

Karakteristik Butir Tes dan Analisisnya

Oleh: Heri Retnawati

Pada suatu pengukuran baik di dunia pendidikan maupun sosial, diperlukan instrument yang baik. Untuk memperoleh suatu instrumen yang baik, prosedur pengembangannya melalui prosedur yang tersatndar. Langkah-langkah mengembangkan instrument pengukuran yang baik, untuk bentuk tes maupun nontes sebagai berikut (Heri Retnawati, 2015).

1. Menentukan tujuan penyusunan instrumen

Pada awal menyusun instrumen, perlu ditetapkan apa tujuan penyusunan instrumen. Tujuan penyusunan ini memandu teori untuk mengkonstruk instrumen, bentuk instrumen, penyekoran sekaligus pemaknaan hasil penyekoran pada intrumen yang akan dikembangkan. Tujuan penyusunan instrumen ini perlu disesuaikan dengan tujuan pengukuran.

2. Mencari teori yang relevan atau cakupan materi

Setelah tujuan penyusunan instrumen, selanjutnya perlu dicari teori atau cakupan materi yang relevan. Teori yang relevan digunakan untuk membuat konstruk, apa saja indicator suatu variabel yang akan diukur. Kaitannya dengan tes, perlu dibatasi juga cakupan materi apa saja yang menjadi bahan menyusun tes. Sebagai contoh untuk mengukur kopetensi guru, cakupan materi apa saja yang akan diukur perlu menjadi bahan pertimbangan.

Disampaikan pada Workshop Penyusunan Instrumen Uji Kompetensi Guru, 30 September-1 Oktober 2015 Di PPP4TK Seni Budaya Yogyakarta.

3. Menyusun indikator butir instrumen/soal

Indikator soal ini ditentukan berdasarkan kajian teori yang relevan pada instrumen nontes. Adapun pada instrumen tes, selain mempertimbangkan kajian teori, perlu dipertimbangkan cakupan dan kedalaman materi. Indikator ini telah bersifat khusus, sehingga dengan menggunakan indikator dapat disusun menjadi butir instrumen. Biasanya aspek yang akan diukur dengan indikatornya disusun menjadi suatu tabel. Tabel tersebut kemudian disebut dengan kisi-kisi (*blue print*). Penyusunan kisi-kisi ini mempermudah peneliti menyusun butir soal.

4. Menyusun butir instrumen

Langkah selanjutnya adalah menyusun butir-butir instrumen. Penyusunan butir ini dilakukan dengan melihat indikator yang sudah disusun pada kisi-kisi. Pada penyusunan butir ini, peneliti perlu mempertimbangkan bentuknya. Misal untuk nontes akan menggunakan angket, angket jenis yang mana, dan seterusnya.

Jika peneliti akan menggunakan instrumen berupa tes, perlu dipikirkan apakah akan menggunakan bentuk objektif atau menggunakan bentuk uraian (*construted response*). Pada penyusunan butir ini, peneliti telah mempertimbangkan penskoran untuk tiap butir, sehingga memudahkan analisis. Jika perlu, pedoman penskoran disusun setelah peneliti menyelesaikan penyusunan butir instrumen.

5. Validasi isi (*expert judgement*)

Setelah butir-butir soal tersusun, langkah selanjutnya adalah validasi. Validasi ini dilakukan dengan menyampaikan kisi-kisi, butir instrumen, dan lembar diberikan kepada ahli untuk ditelaah secara kuantitatif dan kualitatif. Tugas ahli adalah melihat kesesuaian indikator dengan tujuan pengembangan instrumen, kesesuaian indikator dengan cakupan materi atau kesesuaian teori, melihat kesesuaian instrumen dengan indikator butir, melihat kebenaran konsep butir soal, melihat kebenaran isi, kebenaran kunci (pada tes), bahasa dan budaya. Proses ini disebut dengan validasi isi dengan ahli (*expert judgement*).

Jika validasi isi akan dikuantifikasi, peneliti dapat meminta ahli mengisi lembar penilaian validasi. Paling tidak, ada 3 ahli yang dilibatkan untuk proses validasi instrumen

penelitian. Berdasarkan isian 3 ahli, selanjutnya penelitian menghitung indeks kesepakatan ahli atau kesepakatan validator dengan menggunakan indeks Aiken atau indeks Gregory.

6. Revisi berdasarkan masukan validator

Biasanya validator memberikan masukan. Masukan-masukan ini kemudian digunakan peneliti untuk merevisinya. Jika perlu, peneliti perlu mengkonsultasikan lagi hasil perbaikan tersebut, sehingga diperoleh instrumen yang benar-benar valid.

7. Melakukan ujicoba kepada responden yang bersesuaian untuk memperoleh data respons peserta

Setelah revisi, butir-butir instrumen kemudian disusun lengkap (dirakit) dan siap diujicobakan. Ujicoba ini dilakukan dalam rangka memperoleh bukti empiris. Ujicoba ini dilakukan kepada responden yang bersesuaian dengan subjek penelitian. Peneliti dapat pula menggunakan anggota populasi yang tidak menjadi anggota sampel.

8. Melakukan analisis (reliabilitas, tingkat kesulitan, dan daya pembeda)

Setelah melakukan ujicoba, peneliti memperoleh data respons peserta ujicoba. Dengan menggunakan respons peserta, peneliti kemudian melakukan penskoran tiap butir. Selanjutnya hasil penskoran ini digunakan untuk melakukan analisis reliabilitas skor perangkat tes dan juga analisis karakteristik butir. Analisis karakteristik butir dapat dilakukan dengan pendekatan teori tes klasik maupun teori respons butir. Analisis pada kedua pendekatan ini akan dibahas pada bab-bab selanjutnya.

9. Merakit instrumen

Setelah karakteristik butir diketahui, peneliti dapat merakit ulang perangkat instrumen. Pemilihan butir-butir dalam merakit perangkat ini mempertimbangkan karakteristik tertentu yang dikehendaki peneliti, misalnya tingkat kesulitan butir. Setelah diberi instruksi pengerjaan, peneliti kemudian dapat mempergunakan instrumen tersebut untuk mengumpulkan data penelitian.

Untuk mendapatkan instrumen berkualitas tinggi, selain dilakukan analisis secara teori (telaah butir berdasarkan aspek isi, konstruksi, dan bahasa) perlu juga dilakukan analisis butir secara empirik. Analisis butir secara empirik ini dapat dibedakan menjadi dua, yaitu dengan pendekatan teori tes klasik dan teori respons butir (*Item Response Theory, IRT*). Pada artikel ini, disajikan analisis untuk mengetahui karakteristik butir berdasarkan teori tes klasik, dan analisis berdasarkan teori respons butir.

Teori tes klasik atau disebut teori skor murni klasik (Allen & Yen, 1979) didasarkan pada suatu model aditif, yakni skor amatan merupakan penjumlahan dari skor sebenarnya dan skor kesalahan pengukuran. Kesalahan pengukuran yang dimaksudkan dalam teori ini merupakan kesalahan yang tidak sistematis atau acak. Kesalahan ini merupakan penyimpangan secara teoritis dari skor amatan yang diperoleh dengan skor amatan yang diharapkan. Kesalahan pengukuran yang sistematis dianggap bukan merupakan kesalahan pengukuran. Karakteristik butir dan instrumen berdasarkan teori tes klasik meliputi indeks tingkat kesukaran, daya pembeda, dan reliabilitas.

Tingkat kesukaran suatu butir soal, yang disimbolkan dengan p_i , merupakan salah satu parameter butir soal yang sangat berguna dalam penganalisisan suatu tes. Hal ini disebabkan karena dengan melihat parameter butir ini, akan diketahui seberapa baiknya kualitas suatu butir soal. Jika p_i mendekati 0, maka soal tersebut terlalu sukar, sedangkan jika p_i mendekati 1, maka soal tersebut terlalu mudah, sehingga perlu dibuang. Hal ini disebabkan karena butir tersebut tidak dapat membedakan kemampuan seorang siswa dengan siswa lainnya.

Allen dan Yen (1979) menyatakan bahwa secara umum indeks kesukaran suatu butir sebaiknya terletak pada interval 0,3 – 0,7. Pada interval ini, informasi tentang kemampuan siswa akan diperoleh secara maksimal. Dalam merancang indeks kesukaran suatu perangkat tes, perlu dipertimbangkan tujuan penyusunan perangkat tes tersebut.

Untuk menentukan indeks kesukaran dari suatu butir pada instrumen dengan penskoran dikotomi, digunakan persamaan sebagai berikut :

$$p_i = \frac{\sum B}{N} \quad (1)$$

dengan :

p = proporsi menjawab benar pada butir soal tertentu.

ΣB = banyaknya peserta tes yang menjawab benar.

N = jumlah peserta tes yang menjawab.

Indeks daya pembeda Untuk menentukan daya pembeda, dapat digunakan indeks diskriminasi, indeks korelasi biserial, indeks korelasi *point biserial*, dan indeks keselarasan. Formula indeks korelasi *point biserial* sebagai berikut.

$$r_{pbis} = \left[\frac{\bar{X}_1 - \bar{X}}{s_x} \right] \sqrt{\frac{p_1}{1-p_1}} \quad (2)$$

dengan r_{pbis} = koefisien korelasi point biserial, X_i merupakan variabel kontinu, \bar{X}_1 merupakan rerata skor X untuk peserta tes yang menjawab benar butir tersebut, \bar{X} merupakan rerata skor X , s_x merupakan standar deviasi dari skor X , dan p_1 merupakan proporsi peserta tes yang menjawab benar butir tersebut.

Pada suatu butir soal, indeks daya beda dikatakan baik jika lebih besar atau sama dengan 0,3. Indeks daya pembeda suatu butir yang kecil nilainya akan menyebabkan butir tersebut tidak dapat membedakan siswa yang kemampuannya tinggi dan siswa yang kemampuannya rendah. Pada analisis tes dengan *Content-Referenced Measures*, indeks daya pembeda butir tidak terlalu perlu menjadi perhatian, asalkan tidak negatif (Frisbie, 2005). Jika nilainya kecil, menunjukkan bahwa kemencengan distribusi skor dari populasi, yang juga mengakibatkan validitas tes menjadi rendah.

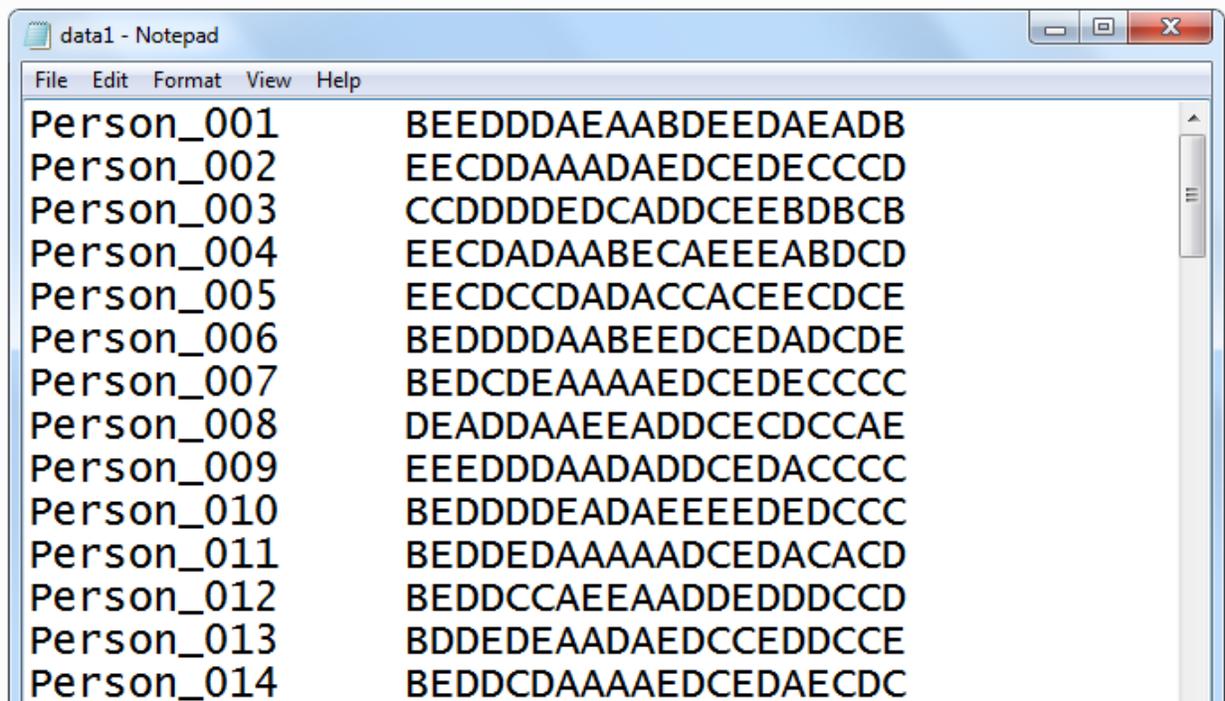
Analisis terkait dengan data empiris yang penting dan perlu menjadi perhatian terkait dengan instrumen pengukuran adalah reliabilitas. Salah satu rumus untuk menganalisis reliabilitas instrumen adalah rumus alpha. Formulanya sebagai berikut.

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right) \quad (3)$$

Dengan α = koefisien reliabilitas instrument; k = banyaknya butir pertanyaan dalam instrument; $\sum \sigma_i^2$ = jumlah varians butir instrument; σ_t^2 = varians skor total.

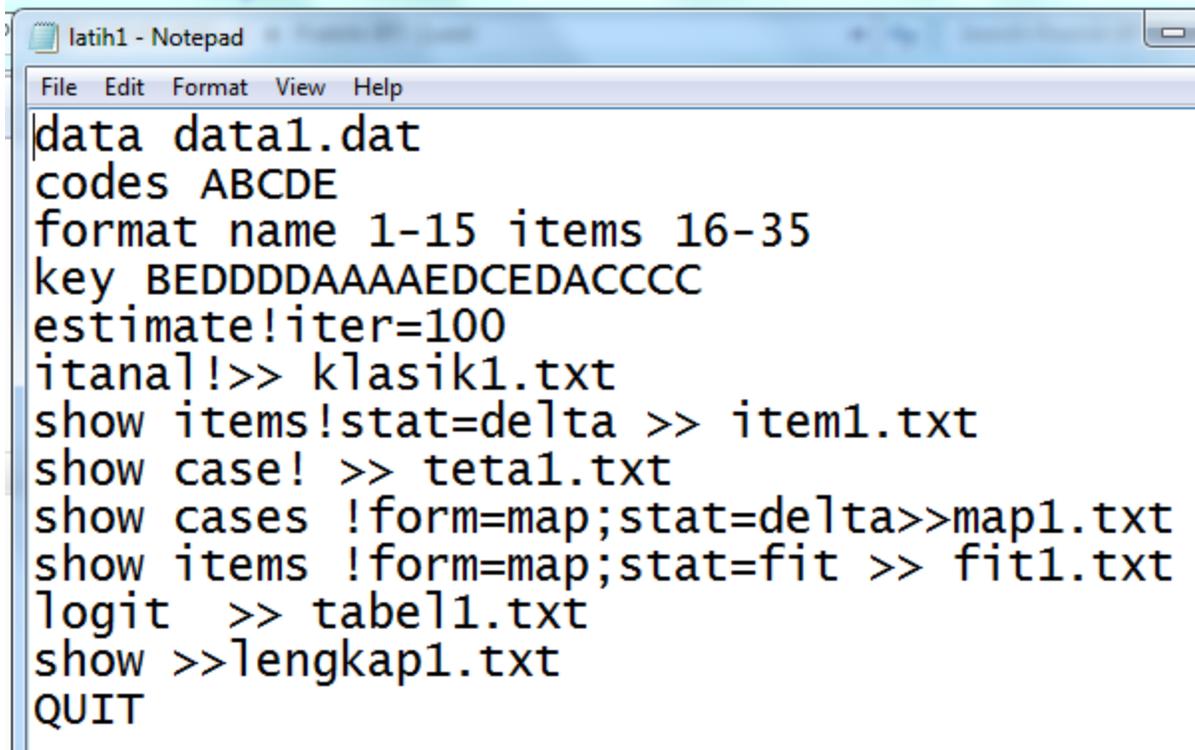
Analisis karakteristik butir berdasarkan teori tes klasik dapat dilakukan dengan program Excel, terutama untuk analisis tingkat kesulitan butir. Untuk analisis daya pembeda, program Excel dapat digunakan namun fungsinya edikit agak rumit. Pada analisis ini, diberikan contoh tingkat kesulitan butir dengan penskoran dikotomi dengan menggunakan Excel, tingkat kesulitan butir politomi dengan menggunakan Excel, tingkat kesulitan dan daya pembeda butir dikotomi dengan program QUEST (Adam & Kho. 1993), dan tingkat kesulitan dan daya pembeda butir politomi dengan program QUEST. Contoh input data, sintaks, dan output sebagai berikut.

Input data dapat melalui Notepad, pada kasus ini 10 karakter pertama merupakan identitas peserta, 20 butir berikutnya merupakan butir yang dianalisis.



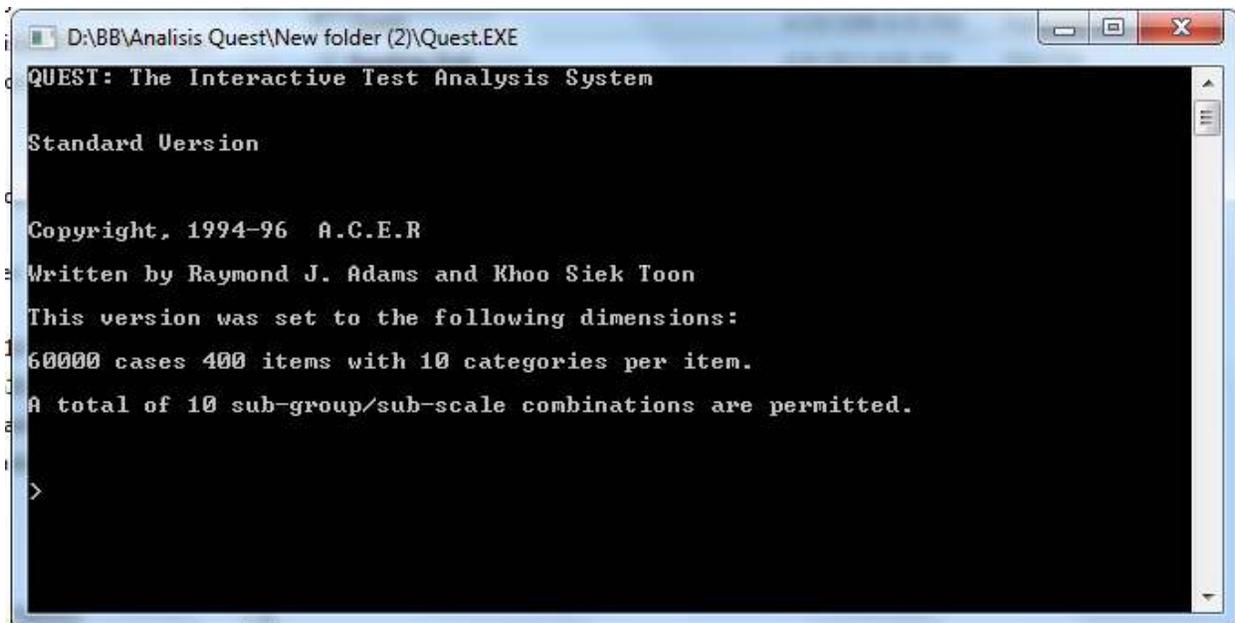
```
data1 - Notepad
File Edit Format View Help
Person_001      BEEDDDAEAABDEEDAADB
Person_002      EECDDAAADAEDCEDECCCD
Person_003      CCDDDEDCADDCEEBCBCB
Person_004      EECDADAABECAEEEEABDCD
Person_005      EECDCCDADACCACEECDCE
Person_006      BEDDDDAABEEDCEDADCDE
Person_007      BEDCDEAAAAEDCEDECCCC
Person_008      DEADDAAEEADDCECDCCAE
Person_009      EEEDDDAADADDCEDACCCC
Person_010      BEDDDDEADAEEEEDEDCCC
Person_011      BEDDEDAAAAADCEDACACD
Person_012      BEDDCCAEEAADDEDDDCCD
Person_013      BDEDEAADAEDCCEDDCCE
Person_014      BEDDCDAAAEDCEDAECDC
```

Sintaks atau baris perintah dapat diketikkan pula dengan Notepad. Contohnya sebagai berikut.



```
data data1.dat
codes ABCDE
format name 1-15 items 16-35
key BEDDDDAAAAEDCEDACCCC
estimate!iter=100
itana!>> klasik1.txt
show items!stat=delta >> item1.txt
show case! >> teta1.txt
show cases !form=map;stat=delta>>map1.txt
show items !form=map;stat=fit >> fit1.txt
logit >> tabel1.txt
show >>lengkap1.txt
QUIT
```

Kemudian data, sintaks dan program QUEST disimpan dalam satu folder. Program QUEST diklik dua kali, kemudian akan muncul tampilan sebagai berikut.



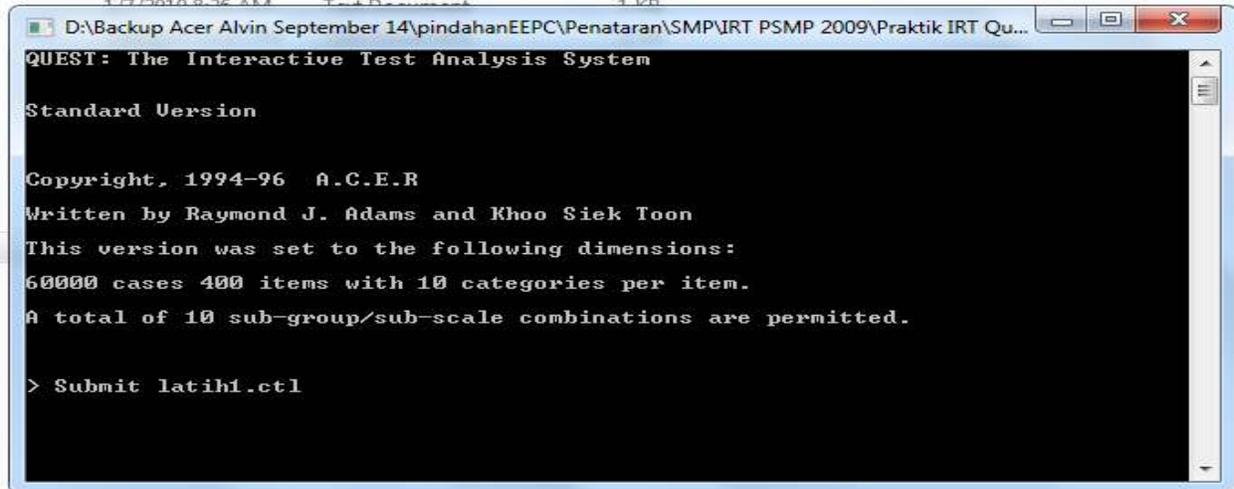
```
D:\BB\Analisis Quest\New folder (2)\Quest.EXE
QUEST: The Interactive Test Analysis System

Standard Version

Copyright, 1994-96 A.C.E.R
Written by Raymond J. Adams and Khoo Siek Toon
This version was set to the following dimensions:
60000 cases 400 items with 10 categories per item.
A total of 10 sub-group/sub-scale combinations are permitted.

>
```

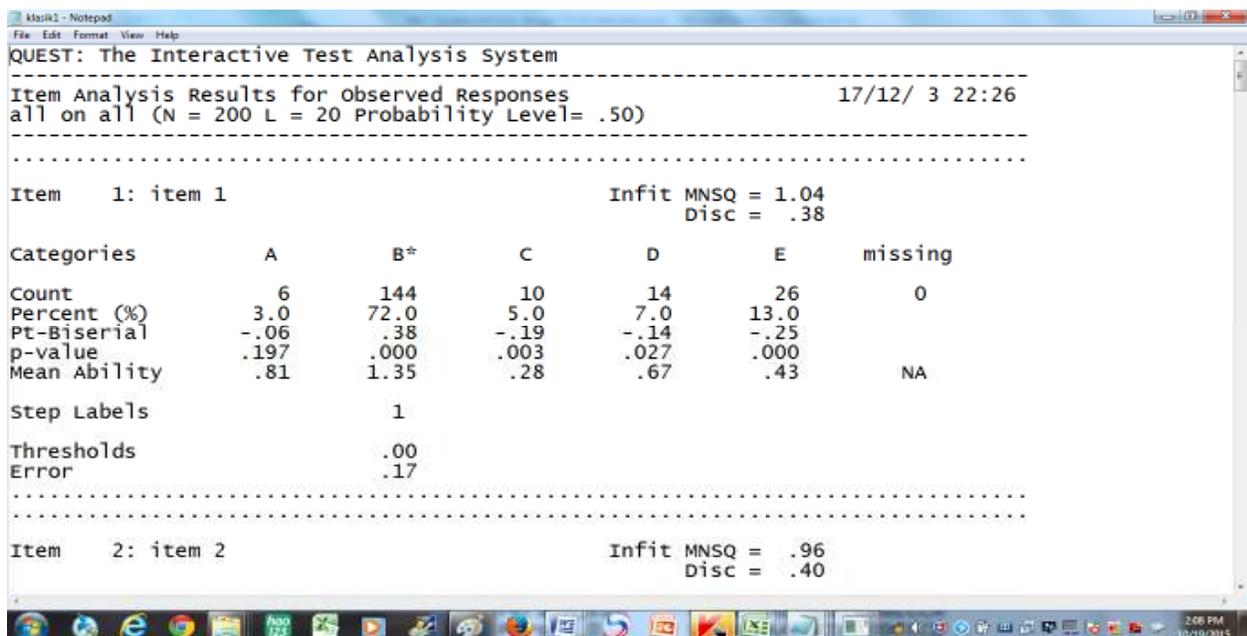
Kemudian diketikkan Submit, spasi nama file sitaksnya berikut ekstensinya.



```
QUEST: The Interactive Test Analysis System
Standard Version
Copyright, 1994-96 A.C.E.R
Written by Raymond J. Adams and Khoo Siek Toon
This version was set to the following dimensions:
60000 cases 400 items with 10 categories per item.
A total of 10 sub-group/sub-scale combinations are permitted.
> Submit latih1.ctl
```

Setelah itu klik Enter, pada folder tempat kita menganalisis file akan bertambah.

Output Klasik.txt merupakan hasil analisis teori tes klasik, Hasilnya sebagai berikut.



```
QUEST: The Interactive Test Analysis System
Item Analysis Results for Observed Responses
all on all (N = 200 L = 20 Probability Level= .50)
17/12/ 3 22:26
Item 1: item 1
Infit MNSQ = 1.04
Disc = .38
Categories A B* C D E missing
Count 6 144 10 14 26 0
Percent (%) 3.0 72.0 5.0 7.0 13.0
Pt-Biserial -.06 .38 -.19 -.14 -.25
p-value .197 .000 .003 .027 .000
Mean Ability .81 1.35 .28 .67 .43 NA
Step Labels 1
Thresholds .00
Error .17
Item 2: item 2
Infit MNSQ = .96
Disc = .40
```

Hal tersebut dapat diinterpretasikan sebagai berikut. Untuk butir 1, B* menunjukkan bahwa butir nomor 1 kuncinya B. Tingkat kesulitan butir nomor 1 adalah 72% atau 0,72. Untuk pilihan lainnya yang bukan kunci, pilihan butir baik jika proporsi nilainya tidak nol atau dengan kata lain

ada yang memilih. Daya pembedanya (Pt-Biserial) sebesar 0,38. Korelasi point biserial ini harganya negatif untuk bukan kunci (distaktor).

Daftar Pustaka

Adams, R. J., & Khoo, S. T. (1993). *Quest: The interactive test analysis system*. Hawthorn: Australian Council for Educational Research.

Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.

Frisbie, D. A. 2005. Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice*, 25 (3), 21-28.

Heri Retnawati. (2015). Analisis kuantitatif instrumen penelitian. Yogyakarta: Parama.



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI YOGYAKARTA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
Jl. Colombo No 1 Yogyakarta 55281 Telp. 0274-588168 Fax 217, 0274-565411(TL)
0274-550217(Dekan), Fax. 0274-548201, Website: <http://www.unj.ac.id/facul>, humas@unijam.unj.ac.id

SURAT IZIN

NO. : 2693/UN34.13/KP/2015

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta memberikan izin kepada :

No	Nama	NIP	Pangkat/Gol.	Jabatan
1.	Dr. Hari Retnowati	19770503 200003 2 001	Penata / IIIa	Lektor

Kepertanian : Sebagai Minisumber Penyusunan Soal Uji Kompetensi Guru (UKG) Mata Pelajaran Seri dan Budaya

Tanggal : 30 September – 01 Oktober 2015

Tempat : Ruang Pantadawa (Micro Teaching) MTK Seri dan Budaya Yogyakarta

Keterangan : Berdasarkan surat dari Kabag Urutan, No. : 1658/013.2/KP/2015,
Tanggal 28 September 2015

Surat izin ini diberikan untuk dilaksanakan sebaik-baiknya dan mohon melaporkan hasilnya kepada Dekan.



Yogyakarta, 12 Oktober 2015

Dekan,

Dr. Harsono
NIP. 19620329 198702 1 002

Tambahan :

1. Wakil Dekan I FMIPA
2. Kajurdi Matematika FMIPA
3. Kabag UKP FMIPA
4. Yang bersangkutan



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
PUSAT PENGEMBANGAN DAN PEMBERDAYAAN
PENDIDIK DAN TENAGA KEPENDIDIKAN SENI DAN BUDAYA**

Jalan Kaliurang Km. 12,5, Klidon, Sukoharjo, Ngaglik, Sleman, Yogyakarta - 55581
Telp. (0274) 895803, 895804, 895805 / Fax (0274) 895804, 895805
Laman : www.pppptk.com Email : puasa@ppppptk.com

Form: SRT 02.1

Nomor : 1639/013.2/KP/2015
Hal : Permohonan menjadi narasumber

28 September 2015

Yth. Dekan Fakultas MIPA
Universitas Negeri Yogyakarta
di Yogyakarta

Dibertahukan dengan hormat, bahwa PPPPTK Seni dan Budaya Yogyakarta sedang menyelenggarakan kegiatan Penyusunan Soal uji Kompetensi Guru (UKG) mata pelajaran Seni dan Budaya.

Selubungan dengan hal tersebut, kami mohon Saudara berkenan memberikan izin sekaligus menunjuk Dr. Ieri Retnowati untuk menjadi narasumber dalam kegiatan dimaksud, pada :

Hari/tanggal : Rabu s.d Kamis, 30 September s.d 01 Oktober 2015
Waktu : Pukul 13.00 – 15.00 WIB
Tempat : Ruang Purnadewa (After Teaching) PPPPTK Seni dan Budaya Yogyakarta
Jalan Kaliurang Km. 12,5, Klidon, Sukoharjo, Ngaglik, Sleman, Yogyakarta

Atas perhatian dan kerjasman yang baik kami ucapkan terima kasih.



Tembusan :
- Kepala PPPPTK Seni dan Budaya,

