



**KEMENTERIAN RISET, TEKNOLOGI, DAN PENDIDIKAN TINGGI**  
**UNIVERSITAS NEGERI YOGYAKARTA**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

*SERTIFIKAT*

No. 17511/UN34.13/TU/2017

Diberikan kepada:

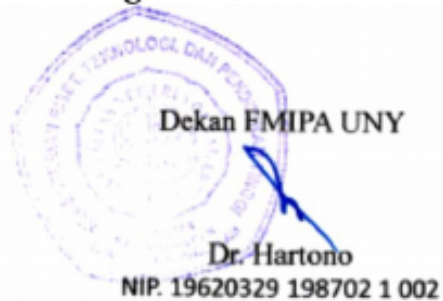
**Didik Setyawarno**

Atas Partisipasinya sebagai:

**Pemateri**

Pada kegiatan pelatihan Guru IPA SMP Sleman Yogyakarta dengan judul kegiatan “WORKSHOP PEMANFAATAN APLIKASI BERBASIS KOMPUTER “CMI-SIPMA SISTEM INFORMASI PENILAIAN SEKOLAH” BAGI GURU IPA TINGKAT SMP DI SLEMAN, YOGYAKARTA” di Laboratorium Komputer IPA

Mengetahui,



Yogyakarta, 9 September 2017

Kajurdik. IPA

Dr. Dadan Rosana, M.Si  
NIP. 19690202 199303 1 002

**UPAYA PENINGKATAN KUALITAS BUTIR SOAL DENGAN  
ANALISIS APLIKASI QUEST**

*Makalah PPM bagi guru IPA SMP di Sleman disampaikan pada hari  
Sabtu, 9 September 2017 di Lab. Komputer IPA*



**Oleh:**

**Didik Setyawarno, M.Pd**

**JURUSAN PENDIDIKAN IPA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**2017**

## BAB I PENDAHULUAN

Aplikasi komputer dalam bidang evaluasi pembelajaran yang akan di bahas pada bagian ini adalah *Quest*. Aplikasi tersebut sangat familiar dikalangan pendidik yang dapat digunakan untuk menganalisis butir soal, sehingga mengetahui kualitas butir soal tersebut dan mampu meningkatkan mutu soal yang telah ditulis. Penelaahan soal secara kuantitatif maksudnya adalah penelaahan butir soal didasarkan pada data empirik dari butir soal yang bersangkutan. Data empirik ini diperoleh dari soal yang telah diujikan. Aspek yang di analisis dalam butir soal meliputi: tingkat kesukaran butir, daya pembeda butir, dan penyebaran pilihan jawaban (untuk soal bentuk obyektif) atau frekuensi jawaban pada setiap pilihan jawaban.

Teori respon butir merupakan teori pengukuran modern yang biasanya digunakan dalam analisis butir soal. Nama lain teori respon butir adalah *latent trait theory (LTT)* atau *characteristics curve theory (CCT)*. Pengembangan teori respon butir didasarkan kepada dua postulat. Postulat pertama adalah bahwa kemampuan subyek (*examinee*) pada suatu butir dapat diprediksi oleh seperangkat faktor yang disebut *traits, latent traits* atau *abilities*. *Trait* adalah dimensi kemampuan seseorang seperti kemampuan verbal, kemampuan psikomotor, kemampuan kognitif, dan sebagainya. Postulat kedua adalah bahwa hubungan antara kemampuan peserta tes pada suatu butir dan perangkat kemampuan laten yang mendasarinya dapat digambarkan melalui kurva karakteristik butir atau *item characteristics curve (ICC)*.

Ada tiga model logistik yang sering digunakan saat ini yaitu: model logistik satu parameter (*model rasch*) atau *item response theory 1-parameter logistic (IRT 1PL)* yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran. Model logistik dua parameter atau *item response theory 2-parameter logistic (IRT 2PL)* yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran dan daya pembeda soal. Model logistik tiga parameter atau *item response theory 3-parameter logistic (IRT 3PL)* yaitu untuk menganalisis data yang menitikberatkan pada parameter tingkat kesukaran soal, daya pembeda soal, dan peluang menebak (*guessing*).

Setelah mempelajari panduan ini, peserta pelatihan memiliki sikap kritis dalam menganalisis data penelitian khususnya analisis butir soal dengan pendekatan teori

test modern/IRT, dan memiliki kemandirian dalam menyelesaikan analisis butir soal secara empiris dengan aplikasi quest.

## BAB II PEMBAHASAN

### A. Quest

Aplikasi *Quest* memiliki keunggulan yang tidak dimiliki oleh *IteMan* dan *AnBuso* yakni menawarkan suatu lingkup analisis kuesioner dan tes secara komprehensif dengan menyediakan suatu analisis data dengan mengarah pada pengembangan yang terbaru dalam teori pengukuran *Rasch*. Selain itu, aplikasi dengan *Quest* dapat digunakan untuk mengkonstruksi dan memvalidasi variabel yang berbasis observasi dikotomis dan politomis, misalnya tes pilihan ganda. Hasil dari analisis dengan aplikasi *Quest* menyajikan estimasi butir, estimasi kemampuan responden, dan fit statistik dapat diakses melalui berbagai tabel dan map yang informatif. Analisis tambahan melaporkan *count*, *presentase*, dan *point biserial* untuk masing-masing respon terhadap masing-masing butir. Output *Quest* juga dapat menghasilkan perbandingan tingkat kesukaran peserta dengan model pada masing-masing butir.

### B. Prosedur Penggunaan Aplikasi Quest

Langkah analisis butir soal pilihan ganda dengan data huruf sebagai berikut.

- a. Ketik semua jawaban peserta tes ke *notepad* dengan format sebagai berikut.

```
001 CCBABCCBBACCBBABACBAAACDBBCDCBDABDDDBBADBAACDCCBCD
002 CCBABCCBBACCBBABACBAAACDBBCDCBDABDDDBBADBAACDCCBCD
003 CCBABCCBBACCBBABACBAAACDBBCDCBDABDDDBBADBAACDCCBCD
004 CCBABCCBBACCBBABACBAAACDBBCDCBDABDDDBBADBAACDCCBCD
... DST
```

- b. Simpan file notepad tersebut satu folder dengan aplikasi *Quest* dengan nama misalnya "ciparay.txt".
- c. Buka aplikasi *Quest* dan masukkan perintah sebagai berikut.

```

title SMA Ciparay (50 ITEM PG dengan 4 alternatif)
data_file ciparay.txt
codes 0ABCD9
format id 1-4 items 5-54
key
CCBABCCBBACCBBABACBAAACDBBCDCBDABDDBBADDBAACDCCCBCD
set width=107 ! page
estimate
show >> ouput1.txt
show items >>output2.txt
show cases >> output3.txt
itanal >> output4.txt
quit

```

### Keterangan

- *title* (judul), tuliskan judul dari file yang akan dianalisis misal ciparay, ujian, test, UAS, atau yang lainnya.
- *data\_file*, nama file yang akan di analisis (format txt) misal ciparay.txt, ujian.txt, test.txt, UAS.txt, atau yang lainnya. Dalam hal ini dapat pula diberi nama dengan ekstensi “namafilename.dat” bila komputer tidak berisi program *adobe flash (macromedia)*.
- *codes*, berisi kode jawaban peserta tes, misalnya tuliskan kode 0 di awal bila dilewati dan 9 bila tidak dikerjakan (*omit*) dan pilihan jawaban misal ABCD atau 1234 di antara 0 dan 9.
- *format id*, identitas peserta tes (misal karakter/spasi dari 1-4), *items* untuk karakter jawaban peserta tes (misal karakter/spasi dari 5-54 ) dalam hal ini hanya menggunakan nomor), dan spasi 5 sampai 54 adalah untuk data sebanyak 50 item
- *key*, kunci jawaban butir soal missalnya CCBABCCBBACCBBABACBAAACDBBCDCBDABDDBBADDBAACDCCC BCD, atau 33212332213322121321113422343241244221421134333234, sesuai bentuk soal.
- *set width*, untuk lebar halaman kertas misal 107 ! page
- *estimate*, perintah untuk menganalisis secara otomatis menurut program *Quest*

- *show ! scale=all >> output1.txt* → untuk menunjukkan hasil analisis secara simultan.
- *show items >> output2.txt* → untuk menunjukkan hasil analisis yang menyajikan informasi tentang item secara singkat (estimasi tingkat kesukaran, nilai *INFIT MNSQ*, nilai *INFIT t*)
- *show cases >> output3.txt* → untuk menunjukkan hasil analisis yang menyajikan informasi testi (skor mentah, estimasi skor kalibrasi, nilai *INFIT MNSQ*, nilai *INFIT t*)
- *itanal ! scale=all >> output4.txt* → untuk menunjukkan hasil analisis yang menyajikan informasi tentang item secara lengkap hasil analisis menurut *Classical Test Theory (CTT)* dan *Item Respon Theory (IRT)*.
- *quit* → kode perintah diakhiri.
- Keterangan: output1, output2, output3, dan output4 adalah nama file hasil analisis yang dapat diganti dengan nama lain, missal hasil1, hasil2, hasil3, dan hasil4, atau terkadang diberi akhiran sh, it, ca, dan tn yang masing-masing menunjukkan output dari perintah show, show items, show cases, dan itanal.

Berikut contoh *running* aplikasi analisis butir soal dengan program Quest.

Prosedur/langkah yang dilakukan untuk menganalisis di Quest yaitu mengetik sintak/perintah di notepad dan disimpan dengan ekstensi **.txt**. Prosedur analisis data tersebut sebagai berikut.

- Buka aplikasi Quest
- Ketik **SUBMIT** spasi nama file perintahnya lengkap dengan kode ekstensinya. Jika nama file perintahnya **analisis.txt** maka perintah yang diketik di Quest sebagai berikut.  
>SUBMIT analisis.txt  
Atau  
>Submit analisis.txt

```
D:\Kuliah Didik 2016\Praktikum Analisis Butir Soal CTT dan IRT Pasca UNY\Quest\quest90_standardv...
QUEST90: The Interactive Test Analysis System
Standard Version
Copyright, 1994-99 A.C.E.R
Written by Raymond J. Adams and Khoo Siek Toon
This version was set to the following dimensions:
100000 cases 1200 items with 10 categories per item.
A total of 60 sub-group/sub-scale combinations are permitted.
> Submit analisis.txt_
```

- Tekan enter
- Hasil analisis dapat dilihat pada satu folder yang sama dimana aplikasi Quest di simpan.

Output analisis yang dibahas pada bagian ini adalah output1, output2, output3, dan output4 (tergantung saat penulisan perintah pada program Quest). Masing-masing file mempunyai penafsiran sendiri-sendiri yang lebih komprehensif dari pada analisis dengan pendekatan klasik.

*Output1* atau hasil keluaran dari perintah *show* untuk menampilkan hasil analisis yang merupakan rangkuman atau garis besar hasil analisis Quest. *Output1* terdapat lima bagian hasil analisis.

Bagian 1 dari *Output1* sebagai berikut.

```
UN SMA Ciparay
-----
Current System Settings      8/12/16 6:58
all on all (N = 130 L = 40 Probability Level= .50)
-----

Data File   = ciparay.txt
Data Format  = id 1-4 items 5-44

Log file    = LOG not on

Page Width  = 80
Page Length = 65
Screen Width = 78
Screen Length = 24

Probability level = .50

Maximum number of cases set at 60000

VALID DATA CODES   0 A B C D E 9

GROUPS

1 all          ( 130 cases ) : All cases

SCALES

1 all          ( 40 items ) : All items

DELETED AND ANCHORED CASES:

No case deletes or anchors

DELETED AND ANCHORED ITEMS:

No item deletes or anchors

RECODES

SCORING KEYS

Score = 1   EACDBEEBCCBCCDDABAAEBEDACEBBBAADEEEAABECC
=====
```

**Keterangan:**

- $N = 130$   $L = 40$  menunjukkan banyak peserta tes 130 orang dengan jumlah item soal 40. Jika jumlah peserta tes terlalu banyak (lima digit atau  $\geq 10.000$ ) akan muncul tanda \*\*\*.
- *Data File* = *ciparay.txt* menunjukkan nama file yang dianalisis, dalam hal ini file yang dianalisis adalah “ciparay.txt”.
- *Data Format* = *id 1-4 items 5-44* menunjukkan bahwa data yang dianalisis karakter/spasi satu sampai empat menunjukkan identitas peserta tes,



sedangkan karakter/spasi lima sampai empat puluh empat menunjukkan jawaban peserta tes.

- *Probability level = .50* menunjukkan bahwa peluang benar atau salah sebesar 0,5. Nilai tersebut didasarkan pada metode kemungkinan maksimum atau *Likelihood Maximum*.
- *No case deletes or anchors* dan *No item deletes or anchors* menunjukkan bahwa tidak ada *case* (*testee* atau peserta tes) maupun *items* (*anchor*) yang dihapus atau tidak disertakan dalam analisis, sehingga semua *case* dan *items* dianalisis oleh program Quest.
- *Score = 1 EACD....* menunjukkan bahwa peserta tes yang menjawab sesuai dengan kunci akan mendapat skor satu.

Bagian 2 dari *Output1* sebagai berikut.

UN SMA Ciparay			
Item Estimates (Thresholds)		8/12/16 6:58	
all on all (N = 130 L = 40 Probability Level= .50)			
Summary of item Estimates			
Mean	.00		
SD	1.25		
SD (adjusted)	1.23		
Reliability of estimate	.96		
Fit Statistics			
Infit Mean Square		Outfit Mean Square	
Mean	1.00	Mean	1.03
SD	.12	SD	.30
Infit t		Outfit t	
Mean	-.10	Mean	.01
SD	1.40	SD	1.20
0 items with zero scores			
0 items with perfect scores			

#### Keterangan:

- *Reliability of item estimate* .96 menunjukkan bahwa reliabilitas tes atau butir soal bernilai 0,96. Nilai reliabilitas berdasarkan *estimasi item* disebut juga reliabilitas sampel. Semakin tinggi nilainya semakin banyak item yang fit atau cocok dengan model yang diujikan. Sebaliknya semakin rendah rendah nilainya semakin banyak item yang tidak fit atau cocok dengan model yang diujikan, sehingga tidak memberi informasi yang diharapkan.
- Penetapan *fit item* secara keseluruhan dengan model dalam program QUEST didasarkan pada besarnya nilai rata-rata *INFIT Mean of Square (INFIT MNSQ)* beserta simpangan bakunya atau nilai rata-rata *INFIT Mean of INFIT t*. Penetapan fit tiap item dengan model dalam program QUEST didasarkan pada besarnya nilai *INFIT MNSQ* atau *nilai INFIT t* item yang bersangkutan. Besarnya kuadrat tengah yang tidak tertimbang (*Unweighted Mean Square*) dalam program QUEST disingkat *OUTFIT MNSQ* maupun kuadrat tengah yang tertimbang (*Wighted Mean Square*) yang diharapkan adalah sebesar 1 dan varians sebesar 0. Sementara besarnya nilai harapan *Mean INFIT t* sama

dengan 0 dengan varians sama dengan 1. Ketentuan *INFIT MNSQ* untuk *Rash Modal* sebagai berikut.

Nilai <i>INFIT MNSQ</i>	Keterangan
> 1,33	Tidak cocok dengan model <i>Rasch</i>
0,77 – 1,33	Cocok dengan model <i>Rasch</i>
< 0,77	Tidak cocok dengan model <i>Rasch</i>

Ketentuan *Outfit t* untuk *Rash Modal* sebagai berikut.

Nilai <i>OUTFIT T</i>	Keterangan
$OUTFIT T \leq 2,00$	Cocok dengan model <i>Rasch</i>
$OUTFIT T \geq 2,00$	Tidak cocok dengan model <i>Rasch</i>

- Hasil analisis Quest diperoleh nilai rata-rata *Infit Mean Square (Infit MNSQ)* adalah 1,00 dengan standar deviasi 0,12 dan nilai rata-rata *Outfit Mean Square* adalah 1,03 dengan standar deviasi 0,30. Jika dilihat lebih teliti dengan mengaitkan standar deviasi, nilai ***Infit Mean Square* adalah  $1,00 \pm 0,12$  atau  $1,00 - 0,12 = 0,88$  sampai  $1,00 + 0,12 = 1,12$** , sedangkan nilai *Outfit Mean Square* adalah  $1,03 \pm 0,30$  atau  $1,03 - 0,30 = 0,70$  sampai  $1,03 + 0,30 = 1,33$ .
- Hasil analisis Quest diperoleh nilai *INFIT Mean of INFIT t* adalah -0,10 dengan standar deviasi 1,40 dan nilai *OUTFIT Mean of OUTFIT t* adalah 0,01 dengan standar deviasi 0,12. Jika dilihat lebih teliti dengan mengaitkan standar deviasi, nilai *INFIT Mean of INFIT t* adalah  $-0,10 \pm 0,14$  atau  $-0,10 - 0,14 = -0,24$  sampai  $-0,10 + 0,14 = 0,04$ , sedangkan nilai ***OUTFIT Mean of OUTFIT t* adalah  $0,01 \pm 0,12$  atau  $0,01 - 0,12 = -0,11$  sampai  $0,01 + 0,12 = 0,13$** .
- Kesimpulan: berdasarkan hasil analisis diperoleh *infit meansquare* berada dalam rentang 0,77 -1,30 dan nilai *outfit t*  $\leq 2$ , sehingga secara keseluruhan items sesuai dengan *Rash Model*.

Bagian 3 dari *Output1* sebagai berikut.

UN SMA Ciparay			
Case Estimates		8/12/16 6:58	
all on all (N = 130 L = 40 Probability Level= .50)			
Summary of case Estimates			
Mean	.85		
SD	.88		
SD (adjusted)	.78		
Reliability of estimate	.78		
Fit Statistics			
Infit Mean Square		Outfit Mean Square	
Mean	.97	Mean	1.03
SD	.21	SD	.42
Infit t		Outfit t	
Mean	-.01	Mean	.15
SD	1.04	SD	.97
0 cases with zero scores			
0 cases with perfect scores			

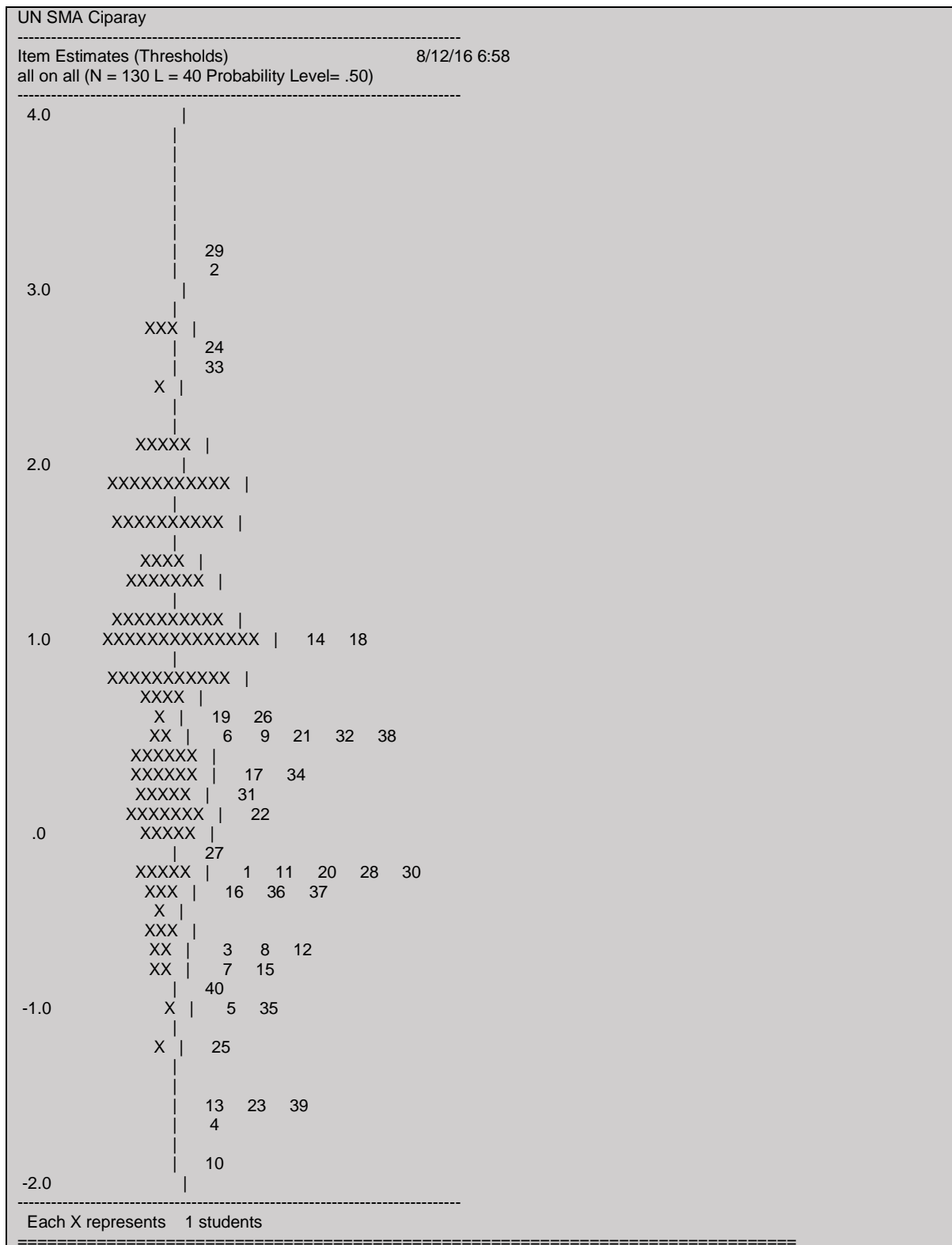
#### Keterangan:

- *Reliability of case estimate* .78 menunjukkan bahwa reliabilitas tes untuk tes buatan guru bernilai 0,78. Nilai tersebut termasuk tinggi artinya jika dilakukan pengulangan tes akan menghasilkan hasil yang stabil. Semakin tinggi nilainya semakin meyakinkan bahwa pengukuran memberikan hasil yang konsisten. Peserta tes atau *testee* yang mendapatkan skor tinggi akan benar pada setiap skor, dan peserta tes atau *testee* yang skor rendah akan salah pada setiap skor. Nilai reliabilitas rendah jika peserta tes tidak konsisten. Artinya pada suatu item tertentu ada banyak siswa yang berkemampuan tinggi justru salah mengerjakannya, dan sebaliknya banyak siswa yang berkemampuan rendah yang benar mengerjakannya.
- Penetapan *fit case* secara keseluruhan dengan model dalam program QUEST didasarkan pada besarnya nilai *INFIT Mean of Square (INFIT MNSQ)* beserta simpangan bakunya. Ketentuan *INFIT MNSQ* untuk *Rash Modal* sebagai berikut.

Nilai <i>INFIT MNSQ</i>	Keterangan
> 1,33	Tidak cocok dengan model <i>Rasch</i>
0,77 – 1,33	Cocok dengan model <i>Rasch</i>
< 0,77	Tidak cocok dengan model <i>Rasch</i>

- Hasil analisis Quest diperoleh nilai rata-rata *Infit Mean Square (Infit MNSQ)* adalah 0,97 dengan standar deviasi 0,21. Jika dilihat lebih teliti dengan mengaitkan stndar deviasi, nilai ***Inffit Mean Square*** adalah  $0,97 \pm 0,21$  atau  $0,97 - 0,21 = 0,76$  sampai  $0,97 + 0,21 = 1,18$ .
- Kesimpulan: berdasarkan hasil analisis diperoleh *infit meansquare* 0,76-1,18 sehingga ada sebagian tidak berada dalam rentang 0,77 -1,33, sehingga secara ada sebagian *testee* tidak sesuai dengan *Rash Model*.

Bagian 4 dari *Output1* sebagai berikut.

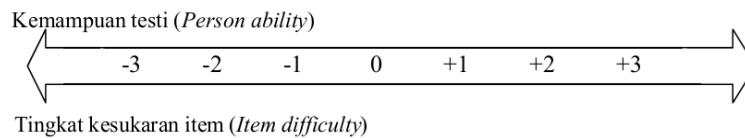


**Keterangan:**

- Peta distribusi tingkat kesulitan butir dan kemampuan responden dapat dilihat pada file map (*Item Estimates (Thresholds)*). File ini menyajikan persebaran

responden menurut tingkat kesukarannya dalam skala logit -4,0 sampai +4,0. Dari peta ini dapat dilihat tingkat kesukaran butir jika dibandingkan dengan kemampuan respondennya.

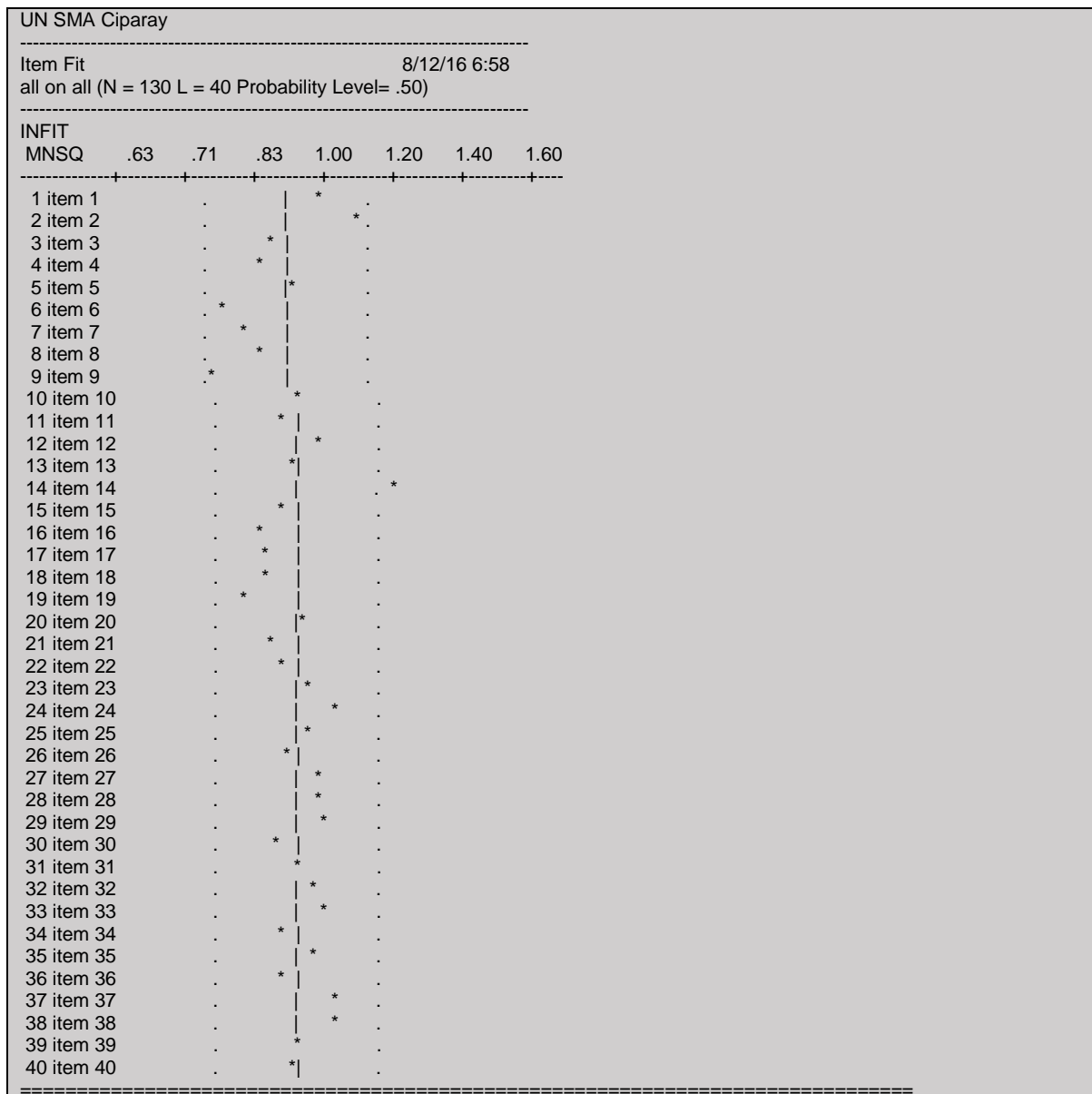
- Tingkat kemampuan testi maupun tingkat kesukaran item dalam *Rasch Model (RM)* diekspresikan pada satu garis berupa absis pada grafik dengan satuan berupa logit (*logg -odd unit*). Garis tersebut terbentang dari  $-\infty$  sampai dengan  $+\infty$ , sebagaimana gambar berikut.



Gambar grafik kemampuan peserta tes yang ideal menunjukkan kurva normal artinya peserta tes dengan kemampuan rendah dan tinggi berjumlah sedikit, dan kemampuan sedang adalah yang berjumlah banyak.

- Tingkat kesukaran butir atau threshold menunjukkan bahwa butir soal nomor 29 adalah yang paling sukar, sedangkan nomor 10 adalah yang paling mudah.

Bagian 5 dari *Output1* sebagai berikut.



**Keterangan:**

Butir soal yang fit dengan model *Rash* berada pada rentang 0,77 -1,33. Berdasarkan grafik diatas dari 40 butir soal terlihat bahwa butir nomor 14 berada di luar rentang 0,77 -1,33, terlihat dari butir tersebut di luar garis. Sedangkan 39 butir soal lain berada di dalam garis tersebut, artinya butir soal tersebut fit dengan model *Rash*.





Nilia Theshold	Keterangan
$b > 2$	Sangat Sukar
$1 < b \leq 2$	Sukar
$-1 < b \leq 1$	Sedang
$-1 > b \geq 2$	Mudah
$b < -2$	Sangat Mudah

- *INFT MNSQ*, *OUTFT MNSQ*, *INFT t*, dan *OUTFT t* menunjukkan penetapan *fit item* setiap butir atau items dengan model. Ketentuan *INFIT MNSQ* untuk *Rash Modal* sebagai berikut.

Nilai <i>INFIT MNSQ</i>	Keterangan
$> 1,33$	Tidak cocok dengan model <i>Rasch</i>
$0,77 - 1,33$	Cocok dengan model <i>Rasch</i>
$< 0,77$	Tidak cocok dengan model <i>Rasch</i>

Ketentuan *Outfit t* untuk *Rash Modal* sebagai berikut.

Nilai <i>OUTFIT T</i>	Keterangan
$OUTFIT T \leq 2,00$	Cocok dengan model <i>Rasch</i>
$OUTFIT T \geq 2,00$	Tidak cocok dengan model <i>Rasch</i>

*Output3* merupakan hasil keluaran dari perintah *show cases* yang menyajikan informasi tentang peserta tes atau case secara singkat (skor, nilai *INFIT MNSQ*, nilai *INFIT t*).

Berikut disajikan hasil *Output3* sebagai berikut.

UN SMA Ciparay									
Case Estimates In input Order					8/12/16 6:58				
all on all (N = 130 L = 40 Probability Level= .50)									
NAME	SCORE	MAXSCR	ESTIMATE	ERROR	INFT	OUTFT	INFT	OUTFT	
			MNSQ	MNSQ	t	t			
1 001	36	40	2.82	.61	.63	.28	-.89	-.93	
2 002	36	40	2.82	.61	.63	.28	-.89	-.93	
3 003	35	39	2.80	.61	.65	.36	-.81	-.72	
4 004	34	40	2.20	.52	.93	1.25	-.08	.58	
5 005	34	40	2.20	.52	.60	.54	-1.18	-.70	
6 006	34	40	2.20	.52	.60	.50	-1.20	-.80	
7 007	34	39	2.45	.56	.46	.23	-1.63	-1.37	
8 008	34	40	2.20	.52	1.23	1.20	.71	.51	
9 009	34	40	2.20	.52	.60	.55	-1.18	-.66	
-----dan seterusnya-----									
126 126	15	40	-.72	.36	1.04	.93	.31	-.04	
127 127	14	40	-.85	.36	1.04	1.01	.32	.15	
128 128	14	40	-.85	.36	1.37	1.36	2.38	.92	
129 129	13	40	-.98	.37	.96	1.83	-.23	1.67	
130 130	11	40	-1.27	.38	.97	.84	-.12	-.16	
Mean			.85	.97	1.03	-.01	.15		
SD			.88	.21	.42	1.04	.97		

**Keterangan:**

- *SCORE-MAXSCR* berturut-turut menunjukkan skor yang diperoleh oleh setiap peserta dari skor maksimal. Misal peserta pertama dengan nama 001 menunjukkan “36 40” menunjukkan bahwa peserta tes tersebut mendapatkan skor 36 dari 40, atau dengan bahasa lain dari 40 butir soal peserta tersebut mampu menjawab benar 36 butir soal.
- *ESTIMATE ERROR* menunjukkan kemampuan rata-rata peserta tes atau *testee*. Estimasi kemampuan peserta tes dapat dilihat pada tabel berikut.

Nilai <i>Estimate</i>	Keterangan
> 1,00	Kemampuan Tinggi
-1,00 s.d +1,00	Kemampuan Sedang
< -1,00	Kemampuan Rendah

- *INFT MNSQ, OUTFT MNSQ, INFT t, dan OUTFT t* menunjukkan penetapan *fit case* setiap peserta tes atau *testee* dengan model. Ketentuan *INFT MNSQ* untuk *Rash Modal* sebagai berikut.

Nilai <i>INFT MNSQ</i>	Keterangan
> 1,33	Tidak cocok dengan model <i>Rasch</i>
0,77 – 1,33	Cocok dengan model <i>Rasch</i>
< 0,77	Tidak cocok dengan model <i>Rasch</i>

Ketentuan *Outfit t* untuk *Rash Modal* sebagai berikut.

Kreteria	Keterangan Peserta Tes/ <i>Testee</i>
<i>Outfit t</i> ≤ 2,00	Peserta Lolos
<i>Outfit t</i> ≥ 2,00	Peserta Gugur

*Output4* merupakan hasil analisis dari perintah *itanal* yang menunjukkan hasil analisis yang menyajikan informasi tentang item secara lengkap hasil analisis menurut *Classical Test Theory* (CTT) dan *Item Respon Theory* (IRT).

Berikut disajikan hasil *Output4* sebagai berikut.

UN SMA Ciparay						
Item Analysis Results for Observed Responses all on all (N = 130 L = 40 Probability Level= .50)						8/12/16 6:58
.....						
Item 1: item 1	Infit MNSQ = 1.12 Disc = .21					
Categories	0	A	B	C	D	E*
Count	0	10	12	5	10	93
Percent (%)	.0	7.7	9.2	3.8	7.7	71.5
Pt-Biserial	NA	-.12	.15	-.31	-.17	.21
p-value	NA	.082	.045	.000	.024	.008
Mean Ability	NA	.48	1.28	-.40	.34	.96
Step Labels	1					
Thresholds	-.21					
Error	.21					
.....						
.....dan seterusnya.....						
.....						
Item 40: item 40	Infit MNSQ = .98 Disc = .34					
Categories	0	A	B	C*	D	E
Count	0	8	1	108	11	2
Percent (%)	.0	6.2	.8	83.1	8.5	1.5
Pt-Biserial	NA	-.33	.01	.34	-.15	-.03
p-value	NA	.000	.454	.000	.041	.371
Mean Ability	NA	-.21	.85	.98	.41	.60
Step Labels	1					
Thresholds	-.95					
Error	.25					
.....						
Mean test score	26.32					
Standard deviation	5.67					
Internal Consistency	.80					
The individual item statistics are calculated using all available data.						
The overall mean, standard deviation and internal consistency indices assume that missing responses are incorrect. They should only be considered useful when there is a limited amount of missing data.						

**Keterangan:**

- *Count* menunjukkan jumlah peserta tes yang memilih jawaban
- *Percent* menunjukkan indeks kesukaran butir soal
- *Point Biserial* menunjukkan indeks daya beda butir soal
- *Mean Ability* menunjukkan estimasi kemampuan responden dalam menjawab butir soal. Kemampuan responden dilihat dari banyaknya butir yang dapat dijawab dengan benar. Semakin banyak butir yang dijawab dengan benar maka

kemampuan responden semakin tinggi. Estimasi kemampuan peserta tes dapat dilihat pada tabel berikut.

Nilai <i>Estimate</i>	Keterangan
> 1,00	Kemampuan Tinggi
-1,00 s.d +1,00	Kemampuan Sedang
< -1,00	Kemampuan Rendah

- *Thresholds* = parameter tingka kesukaran butir dalam skala logit (dalam program lain juga disebut *Delta*)
- Hasil analisis menurut teori tes klasik menunjukkan untuk item nomor 1 dengan kunci jawaban E memiliki indek kesukaran 0,72 (atau 71,5%), sehingga soal tergolong mudah. Indeks daya beda item nomor 1 sebesar 0,21, sehingga soal tersebut perlu diperbaiki.
- Perhatikan bagian akhir analisis dari *Output4* tersebut. Pada bagian akhir akhir analisis disajikan nilai *internal consistency* sebesar 0,80. Nilai ini adalah nilai reliabilitas tes menurut teori tes klasik yang dihitung menurut Indeks Reliabilitas Kuder-Richardson-20 (KR-20).

### BAB III PENUTUP

Aplikasi *Quest* merupakan aplikasi analisis butir soal yang dikembangkan berdasarkan statistic terapan dengan berlandaskan teori respon butir. Teori respon butir merupakan teori pengukuran modern yang biasanya digunakan dalam analisis butir soal. Nama lain teori respon butir adalah *latent trait theory (LTT)* atau *characteristics curve theory (CCT)*. Pengembangan teori respon butir didasarkan kepada dua postulat. Postulat pertama adalah bahwa kemampuan subyek (*examinee*) pada suatu butir dapat diprediksi oleh seperangkat faktor yang disebut *traits, latent traits* atau *abilities*. *Trait* adalah dimensi kemampuan seseorang seperti kemampuan verbal, kemampuan psikomotor, kemampuan kognitif, dan sebagainya. Postulat kedua adalah bahwa hubungan antara kemampuan peserta tes pada suatu butir dan perangkat kemampuan laten yang mendasarinya dapat digambarkan melalui kurva karakteristik butir atau *item characteristics curve (ICC)*. Model logistik dipelajari dalam kegiatan PPM ini yaitu: model logistik satu parameter (*model rasch*) atau *item response theory 1-parameter logistic (IRT 1PL)* yaitu untuk menganalisis data yang

hanya menitikberatkan pada parameter tingkat kesukaran. Setelah mempelajari panduan ini, peserta pelatihan memiliki sikap kritis dalam menganalisis data penelitian khususnya analisis butir soal dengan pendekatan teori test modern/IRT, dan memiliki kemandirian dalam menyelesaikan analisis butir soal secara empiris dengan aplikasi quest.

#### **DAFTAR PUSTAKA**

- Adam, R.J. & Kho, Seik-Tom. 1996. *Acer Quest Version 2.1*. Camberwell, Victoria: The Austrian Council for Education Research.
- Bambang Subali & Pujiati Suyata. 2012. *Pengembangan Item Tes Konvergen dan Divergen*. Yogyakarta: Diandra.
- Dadan Rosana & Didik Setyawarno. 2017. *Statistik Terapan untuk Penelitian Pendidikan*. Yogyakarta: UNY Press.
- Dadan Rosana. 2014. *Evaluasi Pembelajaran Sains (Asesmen Pendekatan Saintifik Pembelajaran Terpadu)*. Yogyakarta: UNY Press.
- Hambleton, R.K. & Swaminathan, H. 1985. *Item Response Theory*. Boston: Kluwer Nijhoff Publiser.
- I Wayan Koyan. 2012. *Konstruksi Tes*. Bali: Undiksha Press.

## UJI KOMPETENSI

Berikut disajikan jawaban siswa untuk 25 butir soal pada UAS SMP kelas IX untuk 25 siswa (beberapa butir tidak di jawab oleh siswa).

NAMA LENGKAP	BUTIR SOAL																								
	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
Person 01	A	B	D	C	E	B	D	E	A	B	C	D	E	B	C	D	E	E	A	A	E	A	D	A	C
Person 02	E	C	D	E	B	C	D	E	A	B	E	A	C	B	E	E	C	E	A	C	E	B	A	E	C
Person 03	B	B	B	D	E	C	C	B	A	D	D	E	A			A	B	C	B	E	E	D	A	C	A
Person 04	B	A	D	C	E	B	B	C	B	B	B	D	B	B	A	D	E	A	B	C	C	C	B	A	A
Person 05	A	B	D	C	A	B	E	E	A	B	B	A	B	B	C	C	D	B	A	A	E	C	C	A	D
Person 06	B	B	D	C	E	B	B	E	B	B	A	D	A	B	C	D	E	B	C	A	E	C	C	A	C
Person 07	A	C	D	D	E		E	E	A	B	C	D	E	B	E	D	E	D	C	C	E	D	D	A	E
Person 08	B	C	B	A	A	A	C	A	B	D	C	A	D	B	C	A	B	D	A	B	E	D	A	D	B
Person 09	D	B	D	C	E	B	B	E	A	B	C	D	C	B	C	D	E	B	A	A	E	C	C	D	C
Person 10	C	B	D	C	E	B	E	E	A	B	C	D	C	C	C	D	B	B	A	A	D	C	C	A	C
Person 11	C	B	B	C	E	B	D	E	B	B	B	D	A	B	E	D	E	A	C	B	E	C	E	A	C
Person 12	C	B	B	C	E	B	A	A	B	C	C	B	D	B	E	D	E	B	B	B	E	C	C	D	D
Person 13	E	B	B	A	D	A	E	E	A	B	C	D	A	B	E	D	E	C	E	A	E	D	A	A	C
Person 14	B	B	B	C	C	D	E	E	A	B	C	D	B	B	E	A	A	C	D	A	E	D	D	A	A
Person 15	A	C	B	C	D	B	E	E	A	B	C	D	E	B	C	C	E	B	C	A	A	D	D	A	A
Person 16	B	B	D	C	E	B	D	E	A	B	C	D	A	B	C	D	E	E	C	B	E	C	D	A	C
Person 17	B	B	B	C	E	E	E	E	A	B	C	D	C	B	E	D	E	C	B	A	E	D	D	A	A
Person 18	A	B	C	C	E	B	E	E	A	B	C	D	E	B	A	D	E	D	D	A	E	D	D	A	C
Person 19	B	B	D	C	D	B	C	E	B	D	B	D	C	D	E	D	C	A	D	A	E	C	A	A	C
Person 20	A	B	D	C	E	C	B	E	B	E	B	D	C	C	C	A	E	D	A	A	E	C	B	A	C
Person 21		B	D	C		B	E	E		B	C	D		B	C	D	E	E		A	E	D	D	A	C
Person 22	B	B	B	E	E	B	A	B	B	D	A	A	A	C	C	D	C	E	C	B	E	B	D	D	E



Person 23	A	A	D	C	E	D	C	E	A	B	D	D	A	D	A	B	D	D	C	C	E	D	C	A	C
Person 24	B	B	B	A	B	B	C	A	B	C	C	D	A	B	E	D	E	A	B	B	E	C	D	A	B
Person 25	B	B	E	C	A	A	C	B	E	A	D	B	E	D	D	C	A	A	E	B	E	B	B	A	A
Person 26	D	B	B	A	A	B	C	E	B	B	C	A	C	C	E	B	B	B			E	D	B	A	C
Person 27	B	D	B	C	E	B	D	C	B	A	B	D	A	C	E	D	E	C	C	C	C	D	D	D	C
Person 28	B	D	C	A	E	E	A	E	E	B	D	B	C	B	E	D	E	B	B	A	E	D	D	C	C
Person 29	B	A	D	C	E	E	D	E	A	E	D	D	A	B	C	C	A	B	A	A	E	C	C	A	C
Person 30	C	B	E	C	E	B	E	E	A	B	A	D	A	B	C	A	E	B	A	A	E	C	C	A	C

Kunci jawaban dari soal yang digunakan sebagai berikut.

Soal	Kunci
Soal1	C
Soal2	B
Soal3	D
Soal4	C
Soal5	E
Soal6	B
Soal7	E
Soal8	E
Soal9	A
Soal10	B

Soal	Kunci
Soal11	C
Soal12	D
Soal13	A
Soal14	B
Soal15	C
Soal16	D
Soal17	E
Soal18	B
Soal19	A
Soal20	A

Soal	Kunci
Soal21	E
Soal22	C
Soal23	C
Soal24	A
Soal25	C
Soal26	D
Soal27	A
Soal28	C
Soal29	B
Soal30	C

Soal	Kunci
Soal31	A
Soal32	D
Soal33	E
Soal34	D
Soal35	C
Soal36	D
Soal37	B
Soal38	C
Soal39	D
Soal40	E

Analisis butir soal di atas dengan pendekatan CTT dan IRT dengan aplikasi Quest dan buatlah grafik hubungan kemampuan *testee* terhadap peluang menjawab dengan model IRT-1PL