soal/ satu mapel	Apakah item ini mau dianalisis atau <b>diexcludekan</b> - Y = yes - N = No - P= Pretest	M : multiple choice (pilihan ganda) R : rating scale (0,1) P : partial scale (numeric responses)

Item01	B	4	1	Y	M
--------	---	---	---	---	---

Keterangan :

Item01 : Nomor Butir Soal

A : Kunci Jawaban

4 : Alternatif opsi jawaban (A, B, C, D)

1 : Domain (bisa isi Matk/ mapel)

Y : Ketentuan Perintah analisis masing-masing butir

(Y = Yes; N=No)

M : Jenis analisis (M= Multiple choice,

R= Rating scale, P = Partial Scale)

3. Isi excel dengan kunci jawaban examinee.

Silahkan Entry kunci jawaban soal pilihan, pada microsoft excel.

Model entry dibuat seperti pada tampilan gambar di bawah ini.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Item01	B	4	1	Y	M							
2	Item02	A	4	1	Y	M							
3	Item03	A	4	1	Y	M							
4	Item04	C	4	1	Y	M							
5	Item05	C	4	1	Y	M							
6	Item06	D	4	1	Y	M							
7	Item07	B	4	1	Y	M							
8	Item08	A	4	1	Y	M							
9	Item09	C	4	1	Y	M							

Gambar. 3.8. Entri Kunci dalam Excel

4. Lalu lakukan “Save As”, dengan memilih option “Text (Tab Delimited).txt”
5. Lalu dibuka file “Text (Tab Delimited).txt” akan muncul seperti tampilan berikut

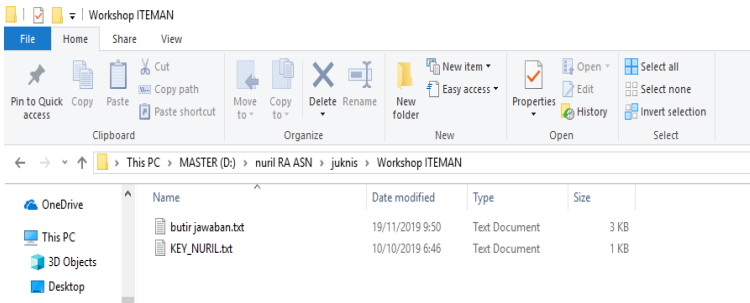
```

KEY_NURIL.txt - Notepad
File Edit Format View Help
item61 B      4      1      Y      M
item62 A      4      1      Y      M
item63 A      4      1      Y      M
item64 C      4      1      Y      M
item65 C      4      1      Y      M
item66 D      4      1      Y      M
item67 B      4      1      Y      M
item68 A      4      1      Y      M
item69 C      4      1      Y      M
item70 D      4      1      Y      M

```

Gambar. 3.9. Kunci dalam txt

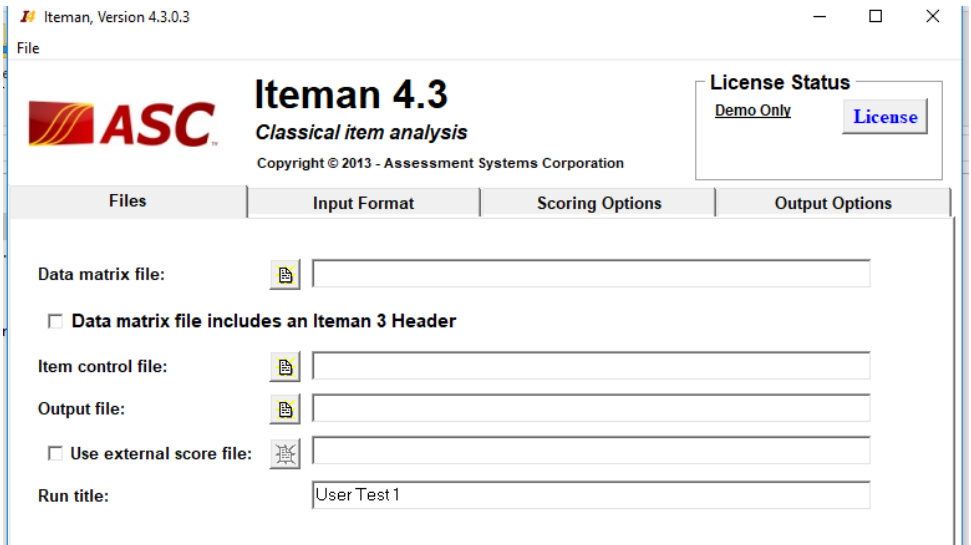
6. Menyiapkan terakhir penggunaan software aplikasi  
ITEMAN 4.3



Gambar. 3.10. Data yang Diperlukan untuk Analisis

## B. Tahap Menjalankan Software Aplikasi ITEMAN 4.3


### 1. Membuka software aplikasi ITEMAN 4.3



## 2. Memasukkan butir jawaban siswa

Iteman, Version 4.3.0.3

File



# Iteman 4.3

*Classical item analysis*

Copyright © 2013 - Assessment Systems Corporation

**License Status**  
[Demo Only](#) [License](#)

Files	Input Format	Scoring Options	Output Options
<b>Data matrix file:</b>	<input type="file"/>	<input type="text" value="D:\ITEMAN 4\butir jawaban.txt"/>	
<input type="checkbox"/> <b>Data matrix file includes an Iteman 3 Header</b>			
<b>Item control file:</b>	<input type="file"/>	<input type="text"/>	
<b>Output file:</b>	<input type="file"/>	<input type="text"/>	
<input type="checkbox"/> <b>Use external score file:</b>	<input type="file"/>	<input type="text"/>	
<b>Run title:</b>	<input type="text" value="User Test 1"/>		

Data matrix file, pada bagian ini dimasukkan data respon jawaban dari siswa yang sudah disimpan dengan format txt seperti gambar di atas.

### 3. Memasukkan kunci jawaban

Iteman, Version 4.3.0.3

File

**ASC** **Iteman 4.3**  
*Classical item analysis*  
Copyright © 2013 - Assessment Systems Corporation

License Status  
[Demo Only](#) [License](#)

Files | Input Format | Scoring Options | Output Options

Data matrix file:

Data matrix file includes an IteMan 3 Header

Item control file:

Output file:

Use external score file:

Run title:

Item control file, pada bagian ini dimasukkan data kunci jawaban yang sudah disimpan dengan format txt seperti gambar di atas.

#### 4. Menuliskan judul Output file

Iteman, Version 4.3.0.3

File

**ASC** **Iteman 4.3**  
*Classical item analysis*  
Copyright © 2013 - Assessment Systems Corporation

License Status  
Demo Only [License](#)

Files | Input Format | Scoring Options | Output Options

Data matrix file: D:\ITEMAN 4\butir jawaban.txt

Data matrix file includes an IteMan 3 Header

Item control file: D:\ITEMAN 4\KEY\_NURIL.txt

Output file: Hasil Analisis

Use external score file:

Run title: Latihan 1

Outfile, pada bagian ini diisikan nama file yang diinginkan untuk hasil dari analisis.

## 5. Informasi format dalam ITEMAN 4.3

The screenshot shows the Iteman 4.3 software interface. The title bar reads "Iteman, Version 4.3.0.3". The main window has a menu bar with "File". Below the menu bar is the ASC logo and the text "Iteman 4.3 Classical item analysis Copyright © 2013 - Assessment Systems Corporation". On the right, there is a "License Status" box with "Demo Only" and "License" buttons. Below the menu bar are four tabs: "Files", "Input Format", "Scoring Options", and "Output Options". The "Input Format" tab is active and contains two sections: "Fixed Width Data" and "Delimited Data".

**Fixed Width Data:**

- Number of examinee ID columns: 6
- Examinee IDs begin in column: 1
- Item responses begin in column: 7
- Omit character: O
- Not administered character: N

**Delimited Data:**

- The data matrix file is delimited by a:
  - Comma
  - Tab
- Response matrix includes examinee ID in first column

**Test for differential item functioning**

- Group status appears in column: 0 Create 6 ability levels for the DIF test
- Group 1 code: 1 Group 2 code: 2
- Group 1 label: Reference Group 2 label: Focal

Pada Input Format terdapat beberapa point yang harus diisi sesuai dengan data yang disimpan pada notepad. Contoh;

- Number of examinee ID columns merupakan jumlah kolom nomer/nama peserta tes.
- Examinee ID begin in column merupakan kolom dimulainya nomer/nama peserta tes.
- Item response begin in column merupakan kolom dimulainya respon jawaban dari peserta tes.

Iteman, Version 4.3.0.3

File

**ASC** **Iteman 4.3**  
*Classical item analysis*  
 Copyright © 2013 - Assessment Systems Corporation

License Status  
 Demo Only [License](#)

Files | **Input Format** | Scoring Options | Output Options

**Input Format**

**Fixed Width Data:**

Number of examinee ID columns:

Examinee IDs begin in column:

Item responses begin in column:

Omit character:

Not administered character:

**Delimited Data:**

The data matrix file is delimited by a:

Comma  Tab

Response matrix includes examinee ID in first column

Test for differential item functioning

Group status appears in column:  Create  ability levels for the DIF test

Group 1 code:  Group 2 code:

Group 1 label:  Group 2 label:

**Run** [Help](#)

## 6. Scaled scoring

Iteman, Version 4.3.0.3

File

**ASC** **Iteman 4.3**  
*Classical item analysis*  
 Copyright © 2013 - Assessment Systems Corporation

License Status  
 Demo Only [License](#)

Files | **Input Format** | **Scoring Options** | Output Options

**Scaled Scoring**

Compute scaled scores for the scored items for:

Total score

Each domain separately

Scaling function:

Linear, with a slope of  and intercept of

Standardized, with a mean of  and SD of

**Classification**

Perform dichotomous classification using:

Scored Items (Number Correct)

Scaled Scores

Use a cutpoint of  for classification


Low group label:  High group label:

**Run** [Help](#)

## 7. Item statistic

Iteman, Version 4.3.0.3

File

 **Iteman 4.3**  
*Classical item analysis*  
Copyright © 2013 - Assessment Systems Corporation

License Status  
[Demo Only](#) [License](#)

Files | Input Format | Scoring Options | Output Options

**Item Statistics**

Acceptable P (difficulty) range:  to

Acceptable item mean range:  to

Acceptable item correlation range:  to

Exclude omits from option statistics

Correct the item-total score correlations for spuriousness

**Output Options**

Use  decimal places of precision in the tables

Produce quantile plots for each item

Produce the quantile plot data table for each item

Create  score groups for the quantile plots / summary table

Produce collusion index matrix (multiple-choice items only)

Save scored item responses  Save item control file

Include Omit codes in the scored matrix

Include Not Administered codes in the scored matrix

**Flags**

Key Flag

	Low Flag	High Flag
P-value	<input type="text" value="LP"/>	<input type="text" value="HP"/>
R-pbis	<input type="text" value="LR"/>	<input type="text" value="HR"/>
Item Mean	<input type="text" value="LM"/>	<input type="text" value="HM"/>

DIF Flag

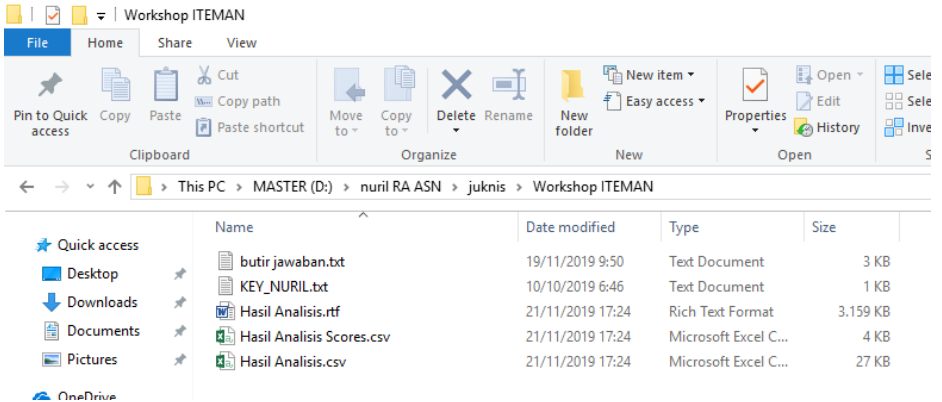
**Run**  **Help**

## 8. Menjalankan software aplikasi ITEMAN 4.3

The screenshot displays the ITEMAN 4.3 software interface. At the top, the title bar reads "Iteman, Version 4.3.0.3". The main window features the ASC logo and the text "Iteman 4.3 Classical item analysis Copyright © 2013 - Assessment Systems Corporation". A "License Status" box in the top right corner shows "Demo Only" and a "License" button. Below this is a menu bar with "Files", "Input Format", "Scoring Options", and "Output Options". The main workspace contains several input fields: "Data matrix file:" with a file explorer icon and the path "D:\ITEMAN 4\butir jawaban.txt"; "Data matrix file:" with an unchecked checkbox; "Item control file:"; "Output file:"; "Use external scor:" with an unchecked checkbox; and "Run title:". A "Confirm" dialog box is overlaid on the workspace, containing the following text: "The run is complete. The summary output file can be found at the following Windows path: Hasil Analisis.rtf. Do you want to open the directory for the output file?". The dialog box has "Yes" and "No" buttons. At the bottom of the software window, there is a "Run" button and a "Help" button.

Run title, pada bagian ini diisi nama analisis, yang akan muncul pada hasil output.

## 9. Hasil Output dari ITEMAN 4.3



The screenshot shows a Windows File Explorer window titled 'Workshop ITEMAN'. The address bar indicates the path: This PC > MASTER (D:) > nuril RA ASN > juknis > Workshop ITEMAN. The ribbon includes 'File', 'Home', 'Share', and 'View'. The ribbon tabs are 'Clipboard', 'Organize', 'New', and 'Open'. The main area displays a list of files and folders:

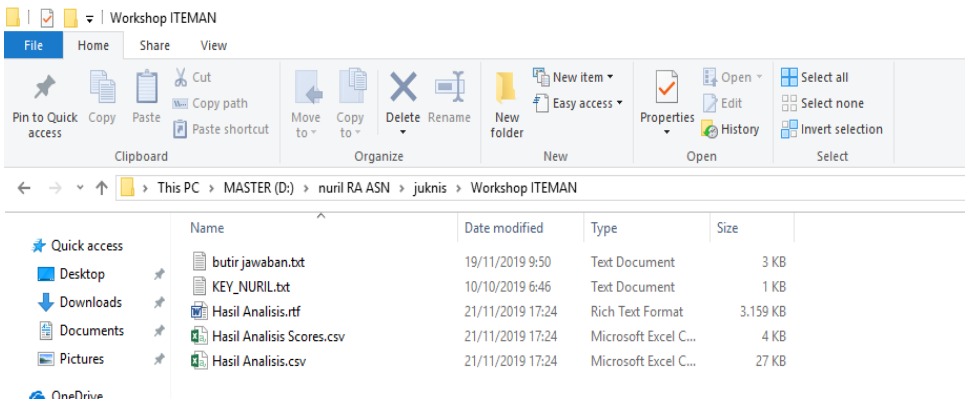
Name	Date modified	Type	Size
butir jawaban.txt	19/11/2019 9:50	Text Document	3 KB
KEY_NURIL.txt	10/10/2019 6:46	Text Document	1 KB
Hasil Analisis.rtf	21/11/2019 17:24	Rich Text Format	3.159 KB
Hasil Analisis Scores.csv	21/11/2019 17:24	Microsoft Excel C...	4 KB
Hasil Analisis.csv	21/11/2019 17:24	Microsoft Excel C...	27 KB

# INTERPRETASI OUTPUT

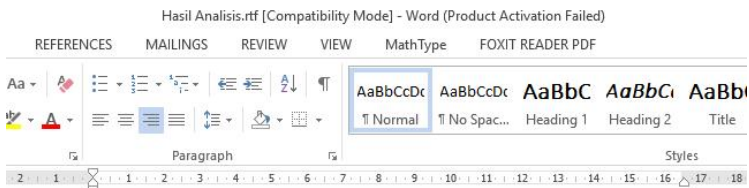
## ANALISIS ITEMAN 4.3

### A. Output dari ITEMAN 4.3

#### 1. Membuka file hasil analisis yang berupa “word”



2. Akan muncul tampilan seperti berikut



## ***Classical Item and Test Analysis Report***

### ***Latihan 1***

***Report created on 21/11/2019***

***This report was produced by the demo version of Iteman 4.3,  
which is limited to 50 items and 50 examinees.***

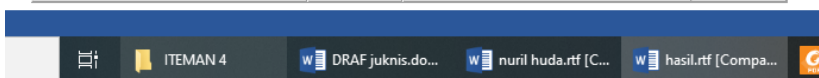
***Iteman: Software for Classical Analysis***

Pada dasarnya ITEMAN 4.3 adalah perangkat lunak komputer (software) yang dapat menyajikan ringkasan analisis secara keseluruhan (analisis tes) dan analisis tiap butir soal (analisis item).

3. Merekap hasil output dari reliabilitas, tingkat kesukaran, daya beda, efektivitas pengecoh tiap soal

**Table 1: Specifications**

Specification	Value	Specification	Value
Number of examinees	80	Total Items	30
Scored Items	30	Pretest Items	0
Multiple Choice Items	30	Polytomous Items	0
Number of domains	1	External scores	No
Minimum P	0,00	Maximum P	1,00
Minimum item mean	0,00	Maximum item mean	15,00
Minimum item correlation	0,00	Maximum item correlation	1,00
ITEMAN 3.0 Header	No	Exclude omits from option statistics	No
Number of ID columns	6	ID begins in column	1
Responses begin in column	7	Omit character	O
Not Admin character	N	Produce quantile tables	Yes
Correct for spuriousness	Yes	Produce quantile plots	Yes
Save data matrix	No	Include omit codes in matrix	N/A
Include Not Admin codes in matrix	N/A	Include scaled scores for	N/A
Scaling function	N/A	Scaled score setting 1	N/A
Scaled score setting 2	N/A	Dichotomous Classification	No
Classify based on	N/A	Cutpoint	N/A
Low group label	Low	High group label	High
Data is delimited by	N/A	Test for DIF	No
Group status is in column	N/A	Ability levels for DIF	N/A
Group 1 code	N/A	Group 2 code	N/A



**Gambar 4. 1** Statistic dari Hasil Analisis Butir Soal

Table 2 presents the summary statistics of the test, for the scored items. Definitions of these statistics are found in the Ite-man manual.

**Table 2: Summary statistics**

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
Scored Items	30	18,575	6,164	6	30	0,619	0,395

Table 3 presents a reliability analysis of the tests. Alpha (also known as KR-20) is the most commonly used index of reliability, and is therefore used to calculate the standard error of measurement (SEM) on the raw score scale. Also presented are three configurations of split-half reliability, first as uncorrected correlations, and then as Spearman-Brown (S-B) corrected correlations. This is because an uncorrected split-half correlation is referenced to a "test" that only contains half as many items as the full test, and therefore underestimates reliability.

**Table 3: Reliability**

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0,865	2,264	0,841	0,660	0,806	0,914	0,795	0,892

**Item information**

Seq.	ID	Key	Scored	Num Options	Domain	Flags
1	item61	B	Yes	4	1	

### Item statistics

N	P	Total Rpbis	Total Rbis	Alpha w/o
80	0,675	0,212	0,276	0,866

### Option statistics

Option	N	Prop	Rpbis	Rbis	Mean	SD	Color	
A	18	0,225	-0,101	-0,141	16,778	5,526	Maroon	
B	54	0,675	0,212	0,276	19,778	6,037	Green	**KEY**
C	6	0,075	-0,186	-0,346	14,000	7,183	Blue	
D	2	0,025	-0,051	-0,136	16,000	8,485	Olive	
Omit	0							
Not Admin	0							

4. Menginterpretasi hasil output dari reliabilitas, tingkat kesukan, daya beda, efektivitas pengecoh tiap soal sesuai dengan kriteria dari pendekatan teori tes klasik

Tabel 5. Kriteria Koefisien Reliabilitas

Koefisien Reliabilitas	Makna
$r \leq 0,20$	Sangat Rendah
$0,20 < r \leq 0,40$	Rendah
$0,40 < r \leq 0,60$	Sedang
$0,60 < r \leq 0,80$	Tinggi
$0,80 < r \leq 1,00$	Sangat Tinggi

Tabel 6. Kriteria Tingkat Kesukaran

Kriteria $P$	Keterangan
$P > 0,70$	Mudah
$0,30 \leq P \leq 0,70$	Sedang
$P < 0,30$	Sukar

Tabel 7. Kriteria Daya Beda

Kriteria	Keterangan
0,40 – 1,00	Sangat Memuaskan
0,30 – 0,39	Memuaskan
0,20 – 0,29	Belum Memuaskan
Negatif – 0,19	Jelek

Pengecoh (distraktor) dikatakan berfungsi efektif, apabila dipilih minimal 5% dari seluruh peserta ujian (Djemari Mardapi, 2008:143).

- Menggolongkan butir-butir yang baik sesuai dengan kriteria dari pendekatan teori tes klasik.

Kriteria butir soal yang baik adalah memiliki tingkat kesukaran **sedang**, daya beda **minimal memuaskan**, dan pengecoh/ opsi jawaban dari butir soal **efektif**. Ditambah nilai reliabilitas **minimal tinggi**.

## B. Interpretasi Hasil Analisis ITEMAN 4.4

Terdapat 2 hasil analisis yaitu analisis keseluruhan dan analisis tiap butir soal.

### 1. Ringkasan Analisis Keseluruhan (analisis tes)

**Table 2: Summary statistics**

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
Scored Items	30	18,575	6,164	6	30	0,619	0,395

- Berdasarkan tabel 2 di atas menjelaskan item yang dianalisis ada 30 butir soal dan rata-rata 18,575 dan standar baku 6,164

serta Skor minimal 6 dan skor maksimal 30.

- Tingkat kesukaran (Mean P)

Pada tabel di atas didapat Mean P sebesar 0,619 berdasarkan tabel 6 nilai tersebut soal yang dianalisis memiliki tingkat kesukaran sedang.

- Daya beda (Rpbis)

Pada tabel di atas didapat Mean Rpbis sebesar 0,395, berdasarkan tabel 7 nilai tersebut soal yang dianalisis memiliki daya beda memuaskan.

**Table 3: Reliability**

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0,865	2,264	0,841	0,660	0,806	0,914	0,795	0,892

- Reliabilitas (Alpha)

Pada tabel di atas didapat Alpha sebesar 0,865, berdasarkan tabel 5 nilai tersebut soal yang dianalisis memiliki reliabilitas sangat tinggi dengan standard error of measurement 2,264

## 2. Ringkasan analisis tiap butir soal (analisis item)

Untuk ringkasan analisis tiap butir soal diambil 1 item sampel dari 30 item yang di analisis, yaitu item nomor 61. Berikut hasil interpretasinya;

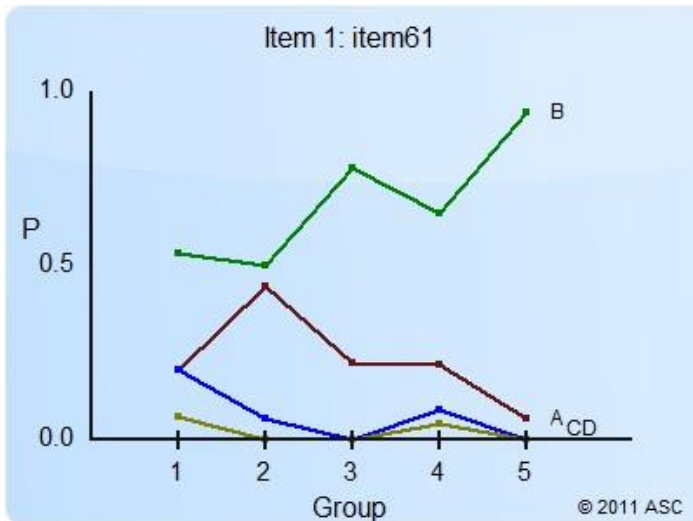


Diagram 4.1 Hasil Analisis Soal No. 61.

### Item information

Seq.	ID	Key	Scored	Num Options	Domain	Flags
1	item61	B	Yes	4	1	

### Item statistics

N	P	Total Rpbis	Total Rbis	Alpha w/o
80	0,675	0,212	0,276	0,866

### Option statistics

Option	N	Prop.	Rpbis	Rbis	Mean	SD	Color	
A	18	0,225	-0,101	-0,141	16,778	5,526	Maroon	
B	54	0,675	0,212	0,276	19,778	6,037	Green	**KEY**
C	6	0,075	-0,186	-0,346	14,000	7,183	Blue	
D	2	0,025	-0,051	-0,136	16,000	8,485	Olive	
Omit	0							
Not Admin	0							

Berdasarkan tabel hasil analisis tes dan tabel kriteria dapat disimpulkan hasil interpretasi dari analisis item soal nomor 11 sebagai berikut;

- Tingkat kesukaran (P) item no. 61.

Pada tabel item statistic didapat P sebesar 0,675, berdasarkan tabel 6. nilai tersebut soal yang dianalisis memiliki tingkat kesukaran sedang.

- Daya beda (Rpbis)

Pada tabel item statistic didapat Rpbis sebesar 0,212, berdasarkan tabel 7. nilai tersebut soal yang dianalisis memiliki daya beda belum memuaskan.

- Prop (Pengecoh)

Pada item ini setiap opsi jawaban dijawab ada yang kurang dari 5% dari seluruh jumlah peserta tes, maka pengecoh tergolong belum efektif.

Rekap lengkap seperti gambar di bawah ini.

Tabel 8. Rekap Analisis Butir Soal

No	Reliabel	Ket	TK (P)	Ket	DP (Total Rpbis)	Ket	Opsis Jawaban			
61	0,866	ST	0,675	SD	0,212	BM	0,225	0,675	0,075	0,025

# STUDI KASUS

## A. Wacana Kasus

Tahun 2019 tepatnya 26 Maret Dinas Pendidikan Dasar Kabupaten Banjarejo melaksanakan try out IPA untuk mengetahui daya serap siswa kelas 9 SMP/MTs di Kabupaten Banjarejo. Pak Sun sebagai TIM Evaluasi pelaksanaan kegiatan tersebut menunjuk stafnya untuk **menganalisis butir soal** yang sudah diujikan kepada seluruh siswa kelas 9 SMP/MTs di Kabupaten Banjarejo. Pak Alim adalah staf yang ditunjuk oleh pak Sun mengambil 100 sampel secara acak dari 1239 siswa yang mengikuti try out tersebut. Data sampel yang diambil pak Alim.

Tabel 9. Rekap Jawaban Siswa

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
kunci	B	C	B	C	D	B	A	D	B	D	D	D	C	C	D	D	C	A	A	C	B	C	C	A	C	D	B	B	D	A	
ID001	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID002	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID003	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID004	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID005	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID006	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID007	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID008	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID009	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID010	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID011	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	D	A
ID012	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	C	B	C	C	A	C	D	B	B	A	A
ID013	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID014	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID015	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID016	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	D	B	C	C	A	C	D	B	B	A	A
ID017	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID018	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID019	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID020	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID021	B	C	B	C	A	B	A	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A

ID022	B	C	B	C	A	D	A	D	B	D	D	C	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	C	D	A
ID023	B	C	B	C	A	D	A	D	B	D	D	C	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	C	D	A
ID024	B	C	B	C	A	D	A	D	B	D	D	C	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	C	D	A
ID025	B	C	B	C	A	D	A	D	B	D	D	C	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID026	B	C	B	C	A	D	A	D	B	D	D	C	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID027	B	C	B	C	A	B	A	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	B	D	B	A	A	A
ID028	B	C	B	C	A	D	A	D	B	D	D	C	C	C	D	D	C	A	A	B	B	C	C	A	B	D	B	B	D	A
ID029	B	C	B	C	A	D	A	D	B	D	D	D	C	C	D	D	C	A	A	D	B	C	C	A	C	D	B	A	A	A
ID030	B	C	B	C	A	D	A	D	B	D	D	C	C	C	D	D	C	A	A	B	B	C	C	A	B	D	B	B	D	A
ID031	A	C	B	C	C	B	D	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	C	A
ID032	B	C	B	C	A	B	A	D	B	D	D	D	C	C	D	D	D	A	C	B	B	C	C	A	C	D	B	B	A	A
ID033	B	C	B	C	A	B	A	D	B	C	D	D	C	C	D	D	C	A	A	B	B	C	C	A	B	D	B	B	A	A
ID034	A	C	B	C	A	B	A	D	B	A	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID035	B	C	B	C	A	B	A	D	B	D	D	D	C	C	D	D	D	A	C	B	B	C	C	A	C	D	B	B	A	A
ID036	B	C	B	C	A	B	A	D	B	D	D	D	C	C	D	D	D	A	C	B	B	C	C	A	C	D	B	B	A	A
ID037	A	C	B	C	A	B	A	D	B	A	D	D	C	C	D	D	D	A	A	B	B	C	C	A	C	D	B	B	A	A
ID038	A	C	B	D	A	B	A	D	B	A	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID039	B	C	B	C	A	B	A	D	B	D	D	D	C	C	D	D	D	A	A	B	B	C	C	C	B	D	B	B	A	A
ID040	B	C	B	C	A	B	A	D	B	D	D	D	C	C	D	D	D	A	A	B	B	C	C	C	B	D	B	B	A	A
ID041	B	C	B	C	A	B	A	C	B	D	D	D	C	C	D	D	C	A	A	C	A	C	B	A	C	D	B	A	A	A
ID042	B	C	B	C	A	B	A	D	B	A	D	D	C	C	D	D	D	A	A	B	B	C	C	A	C	C	B	B	A	A
ID043	B	C	B	C	A	B	A	D	B	*	D	D	C	C	D	D	D	A	A	B	B	C	C	A	C	C	B	B	A	A
ID044	B	C	B	C	A	B	A	C	B	D	D	D	C	C	D	D	C	A	A	C	A	D	D	A	C	D	B	B	A	A
ID045	B	C	B	C	A	B	A	D	B	A	D	D	C	C	D	D	C	A	A	B	B	C	C	A	B	C	B	B	A	A

ID046	B	C	B	C	A	B	A	D	B	D	D	D	C	C	D	D	C	A	A	B	A	C	D	D	C	D	B	B	A	A
ID047	B	C	B	C	A	B	A	D	A	A	D	D	C	C	D	D	C	A	C	C	B	D	C	A	C	D	B	D	C	A
ID048	B	C	B	C	A	B	A	D	A	A	D	D	C	C	D	D	C	A	C	C	B	D	C	A	C	D	B	D	C	A
ID049	B	C	B	C	A	B	A	D	B	D	D	A	C	C	D	D	C	A	A	C	C	A	C	A	C	D	A	A	A	A
ID050	B	C	B	C	A	B	A	D	B	D	D	A	C	C	D	D	C	A	A	C	C	A	C	A	C	D	A	A	A	A
ID051	A	D	B	C	A	D	D	D	B	D	D	D	C	C	D	D	C	A	A	B	B	C	C	A	C	D	B	B	A	A
ID052	B	C	B	C	A	D	B	C	B	D	D	D	C	C	D	D	D	A	A	C	B	D	C	A	C	D	B	B	B	D
ID053	B	C	D	C	A	B	A	D	B	A	D	D	C	C	B	C	C	A	A	B	B	C	C	A	C	D	B	B	C	D
ID054	A	C	B	C	A	D	A	D	A	D	D	D	B	B	D	D	C	A	C	B	A	C	C	A	C	D	B	C	D	D
ID055	B	C	A	C	A	B	A	D	B	A	D	D	C	D	A	D	A	D	A	C	B	D	C	A	B	D	D	B	C	A
ID056	A	C	B	C	A	B	D	D	B	D	C	C	C	C	A	D	A	A	A	A	B	C	B	A	B	D	B	B	A	A
ID057	A	C	B	C	A	B	D	D	B	D	C	C	C	C	A	D	A	A	A	A	B	C	B	A	B	D	B	B	A	A
ID058	A	C	B	C	A	B	C	D	B	D	C	C	C	C	A	D	A	A	A	A	B	C	B	A	B	D	B	B	A	A
ID059	B	C	C	C	A	B	A	D	B	A	B	C	C	C	D	D	C	A	A	C	B	B	B	A	B	D	B	A	C	D
ID060	B	D	B	C	D	D	A	D	B	C	C	A	C	D	D	D	C	A	C	D	B	C	C	A	B	D	C	C	D	D
ID061	B	D	B	C	D	D	A	D	B	C	D	A	C	D	D	D	C	A	C	D	B	C	C	A	B	D	C	C	C	D
ID062	B	C	B	C	A	B	A	D	B	A	D	D	C	C	D	D	A	B	A	B	C	C	D	A	B	D	C	C	C	D
ID063	A	C	B	A	A	B	C	D	B	B	D	D	C	C	D	D	C	B	A	B	A	C	B	A	D	D	A	B	C	C
ID064	A	B	B	D	B	B	D	A	B	D	D	D	C	C	D	D	B	A	A	D	A	C	D	D	C	D	B	C	B	A
ID065	A	B	B	B	A	C	D	D	B	C	D	D	C	C	D	D	C	A	A	C	A	C	B	A	D	C	D	B	A	A
ID066	A	C	B	C	B	B	D	B	B	A	B	A	C	C	C	D	D	D	A	B	B	C	D	A	C	D	B	B	D	D
ID067	D	D	A	C	B	D	D	D	B	D	C	A	C	C	C	D	C	D	A	B	B	C	B	A	C	D	B	A	B	A
ID068	A	C	B	C	A	B	A	B	B	A	C	C	C	C	D	C	C	A	A	A	A	C	D	C	A	D	B	C	A	A
ID069	A	C	B	C	B	B	A	B	B	A	C	C	C	C	D	D	C	A	A	A	A	D	D	C	A	D	B	C	A	A

ID070	A	D	B	C	D	B	D	B	B	A	D	D	C	C	C	D	C	A	A	A	A	C	C	C	A	D	A	A	A	D	
ID071	D	C	A	C	A	B	A	C	B	A	D	C	D	B	A	D	C	D	A	B	A	C	C	A	B	D	B	B	C	D	
ID072	D	C	A	C	A	B	A	C	B	A	D	C	D	B	A	D	C	D	A	B	A	C	C	A	B	D	B	B	B	C	
ID073	B	A	B	C	B	B	D	B	B	A	A	C	C	B	A	A	C	D	C	B	B	C	C	A	A	D	B	B	C	A	
ID074	B	C	B	C	A	B	A	B	B	C	A	C	D	C	A	A	C	A	A	B	D	A	A	A	C	D	C	B	C	B	
ID075	A	B	C	A	A	B	C	D	B	D	D	D	C	A	D	D	C	B	A	B	A	C	B	A	C	A	C	C	C	D	
ID076	A	B	C	A	A	B	C	C	D	C	B	D	C	C	D	D	C	B	A	B	B	C	C	A	B	D	B	C	C	D	
ID077	A	B	B	A	A	D	A	D	A	B	C	B	C	B	A	C	C	D	A	B	A	C	D	A	A	D	A	B	C	C	
ID078	A	B	B	A	A	D	A	D	D	B	C	B	C	B	A	C	C	D	A	B	A	C	D	A	A	D	A	B	C	C	
ID079	C	C	C	C	A	A	A	A	B	A	B	B	B	B	A	A	A	A	A	B	C	C	C	C	D	D	D	C	B		
ID080	A	B	B	A	A	D	A	D	A	B	C	B	C	B	A	C	C	D	A	B	A	C	D	A	A	D	A	B	C	C	
ID081	B	B	B	D	A	C	D	B	B	B	C	B	B	B	D	D	C	A	A	B	A	B	A	A	B	C	C	C	C	A	
ID082	B	B	D	D	B	C	D	B	B	D	C	B	C	D	C	D	C	C	A	A	A	B	C	D	C	A	B	C	C	B	
ID083	A	B	B	A	A	D	A	D	A	B	C	B	C	B	A	C	C	D	A	B	A	C	D	A	A	D	A	B	C	C	
ID084	B	D	D	D	A	B	D	C	A	A	C	D	B	D	D	D	A	C	A	B	A	D	B	A	D	A	B	C	D	A	
ID085	B	B	C	A	D	C	C	A	D	D	A	B	C	D	D	D	A	D	A	A	A	C	B	C	B	A	D	B	D	D	
ID086	B	C	D	A	A	D	B	C	B	D	B	C	A	C	D	A	B	D	C	C	A	B	B	A	D	B	C	B	A	A	
ID087	B	D	A	D	A	A	D	A	A	B	D	B	C	D	B	A	B	A	C	C	C	B	D	A	A	D	A	B	B	A	
ID088	B	D	B	B	B	B	B	A	A	C	B	C	A	C	D	D	C	C	D	D	B	D	B	A	A	D	D	D	C	D	D
ID089	B	A	B	D	B	B	B	A	A	C	B	C	A	C	D	D	C	C	D	D	B	D	B	A	A	D	D	D	C	D	D
ID090	A	C	A	D	A	B	A	C	A	B	C	D	D	D	A	C	D	D	D	A	D	B	C	A	D	D	D	B	B	D	
ID091	C	C	A	D	A	B	A	C	A	B	C	D	D	D	A	C	D	D	D	A	D	B	C	A	D	D	D	B	B	D	
ID092	C	C	A	D	A	B	A	C	A	B	C	D	D	D	A	C	D	D	D	A	D	B	C	A	D	D	D	B	B	D	
ID093	A	B	B	A	A	C	D	D	C	B	C	C	B	C	A	B	C	B	D	B	A	D	A	A	D	A	A	B	C	C	

ID094	A	B	A	A	A	D	B	A	D	B	D	C	D	C	A	A	B	B	D	B	C	D	A	D	A	A	A	A	D	
ID095	D	B	D	A	A	D	D	D	A	C	B	A	C	B	C	C	A	A	B	D	B	D	D	A	D	C	A	A	A	A
ID096	A	B	B	A	B	A	B	B	D	A	C	C	B	C	D	A	B	C	B	C	B	D	C	B	D	A	C	C	B	D
ID097	D	C	A	B	C	D	B	C	A	B	C	A	C	D	B	C	A	B	C	B	C	D	C	C	B	A	B	B	B	A
ID098	A	B	A	D	A	D	D	A	C	B	C	B	C	B	A	C	C	A	B	B	A	A	B	A	D	B	A	B	B	A
ID099	A	B	B	C	A	C	D	B	C	B	C	D	C	B	C	A	B	D	C	B	D	B	A	B	C	A	D	B	A	D
ID100	A	B	C	D	B	A	A	B	B	A	B	C	D	C	B	A	C	D	B	C	A	A	B	C	D	A	A	B	C	B

## B. Output Hasil Analisis dari ITEMAN 4.3

Table 1 presents the specifications and basic information concerning the analysis. This provides important documentation of the setup of the program for historical purposes.

**Table 1: Specifications**

Specification	Value	Specification	Value
Number of examinees	100	Total Items	30
Scored Items	30	Pretest Items	0
Multiple Choice Items	30	Polytomous Items	0
Number of domains	3	External scores	No
Minimum P	0,00	Maximum P	1,00
Minimum item mean	0,00	Maximum item mean	15,00
Minimum item correlation	0,00	Maximum item correlation	1,00
ITEMAN 3.0 Header	No	Exclude omits from option statistics	No
Number of ID columns	6	ID begins in column	1
Responses begin in column	7	Omit character	O
Not Admin character	N	Produce quantile tables	Yes
Correct for spuriousness	Yes	Produce quantile plots	Yes
Save data matrix	No	Include omit codes in matrix	N/A
Include Not Admin codes in matrix	N/A	Include scaled scores for	Domain Scores
Scaling function	Standardized	Scaled score new mean	0,000
Scaled score new SD	1,000	Dichotomous Classification	No
Classify based on	N/A	Cutpoint	N/A
Low group label	Low	High group label	High
Data is delimited by	N/A	Test for DIF	No
Group status is in column	N/A	Ability levels for DIF	N/A
Group 1 code	N/A	Group 2 code	N/A
Group 1 label	N/A	Group 2 label	N/A

## **Summary statistics**

Table 2 presents the summary statistics of the test, for all items, scored items only, and for each domain (content area). Definitions of these statistics are found in the Iteman manual.

**Table 2: Summary statistics**

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
All items	30	19,400	7,395	6	28	0,647	0,504
Scored Items	30	19,400	7,395	6	28	0,647	0,504
1 (fisika)	12	7,490	3,350	0	11	0,624	0,539
2 (biologi)	13	8,810	3,253	2	13	0,678	0,498
3 (kimia)	5	3,100	1,314	0	5	0,620	0,435
Scaled 1	12	0,000	1,000	-2,236	1,048	-	-
Scaled 2	13	0,000	1,000	-2,094	1,288	-	-
Scaled 3	5	0,000	1,000	-2,359	1,446	-	-
Scaled		0,000	0,000	0,000	0,000	-	-

- Berdasarkan tabel 2 di atas menjelaskan item yang dianalisis ada 30 butir soal dan rata-rata 19,400 dan standar baku 7,395 serta Skor minimal 6 dan skor maksimal 28.
- Tingkat kesukaran (Mean P)  
Pada tabel di atas didapat Mean P sebesar 0,647 berdasarkan tabel 6 nilai tersebut soal yang dianalisis memiliki tingkat kesukaran sedang.
- Daya beda (Rpbis)  
Pada tabel di atas didapat Mean Rpbis sebesar 0,504, berdasarkan tabel 7 nilai tersebut soal yang dianalisis memiliki daya beda sangat memuaskan.

Table 3 presents a reliability analysis of the tests. Alpha (also known as KR-20) is the most commonly used index of reliability, and is therefore used to calculate the standard error of measurement (SEM) on the raw score scale. Also presented are three configurations of split-half reliability, first as uncorrected correlations, and then as Spearman-Brown (S-B) corrected correlations. This is because an uncorrected split-half correlation is referenced to a "test" that only contains half as many items as the full test, and therefore underestimates reliability.

**Table 3: Reliability**

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0,924	2,033	0,911	0,870	0,849	0,953	0,931	0,919
1 (fisika)	0,858	1,260	0,737	0,760	0,758	0,848	0,864	0,862
2 (biologi)	0,832	1,333	0,644	0,721	0,797	0,783	0,838	0,887
3 (kimia)	0,545	0,886	0,405	0,444	0,442	0,576	0,615	0,613

- Reliabilitas (Alpha)

Pada tabel di atas didapat Alpha sebesar 0,924, berdasarkan tabel 5 nilai tersebut soal yang dianalisis memiliki reliabilitas sangat tinggi dengan standard error of measurement 2,033

Table 4 presents the item statistics and flags for the item(s) that were flagged during the analysis

**Table 4: Summary Statistics for the Flagged Items**

Item ID	P / Item Mean	R	Flag(s)
item095	0,040	-0,135	K, LR
item110	0,150	-0,106	K, LR
item119	0,230	0,201	K

Table 9 displays the correlations of domain scores.

**Table 9: Correlations for Domain Scores**

Domain	1	2	3
1	1,000	0,865	0,721
2	0,865	1,000	0,697
3	0,721	0,697	1,000

### C. Rekap Hasil Analisis Butir dari ITEMAN 4.3

Item	TK	Ket	DP	Ket	Ops				Ket
item01	0,630	Sedang	0,578	Sangat Memuaskan	0,290	0,630	0,030	0,050	tidak baik
item02	0,720	Mudah	0,699	Sangat Memuaskan	0,020	0,180	0,720	0,080	tidak baik
item03	0,780	Mudah	0,553	Sangat Memuaskan	0,110	0,780	0,060	0,050	tidak baik
item04	0,720	Mudah	0,759	Sangat Memuaskan	0,130	0,030	0,720	0,120	tidak baik
item05	0,040	Sukar	-0,133	Jelek	0,830	0,100	0,030	0,040	tidak baik
item06	0,670	Sedang	0,499	Sangat Memuaskan	0,050	0,670	0,060	0,220	baik
item07	0,700	Sedang	0,608	Sangat Memuaskan	0,700	0,060	0,050	0,190	baik
item08	0,690	Sedang	0,610	Sangat Memuaskan	0,070	0,120	0,120	0,690	baik
item09	0,760	Mudah	0,680	Sangat Memuaskan	0,150	0,760	0,050	0,040	tidak baik
item10	0,540	Sedang	0,521	Sangat Memuaskan	0,210	0,170	0,070	0,540	baik
item11	0,650	Sedang	0,779	Sangat Memuaskan	0,040	0,070	0,240	0,650	tidak baik
item12	0,590	Sedang	0,510	Sangat Memuaskan	0,100	0,100	0,210	0,590	baik

item13	0,860	Mudah	0,525	Sangat Memuaskan	0,010	0,060	0,860	0,070	tidak baik
item14	0,720	Mudah	0,699	Sangat Memuaskan	0,010	0,130	0,720	0,140	tidak baik
item15	0,710	Mudah	0,654	Sangat Memuaskan	0,170	0,050	0,070	0,710	tidak baik
item16	0,760	Mudah	0,749	Sangat Memuaskan	0,090	0,010	0,140	0,760	tidak baik
item17	0,710	Mudah	0,394	Memuaskan	0,110	0,050	0,710	0,130	tidak baik
item18	0,710	Mudah	0,702	Sangat Memuaskan	0,710	0,070	0,030	0,190	tidak baik
item19	0,760	Mudah	0,539	Sangat Memuaskan	0,760	0,050	0,130	0,060	tidak baik
item20	0,150	Sukar	-0,096	Jelek	0,120	0,660	0,150	0,070	tidak baik
item21	0,630	Sedang	0,639	Sangat Memuaskan	0,250	0,630	0,050	0,070	baik
item22	0,730	Mudah	0,535	Sangat Memuaskan	0,050	0,110	0,730	0,110	tidak baik
item23	0,660	Sedang	0,517	Sangat Memuaskan	0,060	0,140	0,660	0,140	baik
item24	0,860	Mudah	0,319	Memuaskan	0,860	0,020	0,090	0,030	tidak baik
item25	0,550	Sedang	0,601	Sangat Memuaskan	0,090	0,210	0,550	0,150	baik
item26	0,820	Mudah	0,483	Sangat Memuaskan	0,100	0,020	0,060	0,820	tidak baik
item27	0,680	Sedang	0,679	Sangat Memuaskan	0,140	0,680	0,080	0,100	baik
item28	0,690	Sedang	0,390	Memuaskan	0,100	0,690	0,180	0,030	tidak baik
item29	0,230	Sukar	0,184	Jelek	0,430	0,110	0,230	0,230	tidak baik
item30	0,680	Sedang	0,581	Sangat Memuaskan	0,680	0,040	0,070	0,210	tidak baik

Berdasarkan tabel di atas ada 8 butir yang baik (baik dari tingkat kesukaran, daya beda, dan efektivitas pengecoh).

Sebagai Staf pak Alim memberikan hasil dari analisis tersebut kepada pak Sun dan memberikan saran untuk **mengadakan pelatihan dalam menyusun soal** yang baik kepada tim penyusun soal try out di Kabupaten Banjarejo.

## DAFTAR PUSTAKA

- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. California: Brooks
- Azwar, Saifuddin. (2012). *Tes Prestasi (fungsi dan pengembangan pengukuran prestasi belajar) edisi II*, Yogyakarta: Pustaka Pelajar Offset.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Wiston, Inc.
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement*. Bergen: Prentice Hall, Inc
- Guyer, R., & Thompson, N.A., 2013. *User's Manual for Iteman 4.3*. Woodbury, MN: Assessment Systems Corporation.
- Huda, N., & Wahyuni, T. S. (2019). Penggunaan Aplikasi Item and Test Analysis (ITEMAN) Pada Soal Try Out UN IPA SMP Tahun 2019. *JPS (Jurnal Pembelajaran Sains)* Volume 4, No. 1, Malang: UM page 1-8.
- Linn, R. (1989). *Educational Measurement* (3<sup>rd</sup> ed). New York: American Council on Education and Mcmillan Publishing Company.
- Mardapi, Djemari. (2008). *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendekian Press.
- Mardapi, Djemari. (2012). *Pengukuran Penilaian dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.

- Mehrens, W. A., & Lehmann, I. J. (1973). *Measurement and evaluation in education and psychology*. New York: Holt, Rinehart and Winston, Inc
- Retnawati, Heri. 2014. Membuktikan validitas instrument dalam pengukuran. FMIPA UNY. Diambil tanggal 27 November 2019, <http://evaluation-edu.com/2014/10/06/membuktikan-validitas-instrumen/>
- Retnawati, Heri. 2015. *Reliabilitas*. FMIPA UNY. Diambil tanggal 27 November 2019, <http://evaluation-edu.com/2015/03/28/estimasi-reliabilitas-skor-hasil-pengukuran/>
- Reynolds, C.R., Livingston, R.B., & Wilson, V. (2010). *Measurement and assessment in education* (2<sup>nd</sup> ed). Upper Saddle River: Pearson.
- Setyawarno, Didik. (2017). Penggunaan Aplikasi Software Iteman (Item and Test Analysis) Untuk Analisis Butir Soal Pilihan Ganda Berdasarkan Teori Tes Klasik. *JIFP (Jurnal Ilmu Fisika dan Pembelajarannya)* Volume 1, No. 1, Palembang: UIN Raden Fatah March 17.
- Sutrisno, Nizaruddin, Yanuar Hery M, Muhtarom. (2018). *ITEMAN: Item Klasik dan Analisis Uji*. Semarang: CEP-CCDA Universitas PGRI Semarang.
- Wright, R. J. (2008) *Educational assessment: test and measurement in the age of accountability*. Los Angeles: Sage Publications.