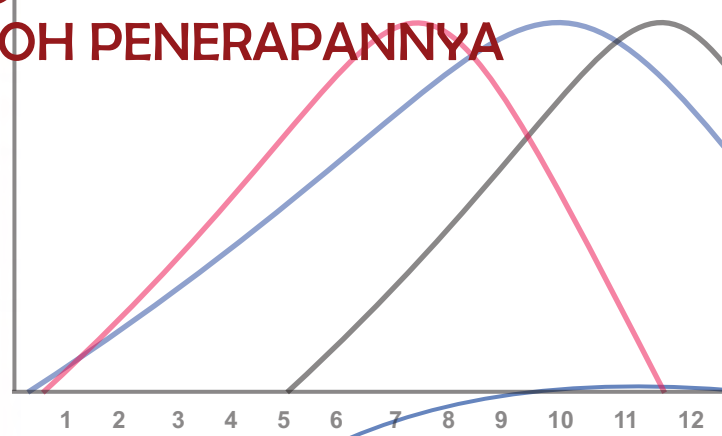


$S_f$ ( $10^{-4}$ )	$S_f$ ( $10^{-4}$ )	$\Delta X$ (m)	$hf$ (m)	$E_2$ (m)
-9-	-10-			
0,544				
0,747	0,648	200	0,019	4,536
1,045	0,896	200	0,018	4,054
1,491	1,268	200	0,025	4,080
2,169	1,830	200	0,037	4,117
3,210	2,689	200	0,054	4,171
4,785	3,997	200	0,080	4,251
7,072	5,822	200	0,118	4,368
9,911	8,200	200	0,168	4,536
13,411	11,171	200	0,221	4,757
17,681	14,785	200	0,263	5,021
22,731	19,097	200	0,286	5,307
28,571	24,169	200		5,622

# TEORI TES KLASIK & TEORI RESPON BUTIR

## KONSEP & CONTOH PENERAPANNYA



Sumaryanta

**TEORI TES KLASIK &  
TEORI RESPON BUTIR**  
KONSEP & CONTOH PENERAPANNYA

Sumaryanta

CV. CONFIDENT  
Unggul & Luhur

**Perpustakaan Nasional RI. Data Katalog dalam Terbitan (KDT)**

**Sumaryanta**

Teori Tes Klasik & Teori Respon Butir: Konsep & Contoh Penerapannya/  
Sumaryanta,; --  
Cirebon: Confident, 2021.

Cetakan Pertama, Juni 2021  
iv + 134 hlm. ; 25 cm

ISBN 978-623-6834-42-8

- 1. TEORI TES KLASIK & TEORI RESPON BUTIR: KONSEP & CONTOH PENERAPANNYA. I. Judul**
- II. Sumaryanta**

---

Penulis : Sumaryanta

Desain Cover : Reza Oktiana Akbar

Tata Letak : Reza Oktiana Akbar

ISBN : 978-623-6834-42-8

Penerbit : CV. Confident (Anggota IKAPI Jabar)

Alamat : Jl. Karang Anyar No.17 Jamblang Kab. Cirebon 45156

---

**Hak cipta dilindungi undang-undang  
Dilarang memperbanyak buku ini dalam bentuk dan dengan cara apapun  
tanpa ijin tertulis dari penulis dan penerbit**

## KATA PENGANTAR

Puji syukur penulis panjatkan atas segala rahmat dan nikmat-NYA sehingga penulisan buku berjudul “*Teori Tes Klasik & Teori Respon Butir: Konsep dan Contoh Penerapannya*” ini bisa terselesaikan. Buku ini disusun memanfaatkan sebagian hasil disertasi penulis saat menyelesaikan Pendidikan Doktoral (S3) di Program Studi Penelitian dan Evaluasi Pendidikan Pascasarjana Universitas Negeri Yogyakarta.

Teori tes klasik dan teori respons butir merupakan teori utama yang banyak digunakan untuk pengembangan tes. Walaupun teori tes klasik dan teori respon butir memiliki perbedaan satu sama lain, tetapi keduanya masih eksis sampai saat ini. Buku ini memuat kajian konseptual terhadap kedua teori tes tersebut disertai dengan contoh penerapannya. Kehadiran buku ini diharapkan dapat menambah wawasan bagi para pembaca sekaligus menginspirasi untuk pemanfaatannya bagi pembaca yang sedang mengembangkan soal.

Penulis menyadari masih terdapat kekurangan dalam penulisan buku ini. Masukan – saran perbaikan dari pembaca sangat penulis harapkan untuk penyempurnaan buku ini. Semoga Allah SWT melimpahkan petunjuk dan hidayah-NYA untuk kita semua.

Yogyakarta, 26 Juni 2021

Penulis  
(Sumaryanta)

## **UCAPAN TERIMA KASIH**

*Buku ini tersusun tidak lepas dari kontribusi berbagai pihak. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah berkontribusi mendukung terselesaikannya buku ini. Ucapan terima kasih secara khusus penulis sampaikan kepada:*

- 1. Seluruh keluarga penulis atas segala pengertian dan dukungannya sehingga penulis dapat menyelesaikan studi dan menulis buku ini.*
- 2. Prof. Djemari Mardapi, Ph.D. (alm.), Prof. Dr. Badrun Kartowagiran, M.Pd., dan Prof. Dr. Drs. Sugiman, M.Si. selaku pembimbing disertasi penulis atas segala arahan dan masukannya.*
- 3. Kepala PPPPTK Matematika beserta seluruh jajarannya yang telah mendukung dan memfasilitasi pelaksanaan penelitian disertasi penulis.*
- 4. Para widyaiswara di PPPPTK Matematika atas segala masukan dan kontribusinya selama penulis menyelesaikan disertasi.*
- 5. Teman-teman mahasiswa seangkatan di Prodi Penelitian dan Evaluasi Pendidikan (S3) Program Pascasarjana Universitas Negeri Yogyakarta atas semua kebersamaan dan dukungannya selama penulis menyelesaikan studi dan disertasi.*
- 6. Semua pihak lain yang tidak dapat penulis sebutkan satu persatu yang telah membantu hingga terselesaikannya buku ini.*

*Semoga amal baik tersebut mendapatkan balasan yang lebih baik dari Allah SWT.*

*Yogyakarta, 26 Juni 2021*

*Penulis  
(Sumaryanta)*

## DAFTAR ISI

KATA PENGANTAR.....	i
UCAPAN TERIMA KASIH .....	ii
DAFTAR ISI .....	iii
BAB I. TEORI TES.....	1
A. Teori Tes Klasik .....	1
B. Teori Respon Butir .....	8
BAB II.TERAPAN TEORI TES KLASIK .....	18
A. Kerangka Analisis dengan Teori Tes Klasik .....	18
B. Terapan Teori Tes Klasik untuk Analisis Butir Tes Kompetensi Guru Matematika Paket I.....	19
1. Analisis Data Ujicoba Tes Paket I.....	20
2. Analisis Butir Soal Paket I Berdasarkan Teori Tes Klasik.....	30
a. Tingkat kesukaran butir tes .....	30
b. Daya beda butir tes .....	31
c. Keberfungsian distraktor .....	33
3. Tindaklanjut Hasil Analisis Paket I .....	34
C. Terapan Teori Tes Klasik untuk Analisis Butir Tes Kompetensi Guru Matematika Paket II .....	41
1. Analisis Data Ujicoba Tes Paket II.....	52
2. Analisis Butir Soal Paket II Berdasarkan Teori Tes Klasik .....	52
a. Tingkat kesukaran butir tes.....	52
b. Daya beda butir tes .....	53
c. Keberfungsian distraktor .....	55
3. Tindaklanjut Hasil Analisis Paket II.....	56
BAB III TERAPAN TEORI RESPON BUTIR .....	65
A. Kerangka Analisis dengan Teori Respon Butir .....	65
B. Terapan Teori Respon Butir untuk Analisis Tes Kompetensi Guru Matematika Paket I .....	68
1. Uji Asumsi.....	68
a. Asumsi unidimensi .....	68
b. Asumsi independensi lokal .....	70

c. Asumsi invariansi parameter .....	71
2. Uji Kecocokan Model.....	74
3. Parameter Butir Tes .....	88
a. Tingkat kesukaran.....	88
b. Daya beda butir.....	90
c. Kurva karakteristik .....	91
d. Fungsi informasi .....	93
4. Parameter Kemampuan.....	95
C. Terapan Teori Respon Butir untuk Analisis Tes Kompetensi Guru Matematika Paket II.....	96
1. Uji Asumsi.....	96
a. Asumsi unidimensi .....	96
b. Asumsi independensi lokal.....	98
c. Asumsi invariansi parameter .....	99
2. Uji Kecocokan Model.....	103
3. Parameter Butir Tes .....	117
a. Tingkat kesukaran.....	117
b. Daya beda butir.....	118
c. <i>Pseudo guesing</i> .....	120
d. Kurva karakteristik .....	121
e. Fungsi informasi .....	124
4. Parameter Kemampuan.....	126
DAFTAR PUSTAKA .....	130

# **BAB I**

## **TEORI TES**

Teori tes secara umum dibagi dalam dua kategori, yaitu teori tes klasik dan teori respons butir (Hambleton & van der Linden, 1982: 373). Kedua teori tersebut, walaupun memiliki beberapa perbedaan satu sama lain, sampai saat banyak digunakan untuk kepentingan pengembangan tes (Magno, 2009: 1 – 3). Teori tes klasik dan teori respons butir banyak digunakan para pengembang tes untuk menganalisis data empirik hasil ujicoba suatu tes yang sedang dikembangkan.

### **A. Teori Tes Klasik**

Teori tes klasik merupakan teori tes yang sangat terkenal dan luas penggunaannya. Teori tes klasik adalah teori dasar pengukuran kemampuan mental dengan menggambarkan hubungan skor teramati pada tes dan skor sebenarnya yang tidak teramati (Wang & Osterlind, 2013: 31). Dalam teori tes klasik, skor tes dipandang terdiri dari dua komponen, skor sebenarnya dan kesalahan pengukuran (Stage, 2003: 1 – 2). Skor sebenarnya adalah skor yang diperoleh apabila tidak terdapat kesalahan dalam pengukuran, sedangkan kesalahan pengukuran menunjukkan perbedaan antara skor sebenarnya dengan skor amatan (Magno, 2009: 1 – 2). Teori tes klasik ini digunakan secara luas dalam berbagai bidang pengukuran, termasuk pada lingkup pendidikan.

Teori tes klasik didukung dengan beberapa prinsip penting. Prinsip utama dari teori tes klasik adalah bahwa skor yang diperoleh seorang pada ujian yang diberikan, yang dilambangkan ( $X$ ), dapat di dekomposisi ke dalam skor sebenarnya ( $T$ ) dan kesalahan ( $E$ ), dinyatakan dengan:  $X = T + E$  (Champlain, 2010: 110). Skor amatan sebagai jumlahan dari skor sebenarnya



dan kesalahan pengukuran menunjukkan bahwa skor yang diperoleh pada suatu tes terdiri dari dua komponen, yaitu skor sebenarnya dan kesalahan pengukuran (Thissen & Wainer, 2001: 26). Prinsip ini menjadi pijakan kuat yang sangat mewarnai berbagai kajian dalam teori tes klasik.

Teori tes klasik didukung dengan beberapa asumsi. Asumsi pada teori tes klasik adalah bahwa tidak ada hubungan antara skor sebenarnya dan skor kesalahan dan antara skor kesalahan satu dengan lainnya (Novick, 1966: 1). Kesalahan pengukuran yang menunjukkan perbedaan antara skor sebenarnya dengan skor amatan, pada pengukuran yang baik diharapkan kesalahan pengukuran yang diperoleh sekecil mungkin (Thissen & Wainer, 2001: 34). Semakin kecil kesalahan pengukuran menunjukkan bahwa skor amatan yang diperoleh semakin mendekati skor sebenarnya, dan sebaliknya.

Teori tes klasik banyak digunakan sebagai dasar analisis pada suatu instrumen yang sedang dikembangkan. Parameter utama yang dianalisis dengan teori tes klasik antara lain: tingkat kesukaran butir, daya beda butir, dan reliabilitas instrumen (Bichi, 2016: 28). Selain ketiga parameter tersebut, analisis dengan teori tes klasik juga digunakan untuk menentukan keberfungsian distraktor pada butir soal pilihan ganda (Impara & Plake, 1997; Magno, 2009: 2). Parameter-parameter tersebut merupakan informasi sangat penting untuk memahami kualitas tes yang dikembangkan.

Tingkat kesukaran butir merujuk pada proporsi peserta tes menjawab benar dibanding dengan jumlah benar keseluruhan (Crocker & Aigina: 2008: 90). Dengan demikian, indeks kesukaran butir dihitung dengan membagi banyak peserta ujian yang menjawab item dengan benar dengan jumlah total peserta ujian yang menjawab. Tingkat kesukaran butir dapat dibedakan dalam beberapa kategori, yaitu: sukar, sedang dan mudah (Bichi, 2016: 29), seperti tersaji pada Tabel 1.

Tabel 1. Kategori Tingkat Kesukaran Butir

Indeks tingkat kesukaran (p)	Kategori
$p < 0,30$	Sukar
$0,31 \leq p \leq 0,70$	Sedang
$p > 0,7$	Mudah

(Bichi, 2016: 29)

Menurut Shakil (2016: 2), tes pada umumnya sebaiknya memiliki tingkat kesukaran butir sekitar 0,5, walaupun pada masing-masing butir tingkat kesukaran dapat bervariasi dari 0.2 sampai 0.8. Hal senada dinyatakan Ashraf & Jaseem (2020: 399) bahwa biasanya butir dengan tingkat kesukaran sekitar 0.5 menunjukkan butir yang baik, tetapi tidak ada ketentuan ketat untuk menentukan variasi yang diterima dari 0.5. Menurut Kaplan dan Succuzzo (2001, Ashraf & Jaseem, 2020: 399), untuk kebanyakan tes butir dengan tes tingkat kesukaran butir antara 0.3 sampai 0.7 akan memaksimalkan informasi tentang perbedaan individu peserta tes.

Pada tes pilihan ganda, penentuan tingkat kesukaran mempertimbangkan banyaknya pilihan jawaban. Menurut Ashraf & Jaseem (2020: 399), tingkat optimum pada tes pilihan ganda dengan 4 pilihan jawaban adalah 0.65, pada tes dengan 3 pilihan jawaban 0.67, dan pada tes dengan 2 pilihan jawaban 0.75. Menurut Shakil (2016: 2), tes pilihan ganda dengan 2 pilihan jawaban tingkat kesukaran terbaik adalah 0.75, tes dengan 3 pilihan jawaban tingkat kesukaran terbaik adalah 0.67, tes dengan 4 pilihan jawaban tingkat kesukaran terbaik adalah 0.63, dan tes dengan 5 pilihan jawaban tingkat kesukaran terbaik adalah 0.60. Menurut Fareed dan Ashraf (2008, Ashraf & Jaseem, 2020: 399) penentuan tingkat kesukaran butir dapat dilakukan secara

statistik, yaitu  $Z = \frac{\check{p}_i - p_o}{\sqrt{p_o(1-p_o)/n}}$ , dengan  $p_i$  adalah tingkat kesukaran butir ke- $i$ ,  $p_o$  adalah nilai optimum tingkat kesukaran butir, dan  $n$  adalah banyaknya sampel.

Pada tes yang bertujuan untuk seleksi memilih calon terbaik dari peserta tes maka butir soal dengan tingkat kesukaran lebih tinggi lebih disarankan, sedangkan pada tes untuk tes prestasi yang mengacu pada norma yang dirancang untuk membedakan antara peserta ujian yang berkaitan dengan kompetensi maka butir soal biasanya dipilih dalam tingkat kesukaran menengah dan kisaran tingkat kesulitan yang sempit (Bichi, 2016: 31). Pada tes berdasarkan acuan kriteria, yang menekankan pada ketntasan peserta tes, nilai optimum tingkat kesukaran butir soal yang diharapkan 0.9 atau lebih, sedangkan untuk tes dengan acuan norma dengan penekanan pada perbedaan level kemampuan peserta tes maka tingkat kesukaran butir adalah sekitar 0.5.

Berbagai penjelasan penentuan tingkat kesukaran butir di atas menunjukkan bahwa tidak ada ketentuan baku dalam memutuskan batas terbaik tingkat kesukaran butir pada suatu tes. Hal ini selaras dengan pendapat Ashraf & Jaseem (2020: 399) yang menyatakan bahwa memang tidak ada ketentuan baku untuk memutuskan batas nilai tingkat kesukaran butir pada suatu tes. Target tingkat kesukaran butir bergantung pada dari tes yang dilaksanakan (Champlain, 2010: 112). Dalam proses pengembangan tes, penyusun tes dapat menentukan batas tingkat kesukaran butir yang digunakan disesuaikan dengan tujuan dan kebutuhan pengembangan tesnya.

Parameter kedua yang sangat penting dalam teori tes klasik adalah daya beda butir. Daya beda butir merujuk pada kemampuan butir membedakan peserta tes yang memiliki kemampuan tinggi dan kemampuan rendah (Champlain, 2010: 111). Penentuan data daya beda butir salah satunya dapat

dilakukan dengan membagi peserta tes dalam dua kelompok, yaitu kelompok atas dan kelompok bawah. Ada beberapa variasi pengelompokan, antara lain: 50% kelompok atas dan 50% kelompok bawah, 30% kelompok atas dan 30% kelompok bawah, atau 27% kelompok atas dan 27% kelompok bawah (Crocker & Aigina: 2008: 314). Setelah kelompok atas dan bawah telah diidentifikasi, indeks diskriminasi (D) dihitung sebagai  $D = p_u - p_i$ , di mana  $p_u$  adalah proporsi kelompok atas yang menjawab benar dan  $p_i$  adalah proporsi kelompok bawah yang menjawab benar (Bichi, 2016: 29). Nilai indeks diskriminasi (D) inilah yang digunakan sebagai pijakan menetapkan daya beda butir.

Cara lain menentukan daya beda butir adalah menggunakan koefisien korelasi *point biserial*. Hal ini dinyatakan oleh Crocker dan Aigina (2008: 317) bahwa koefisien korelasi *point biserial* dapat digunakan untuk menentukan daya beda butir soal pada suatu proses pengembangan soal. Korelasi *point biserial* ( $\rho_{pbis}$ ) ditentukan dengan  $\rho_{pbis} = \frac{(\mu_+ - \mu_x)}{\sigma_x} \sqrt{p/q}$ , dengan  $\mu_+$  rata-rata skor peserta tes menjawab benar,  $\mu_x$  rata-rata skor seluruh kelompok,  $\sigma_x$  standar deviasi mereka, p kesulitan item, dan q adalah (1 - p). Nilai  $\rho_{pbis}$  menunjukkan indeks daya beda butir.

Nilai daya beda butir (D) berkisar antara -1 sampai 1. Nilai positif D menunjukkan bahwa item tersebut mendukung kelompok atas; nilai negatif menunjukkan bahwa item tersebut adalah diskriminator terbalik, lebih menyukai kelompok yang memiliki skor lebih rendah (Crocker & Aigina: 2008: 314). Penafsiran daya beda butir dapat dilakukan berdasarkan pendapat dari Bichi (2016: 29) yang tersaji pada Tabel 2.

Tabel 2. Kriteria Penafsiran Daya Beda Butir

Indeks data beda (D)	Penafsiran
$D \geq 0,40$	Soal berfungsi dengan memuaskan
$0,30 \leq D < 0,40$	Soal baik, sedikit atau tidak perlu direvisi
$0,20 \leq D < 0,30$	Soal kurang baik, perlu direvisi
$D < 0,20$	Soal jelek, dibuang atau sepenuhnya direvisi

(Bichi, 2016: 29)

Parameter penting lain yang dapat dihasilkan dari analisis menggunakan teori tes klasik adalah keberfungsian distraktor atau pengecoh. Menurut Asamoah dan Ocansey (2019: 23), efektifitas distraktor dapat dilihat dari banyaknya peserta tes yang memilih distraktor tersebut. Menurut Musa, Shaheen, dan Ahmed (2018: 1445), ada dua kriteria yang perlu diperhatikan untuk melihat keberfungsian distraktor, yaitu: (1) terkait dengan keterpilihan distraktor tersebut, dan (2) terkait dengan kemampuan distraktor membedakan antara peserta dengan kemampuan lebih dengan yang kurang. Pertama (1), suatu distraktor dikatakan berfungsi apabila setidaknya ada peserta ujian yang memilihnya, disarankan bahwa setidaknya 5% peserta ujian harus memilih distraktor untuk distraktor tersebut layak dipertahankan. Jika tidak ada yang memilih maka distraktor tersebut dikatakan sebagai tidak berfungsi dan harus diganti. Kedua (2), sebuah distraktor dikatakan memiliki kemampuan membedakan yang baik apabila peserta ujian dengan kemampuan lebih tinggi memilih distraktor lebih sedikit dibanding yang kemampuannya lebih rendah, dan sebaliknya peserta dengan kemampuan lebih rendah memilih lebih banyak distraktor tersebut dibanding yang kemampuannya lebih tinggi. Dengan kata lain, distraktor berkorelasi negatif dengan nilai tes total, sehingga distraktor yang memiliki korelasi positif atau nol dengan skor

tes total dianggap tidak berfungsi dengan baik dan mengurangi kualitas keseluruhan item.

Teori tes klasik sangat bermanfaat dalam pengembangan ilmu pengukuran. Teori klasik terbukti sangat bermanfaat karena memiliki beberapa kelebihan (Bichi, 2016: 31). Beberapa kelebihan teori tes klasik antara lain: sampel yang dibutuhkan relatif kecil, analisis matematika yang digunakan relatif sederhana, model estimasi parameter secara konseptual mudah, dan analisis tidak memerlukan studi ketat untuk memastikan kecocokan model dengan data uji (Hambleton dan Jones, 1993: 44). Kelebihan ini mendorong pemanfaatan yang cukup banyak oleh para pengembang tes.

Teori tes klasik, di samping memiliki sejumlah kelebihan, juga memiliki sejumlah keterbatasan. Menurut Hambleton, Swaminathan, dan Rogers (1991: 4 – 5), teori tes klasik tidak memberikan informasi yang memadai terkait respon peserta tes terhadap butir, sehingga tidak cukup dasar menentukan seberapa baik peserta tes apabila dihadapkan pada suatu butir. Secara lebih spesifik, teori tes klasik tidak dapat memberikan prediksi yang tepat tentang performansi peserta tes pada butir. Hal ini menyulitkan pengembang tes apabila ingin memprediksi karakteristik skor tes pada satu atau lebih populasi, atau mendesain tes dengan karakteristik tertentu untuk suatu populasi tes.

Teori tes klasik juga memiliki kelemahan terkait reliabilitas dan standar *error* pengukuran. Menurut Hambleton, Swaminathan, dan Rogers (1991: 4), pada teori tes klasik, reliabilitas didefinisikan sebagai korelasi skor tes dari perangkat tes yang paralel. Pada faktanya, sangat sulit ditemukan dua perangkat tes yang benar-benar paralel. Koefisien reliabilitas yang ditemukan memberikan perkiraan batas bawah dari reliabilitas atau perkiraan reliabilitas dengan dengan bias tidak diketahui. Terkait standar *error* kesalahan, teori tes klasik mengasumsikan sama untuk semua peserta tes. Padahal, skor tes tidak

persis sama mengukur kemampuan peserta tes, sehingga asumsi bahwa standar *error* kesalahan sama tidak masuk akal.

Kelemahan lain teori tes klasik diungkapkan oleh Champlain. Menurut Champlain (2010: 112 – 113), teori tes klasik terlalu bergantung pada sampel. Hal ini berdampak hasil yang diperoleh sulit digeneralisasikan pada peserta tes lainnya. Teori tes klasik juga tidak menyediakan mekanisme yang mudah untuk menguji peserta tes dengan kemampuan tertentu, sesuatu yang sangat penting dipertimbangkan pada tes.

Kelemahan–kelemahan teori tes klasik mendorong para pakar psikometri mencari alternatif lain yang lebih baik. Menurut Hambleton, Swaminathan, dan Rogers (1991: 5), diperlukan alternatif model baru teori tes yang dapat memberikan informasi tentang: 1) karakteristik tes yang tidak bergantung pada grup peserta tes tertentu, 2) skor yang menggambarkan kemampuan peserta tes yang tidak bergantung pada tes, 3) model yang dinyatakan lebih pada level butir bukan tes, 4) model yang tidak membutuhkan tes paralel untuk mendapatkan reliabilitas, dan 5) model yang dapat menyediakan hasil pengukuran yang tepat untuk setiap skor kemampuan. Berbagai kebutuhan tersebut pada akhirnya dapat dipenuhi dengan teori respons butir.

## **B. Teori Respons Butir**

Teori respon butir saat ini menjadi teori tes yang banyak digunakan. Selama beberapa dekade terakhir teori respons butir telah berkembang pesat menjadi pelengkap penting untuk teori tes klasik dalam pengembangan tes (Stage, 2003: 2). Teori respons butir banyak digunakan dalam banyak pelaksanaan tes (Kolen & Brennan, 1995: 155). Berkembang pesatnya teori

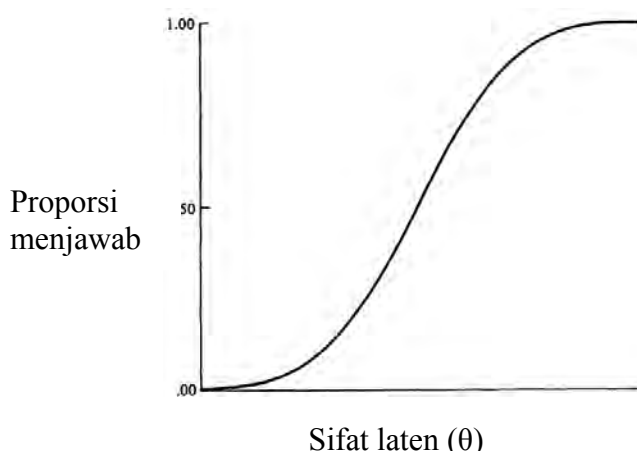
respons butir tidak lepas dari kemampuan teori respons butir dalam mengatasi kelemahan teori tes klasik. Teori respons butir menyediakan kerangka kerja yang bermanfaat dan secara teoritis kuat untuk pengukuran pendidikan (Glas, 2008: 19). Teori respons butir terus berkembang dan dimanfaatkan semakin luas oleh para praktisi pengembang tes dalam proses mengembangkan tes sesuai kebutuhan.

Teori respons butir memiliki beberapa kelebihan dibandingkan dengan teori tes klasik. Menurut (Bichi & Talib, 2018: 149), kelebihan teori respons butir antara lain: statistik butir tidak tergantung pada sampel, kemampuan peserta tes tidak tergantung tingkat kesukaran butir, analisis butir mengakomodasi butir tes yang cocok untuk menguji kemampuan level tertentu, analisis tes tidak membutuhkan tes paralel yang ketat untuk mengestimasi kehandalan tes, dan statistik butir dan kemampuan peserta ujian keduanya dilaporkan pada skala yang sama. Beberapa kelebihan teori respons butir ini mendorong pemanfaatannya semakin luas untuk pengembangan tes.

Teori respons butir menggunakan pendekatan yang berbeda secara mendasar dengan teori tes klasik. Teori respons butir merupakan model non linear yang memberikan probabilitas respon menjawab benar pada butir sebagai fungsi dari karakteristik butir dan kemampuan peserta tes (Champlain, 2010: 113). Teori respons butir didasarkan pada dua postulat utama, yaitu: 1) performansi peserta tes pada suatu butir tes dapat diprediksi atau dijelaskan dengan seperangkat faktor, yaitu: sifat, sifat tersembunyi, atau kemampuan, dan 2) relasi antara performansi peserta tes pada butir dan sifat yang mendasari performansi butir dapat digambarkan dengan kurva karakteristik butir (Hambleton, Swaminathan, dan Rogers, 1991: 7). Hal tersebut menjadikan teori tes respon butir berbeda dengan teori tes klasik.



Pada teori respons butir, hubungan antara kemampuan dan proporsi menjawab benar dapat digambarkan sebagai kurva karakteristik butir. Kurva karakteristik butir menggambarkan kemungkinan merespons dengan benar suatu item sebagai fungsi dari sifat laten (dilambangkan dengan  $\theta$ ) yang mendasari kinerja pada item-item pada tes (Crocker dan Aigina, 2008: 340). Kurva karakteristik butir untuk item  $i$  ( $P_i(\theta)$ ) menunjukkan probabilitas untuk menjawab item  $i$  untuk kelulusan dengan kemampuan  $\theta$  (Kolen & Brennan, 1995: 157). Kurva karakteristik butir diasumsikan memiliki bentuk S, seperti yang diilustrasikan dalam Gambar 1.



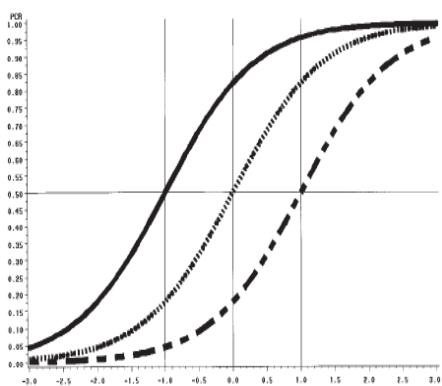
Gambar 1. Hubungan antara Kemampuan dan Respon Butir  
(Sumber: Crocker dan Aigina, 2008: 341)

Gambar 1 menunjukkan bahwa probabilitas menjawab benar tergantung pada kemampuan peserta tes dan karakteristik butir. Gambar 5 juga menunjukkan bahwa ketika skor pada sifat laten meningkat, maka probabilitasnya menjawab dengan benar pada butir sebagai fungsi karakteristik butir. Dengan kata lain, peserta tes dengan kemampuan lebih tinggi mempunyai probabilitas menjawab benar lebih besar dibandingkan peserta dengan kemampuan lebih rendah.

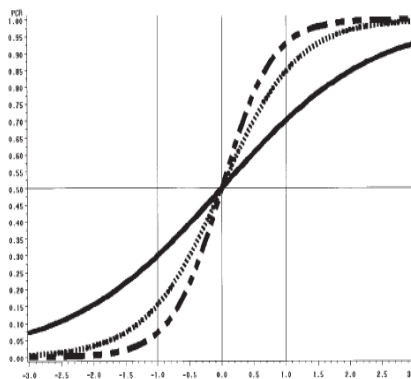
Teori respons butir didasarkan pada sejumlah asumsi penting. Ada tiga asumsi yang perlu dipenuhi pada teori respons butir, yaitu: unidimensi, independensi lokal, dan invariansi parameter (Mahmud, 2017: 259). Asumsi unidimensi menunjuk pada setiap butir tes hanya mengukur satu kemampuan saja (Kolen & Brennan, 1995: 156 – 157). Asumsi independensi lokal menghendaki jawaban peserta benar pada suatu butir tidak mempengaruhi jawaban peserta pada butir yang lain (Bulut, 2015: 315). Ketika asumsi unidimensi terpenuhi, maka asumsi independensi lokal juga diperoleh (Hambleton, Swaminathan, dan Rogers, 1991: 11). Asumsi invariansi parameter merujuk pada parameter butir yang dihasilkan dan parameter kemampuan tidak saling bergantung, yaitu bahwa parameter butir tidak bergantung pada parameter kemampuan peserta tes dan kemampuan peserta tes tidak bergantung pada parameter butir (Kahrahman, 2014: 118). Berbagai asumsi tersebut sangat penting dalam proses pengembangan dan penggunaan teori respons butir.

Banyak parameter menjadi perhatian penting pada teori respons butir. Berdasarkan banyak parameter butir, ada tiga model pada teori respons butir, yaitu: model 1 parameter (1P), 2 parameter (2P), dan 3 parameter (3P) (Ogunsakin & Shogbesani, 2018: 629 – 631). Model 1P menggunakan parameter tunggal, yaitu tingkat kesukaran butir, model 2P menggunakan 2 parameter, yaitu tingkat kesukaran butir dan daya beda butir, dan model 3P menggunakan 3 parameter, yaitu tingkat kesukaran butir, daya beda butir, dan *pseudo guessing* (tebakan semu) (Bichi, *et al*, 2015: 551). Ketiga model tersebut memiliki karakteristik dan representasi grafis yang unik satu dengan yang lain.

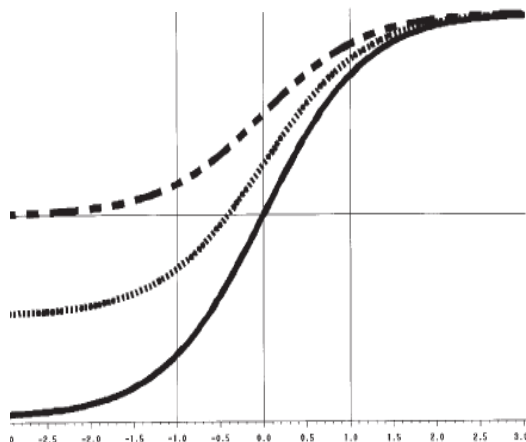
Model 1P, 2P, dan 3P dapat digambarkan dalam kurva karakteristik butir (Harvey & Hammer, 1999: 360 – 364). Gambar 2, Gambar 3, dan Gambar 4 menggambarkan suatu tampilan kurva karakteristik butir dari 1P, 2P, dan 3P.



Gambar 2. Kurva Karakteristik 3 Butir Tes pada Model 1P (Sumber: Harvey & Hammer, 1999: 360)



Gambar 3. Kurva Karakteristik 3 Butir Tes pada Model 2P (Sumber: Harvey & Hammer, 1999: 362)



Gambar 4. Kurva Karakteristik 3 Butir Tes pada Model 3P (Sumber: Harvey & Hammer, 1999: 364)

Sifat laten ( $\theta$ ) dan probabilitas peserta tes menjawab benar suatu butir memiliki hubungan yang spesifik. Hubungan antara sifat laten ( $\theta$ ) dan probabilitas peserta tes menjawab benar suatu butir dapat dimodelkan dengan

fungsi matematika (Erguven, 2013: 25). Besarnya peluang menjawab benar seseorang dengan kemampuan  $\theta$  suatu butir tes  $i$  pada model 1P, 2P, dan 3P dirumuskan dengan persamaan berikut (Ogunsakin & Shogbesani, 2018: 630 – 631).

$$\begin{aligned}
 1-P \quad P_i(\theta) &= \frac{1}{1+\exp(\theta-b_i)} \\
 2-P \quad P_i(\theta) &= \frac{1}{1+\exp[-Da_i(\theta-b_i)]} \\
 3-P \quad P_i(\theta) &= c_i + \frac{1-c_i}{1+\exp[-Da_i(\theta-b_i)]}
 \end{aligned}$$

Keterangan:

- $P_i(\theta)$  = peluang peserta tes dengan kemampuan  $\theta$  dipilih acak dapat menjawab butir  $i$  dengan benar  
 $b_i$  = tingkat kesulitan butir  $i$   
 $a_i$  = daya beda butir  $i$   
 $c_i$  = *pseudo guessing* butir  $i$   
 $n$  = banyaknya butir tes  
 $e$  = bilangan natural (2.718, pembulatan 3 desimal)  
 $\theta$  = tingkat kemampuan  
 $D$  = konstanta, ditetapkan 1,7 atau 1,72

Faktor  $D$  adalah faktor penskalaan yang diberikan untuk membuat fungsi logistik sedekat mungkin dengan fungsi *ogive* normal (Hambleton, Swaminathan, & Rogers, 1991: 15). Hubungan matematis sifat laten ( $\theta$ ) dan probabilitas peserta tes menjawab benar inilah yang secara grafis disajikan pada Gambar 2, Gambar 3, dan Gambar 4.

Parameteri butir menentukan karakteristik masing-masing butir. Menurut Hambleton, Swaminathan, dan Rogers (1991: 13 – 15), parameter tingkat kesukaran ( $b$ ) pada suatu butir adalah suatu titik pada skala kemampuan yang mana peluang menjawab benar adalah 0.5. Ketika nilai kemampuan ditransformasikan sehingga memiliki rata-rata 0 dan standar deviasi 1, maka nilai  $b$  berkisar dari  $-2$  sampai  $+2$ . Nilai  $b$  mendekati  $-2$  berarti

butir sangat mudah, dan apabila nilai  $b$  mendekati  $+2$  berarti butir sangat sulit. Daya beda ( $a$ ), secara teori memiliki rentang nilai  $(-\infty, +\infty)$ . Butir yang memiliki daya beda negatif dikeluarkan dari tes karena ada yang salah dengan butir, seperti peluang menjawab benar berkurang dengan meningkatnya kemampuan peserta tes. Namun begitu, tidak lazim juga mendapatkan nilai  $a$  lebih dari  $2$ . Dengan demikian, interval yang dapat diterima dari  $a$  adalah  $0$  sampai  $2$ . Untuk parameter *pseudo guessing* ( $c$ ), serangkaian penelitian menggunakan berbagai data nyata dan simulasi menunjukkan bahwa nilai parameter  $c$  yang disarankan untuk mendapatkan estimasi butir yang stabil dan akurat adalah tidak lebih dari  $\frac{1}{k}$  dengan  $k$  banyak pilihan jawaban (Han, 2012: 1). Nilai-nilai batas tersebut digunakan sebagai kriteria penting dalam menentukan kualitas suatu butir.

Dalam model teori respons butir, terdapat kajian menarik tentang fungsi informasi butir. Setiap item memiliki fungsi informasi spesifik yang menyediakan informasi tentang seberapa tepat dan tingkat sifat laten yang dianalisis (Zięba, 2013: 89). Persamaan fungsi informasi dari suatu butir ( $j$ )

pada model 1P dinyatakan dengan rumus:  $I_j(\theta) = \frac{e^{\theta-\alpha_j}}{(1+e^{\theta-\alpha_j})^2}$  (Zeiba, 2013:

89), pada model 2P dinyatakan dengan rumus  $I_j(\theta) = \frac{\beta_j \cdot e^{(\theta-\alpha_j) \cdot \beta_j}}{(1+e^{(\theta-\alpha_j) \cdot \beta_j})^2}$  (Zeiba,

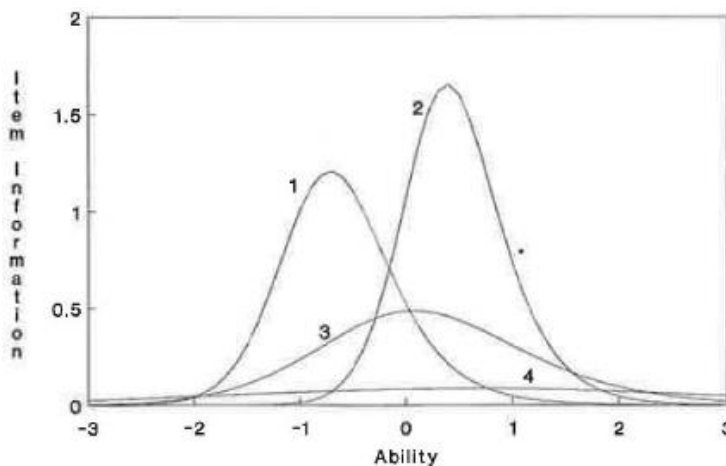
2013: 93), dan pada model 3P dinyatakan dengan rumus  $I_j(\theta) =$

$\frac{2.89a_j^2 \cdot (1-c_i)}{(c_i + e^{1.7\alpha_j(\theta-\alpha_j)}) \left(1 + e^{1.7\alpha_j(\theta-b_j) \cdot \beta_j}\right)^2}$  (Mardapi, 2017: 202). Fungsi informasi

penting diperhatikan dalam pengembangan tes karena semakin besar fungsi informasi akan semakin baik butir tes tersebut.

Besar fungsi informasi tes bergantung pada besar fungsi informasi masing-masing butir. Fungsi informasi suatu tes merupakan penjumlahan dari informasi tiap butir tes, yaitu:  $I(\theta) = \sum_{i=1}^N I_i(\theta)$ ; dengan :  $I(\theta)$  nilai informasi tes, dan  $I_i(\theta)$  nilai informasi butir (Moghadamzadeh, *et al*, 2011: 1361). Hal ini berarti fungsi informasi tes lebih besar dari fungsi informasi setiap butir. Selain itu, fungsi informasi tes sebagai jumlahan fungsi informasi butir berimplikasi fungsi informasi tes tinggi apabila fungsi informasi dari butir-butir tinggi, dan sebaliknya.

Fungsi informasi bermanfaat untuk menentukan ketepatan pengukuran pada tingkat kemampuan tertentu. Fungsi informasi menyediakan representasi grafis dari ketepatan pengukuran pada tingkat kemampuan tertentu, baik untuk masing-masing butir (fungsi informasi butir) maupun untuk keseluruhan tes (fungsi informasi tes) (Champlain, 2010: 115). Ilustrasi grafis dari fungsi informasi empat butir 1, 2, 3, dan 4, dengan butir 2 lebih mudah dari butir 1, dan daya beda butir 3 dan 4 lebih rendah dari butir 1 dan 2, disajikan Hambleton dan Jones (1993: 42) melalui Gambar 5.



Gambar 5. Fungsi Informasi 4 Butir Tes

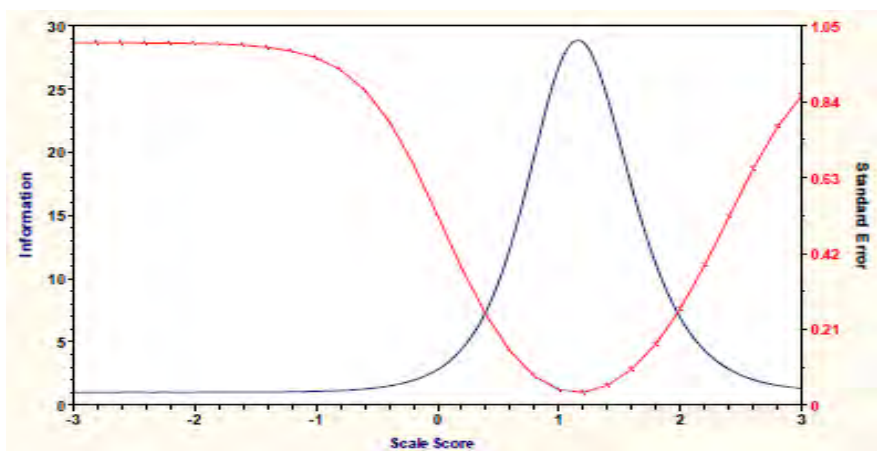
Gambar 5 menunjukkan bahwa butir 2 lebih sulit dibanding butir 1 maka fungsi informasi butir 1 lebih terpusat pada peserta dengan ability lebih rendah dari butir 2. Untuk butir 3 dan 4, karena daya beda kedua butir tersebut lebih rendah dibanding butir 1 dan 2, maka fungsi informasinya lebih rendah dari butir 1 dan 2.

Fungsi informasi memiliki hubungan erat dengan kesalahan pengukuran. Semakin besar informasi item untuk tingkat kemampuan yang diberikan, semakin sedikit kesalahan yang terlibat dalam memperkirakan kemampuan peserta tes (Reid, dkk., 2007: 180). Hambleton dan Jones (1993: 42) menyatakan bahwa fungsi informasi memiliki hubungan berkebalikan dengan besarnya kesalahan pengukuran, dinyatakan secara matematis dengan  $SEM(\hat{\theta}) = \frac{1}{[I(\theta)]^{\frac{1}{2}}}$ . Hal ini berarti bahwa semakin banyak informasi yang diberikan tes pada tingkat kemampuan tertentu, semakin kecil kesalahan yang terkait dengan kemampuan yang diestimasi. Sebaliknya, semakin besar kesalahan pada kemampuan yang diukur, fungsi informasi yang diberikan semakin kecil.

Secara grafis hubungan fungsi informasi dan *Standard Error of Measurement (SEM)* disajikan pada Gambar 9. Gambar 9 menunjukkan *SEM* tinggi saat fungsi informasi rendah, dan sebaliknya *SEM* rendah saat fungsi informasi tinggi. Nilai fungsi informasi yang merujuk pada menentukan tingkat ketepatan pengukuran dan *SEM* yang merujuk pada kesalahan pengukuran, maka Gambar 6 menunjukkan bahwa kemampuan butir dalam memberikan informasi tentang sifat laten yang dianalisis tinggi apabila kesalahan yang terlibat dalam pengukuran rendah, dan sebaliknya.

Gambar 6 juga menunjukkan bahwa pengukuran peserta dengan abilitas di sebelah kiri dan sebelah kanan perpotongan tersebut berakibat kesalahan

pengukuran lebih tinggi dibanding fungsi informasi yang diberikan, dan sebaliknya. Hal ini menunjukkan bahwa tes cocok diberikan pada peserta tes dengan abilitas diantara titik potong sebelah kiri dan kanan dari fungsi informasi dan *SEM*.



Gambar 6. Visualisasi Grafis Fungsi Informasi Tes dan *SEM*  
(Sumber: Gomez & Vance, 2015: 6)

Fungsi informasi pada teori respons butir dapat dimanfaatkan untuk menentukan reliabilitas tes. Pada teori respons butir koefisien reliabilitas tes dinyatakan dengan fungsi informasi tes, yaitu merupakan penjumlahan fungsi informasi semua butir tes (Mardaphi, 2016: 202). Penentuan reliabilitas dengan memanfaatkan fungsi informasi dapat mengatasi kendala penentuan reliabilitas tes dengan teori tes klasik yang mensyaratkan secara ketat adanya tes paralel yang pada praktik riil sulit dipenuhi.



## **BAB II.**

### **TERAPAN TEORI KLASIK**

#### **A. Kerangka Analisis dengan Teori Tes Klasik**

Pada buku ini akan diuraikan contoh penerapan teori tes klasik untuk penentuan tingkat kesukaran butir, daya beda, dan keberfungsian distraktor tes. Tingkat kesukaran butir pada contoh ini dinyatakan dalam bentuk indeks yang menunjukkan proporsi berkisar 0,00 – 1,00. Semakin besar indeks tingkat kesukaran diartikan semakin mudah soal itu. Suatu soal memiliki indeks tingkat kesukaran 0,00 artinya bahwa pada butir tersebut tidak ada testee yang menjawab benar dan bila memiliki indeks tingkat kesukaran 1,00 artinya bahwa pada butir tersebut semua testee menjawab benar. Penafsiran indeks kesukaran didasarkan pada pendapat Bichi (2016: 29) yang tersaji pada Tabel 1 halaman 3. Dalam contoh ini butir dinyatakan baik dan dapat diterima apabila memiliki tingkat kesukaran butir 0.10 – 0.90.

Indeks daya pembeda setiap butir dinyatakan dalam bentuk proporsi yang besarnya berkisar antara  $-1,00$  sampai dengan  $+1,00$ . Penentuan daya beda buku ini menggunakan koefisien *point biserial* ( $r_{pbi}$ ), dimana semakin besar nilai  $r_{pbi}$  maka butir tersebut semakin mampu membedakan peserta tes berkemampuan tinggi dengan yang berkemampuan rendah. Penentuan nilai  $r_{pbi}$  dilakukan dengan menggunakan bantuan program ITEMAN. Kriteria penentuan kualitas butir pada contoh ini didasarkan pada pendapat Bichi (2016: 32), seperti tersaji pada Tabel 2 halaman 6.

Analisis keberfungsian distraktor dimaksudkan untuk mengetahui berfungsi tidaknya distraktor pada alternatif jawaban yang tersedia. Dalam contoh ini keberfungsian distraktor didasarkan pada pendapat Musa, Shaheen,

dan Ahmed (2018: 1445), yaitu berdasarkan: (1) keterpilihan distraktor tersebut, dan (2) kemampuan distraktor membedakan antara peserta dengan kemampuan lebih dengan yang kurang. *Pertama*, suatu distraktor dikatakan berfungsi setidaknya 5% peserta ujian harus memilih distraktor tersebut. *Kedua*, sebuah distraktor dikatakan memiliki kemampuan membedakan yang baik apabila peserta ujian yang memiliki kemampuan lebih tinggi memilih distraktor lebih sedikit dibandingkan dengan yang kemampuannya lebih rendah, dan sebaliknya peserta dengan kemampuan lebih rendah memilih lebih banyak distraktor tersebut dibandingkan peserta dengan kemampuan lebih tinggi

## **B. Terapan Teori Tes Klasik untuk Analisis Butir**

### **Tes Kopetensi Guru Matematika Paket I**

Misalkan suatu tes kompetensi guru matematika Paket 1 berbentuk soal pilihan ganda dengan 4 (empat) pilihan sebanyak 30 butir diujicobakan pada 44 guru matematika. Dari data hasil ujicoba ini dapat diidentifikasi tingkat kesukaran, daya beda, dan keberfungsian distruktornya. Temuan tingkat kesukaran, daya beda, dan keberfungsian distruktur ini selanjutnya dijadikan bahan untuk memperbaiki soal yang disusun.

Pada ujicoba ini, data hasil pengerjaan soal oleh para responden selanjutnya diolah dengan menggunakan program ITEMAN. Selain dengan ITEMAN, analisis hasil ujicoba seperti ini dapat dilakukan dengan aplikasi lain, seperti analtes, excel, atau yang lain. Untuk kasus ujicoba ini, analisis dilakukan dengan memanfaatkan program ITEMAN. Berikut disajikan hasil analisisnya.

# 1. Analisis Data Ujicoba Tes Paket I

## Output Analisis dengan Iteman

MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket I.txt

Page 1

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
1	0-1	0.114	0.232	0.140	A	0.341	0.190	0.147	
					B	0.045	0.540	0.248	?
		CHECK THE KEY			C	0.500	-0.416	-0.332	
		D was specified, B works better			D	0.114	0.232	0.140	*
					Other	0.000	-9.000	-9.000	
2	0-2	0.114	0.111	0.067	A	0.136	0.069	0.044	
					B	0.682	0.130	0.099	?
		CHECK THE KEY			C	0.114	0.111	0.067	*
		C was specified, B works better			D	0.068	-0.628	-0.328	
					Other	0.000	-9.000	-9.000	
3	0-3	0.614	0.383	0.301	A	0.023	-1.000	-0.432	
					B	0.159	-0.660	-0.437	

					C	0.205	0.275	0.193	
					D	0.614	0.383	0.301	*
					Other	0.000	-9.000	-9.000	
4	0-4	0.523	0.518	0.413	A	0.023	0.208	0.075	
					B	0.364	-0.513	-0.400	
					C	0.091	-0.152	-0.087	
					D	0.523	0.518	0.413	*
					Other	0.000	-9.000	-9.000	
5	0-5	1.000	-9.000	-9.000	A	0.000	-9.000	-9.000	
					B	0.000	-9.000	-9.000	
					C	1.000	-9.000	-9.000	*
					D	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
6	0-6	0.818	0.056	0.038	A	0.000	-9.000	-9.000	
					B	0.091	0.061	0.035	
					C	0.818	0.056	0.038	*
					D	0.091	-0.152	-0.087	
					Other	0.000	-9.000	-9.000	
7	0-7	0.091	-0.081	-0.046	A	0.523	-0.008	-0.007	
					B	0.318	-0.048	-0.037	
					C	0.091	-0.081	-0.046	*
					D	0.068	0.256	0.134	?
					Other	0.000	-9.000	-9.000	
					CHECK THE KEY				
					C was specified, D works better				

MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket I.txt

Page 2

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
8	0-8	0.682	0.587	0.449	A	0.023	-0.331	-0.120	
					B	0.045	-0.800	-0.367	
					C	0.250	-0.362	-0.265	
					D	0.682	0.587	0.449	*
					Other	0.000	-9.000	-9.000	
9	0-9	0.818	0.166	0.114	A	0.818	0.166	0.114	*
					B	0.114	-0.162	-0.098	
					C	0.023	-0.547	-0.198	
					D	0.045	0.174	0.080	
					Other	0.000	-9.000	-9.000	
10	0-10	0.818	0.299	0.204	A	0.000	-9.000	-9.000	
					B	0.114	-0.162	-0.098	
					C	0.818	0.299	0.204	*
					D	0.068	-0.363	-0.189	
					Other	0.000	-9.000	-9.000	

11	0-11	0.955	0.983	0.451	A	0.000	-9.000	-9.000	*
					B	0.023	-1.000	-0.432	
					C	0.955	0.983	0.451	
					D	0.023	-0.547	-0.198	
					Other	0.000	-9.000	-9.000	
12	0-12	0.818	0.850	0.582	A	0.045	-0.191	-0.088	*
					B	0.114	-0.978	-0.593	
					C	0.818	0.850	0.582	
					D	0.023	-0.331	-0.120	
					Other	0.000	-9.000	-9.000	
13	0-13	0.614	0.748	0.588	A	0.068	-0.761	-0.397	*
					B	0.614	0.748	0.588	
					C	0.182	-0.144	-0.099	
					D	0.136	-0.677	-0.431	
					Other	0.000	-9.000	-9.000	
14	0-14	0.750	0.765	0.561	A	0.091	-0.472	-0.269	*
					B	0.045	-0.008	-0.004	
					C	0.750	0.765	0.561	
					D	0.114	-0.857	-0.520	
					Other	0.000	-9.000	-9.000	

MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket I.txt

Page 3

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics								
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key				
15	0-15	0.955	-0.296	-0.136	A	0.045	0.296	0.136	?				
					B	0.955	-0.296	-0.136	*				
					CHECK THE KEY				C	0.000	-9.000	-9.000	
					B was specified, A works better				D	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000					
16	0-16	0.477	0.564	0.449	A	0.205	-0.054	-0.038					
					B	0.477	0.564	0.449	*				
					C	0.250	-0.728	-0.534					
					D	0.068	0.168	0.088					
					Other	0.000	-9.000	-9.000					
17	0-17	0.659	0.603	0.467	A	0.250	-0.618	-0.454					
					B	0.045	-0.252	-0.116					
					C	0.045	-0.008	-0.004					
					D	0.659	0.603	0.467	*				

					Other	0.000	-9.000	-9.000	
18	0-18	0.455	0.215	0.171	A	0.000	-9.000	-9.000	
					B	0.455	0.215	0.171	*
					C	0.273	-0.365	-0.272	
					D	0.159	-0.083	-0.055	
					Other	0.114	0.292	0.177	
19	0-19	0.227	-0.148	-0.107	A	0.091	-0.010	-0.006	
					B	0.227	-0.148	-0.107	*
					C	0.182	-0.409	-0.280	
					D	0.136	-0.091	-0.058	
					Other	0.364	0.464	0.362	
20	0-20	0.500	0.708	0.565	A	0.114	-0.403	-0.245	
					B	0.114	-0.585	-0.355	
					C	0.227	-0.283	-0.204	
					D	0.500	0.708	0.565	*
					Other	0.045	-0.069	-0.032	
21	0-21	0.386	0.500	0.393	A	0.136	-0.597	-0.380	
					B	0.182	-0.232	-0.159	
					C	0.091	-0.330	-0.188	
					D	0.386	0.500	0.393	*
					Other	0.205	0.193	0.135	



MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket I.txt

Page 4

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
22	0-22	0.818	0.563	0.386	A	0.068	0.079	0.041	
					B	0.045	-0.678	-0.311	
					C	0.023	-1.000	-0.432	
					D	0.818	0.563	0.386	*
					Other	0.045	-0.313	-0.144	
23	0-23	0.614	0.474	0.372	A	0.091	-0.330	-0.188	
					B	0.159	-0.132	-0.087	
					C	0.114	-0.434	-0.263	
					D	0.614	0.474	0.372	*
					Other	0.023	-0.223	-0.081	
24	0-24	0.636	0.575	0.449	A	0.250	-0.362	-0.265	
					B	0.045	-0.496	-0.227	
					C	0.636	0.575	0.449	*
					D	0.045	-0.557	-0.255	
					Other	0.023	-0.007	-0.003	

25	0-25	0.773	0.438	0.315	A	0.091	-0.543	-0.309	*
					B	0.773	0.438	0.315	
					C	0.114	-0.283	-0.171	
					D	0.023	0.208	0.075	
					Other	0.000	-9.000	-9.000	
26	0-26	0.409	0.581	0.459	A	0.409	0.581	0.459	*
					B	0.091	-0.508	-0.289	
					C	0.455	-0.372	-0.296	
					D	0.045	0.053	0.024	
					Other	0.000	-9.000	-9.000	
27	0-27	0.568	0.626	0.497	A	0.182	-0.144	-0.099	*
					B	0.114	-0.252	-0.153	
					C	0.568	0.626	0.497	
					D	0.136	-0.730	-0.465	
					Other	0.000	-9.000	-9.000	
28	0-28	0.341	0.587	0.454	A	0.364	0.030	0.023	*
					B	0.159	-0.348	-0.230	
					C	0.341	0.587	0.454	
					D	0.136	-0.650	-0.414	
					Other	0.000	-9.000	-9.000	

MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket I.txt

Page 5

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics								
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key				
29	0-29	0.568	-0.144	-0.114	A	0.568	-0.144	-0.114	*				
					B	0.159	0.253	0.167	?				
					CHECK THE KEY				C	0.182	0.142	0.097	
					A was specified, B works better				D	0.091	-0.259	-0.147	
					Other	0.000	-9.000	-9.000					
30	0-30	0.955	0.678	0.311	A	0.000	-9.000	-9.000					
					B	0.000	-9.000	-9.000					
					C	0.000	-9.000	-9.000					
					D	0.955	0.678	0.311	*				
					Other	0.045	-0.678	-0.311					

MicroCAT (tm) Testing System

Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

Item and Test Analysis Program -- ITEMAN (tm)

Version 3.00

Item analysis for data from file Paket I.txt

Page 6

There were 44 examinees in the data file.

Scale Statistics

```
-----  
Scale:           0  
-----  
N of Items       30  
N of Examinees  44  
Mean             18.068  
Variance         15.245  
Std. Dev.       3.905  
Skew            -0.553  
Kurtosis        -0.165  
Minimum         7.000  
Maximum        25.000  
Median         18.000  
Alpha           0.677  
SEM             2.220  
Mean P          0.602  
Mean Item-Tot. 0.303  
Mean Biserial   0.419
```

Berdasarkan hasil analisis di atas selanjutnya dilakukan analisis terhadap tingkat kesukaran, daya beda, dan keberfungsian butir. Hasil analisis terhadap ketika aspek tersebut diuraikan sebagai berikut.

## 2. Analisis Butir Soal Paket I Berdasarkan Teori Tes Klasik

### a. Tingkat kesukaran butir tes Paket I

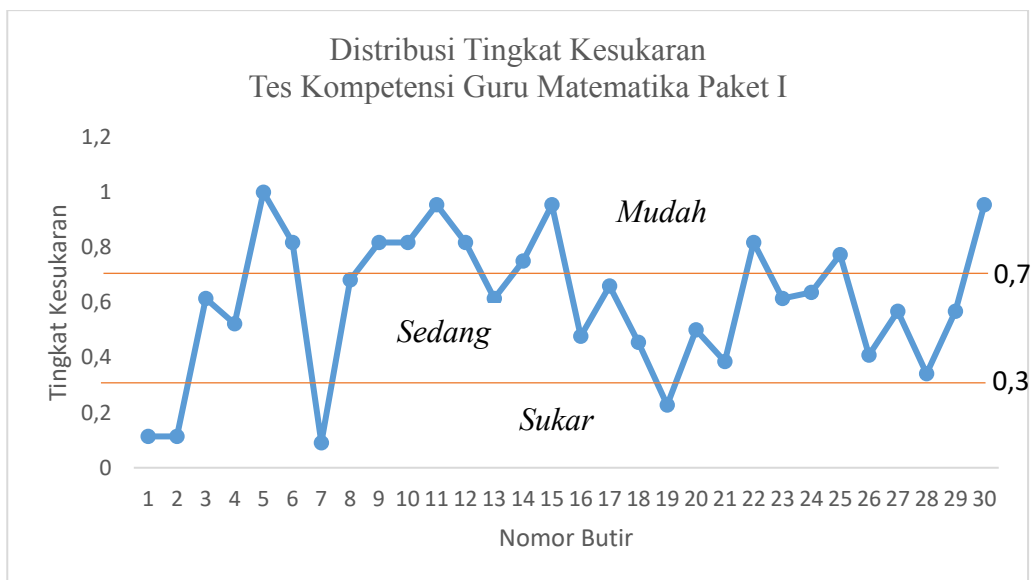
Tingkat kesukaran dari butir tes kompetensi guru matematika Paket I disajikan pada Tabel 3.

Tabel 3. Tingkat Kesukaran Soal Paket I

No Butir	Indeks Tingkat Kesukaran	Kategori	No Butir	Indeks Tingkat Kesukaran	Kategori
1	0,114	sukar	16	0,477	sedang
2	0,114	sukar	17	0,659	sedang
3	0,614	sedang	18	0,455	sedang
4	0,523	sedang	19	0,227	sukar
5	1,000	mudah	20	0,500	sedang
6	0,818	mudah	21	0,386	sedang
7	0,091	sukar	22	0,818	mudah
8	0,682	sedang	23	0,614	sedang
9	0,818	mudah	24	0,636	sedang
10	0,818	mudah	25	0,773	mudah
11	0,955	mudah	26	0,409	sedang
12	0,818	mudah	27	0,568	sedang
13	0,614	sedang	28	0,341	sedang
14	0,750	mudah	29	0,568	Sedang
15	0,955	mudah	30	0,955	Mudah

Tabel 3 menunjukkan bahwa terdapat 4 butir soal yang termasuk kategori sukar, 15 butir soal termasuk kategori sedang, dan 11 butir soal termasuk kategori mudah. Tabel 3 juga menunjukkan bahwa dilihat dari tingkat kesukarannya terdapat 26 butir termasuk kategori baik dan 4 butir soal termasuk kategori tidak baik.

Secara grafis, distribusi tingkat kesukaran butir tes kompetensi guru matematika Paket I disajikan pada Gambar 7.



Gambar 7. Distirbusi Tingkat Kesukaran Butir Tes Kompetensi Guru Matematika Paket I

b. Daya beda butir tes Paket I

Daya beda butir tes kompetensi guru matematika Paket I disajikan pada Tabel 4.

Tabel 4. Daya Beda Butir Tes Paket I

No Butir	Koefisien <i>point biserial</i> ( $r_{pbi}$ )	Kategori	No Butir	Koefisien <i>point biserial</i> ( $r_{pbi}$ )	Kategori
1	0,140	Soal jelek, dibuang atau sepenuhnya direvisi	16	0,449	Soal berfungsi dengan memuaskan
2	0,067	Soal jelek, dibuang atau sepenuhnya direvisi	17	0,467	Soal berfungsi dengan memuaskan
3	0,301	soal baik, sedikit atau tidak perlu perbaikan	18	0,171	Soal jelek, dibuang atau sepenuhnya direvisi

No Butir	Koefisien <i>point biserial</i> ( $r_{pbi}$ )	Kategori
4	0,413	Soal berfungsi dengan memuaskan
5	-9,000	Soal jelek, dibuang atau sepenuhnya direvisi
6	0,038	Soal jelek, dibuang atau sepenuhnya direvisi
7	-0,046	Soal jelek, dibuang atau sepenuhnya direvisi
8	0,449	Soal berfungsi dengan memuaskan
9	0,114	Soal jelek, dibuang atau sepenuhnya direvisi
10	0,204	Soal kurang baik, perlu direvisi
11	0,451	Soal berfungsi dengan memuaskan
12	0,582	Soal berfungsi dengan memuaskan
13	0,588	Soal berfungsi dengan memuaskan
14	0,561	Soal berfungsi dengan memuaskan
15	-0,136	Soal jelek, dibuang atau sepenuhnya direvisi

No Butir	Koefisien <i>point biserial</i> ( $r_{pbi}$ )	Kategori
19	-0,107	Soal jelek, dibuang atau sepenuhnya direvisi
20	0,565	soal sangat baik
21	0,393	soal baik, sedikit atau tidak perlu perbaikan
22	0,386	soal baik, sedikit atau tidak perlu perbaikan
23	0,372	soal baik, sedikit atau tidak perlu perbaikan
24	0,449	Soal berfungsi dengan memuaskan
25	0,315	soal baik, sedikit atau tidak perlu perbaikan
26	0,459	Soal berfungsi dengan memuaskan
27	0,497	Soal berfungsi dengan memuaskan
28	0,454	Soal berfungsi dengan memuaskan
29	-0,114	Soal jelek, dibuang atau sepenuhnya direvisi
30	0,311	soal baik, sedikit atau tidak perlu perbaikan

Daya beda dari butir tes Paket I disajikan pada Tabel 36. Tabel 35, Tabel 36, dan Gambar 19 menunjukkan ada beberapa butir soal memiliki daya beda jelek, bahkan beberapa negatif, yaitu butir 1, 2, 5, 6, 7, 9, 15, 18, 19, 29. Penulis kemudian melakukan kajian terhadap butir-butir tersebut untuk memperoleh temuan sebagai bahan perbaikan.

c. Keberfungsian Distraktor

Selain tingkat kesukaran dan daya beda, analisis juga dilakukan terhadap keberfungsioan distraktor. Analisis keberfungsian distraktor dimaksudkan untuk mengetahui berfungsi tidaknya distraktor pada alternatif jawaban yang tersedia. Setelah dilakukan analisis dengan program ITEMAN, berikut disajikan profil keberfungsian distraktor hasil ujicoba terbatas pada Tabel 5.

Tabel 5. Keberfungsian Distraktor Tes Paket I Hasil Ujicoba Terbatas

No Butir	Kunci	Distraktor		No Butir	Kunci	Distraktor	
		Berfungsi	Tidak berfungsi			Berfungsi	Tidak berfungsi
1	D	A,B,C	-	16	B	A,C,D	-
2	C	A,B,D	-	17	D	A,B,C	-
3	D	A,B,C	-	18	B	C,C	A
4	D	A,B,C	-	19	B	A,C,D	-
5	C	-	A,B,D	20	D	A,B,C	-
6	C	B,D	A	21	D	A,B,C	-
7	C	A,B,D	-	22	D	A,B,C	-
8	D	A,B,C	-	23	D	A,B,C	-
9	A	B,C,D	-	24	C	A,B,D	-
10	C	B,D	A	25	B	A,C,D	-
11	C	B,D	A	26	A	B,C,D	-
12	C	A,B,D	-	27	C	A,B,D	-
13	B	A,C,D	-	28	C	A,B,D	-
14	C	A,B,D	-	29	A	B,C,D	-
15	B	A	C,D	30	D	A,B,C	-



### 3. Tindaklanjut Hasil Analisis Paket I

Setelah dilakukan kajian terhadap butir-butir soal yang telah diujicobakan, terutama butir-butir yang termasuk kategori sukar dan butir-butir yang memiliki daya beda negatif, diperoleh beberapa temuan untuk perbaikan. Beberapa butir soal termasuk kategori sulit dimungkinkan ada beberapa penyebab, antara lain: peserta tes yang tidak menguasai dengan baik konsep/materi yang dipertanyakan pada soal, atau rumusan soal sulit dipahami atau multi tafsir sehingga peserta tes salah memahami soal, atau karena keduanya. Untuk itu, diperlukan kajian seksama terhadap masing butir soal sebagai bagian dari langkah perbaikan terhadap butir soal.

Sebagai gambaran, berikut disajikan beberapa kajian lanjutan terhadap butir-butir soal Paket I.

#### - Butir soal nomor 1

*1 Salah satu kesulitan yang dihadapi siswa dalam menyelesaikan soal kontekstual adalah membentuk model matematika dari masalah kontekstual. Kesulitan tersebut merupakan salah satu masalah yang berkaitan dengan kemampuan .... .*

- A. penalaran*
- B. berpikir kritis*
- C. pemahaman konsep*
- D. pemecahan masalah*

Butir soal nomor 1 sulit dimungkinkan karena banyak peserta tes kurang memahami yang dimaksud dengan penalaran, berpikir kritis, pemahaman konsep, dan pemecahan masalah. Hal ini menyebabkan peserta tes salah ketika ditanya kaitan pembentukan model dari masalah kontekstual dengan keempat hal tersebut. Keempat hal tersebut sebenarnya merupakan aspek penting pembelajaran matematika, sehingga guru seharusnya mengerti dengan baik.

- **Butir soal nomor 2**

- 2 *Seorang guru mengajarkan materi matriks menggunakan model pembelajaran discovery learning. Peserta didik berdiskusi bersama kelompoknya untuk menyelesaikan masalah yang diberikan pada lembar kerja dengan memanfaatkan berbagai informasi yang telah diperoleh. Dalam model pembelajaran discovery learning kegiatan tersebut termasuk pada tahap ....*
- A. *identifikasi masalah*
  - B. *pengolahan data*
  - C. *pengumpulan data*
  - D. *menarik kesimpulan*

Butir soal nomor 2 termasuk kategori sulit dimungkinkan karena banyak guru matematika kurang memahami dengan baik yang dimaksud dengan model pembelajaran, sehingga ketika diberikan suatu ilustrasi sebagian aktivitas pembelajaran menurut model pembelajaran tersebut guru kesulitan menentukan termasuk pada tahap mana aktivitas tersebut dalam model.

Daya beda soal nomor 2 jelek dimungkinkan karena rumusan pokok soal tidak jelas atau menimbulkan multitafsir sehingga guru salah memahami yang dimaksud soal pada saat mengerjakan soal tersebut. Hal ini berdampak justru guru dengan kemampuan lebih tinggi menjawab salah, sedang guru dengan kemampuan lebih rendah justru menjawab dengan benar.

- **Butir soal nomor 7**

- 7 *Meskipun komunikasi antara guru dan peserta didik dalam proses pembelajaran termasuk komunikasi publik atau kelompok, guru sewaktu-waktu bisa mengubahnya menjadi komunikasi antarpersonal; hal ini bisa dilakukan karena ....*
- A. *peserta didik di kelas merupakan bagian komunitas di bawah pengelolaan guru*

- B. peserta didik merespon apapun yang dikomunikasikan guru kepada mereka
- C. proses komunikasi tatap muka di kelas mempunyai kelompok yang relatif kecil
- D. proses komunikasi di kelas bergantung sepenuhnya oleh keputusan guru

Butir soal nomor 7 termasuk kategori sulit dimungkinkan karena banyak guru tidak menguasai kompetensi yang berkaitan dengan strategi komunikasi dalam pembelajaran, sehingga ketika bertemu dengan masalah yang berkaitan dengan komunikasi antara guru dengan peserta didik guru kesulitan menjawab. Analisis lebih jauh penulis juga menemukan bahwa pilihan jawaban A dipilih lebih dari separuh responden, yang berarti bahwa pilihan jawaban A tersebut mengecoh berlebihan.

Butir soal nomor 7 juga memiliki daya beda jelek, bahkan negatif, dimungkinkan karena responden kesulitan memahami maksud soal, atau salah tafsir terhadap soal, sehingga berdampak tidak saja banyak guru salah mengerjakan, tetapi jumlah guru dengan kemampuan tinggi lebih banyak salah dibanding guru dengan kemampuan rendah.

- **Butir soal nomor 19**

- 19 Diketahui akar-akar persamaan kuadrat  $x^2 - (p - 2)x - 6 = 0$  adalah  $\alpha$  dan  $\beta$  yang memenuhi  $\alpha^2 + 2\alpha\beta + \beta^2 = 5$ . Salah satu nilai  $p$  yang memenuhi adalah ....
- A. -2
  - B. -1
  - C. 1
  - D. 2

Butir soal nomor 19 termasuk kategori sulit dimungkinkan karena banyak guru kurang memahami sifat akar persamaan kuadrat untuk menyelesaikan masalah persamaan kuadrat sehingga ketika diberikan soal yang berkaitan dengan sifat akar persamaan kuadrat dan soal tersebut adalah soal yang tidak rutin maka banyak guru salah mengerjakan. Butir soal nomor 19 memiliki daya beda negatif dimungkinkan karena rumusan kalimat soal nomor 19 sulit dipahami atau menimbulkan multitafsir, sehingga berdampak tidak saja banyak guru salah mengerjakan, tetapi jumlah guru dengan kemampuan tinggi lebih banyak salah dibanding guru dengan kemampuan rendah.

- **Butir nomor 5, 6, 9, 18, dan 29**

5 *Berikut disajikan daftar nilai matematika di kelas yang diampu Pak Anton.*

*Daftar Nilai Matematika*

<i>No</i>	<i>Nama Siswa</i>	<i>Nilai</i>
<i>1</i>	<i>Naura</i>	<i>9.25</i>
<i>2</i>	<i>Khayla</i>	<i>9.25</i>
<i>3</i>	<i>Anindya</i>	<i>8.75</i>
<i>.</i>		
<i>36</i>	<i>Zahra</i>	<i>8.75</i>
<i>Rata-rata</i>		

*Jika Pak Anton ingin menemukan rata-rata nilai matematika siswa tersebut dengan menggunakan Microsoft Excel, fungsi pada pada Microsoft Excel yang tepat digunakan adalah ....*

- A. Sum*
- B. STDEV*
- C. Average*
- D. Count numbers*

- 6 Seorang guru memiliki siswa dengan kecerdasan yang beragam di kelasnya. Sebagian siswa memiliki kecerdasan di atas rata-rata dan sebagian yang lain termasuk siswa lambat belajar (*slow learner*). Pada saat mengajar materi dimensi tiga yang merupakan salah satu materi yang dianggap paling sulit oleh sebagian besar siswa, guru tersebut akan menggunakan pembelajaran berkelompok. Berikut ini cara pengelompokan siswa yang paling tepat adalah ....
- Siswa dengan kecerdasan di atas rata-rata dikelompokkan menjadi satu agar mereka tidak terhambat belajarnya oleh siswa lain
  - Siswa yang termasuk lamban belajar dikelompokkan menjadi satu dengan siswa dengan kecerdasan di atas rata-rata agar terbantu belajarnya
  - Siswa dikelompokkan secara heterogen dimana setiap kelompok terdapat siswa dengan kecerdasan di atas rata-rata dan siswa yang lamban belajar
  - Siswa dengan kemampuan rata-rata dibaurkan kelompoknya dengan siswa yang memiliki kecerdasan di atas rata-rata sedangkan siswa lamban belajar di buat kelompok tersendiri
- 9 Berdasarkan hasil analisis penilaian yang dilakukan seorang guru, diperoleh data ketuntasan klasikal kurang dari 50% sehingga peserta didik yang perlu mengikuti pembelajaran remedial lebih dari 50%. Bentuk remedial yang tepat dilakukan guru tersebut adalah ....
- Memberikan pembelajaran ulang dengan metode dan media yang berbeda
  - Memberikan bimbingan secara khusus kepada peserta didik yang belum tuntas
  - Memberikan tugas belajar tambahan bagi peserta didik yang belum tuntas
  - Memberikan penugasan secara berkelompok melibatkan semua peserta didik
- 18 Diketahui fungsi  $f(x) = (p - 4)x^2 + (2p - 20)x + 6$  memiliki sumbu simetri  $x = 2$ . Nilai ekstrim fungsi tersebut adalah ...
- Maksimum  $-6$
  - Maksimum  $-2$

- C. *Minimum 2*
- D. *Minimum 6*

- 29 *Berikut ini yang termasuk kegiatan pengembangan keprofesian berkelanjutan melalui tindakan reflektif adalah ....*
- A. *melaksanakan penelitian tindakan kelas*
  - B. *mengikuti kegiatan kolektif di MGMP*
  - C. *mengikuti diklat fungsional*
  - D. *membuat karya inovatif*

Butir soal nomor 5, 6, 9, 18, dan 29 memiliki daya beda negatif dimungkinkan karena rumusan soal pada butir-butir tersebut sulit dipahami atau dapat menimbulkan multitafsir, sehingga menyebabkan responden salah memahami soal yang menyebabkan banyak guru dengan kemampuan tinggi salah mengerjakan sedangkan guru dengan kemampuan rendah banyak benar mengerjakan. Oleh karena itu diperlukan revisi soal dengan memperbaiki rumusan soal, baik pokok soal, pilihan jawaban, atau keduanya, sehingga lebih mudah dipahami dan tidak menimbulkan multitafsir.

- **Butir soal nomor 15**

- 15 *Diketahui nilai rata-rata dari 20 siswa adalah 60. Jika ditambahkan nilai dari 10 siswa lainnya maka rata-rata nilainya menjadi 65. Nilai rata-rata dari 10 siswa yang ditambahkan adalah ....*
- A. *70*
  - B. *75*
  - C. *80*
  - D. *85*

Untuk butir nomor 15 penulis memandang perlu diganti, karena daya beda negatif ada dua pilihan tidak berfungsi, padahal ketika penulis cermati rumusan soal nomor 15 sudah cukup sederhana dan

jelas, sehingga daya beda negatif tersebut dimungkinkan karena memang butir soal nomor 1 jelek. Selain itu, soal nomor 15 termasuk kategori mudah (hampir semua, yaitu 95%, benar). Apabila rumusan soal diubah penulis menduga justru berdampak lebih tidak baik, karena soal mungkin justru lebih sulit dipahami dan bisa menimbulkan salah tafsir. Dengan pertimbangan ini, penulis memutuskan untuk mengganti butir soal nomor 15.

Berdasarkan hasil ujicoba terbatas di atas selanjutnya penulis merevisi soal Paket I. Revisi dilakukan dengan cara mengkaji kembali rumusan pertanyaan soal/stimulus dan alternatif jawaban, kemudian memperbaiki rumusan soal (pertanyaan dan/atau alternatif jawaban). Hal ini dilakukan karena dimungkinkan soal tidak baik karena sulit dipahami atau multitafsir. Selain itu, revisi dilakukan dengan mengecek data–data di soal dan kunci jawaban dengan pertimbangan kemungkinan salah data atau tidak ada kunci jawaban yang benar.

Secara operasional, tindak lanjut hasil ujicoba terbatas ada 7 (tujuh) macam langkah, yaitu: mengganti soal, merevisi pokok soal, merevisi alternatif jawaban, merevisi pokok soal dan alternatif jawaban, mengganti alternatif jawaban, dan merevisi data soal. Rincian revisi soal Paket I disajikan pada Tabel 6. Setelah revisi dilakukan diperoleh soal yang siap untuk diujicobakan dengan skala lebih luas.

Soal hasil perbaikan yang dilakukan berdasarkan hasil ujicoba di atas selanjutnya dilakukan ujicoba berikutnya dengan skala yang lebih luas. Pembahasan hasil ujicoba skala luas untuk soal Paket 1 akan dibahas di bagian selanjutnya pada buku ini.

Tabel 6. Revisi Butir Soal Paket I

<b>Tindaklanjut hasil ujicoba terbatas</b>	<b>Butir yang direvisi</b>
Soal tetap/tidak direvisi	4,8,12,13,14,16,17,20,22,23,24,25,26,27,28
Soal diganti	15
Revisi pokok soal	1,2,9,19,29
Revisi alternatif jawaban	30
Revisi pokok soal dan alternatif jawaban	3,5,6,7,10,18
Alternatif jawaban diganti	11
Revisi data soal	21

### **C. Analisis Butir Tes Kompetensi Guru Matematika Paket II dengan Teori Klasik**

Misalkan suatu tes kompetensi guru Paket 2 berbentuk soal pilihan ganda dengan 4 (empat) pilihan sebanyak 30 butir dengan diujikan pada 55 guru. Dari data hasil ujicoba ini juga akan diidentifikasi tingkat kesukaran, daya beda, dan keberfungsian distrukturanya. Temuan tingkat kesukaran, daya beda, dan keberfungsian distruktur ini selanjutnya dijadikan bahan untuk memperbaiki soal yang disusun.

Pada ujicoba ini, data hasil pengerjaan tes kompetensi guru Paket 2 oleh para responden selanjutnya diolah dengan menggunakan program ITEMAN. Berikut disajikan hasil analisisnya.



# 1. Analisis Data Tes Paket II

## Output Analisis dengan Iteman

MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket II.txt

Page 1

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics								
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key				
1	0-1	0.909	0.151	0.086	A	0.018	0.426	0.142	?				
					B	0.018	-0.813	-0.272					
					CHECK THE KEY				C	0.055	-0.067	-0.033	
					D was specified, A works better				D	0.909	0.151	0.086	*
					Other	0.000	-9.000	-9.000					
2	0-2	0.345	0.543	0.421	A	0.127	-0.533	-0.333					
					B	0.436	-0.187	-0.148					
					C	0.073	0.050	0.027					
					D	0.345	0.543	0.421	*				
					Other	0.018	-0.503	-0.168					
3	0-3	0.527	0.159	0.126	A	0.164	-0.091	-0.060					
					B	0.091	-0.405	-0.230					

					C	0.218	0.086	0.062	
					D	0.527	0.159	0.126	*
					Other	0.000	-9.000	-9.000	
4	0-4	0.782	0.273	0.195	A	0.018	-0.813	-0.272	
					B	0.000	-9.000	-9.000	
					C	0.200	-0.158	-0.111	
					D	0.782	0.273	0.195	*
					Other	0.000	-9.000	-9.000	
5	0-5	0.727	0.084	0.063	A	0.091	0.187	0.106	?
					B	0.127	-0.290	-0.181	
					C	0.727	0.084	0.063	*
					D	0.055	0.016	0.008	
					Other	0.000	-9.000	-9.000	
					CHECK THE KEY				
					C was specified, A works better				
6	0-6	0.400	0.165	0.130	A	0.091	-0.320	-0.182	
					B	0.400	0.165	0.130	*
					C	0.327	0.042	0.032	
					D	0.182	-0.100	-0.068	
					Other	0.000	-9.000	-9.000	
7	0-7	0.527	0.367	0.293	A	0.000	-9.000	-9.000	
					B	0.236	-0.020	-0.015	
					C	0.236	-0.454	-0.329	
					D	0.527	0.367	0.293	*
					Other	0.000	-9.000	-9.000	

MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket II.txt

Page 2

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
8	0-8	0.218	0.258	0.184	A	0.255	-0.104	-0.076	
					B	0.218	0.258	0.184	*
					C	0.327	-0.047	-0.036	
					D	0.200	-0.092	-0.065	
					Other	0.000	-9.000	-9.000	
9	0-9	0.800	0.174	0.122	A	0.036	-0.274	-0.117	
					B	0.800	0.174	0.122	*
					C	0.109	-0.252	-0.151	
					D	0.055	0.183	0.089	
					Other	0.000	-9.000	-9.000	
10	0-10	0.673	0.622	0.478	A	0.164	-0.352	-0.235	
					B	0.127	-0.754	-0.472	
					C	0.036	0.246	0.105	
					D	0.673	0.622	0.478	*
					Other	0.000	-9.000	-9.000	

11	0-11	0.836	0.016	0.011	A	0.836	0.016	0.011	*
					B	0.036	-0.043	-0.018	
					C	0.109	0.117	0.070	?
					D	0.000	-9.000	-9.000	
					Other	0.018	-0.503	-0.168	
CHECK THE KEY									
A was specified, C works better									
12	0-12	0.782	0.555	0.396	A	0.145	-0.484	-0.313	
					B	0.782	0.555	0.396	*
					C	0.055	-0.193	-0.094	
					D	0.018	-0.709	-0.237	
					Other	0.000	-9.000	-9.000	
13	0-13	0.691	0.443	0.337	A	0.091	-0.433	-0.246	
					B	0.691	0.443	0.337	*
					C	0.109	-0.326	-0.196	
					D	0.055	0.058	0.028	
					Other	0.055	-0.276	-0.134	
14	0-14	0.655	0.457	0.355	A	0.200	-0.372	-0.260	
					B	0.091	-0.489	-0.278	
					C	0.655	0.457	0.355	*
					D	0.036	0.188	0.080	
					Other	0.018	0.013	0.004	

MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket II.txt

Page 3

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
15	0-15	0.491	0.698	0.557	A	0.036	-0.563	-0.240	
					B	0.036	-0.737	-0.314	
					C	0.436	-0.444	-0.353	
					D	0.491	0.698	0.557	*
					Other	0.000	-9.000	-9.000	
16	0-16	0.818	0.449	0.307	A	0.036	-0.852	-0.363	
					B	0.818	0.449	0.307	*
					C	0.127	-0.091	-0.057	
					D	0.018	-0.709	-0.237	
					Other	0.000	-9.000	-9.000	
17	0-17	0.655	0.657	0.510	A	0.091	-0.151	-0.086	
					B	0.655	0.657	0.510	*
					C	0.236	-0.633	-0.459	
					D	0.000	-9.000	-9.000	
					Other	0.018	-0.503	-0.168	

18	0-18	0.636	0.705	0.550	A	0.636	0.705	0.550	*				
					B	0.236	-0.648	-0.470					
					C	0.055	-0.151	-0.073					
					D	0.073	-0.350	-0.186					
					Other	0.000	-9.000	-9.000					
19	0-19	0.200	0.138	0.097	A	0.055	-0.109	-0.053					
					B	0.200	0.138	0.097	*				
					C	0.455	0.060	0.048					
					D	0.182	-0.449	-0.307					
					Other	0.109	0.364	0.218					
20	0-20	0.200	-0.092	-0.065	A	0.200	-0.092	-0.065	*				
					B	0.345	0.255	0.198	?				
					CHECK THE KEY				C	0.400	-0.146	-0.115	
					A was specified, B works better				D	0.055	-0.109	-0.053	
					Other	0.000	-9.000	-9.000					
21	0-21	0.636	0.509	0.397	A	0.636	0.509	0.397	*				
					B	0.109	-0.302	-0.181					
					C	0.182	-0.327	-0.224					
					D	0.073	-0.350	-0.186					
					Other	0.000	-9.000	-9.000					

MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket II.txt

Page 4

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
22	0-22	0.727	0.653	0.487	A	0.036	-0.390	-0.166	
					B	0.145	-0.504	-0.327	
					C	0.727	0.653	0.487	*
					D	0.091	-0.433	-0.246	
					Other	0.000	-9.000	-9.000	
23	0-23	0.327	0.553	0.425	A	0.327	0.553	0.425	*
					B	0.036	-0.274	-0.117	
					C	0.618	-0.504	-0.395	
					D	0.018	0.323	0.108	
					Other	0.000	-9.000	-9.000	
24	0-24	0.709	0.604	0.456	A	0.709	0.604	0.456	*
					B	0.145	-0.403	-0.261	
					C	0.073	-0.050	-0.026	
					D	0.055	-0.901	-0.439	
					Other	0.018	-0.193	-0.065	

25	0-25	0.727	0.417	0.311	A	0.145	-0.141	-0.091	*
					B	0.727	0.417	0.311	
					C	0.036	-0.390	-0.166	
					D	0.073	-0.616	-0.328	
					Other	0.018	0.220	0.073	
26	0-26	0.545	0.591	0.470	A	0.545	0.591	0.470	*
					B	0.036	0.073	0.031	
					C	0.055	-0.693	-0.337	
					D	0.364	-0.435	-0.340	
					Other	0.000	-9.000	-9.000	
27	0-27	0.182	0.284	0.194	A	0.309	0.015	0.012	*
					B	0.036	-0.274	-0.117	
					C	0.473	-0.147	-0.117	
					D	0.182	0.284	0.194	
					Other	0.000	-9.000	-9.000	
28	0-28	0.200	0.352	0.247	A	0.073	-0.216	-0.115	*
					B	0.273	-0.251	-0.187	
					C	0.200	0.352	0.247	
					D	0.455	0.037	0.029	
					Other	0.000	-9.000	-9.000	



MicroCAT (tm) Testing System  
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket II.txt

Page 5

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
29	0-29	0.309	0.251	0.191	A	0.164	-0.315	-0.210	
					B	0.309	0.251	0.191	*
					C	0.200	-0.125	-0.088	
					D	0.327	0.068	0.052	
					Other	0.000	-9.000	-9.000	
30	0-30	0.636	0.349	0.273	A	0.255	-0.391	-0.288	
					B	0.636	0.349	0.273	*
					C	0.091	0.046	0.026	
					D	0.018	-0.297	-0.099	
					Other	0.000	-9.000	-9.000	

MicroCAT (tm) Testing System

Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation

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

Item analysis for data from file Paket II.txt  
Page 6

There were 55 examinees in the data file.

Scale Statistics

-----

Scale:	0
	-----
N of Items	30
N of Examinees	55
Mean	16.873
Variance	15.566
Std. Dev.	3.945
Skew	-0.245
Kurtosis	-0.583
Minimum	9.000
Maximum	26.000
Median	17.000
Alpha	0.639
SEM	2.370
Mean P	0.562
Mean Item-Tot.	0.287
Mean Biserial	0.380

Berdasarkan hasil analisis di atas selanjutnya dilakukan analisis terhadap tingkat kesukaran, daya beda, dan keberfungsian butir. Hasil analisis terhadap ketika aspek tersebut diuraikan sebagai berikut.

## 2. Analisis Butir Soal Paket II Berdasarkan Teori Tes Klasik

### a. Tingkat kesukaran butir tes

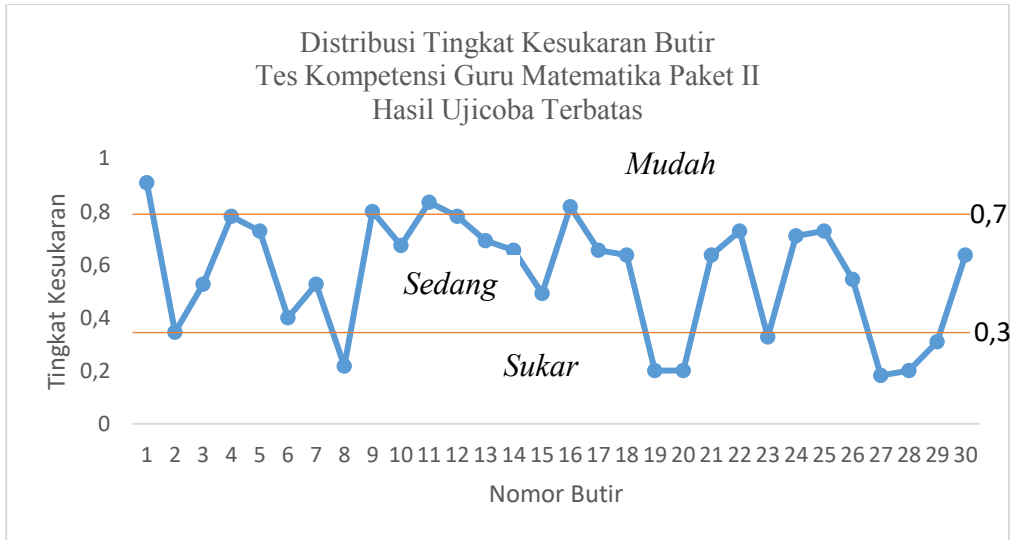
Tingkat kesukaran dari butir tes kompetensi guru matematika Paket II disajikan pada Tabel 7.

Tabel 7. Tingkat Kesukaran Soal Paket II Hasil Ujicoba Terbatas

No Butir	Prop. Correct	Tingkat Kesukaran		No Butir	Prop. Correct	Tingkat Kesukaran	
		Kategori	Kualitas			Kategori	Kualitas
1	0,909	mudah	tidak baik	16	0,818	mudah	baik
2	0,345	sedang	baik	17	0,655	sedang	baik
3	0,527	sedang	baik	18	0,636	sedang	baik
4	0,782	mudah	baik	19	0,200	sukar	baik
5	0,727	mudah	baik	20	0,200	sukar	baik
6	0,400	sedang	baik	21	0,636	sedang	baik
7	0,527	sedang	baik	22	0,727	mudah	baik
8	0,218	sukar	baik	23	0,327	sedang	baik
9	0,800	mudah	baik	24	0,709	mudah	baik
10	0,673	sedang	baik	25	0,727	mudah	baik
11	0,836	mudah	baik	26	0,545	sedang	baik
12	0,782	mudah	baik	27	0,182	sukar	baik
13	0,691	sedang	baik	28	0,200	sukar	baik
14	0,655	sedang	baik	29	0,309	sedang	baik
15	0,491	sedang	baik	30	0,636	sedang	baik

Tabel 7 menunjukkan bahwa terdapat 5 butir soal yang termasuk kategori sukar, 15 butir soal termasuk kategori sedang, dan 10 butir soal termasuk kategori mudah. Tabel 35 juga menunjukkan bahwa dilihat dari tingkat kesukarannya terdapat 29 butir termasuk kategori baik dan 1 butir soal termasuk kategori tidak baik.

Secara grafis, distribusi tingkat kesukaran butir dan daya beda tes kompetensi guru matematika Paket II disajikan pada Gambar 8.



Gambar 8. Distirbusi Tingkat Kesukaran Butir Tes Kompetensi Guru Matematika Paket II Hasil Ujicoba Terbatas

b. Daya Bada butir tes Paket II

Daya beda dari butir tes kompetensi guru matematika SMA Paket II disajikan pada Tabel 8.

Tabel 8. Daya Bada Butir Tes Paket II Hasil Ujicoba Terbatas

No Butir	Koefisien point biserial ( $r_{pbi}$ )	Kategori	No Butir	Koefisien point biserial ( $r_{pbi}$ )	Kategori
1	0,086	Soal jelek, dibuang atau sepenuhnya direvisi	16	0,307	Soal baik, sedikit atau tidak perlu perbaikan
2	0,421	Soal berfungsi dengan memuaskan	17	0,510	Soal berfungsi dengan memuaskan
3	0,126	Soal jelek, dibuang atau sepenuhnya direvisi	18	0,550	Soal berfungsi dengan memuaskan
4	0,195	Soal jelek, dibuang atau sepenuhnya direvisi	19	0,097	Soal jelek, dibuang atau sepenuhnya direvisi

No Butir	Koefisien <i>point biserial</i> ( $r_{pbi}$ )	Kategori
5	0,063	Soal jelek, dibuang atau sepenuhnya direvisi
6	0,130	Soal jelek, dibuang atau sepenuhnya direvisi
7	0,293	Soal kurang baik, perlu direvisi
8	0,184	Soal jelek, dibuang atau sepenuhnya direvisi
9	0,122	Soal jelek, dibuang atau sepenuhnya direvisi
10	0,478	Soal berfungsi dengan memuaskan
11	0,011	Soal jelek, dibuang atau sepenuhnya direvisi
12	0,396	soal baik, sedikit atau tidak perlu perbaikan
13	0,337	soal baik, sedikit atau tidak perlu perbaikan
14	0,355	soal baik, sedikit atau tidak perlu perbaikan
15	0,557	Soal berfungsi dengan memuaskan

No Butir	Koefisien <i>point biserial</i> ( $r_{pbi}$ )	Kategori
20	-0,065	Soal jelek, dibuang atau sepenuhnya direvisi
21	0,397	soal baik, sedikit atau tidak perlu perbaikan
22	0,487	Soal berfungsi dengan memuaskan
23	0,425	Soal berfungsi dengan memuaskan
24	0,456	Soal berfungsi dengan memuaskan
25	0,311	soal baik, sedikit atau tidak perlu perbaikan
26	0,470	Soal berfungsi dengan memuaskan
27	0,194	Soal jelek, dibuang atau sepenuhnya direvisi
28	0,247	Soal kurang baik, perlu direvisi
29	0,191	Soal jelek, dibuang atau sepenuhnya direvisi
30	0,273	Soal kurang baik, perlu direvisi

Tabel 8 menunjukkan bahwa terdapat beberapa butir soal memiliki daya beda jelek, bahkan ada yang negatif, yaitu butir 1, 1, 2, 3, 4, 5, 6, 8, 9, 11, 19, 20, 27, 29. Penulis kemudian melakukan kajian terhadap butir-butir tersebut untuk memperoleh temuan sebagai bahan perbaikan.

c. Keberfungsian Distraktor

Selain tingkat kesukaran dan daya beda, analisis juga dilakukan terhadap keberfungsioan distraktor. Analisis keberfungsian distraktor dimaksudkan untuk mengetahui berfungsi tidaknya distraktor pada alternatif jawaban yang tersedia. Setelah dilakukan analisis dengan program ITEMAN, berikut disajikan profil keberfungsian distraktor hasil ujicoba terbatas pada Tabel 9.

Tabel 9. Keberfungsian Distraktor Soal Paket II Hasil Ujicoba Terbatas

No Butir	Kunci	Distraktor		No Butir	Kunci	Distraktor	
		Berfungsi	Tidak berfungsi			Berfungsi	Tidak berfungsi
1	D	A,B,C	–	16	B	A,C,D	–
2	D	A,B,C	–	17	B	A,C	D
3	D	A,B,C	–	18	A	B,C,D	–
4	D	A,C	B	19	B	A,C,D	–
5	C	A,B,D	–	20	A	B,C,D	–
6	B	A,C,D	–	21	A	B,C,D	–
7	D	B,C	A	22	C	A,B,D	–
8	B	A,C,D	–	23	A	B,C,D	–
9	B	A,C,D	–	24	A	B,C,D	–
10	D	A,B,C	–	25	B	A,C,D	–
11	A	B,C	D	26	A	B,C,D	–
12	B	A,C,D	–	27	D	A,B,C	–
13	B	A,C,D	–	28	C	A,B,D	–
14	C	A,B,D	–	29	B	A,C,D	–
15	D	A,B,C	–	30	B	A,C,D	–

Tabel 9 menunjukkan bahwa terdapat beberapa soal yang distraktornya tidak berfungsi dengan baik karena tidak ada yang memilih atau dipilih kurang dari 5% peserta tes, sehingga soal tersebut perlu dicermati dan diperbaiki lagi.

### 3. Tindaklanjut Hasil Analisis Paket II

Berdasarkan hasil analisis tingkat kesukaran, daya beda, dan keberfungsian distraktor, selanjutnya dilakukan kajian terhadap butir-butir yang belum termasuk kategori baik. Setelah dilakukan kajian terhadap butir-butir soal yang telah diujicobakan, terutama terkait butir-butir yang termasuk kategori sukar dan butir-butir yang memiliki daya beda negatif atau bahkan negatif, penulis memperoleh beberapa temuan. Beberapa butir soal termasuk kategori sulit dimungkinkan ada beberapa penyebab, yaitu: banyak guru yang tidak menguasai dengan baik konsep/materi yang dipertanyakan pada soal, atau rumusan soal sulit dipahami atau multi tafsir sehingga guru salah memahami soal, atau karena keduanya. Untuk itu, diperlukan kajian seksama terhadap masing butir soal sebagai bagian dari langkah perbaikan terhadap butir soal.

Sebagai gambaran, berikut disajikan beberapa kajian lanjutan terhadap butir-butir soal Paket II.

#### - Butir soal nomor 8

- 8 *Setelah dilakukan remedial, peserta didik berhak mendapatkan nilai sebagai gambaran penguasaan kompetensinya. Berikut ini yang tidak termasuk alternatif cara pemberian nilai peserta didik yang telah mengikuti remedial adalah ....*
- A. *Peserta didik diberi nilai sesuai capaian yang diperoleh peserta didik setelah mengikuti remedial.*
  - B. *Peserta didik diberikan nilai paling tinggi sama dengan nilai paling rendah peserta didik yang tidak mengikuti program remedial*
  - C. *Peserta didik diberi nilai sama dengan KKM, berapapun nilai yang dicapai peserta didik tersebut telah melampaui nilai KKM*
  - D. *Peserta didik diberi nilai dengan cara merata-rata nilai capaian awal (sebelum mengikuti remedial) dan capaian akhir (setelah mengikuti remedial)*

Butir soal nomor 8 termasuk kategori sukar dimungkinkan karena banyak guru kurang menguasai berbagai alternatif bentuk remedial, sehingga ketika diberi soal terkait alternatif tindak lanjut remedial banyak guru salah menjawab.

Butir soal nomor 8 memiliki daya beda jelek dimungkinkan karena rumusan soal kurang mudah dipahami atau multi tafsir sehingga kurang mampu membedakan guru dengan kemampuan tinggi dengan kemampuan rendah. Hal ini juga mungkin menjadi penyebab banyak guru salah menjawab, sehingga diperlukan perbaikan rumusan soal.

- **Butir soal nomor 19**

19. Diketahui vektor posisi titik  $A = 2i + j + 3k$ ,  $B = 3i + 2j + k$ , dan  $C = 2i + 3j + 5k$ . Luas segitiga ABC adalah ....
- A.  $\frac{1}{2}\sqrt{2}$
  - B.  $\sqrt{3}$
  - C.  $2\sqrt{3}$
  - D.  $3\sqrt{2}$

Butir soal nomor 19 termasuk kategori sulit dimungkinkan karena banyak guru kurang memahami penerapan vektor untuk menyelesaikan masalah yang berkaitan dengan materi matematika yang lain. Hal ini menyebabkan ketika guru diberikan soal terapan vektor dalam dimensi tiga banyak salah. Butir soal nomor 19 memiliki daya beda jelek dimungkinkan karena redaksi soal tidak baik, sulit dipahami atau menimbulkan multi tafsir. Hal ini juga yang dapat menjadi penyebab banyak guru salah menjawab butir soal nomor 19.



- **Butir soal nomor 20**

20 *Teorema Pythagoras adalah suatu keterkaitan dalam geometri Euklides antara tiga sisi sebuah segitiga siku-siku. Teorema ini dinamakan Teorema Pythagoras karena ....*

- A. *Pythagoras yang pertama memberi sebuah bukti (secara geometris)*
- B. *Pythagoras matematikawan pertama yang menemukan keterkaitan tiga sisi sebuah segitiga siku-siku*
- C. *Pythagoras yang menyatakan keterkaitan tiga sisi sebuah segitiga siku-siku dalam suatu teorema*
- D. *Pythagoras matematikawan utama yang mempublikasikan keterkaitan tiga sisi sebuah segitiga siku-siku*

Butir soal nomor 20 termasuk kategori sukar dimungkinkan banyak guru yang kurang memahami keterkaitan Teorema Pythagoras dengan Geometri Euclides sehingga ketika ditanyakan keterkaitan keduanya banyak guru yang salah.

Butir soal nomor 20 memiliki daya beda negatif, dimungkinkan karena rumusan soal tidak jelas atau menimbulkan salah penafsiran, sehingga guru dengan kemampuan tinggi justru lebih banyak salah mengerjakan dibandingkan guru dengan kemampuan rendah. Rumusan soal yang tidak jelas atau multitafsir ini juga yang menyebabkan banyak guru salah mengerjakan.

- **Butir soal nomor 27**

27 *Perhatikan salah satu kompensi dasar ranah keterampilan kelas X SMA berikut.*

*KD 4.9 Menyelesaikan masalah yang berkaitan dengan aturan sinus dan cosinus. Berkaitan dengan kompetensi dasar di atas pernyataan berikut tepat, **kecuali** ....*

- A. *Kemampuan yang menjadi target utama pembelajaran adalah kemampuan dalam hal menyelesaikan masalah*
- B. *Masalah-masalah yang disajikan dalam pembelajaran dapat berupa masalah kontekstual yang berkaitan dengan aturan sinus dan cosinus*

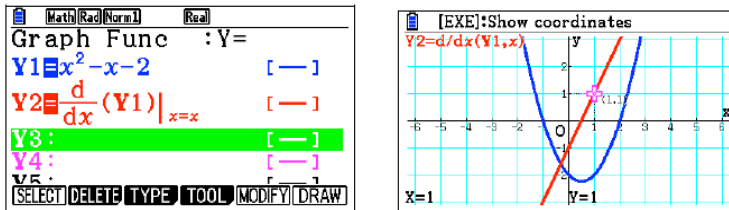
- C. *Penguasaan guru tentang materi aturan sinus dan cosinus penting karena dibutuhkan dalam mengajarkan penyelesaian masalah tersebut pada peserta didik*
- D. *Fokus utama pembelajaran adalah memfasilitasi peserta didik menguasai konsep aturan sinus dan cosinus sehingga dapat mendukung pembelajaran pemecahan masalah*

Butir soal nomor 27 termasuk kategori sukar dimungkinkan karena banyak guru tidak memiliki kemampuan menganalisis kompetensi dasar sehingga ketika diberikan suatu kompetensi dasar dan diminta memilih pernyataan yang tepat terkait kompetensi dasar tersebut banyak guru salah menjawab. Butir soal nomor 27 memiliki daya beda jelek dimungkinkan karena redaksi soal tidak baik, sulit dipahami atau menimbulkan multi tafsir. Hal ini juga yang dapat menjadi penyebab banyak guru salah menjawab butir soal nomor 27.

- Butir soal nomor 28 termasuk kategori sulit dimungkinkan karena banyak guru kurang memiliki kemampuan menentukan masalah kontekstual yang tepat digunakan untuk mendukung pembelajaran, sehingga ketika diberikan beberapa masalah kontekstual banyak guru tidak dapat menentukan dengan benar mana dari masalah tersebut yang tepat digunakan pada suatu materi.
- Butir soal nomor 1, 2, 3, 4, 5, 6, 9, 11, dan 29
  - 1 *Salah satu peserta didik di suatu kelas memiliki kecerdasan di atas rata-rata. Peserta didik tersebut dapat didiberdayakan guru untuk meningkatkan prestasi belajar matematika sekolah melalui beberapa cara berikut, **kecuali** ....*
    - A. *Menjadikan peserta didik tersebut sebagai tutor sebaya*
    - B. *Mengikutsertakan peserta didik tersebut dalam ajang kompetisi matematika*
    - C. *Menjadikan peserta didik tersebut sebagai role model untuk memotivasi peserta didik lain*
    - D. *Memberi layanan pembelajaran sama dengan yang lain untuk menunjukkan keadilan perlakuan di kelas*

- 2 *Pembelajaran matematika kadang membosankan bagi peserta didik karena dianggap tidak bermakna bagi mereka. Hal ini perlu dicari penyelesaiannya agar pembelajaran dapat mendorong peserta didik belajar optimal.  
Apabila masalah tersebut dicarikan solusi dengan merujuk pada teori belajar Ausubel, langkah tepat menjadikan pembelajaran matematika bermakna adalah ....*
- disesuaikan tingkat perkembangan kognitif peserta didik*
  - bertahap dari kongkrit, semi kongkrit, dan terakhir baru tahap abstrak*
  - menghargai kemajuan belajar peserta didik melalui pemberian penghargaan*
  - menghubungkan yang dipelajari dengan struktur pengertian yang sudah dimiliki peserta didik*
- 3 *Salah satu indikator pencapaian kompetensi yang hendak dicapai peserta didik adalah "menemukan peluang empirik dari data luaran berdasarkan sekelompok data nyata". Berikut ini penulisan tujuan pembelajaran yang tepat sesuai dengan indikator tersebut adalah ....*
- Peserta didik dapat menemukan peluang empirik dari data luaran berdasarkan sekelompok data nyata*
  - Peserta didik dapat menemukan peluang empirik dari data luaran berdasarkan sekelompok data nyata dengan tepat*
  - Melalui problem based learning peserta didik dapat menemukan peluang empirik dari data luaran berdasarkan sekelompok data nyata*
  - Melalui problem based learning peserta didik dapat menemukan peluang empirik dari data luaran berdasarkan sekelompok data nyata dengan tepat*
- 4 *Perhatikan beberapa aktivitas pembelajaran berikut.*
- menjelaskan kompetensi dasar yang akan dipelajari*
  - menjelaskan tujuan pembelajaran yang akan dicapai*
  - menyampaikan teknik penilaian yang akan digunakan*
  - menyampaikan materi pelajaran kepada para peserta didik*
- Keempat aktivitas di atas tepat dilakukan pada kegiatan pendahuluan pembelajaran, **kecuali** ....*
- I*
  - II*
  - III*
  - IV*

- 5 Dengan menggunakan kalkulator guru bermaksud membelajarkan kaitan fungsi kuadrat dan turunannya secara grafis. Guru memberikan suatu persamaan fungsi kuadrat, kemudian peserta didik diminta menemukan bentuk grafis dari turunan fungsi tersebut dengan kalkulator, seperti terlihat pada tampilan berikut.



Grafik di atas menunjukkan bahwa fungsi kuadrat berbentuk parabola dan fungsi turunannya berbentuk linear. Siswa dapat belajar banyak tentang sifat derivatif dengan mendiskusikan dengan siswa lain terhadap fungsi kuadrat yang lain.

Pembelajaran di atas menunjukkan bahwa guru tersebut sedang memanfaatkan kalkulator saintifik untuk ....

- merepresentasikan beberapa konsep matematika di layar untuk dipelajari peserta didik
  - menyimbolkan bahwa pembelajaran matematika telah mengintegrasikan perkembangan teknologi
  - memfasilitasi peserta didik melakukan eksplorasi untuk menemukan pemahaman konseptual matematik
  - sarana afirmasi untuk mengecek kebenaran hasil perhitungan manual yang telah dilakukan siswa sebelumnya
- 6 Pada suatu kelas terdapat beberapa peserta didik yang sangat tertarik ketika diberikan soal-soal HOTS. Mereka memiliki kemampuan lebih tinggi dibanding teman-teman sekelasnya dalam menyelesaikan soal-soal HOTS. Guru dapat memberikan pengayaan bagi peserta didik tersebut melalui pemberian pengalaman penyelesaian soal HOTS berikut, **kecuali** ....
- soal HOTS dikaitkan dengan konteks kehidupan nyata yang lebih luas
  - soal HOTS berkaitan materi selanjutnya dengan level berpikir lebih tinggi
  - soal HOTS sejenis yang diberikan secara klasikal dengan tingkat kesulitan lebih tinggi
  - soal HOTS bentuk lain dari yang diberikan secara klasikal dengan kompleksitas lebih tinggi

- 9 *Penilaian proses dan hasil belajar peserta didik dilakukan guru dengan diawali terlebih dahulu pengumpulan data melalui penilaian capaian belajar siswa oleh guru. Perhatikan pelaksanaan penilaian 3 (tiga) orang guru matematika SMA berikut.*
- I. *Pak Ahmad melakukan penilaian pada setiap pembelajaran di kelas melalui tes akhir sebagai bentuk uji untuk mengetahui pencapaian kompetensi peserta didik*
  - II. *Pak Badu melakukan penilaian pada setiap pembelajaran yang diampu melalui berbagai teknik penilaian yang dilakukan baik selama proses maupun setelah selesai pembelajaran*
  - III. *Pak Caca melakukan penilaian pada setiap pembelajaran di kelas melalui tes akhir dan penilaian antar teman sebagai bentuk uji untuk mengetahui pencapaian kompetensi peserta didik*
  - IV. *Pak Dadu melakukan penilaian pada setiap pembelajaran yang diampu melalui penilaian antar peserta didik untuk mengetahui pencapaian hasil belajar peserta didik*

*Dari keempat kegiatan penilaian di atas, yang termasuk telah menerapkan pendekatan *assessment for learning* (penilaian untuk pembelajaran) adalah penilaian yang dilakukan oleh ....*

- A. *Pak Adam*
  - B. *Pak Badu*
  - C. *Pak Caca*
  - D. *Pak Dadu*
- 11 *Jarak kedua titik potong kurva  $y = 3^{2x+1} - 12 \cdot 3^x + 9$  dengan sumbu-x adalah ....*
- A. *1*
  - B. *1.5*
  - C. *2*
  - D. *2.5*
- 29 *Berikut adalah bagian yang tepat dituliskan dalam latar belakang laporan penelitian tindakan kelas, **kecuali** ....*
- A. *Uraian fokus penelitian (masalah riil yang terjadi di kelas)*
  - B. *Rencana pelaksanaan tindakan yang akan dilakukan dalam memecahkan masalah*
  - C. *Faktor-faktor (perilaku yang tampak) dari pengamatan guru yang mendukung permasalahan*
  - D. *Uraian tindakan yang dipilih untuk upaya perbaikan dan alasan yang mendasari pemilihan tindakan*

Butir soal nomor 1, 2, 3, 4, 5, 6, 9, 11, dan 29 memiliki daya beda jelek, dimungkinkan karena ada masalah pada rumusan soal. Rumusan pada soal-soal tersebut mungkin kurang jelas, sulit dipahami, atau menimbulkan multitafsir. Oleh karena itu perlu dilakukan revisi pada butir-butir soal tersebut agar soal lebih jelas, mudah dipahami, dan tidak menimbulkan multitafsir.

Pada tes kompetensi guru matematika Paket II, berdasarkan output analisis dengan iteman juga diperoleh catatan pada beberapa butir untuk dilakukan pengecekan ulang distraktor karena distraktor lain dianggap kurang berfungsi dengan baik, yaitu pada butir: 1, 5, 11, dan 20. Setelah dilakukan pengecekan ternyata butir soal nomor 1 dan 5 memiliki daya beda kurang dari 0.20 yang menunjukkan bahwa peserta tes dengan kemampuan lebih baik lebih sedikit menjawab benar butir soal tersebut, dan sebaliknya peserta tes dengan kemampuan lebih rendah menjawab benar lebih banyak pada butir tersebut. Dengan daya beda sangat rendah kurang dari 0.2 berarti soal tersebut jelek sehingga perlu dibuang atau sepenuhnya direvisi. Pada butir soal no 11 dilihat dari daya bedanya masih di atas 0.2 tetapi dilihat tingkat kesukaran ternyata sangat mudah dengan tingkat kesukaran di atas 0,90, yaitu 0,955 sehingga termasuk kategori soal yang tidak baik dan perlu diperbaiki. Untuk butir soal nomor 20, ketika dicermati ternyata daya beda dan tingkat kesukarannya termasuk kategori baik, yaitu daya beda 0,565 dan tingkat kesukaran 0,500, tetapi setelah dicermati lebih banyak dipilih oleh peserta didik dengan kemampuan rendah dibanding dipilih oleh peserta dengan kemampuan tinggi. Hal ini menunjukkan bahwa distraktor tersebut tidak berfungsi sebagai pengecek dengan baik karena justru peserta tes dengan kemampuan tinggi lebih banyak memilih distraktor tersebut dibandingkan

peserta dengan kemampuan rendah. Dengan demikian, butir soal nomor 1, 5, 11, dan 20 perlu dicermati dan diperbaiki lagi.

Berdasarkan hasil ujicoba terbatas penulis juga merevisi soal Paket II. Seperti halnya revisi soal Paket I, revisi soal Paket II juga dilakukan dengan cara mengkaji kembali rumusan pertanyaan soal/stimulus dan alternatif jawaban, kemudian memperbaiki rumusan soal (pertanyaan dan/atau alternatif jawaban). Hal ini dilakukan karena dimungkinkan soal tidak baik karena sulit dipahami atau multitafsir. Selain itu, revisi dilakukan dengan mengecek data–data di soal dan kunci jawaban dengan pertimbangan kemungkinan salah data soal atau tidak ada kunci jawaban yang benar.

Secara operasional, tindak lanjut hasil ujicoba terbatas ada 5 (lima) macam langkah, yaitu: soal tidak tetap tidak direvisi, merevisi pokok soal, merevisi alternatif jawaban, merevisi pokok soal dan alternatif jawaban, dan mengganti alternatif jawaban. Rincian revisi soal Paket I disajikan pada Tabel 10. Setelah revisi dilakukan diperoleh soal yang siap untuk diujicobakan dengan skala lebih luas.

Tabel 10. Revisi Butir Soal Paket II

<b>Tindak lanjut hasil ujicoba terbatas</b>	<b>Butir yang direvisi</b>
Soal tetap/tidak direvisi	2,10,12,13,14,15,16,18,21,22,23,24,25,26
Revisi pokok soal	3,4,5,6,8,9,19,20,27,28,30
Revisi alternatif jawaban	29
Revisi pokok soal dan alternatif jawaban	1,7,11
Alternatif jawaban diganti	17

Soal hasil perbaikan Soal hasil revisi berdasarkan ujicoba terbatas ini selanjutnya diujicobakan pada ujicoba skala lebih luas.

## BAB IV

### TERAPAN TEORI RESPON BUTIR

#### A. Kerangka Analisis dengan Teori Respon Butir

Teori respon butir memiliki banyak penerapan dalam proses penyusunan soal, antara lain: penentuan karakteristik butir, pengembangan bank soal, penyetaraan (*equiting*), penentuan bias, *standard setting*, dan lain-lain. Terapan teori respon butir pada buku ini difokuskan untuk analisis karakteristik butir. Butir tes yang dianalisis adalah tes kompetensi guru matematika SMA Paket I dan Paket II, yang sebelumnya telah dibahas di bagian sebelumnya untuk melihat contoh terapan teori tes klasik.

Analisis karakteristik tes dalam contoh terapan teori respon butir ini meliputi: tingkat kesukaran butir, daya beda butir, *pseudo-guessing*, kecocokan model, fungsi informasi, dan estimasi kemampuan peserta tes yang cocok dengan tes yang dikembangkan. Analisis dengan teori respon butir ini dilakukan dalam 2 tahap, yaitu tahap uji asumsi dan tahap analisis tesnya.

##### 1. Uji Asumsi

Sebelum analisis dengan teori respon butir ini dilakukan, terlebih dahulu dilakukan uji asumsi, baik untuk tes Paket I maupun Paket II. Uji asumsi yang dilakukan meliputi: undimensi, independensi lokal, dan invariansi parameter, yang penentuannya didasarkan pada pendapat Retnawati (2016: 4 – 10) dan Retnawati (2017: 140 – 146), sebagai berikut.

- a. Uji asumsi unidimensi dilakukan dengan analisis faktor melalui *scree-plot* dan nilai *eigen*. Asumsi unidimensi dikatakan sebagai terpenuhi apabila pada *scree-plot* terdapat satu komponen dominan ditandai dengan curaman tajam pada *scree-plot*, dan yang lain landai. Atau,



dilihat perbandingan nilai eigen dimana asumsi unidimensi dikatakan terpenuhi apabila terdapat satu komponen pertama yang dominan di lihat dari persentase lebih dari 20%.

- b. Uji asumsi independensi lokal dalam buku ini dilakukan dengan mencermati hasil uji asumsi unidimensi. Asumsi independensi lokal dikatakan terpenuhi apabila telah berhasil ditunjukkan bahwa asumsi unidimensi terpenuhi.
- c. Uji invariansi parameter dilakukan terhadap parameter butir dan parameter kemampuan. Invariansi parameter butir meliputi: tingkat kesukaran, daya beda, dan *pseudo-guessing*. Asumsi invariansi parameter butir dilakukan dengan mengestimasi parameter butir pada kelompok genap dan ganjil. Hasil estimasi parameter selanjutnya disajikan dalam diagram pencar. Jika titik–titik pada diagram pencar mendekati garis yang melewati titik asal dan bergradien 1, maka dianggap parameter tersebut invarian.

Invariansi parameter kemampuan dibuktikan dengan terlebih dahulu mengestimasi parameter kemampuan menggunakan butir nomor genap dan ganjil. Hasil estimasi parameter kemampuan kemudian disajikan pada diagram pencar. Jika titik–titik pada diagram pencar mendekati garis yang melewati titik asal dan bergradien 1, maka dianggap parameter tersebut invarian.

## 2. Penentuan Kecocokan Model

Penentuan kecocokan model dalam buku ini didasarkan pada pendapat Retnawati (2015: 25) yang menyatakan bahwa model yang cocok yang selanjutnya digunakan untuk menganalisis data adalah model yang memiliki butir cocok paling banyak. Langkah awal penentuan kecocokan

model adalah mengidentifikasi kecocokan setiap butir dengan model 1P, 2P, dan 3P, kemudian pada ketika model tersebut diidentifikasi model mana yang memiliki butir cocok paling banyak. Model yang memiliki butir cocok paling banyak tersebut dinyatakan sebagai model yang cocok yang kemudian digunakan untuk analisis lanjutan.

### 3. Analisis Parameter Butir

Analisis karakteristik tes dilakukan dengan menggunakan program analisis *BILOG MG*. Penafsiran karakteristik butir tes dalam buku ini didasarkan pendapat Hambleton dan Swaminathan (1985; Retnawati, 2015: 17 – 18), dan Hullin (1983; Retnawati, 2015: 18) , sebagai berikut:

- a. Parameter tingkat kesukaran butir dikatakan baik apabila tingkat kesukaran butir berkisar antara  $-2$  dan  $+2$ . Jika kesukaran butir mendekati  $-2$ , maka indeks kesukaran butir sangat rendah, sedangkan jika tingkat kesukaran butir mendekati  $+2$  maka indeks kesukaran butir sangat tinggi
- b. Parameter daya beda butir dikatakan baik apabila butir mempunyai hubungan positif dengan responden, dan terletak diantara 0 dan 2.
- c. Parameter *pseudo-guessing* dikatakan baik apabila nilainya tidak lebih dari  $\frac{1}{k}$ , dengan k adalah banyaknya pilihan.

### 4. Fungsi Informasi dan Estimasi Kemampuan Peserta Tes

Fungsi informasi dan estimasi kemampuan peserta tes yang cocok dilakukan dengan menganalisis tampilan grafis *output IRT Graphics* pada *BILOG MG*. Estimasi kemampuan peserta yang cocok untuk tes yang dikembangkan dilihat dari kurva informasi tes, yang ditentukan dengan melihat titik potong antara grafik fungsi informasi dan grafik fungsi *SEM*.

## B. Terapan Teori Respon Butir untuk Analisis Tes Kompetensi Guru Matematika Paket I

Misalkan suatu tes kompetensi guru matematika Paket 1 berbentuk soal pilihan ganda dengan 4 (empat) pilihan sebanyak 30 butir revisi dari hasil ujicoba terbatas diujikan dengan skala lebih luas, dengan responden sebanyak 152 orang guru matematika SMA. Dari ujicoba ini selanjutnya dilakukan analisis dengan teori respon butir.

### 1. Uji Asumsi

#### a. Asumsi unidimensi

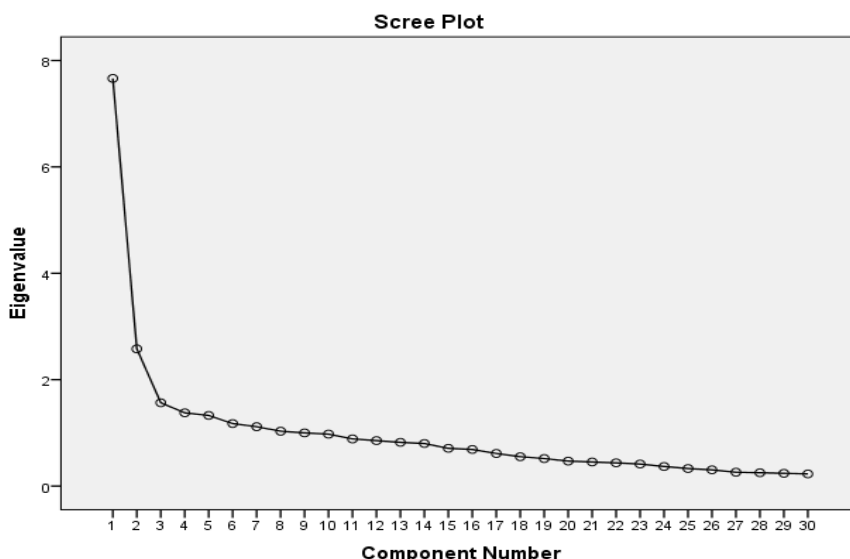
Asumsi unidimensi diuji dengan menggunakan analisis faktor, yaitu dengan melihat nilai eigen pada matriks varians kovarians inter – butir. Analisis faktor menguji asumsi unidimensi dilakukan dengan bantuan program pengolah data SPSS. Uji asumsi dengan analisis faktor ini diawali dengan melakukan analisis kecukupan sampel melalui uji *KMO* dan *Bartlett*, hasilnya disajikan pada Tabel 11.

Tabel 11. Hasil Uji *KMO* dan *Bartlett* Paket I

<i>KMO and Bartlett's Test</i>		
<i>Kaiser–Meyer–Olkin Measure of Sampling Adequacy.</i>	0,823	
<i>Bartlett's Test of Sphericity</i>	<i>Approx. Chi–Square</i>	1206,557
	<i>df</i>	435
	<i>Sig.</i>	0,000

Tabel 11 menunjukkan bahwa nilai *KMO* 0,823 dan nilai *Khi–kuadrat* pada uji *Bartlett* sebesar 1206,557 dengan derajat kebebasan 435 dan nilai sig. 0,00 (kurang dari 0,01). Hasil ini menunjukkan bahwa sampel sebesar 121 yang digunakan dalam proses ini cukup.

Setelah diperoleh hasil uji kecukupan sampel, langkah berikutnya adalah melakukan analisis faktor untuk memperoleh output *scree-plot* dan nilai eigen. Berdasarkan analisis yang dilakukan dihasilkan *scree-plot* yang disajikan pada Gambar 9.



Gambar 9. Output *Scree-Plot* Tes Paket I

Gambar 9 menunjukkan bahwa terdapat terdapat satu curaman dominan, kemudian agak melandai mulai nilai eigen kedua, dan berikutnya semakin melandai mulai nilai eigen ketiga.. Hal ini menunjukkan bahwa tes kompetensi guru matematika Paket I mengukur 1 faktor dominan. Untuk lebih menguatkan hasil analisis terhadap *scree-plot* di atas, penulis melakukan analisis terhadap nilai eigen. *Output* nilai eigen disajikan pada Tabel 12.

Tabel 12 menunjukkan bahwa nilai eigen komponen 1 cukup besar dengan persentase 25,546% terhadap total. Menurut Wells & Purwono, 2008; Retnawati, 2015: 144), persentase varians lebih besar dari 20% dikatakan bahwa perangkat mengukur dimensi utama.

Dengan demikian dapat dikatakan bahwa asumsi unidimensi terpenuhi. Hal ini selaras dengan hasil analisis *scree-plot* di atas.

Tabel 12. *Output* Nilai Eigen Tes Paket I

Komponen	Nilai Eigen	% varians	Kumulatif % Varians
1	7,664	25,546	25,546
2	2,579	8,595	34,141
3	1,566	5,220	39,361
4	1,377	4,590	43,951
5	1,327	4,425	48,376
6	1,174	3,915	52,291
7	1,115	3,718	56,009
8	1,031	3,437	59,446
9	0,999	3,330	62,776
10	0,976	3,254	66,030
Dst sampai 30			

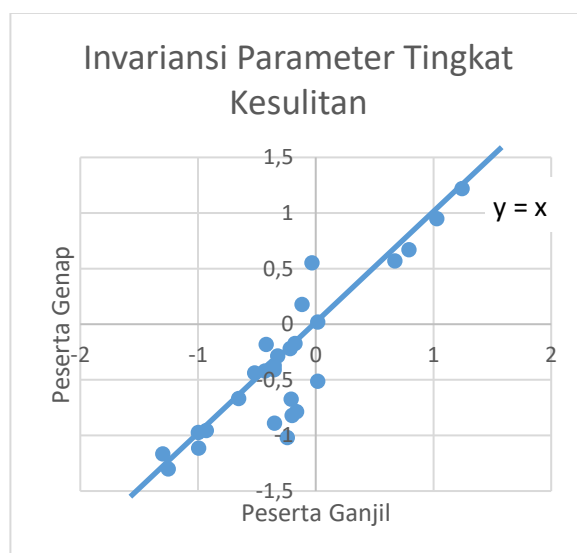
b. Asumsi independensi lokal

Asumsi independensi lokal dalam proses ini tidak diuji secara tersendiri tetapi didasarkan pada hasil uji unidimensi yang telah dilakukan. Hal ini didasarkan pendapat Hambleton, Swaminathan, dan Rogers (1991: 11) yang menyatakan bahwa asumsi independensi lokal butir terpenuhi bila bersifat unidimensi. Hal senada dinyatakan Retnawati (2015: 7) bahwa asumsi independensi lokal otomatis terbukti setelah dibuktikan dengan unidimensionalitas data respon terhadap tes. Oleh karena itu, dengan telah terbuktinya asumsi unidimensi di atas, disimpulkan bahwa asumsi independensi lokal juga terpenuhi.

c. Asumsi invariansi parameter

Ada dua parameter macam yang diuji asumsinya pada tes kompetensi guru matematika Paket I, parameter butir dan parameter kemampuan. Parameter butir meliputi tingkat kesulitan dan daya beda (untuk parameter butir *pseudo guesing* tidak diuji karena model yang cocok untuk tes kompetensi guru matematika Paket I adalah model 2P – uji kecocokan model akan diuraikan di bagian bawah).

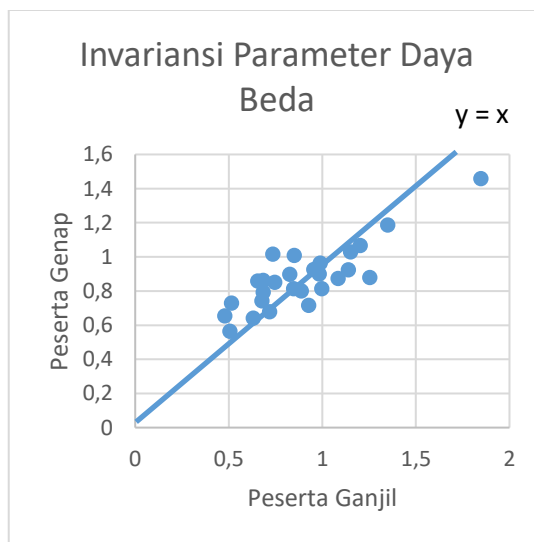
Uji asumsi invariansi parameter tingkat kesukaran butir dibuktikan dengan terlebih dahulu melakukan estimasi parameter tingkat kesukaran butir pada peserta kelompok ganjil dan genap, yang hasilnya kemudian disajikan dalam sebuah gambar diagram pencar (*scatter plot*). Diagram pencar hasil estimasi parameter tingkat kesukaran butir pada peserta kelompok ganjil dan genap disajikan pada Gambar 10.



Gambar 10. Invariansi Paramter Tingkat Kesukaran Butir Tes Paket I

Gambar 10 menunjukkan bahwa masing–masing titik berada relatif dekat dengan garis  $y = x$ . Hal ini menunjukkan bahwa tidak terjadi variasi parameter hasil estimasi pada peserta kelompok genap dan ganjil. Dengan demikian dapat dikatakan invariansi tingkat kesukaran butir terpenuhi.

Uji asumsi invariansi parameter daya beda butir juga dibuktikan dengan terlebih dahulu melakukan estimasi parameter daya beda butir pada peserta kelompok ganjil dan genap, yang hasilnya disajikan dalam gambar diagram pencar (*scatter plot*). Diagram pencar hasil estimasi parameter daya beda butir pada peserta kelompok ganjil dan genap disajikan pada Gambar 11.

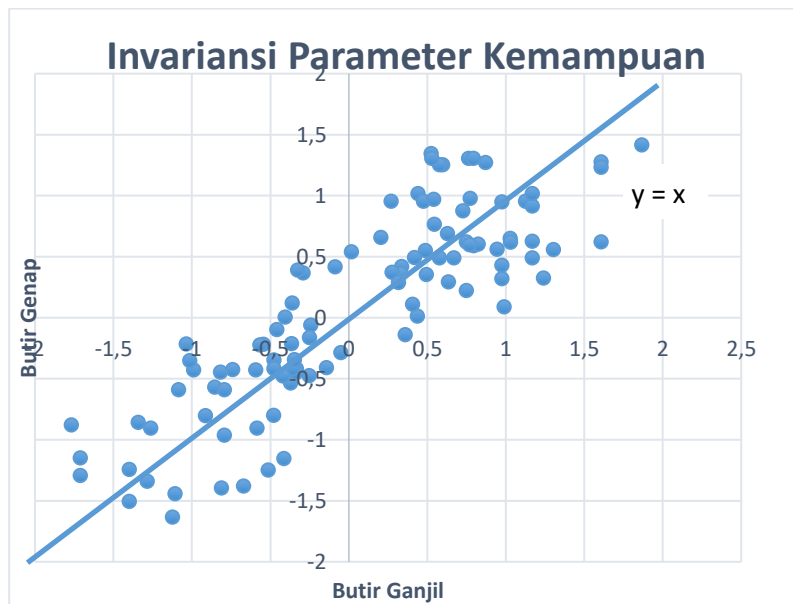


Gambar 11. Invariansi Paramter Daya Beda Butir Tes Paket I

Gambar 11 menunjukkan bahwa masing–masing titik berada relatif dekat dengan garis  $y = x$ . Hal ini menunjukkan bahwa tidak terjadi variasi parameter hasil estimasi pada peserta kelompok genap dan

ganjil. Dengan demikian dapat dikatakan bahwa invariansi daya beda butir terpenuhi.

Uji asumsi parameter berikutnya dilakukan terhadap parameter kemampuan. Asumsi invariansi parameter kemampuan dibuktikan dengan terlebih dahulu melakukan estimasi parameter kemampuan pada butir ganjil dan genap, yang hasilnya kemudian disajikan dalam sebuah gambar diagram pencar (*scatter plot*). Diagram pencar hasil estimasi parameter kemampuan pada butir ganjil dan genap disajikan pada Gambar 12.



Gambar 12. Invariansi Parameter Kemampuan Mengerjakan Butir Ganjil dan Genap

Gambar 12 menunjukkan bahwa masing–masing titik berada relatif dekat dengan garis  $y = x$ . Hal ini menunjukkan bahwa tidak terjadi variasi parameter hasil estimasi pada butir genap dan ganjil. Dengan



demikian dapat dikatakan bahwa invariansi parameter kemampuan terpenuhi.

Berdasarkan uji asumsi di atas dapat disimpulkan bahwa asumsi–asumsi untuk diterapkannya teori respons butir, yaitu asumsi parameter butir dan asumsi parameter kemampuan pada tes kompetensi guru matematika Paket I, dapat terpenuhi.

## 2. Uji Kecocokan Model

Ada 3 model yang diuji kecocokan modelnya pada tes kompetensi guru matematika SMA Paket I, yaitu model 1P, 2P, dan 3P. Pada contoh ini, uji kecocokan model dilakukan melalui kecocokan secara statistik yang didasarkan pada nilai probabilitas (signifikansi, *sig*). Butir dikatakan tidak cocok dengan model apabila nilai  $sig < \alpha$  ( $\alpha = 0,05$ ). Model logistik yang selanjutnya digunakan untuk analisis data adalah model yang memiliki butir cocok paling banyak.

Uji kecocokan model pada contoh ini dianalisis dengan menggunakan bantuan program *BILOG MG*. Analisis dilakukan untuk model 1P, 2P, dan 3P, yang hasilnya kemudian dijadikan dasar untuk menentukan model mana yang paling cocok. Berikut disajikan output analisis dengan program *BILOG MG output fase 2* dari tes Paket 1, baik untuk model 1P, 2P, maupun 3P.

### Output Analisis untuk model 1P

75

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
BUTIR01	-0.677 0.119*	0.765 0.030*	0.885 0.155*	0.608 0.024*	0.000 0.000*	10.4 (0.0654)	5.0
BUTIR02	-0.706 0.136*	0.765 0.030*	0.923 0.178*	0.608 0.024*	0.000 0.000*	2.3 (0.6720)	4.0
BUTIR03	0.272 0.118*	0.765 0.030*	-0.355 0.154*	0.608 0.024*	0.000 0.000*	6.3 (0.3956)	6.0
BUTIR04	0.325 0.123*	0.765 0.030*	-0.425 0.161*	0.608 0.024*	0.000 0.000*	12.6 (0.0496)	6.0
BUTIR05	0.784 0.134*	0.765 0.030*	-1.025 0.176*	0.608 0.024*	0.000 0.000*	4.8 (0.5755)	6.0
BUTIR06	0.849 0.134*	0.765 0.030*	-1.110 0.175*	0.608 0.024*	0.000 0.000*	4.5 (0.6091)	6.0
BUTIR07	-0.535 0.118*	0.765 0.030*	0.700 0.154*	0.608 0.024*	0.000 0.000*	6.9 (0.2244)	5.0
BUTIR08	0.352	0.765	-0.460	0.608	0.000	7.3	7.0

	0.121*	0.030*	0.158*	0.024*	0.000*	(0.4026)		
BUTIR09	0.662	0.765	-0.865	0.608	0.000	3.8	7.0	
	0.133*	0.030*	0.174*	0.024*	0.000*	(0.8012)		
BUTIR10	0.784	0.765	-1.025	0.608	0.000	7.0	6.0	
	0.130*	0.030*	0.169*	0.024*	0.000*	(0.3219)		
BUTIR11	0.662	0.765	-0.865	0.608	0.000	9.8	5.0	
	0.138*	0.030*	0.181*	0.024*	0.000*	(0.0799)		
BUTIR12	0.662	0.765	-0.865	0.608	0.000	7.2	5.0	
	0.145*	0.030*	0.189*	0.024*	0.000*	(0.2055)		
BUTIR13	0.379	0.765	-0.495	0.608	0.000	9.1	5.0	
	0.135*	0.030*	0.176*	0.024*	0.000*	(0.1042)		
BUTIR14	0.064	0.765	-0.084	0.608	0.000	4.7	6.0	
	0.130*	0.030*	0.169*	0.024*	0.000*	(0.5848)		
BUTIR15	0.489	0.765	-0.639	0.608	0.000	6.8	5.0	
	0.135*	0.030*	0.176*	0.024*	0.000*	(0.2379)		
BUTIR16	0.272	0.765	-0.355	0.608	0.000	4.1	6.0	
	0.131*	0.030*	0.171*	0.024*	0.000*	(0.6689)		
BUTIR17	0.461	0.765	-0.602	0.608	0.000	9.2	5.0	
	0.132*	0.030*	0.173*	0.024*	0.000*	(0.1000)		

BUTIR18	0.325 0.130*	0.765 0.030*	-0.425 0.170*	0.608 0.024*	0.000 0.000*	5.5 (0.4827)	6.0
BUTIR19	-0.736 0.128*	0.765 0.030*	0.962 0.167*	0.608 0.024*	0.000 0.000*	1.7 (0.7996)	4.0
BUTIR20	0.116 0.124*	0.765 0.030*	-0.151 0.163*	0.608 0.024*	0.000 0.000*	5.8 (0.4424)	6.0
BUTIR21	0.193 0.125*	0.765 0.030*	-0.253 0.164*	0.608 0.024*	0.000 0.000*	9.1 (0.1669)	6.0
BUTIR22	0.603 0.139*	0.765 0.030*	-0.788 0.182*	0.608 0.024*	0.000 0.000*	3.7 (0.7144)	6.0
BUTIR23	0.246 0.129*	0.765 0.030*	-0.321 0.169*	0.608 0.024*	0.000 0.000*	1.3 (0.9726)	6.0
BUTIR24	0.141 0.127*	0.765 0.030*	-0.185 0.167*	0.608 0.024*	0.000 0.000*	1.5 (0.9589)	6.0
BUTIR25	0.141 0.133*	0.765 0.030*	-0.185 0.174*	0.608 0.024*	0.000 0.000*	5.6 (0.4639)	6.0
BUTIR26	0.272 0.138*	0.765 0.030*	-0.356 0.180*	0.608 0.024*	0.000 0.000*	7.6 (0.2652)	6.0

BUTIR27	0.013	0.765	-0.017	0.608	0.000	8.0	6.0
	0.131*	0.030*	0.171*	0.024*	0.000*	(0.2362)	
BUTIR28	-0.454	0.765	0.593	0.608	0.000	4.7	5.0
	0.132*	0.030*	0.173*	0.024*	0.000*	(0.4502)	
BUTIR29	0.272	0.765	-0.355	0.608	0.000	7.0	6.0
	0.129*	0.030*	0.168*	0.024*	0.000*	(0.3254)	
BUTIR30	0.406	0.765	-0.531	0.608	0.000	5.2	7.0
	0.130*	0.030*	0.170*	0.024*	0.000*	(0.6395)	

\* STANDARD ERROR

LARGEST CHANGE = 0.000735

183.5 171.0  
(0.2425)

### Output Analisis untuk model 2P

79

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
BUTIR01	-0.572 0.123*	0.464 0.119*	1.233 0.362*	0.421 0.108*	0.000 0.000*	18.8 (0.0090)	7.0
BUTIR02	-0.726 0.159*	0.822 0.192*	0.883 0.194*	0.635 0.149*	0.000 0.000*	1.7 (0.9423)	6.0
BUTIR03	0.247 0.116*	0.541 0.121*	-0.456 0.231*	0.476 0.106*	0.000 0.000*	7.7 (0.3608)	7.0
BUTIR04	0.308 0.120*	0.639 0.142*	-0.482 0.213*	0.539 0.119*	0.000 0.000*	8.4 (0.1377)	5.0
BUTIR05	0.751 0.145*	0.682 0.158*	-1.101 0.258*	0.564 0.130*	0.000 0.000*	3.8 (0.5848)	5.0
BUTIR06	0.787 0.144*	0.622 0.150*	-1.266 0.301*	0.528 0.127*	0.000 0.000*	2.9 (0.7081)	5.0
BUTIR07	-0.453 0.121*	0.479 0.118*	0.944 0.295*	0.432 0.107*	0.000 0.000*	12.0 (0.0998)	7.0
BUTIR08	0.323	0.578	-0.559	0.500	0.000	4.4	7.0

08

	0.119*	0.124*	0.230*	0.107*	0.000*	(0.7294)		
BUTIR09	0.643	0.707	-0.909	0.577	0.000	4.0	7.0	
	0.137*	0.147*	0.227*	0.120*	0.000*	(0.7793)		
BUTIR10	0.703	0.552	-1.272	0.484	0.000	6.7	6.0	
	0.132*	0.131*	0.334*	0.115*	0.000*	(0.3488)		
BUTIR11	0.765	1.021	-0.750	0.714	0.000	2.4	5.0	
	0.188*	0.227*	0.148*	0.159*	0.000*	(0.7959)		
BUTIR12	0.826	1.158	-0.713	0.757	0.000	3.1	5.0	
	0.189*	0.240*	0.141*	0.157*	0.000*	(0.6806)		
BUTIR13	0.448	1.046	-0.428	0.723	0.000	5.6	5.0	
	0.158*	0.203*	0.134*	0.140*	0.000*	(0.3492)		
BUTIR14	0.074	0.923	-0.080	0.678	0.000	3.6	6.0	
	0.132*	0.166*	0.145*	0.122*	0.000*	(0.7261)		
BUTIR15	0.545	0.949	-0.575	0.688	0.000	4.9	6.0	
	0.152*	0.186*	0.151*	0.135*	0.000*	(0.5611)		
BUTIR16	0.292	0.868	-0.337	0.656	0.000	5.5	6.0	
	0.132*	0.160*	0.159*	0.121*	0.000*	(0.4834)		
BUTIR17	0.494	0.876	-0.564	0.659	0.000	4.9	6.0	
	0.144*	0.158*	0.158*	0.119*	0.000*	(0.5574)		

BUTIR18	0.350 0.137*	0.875 0.161*	-0.400 0.153*	0.659 0.121*	0.000 0.000*	3.8 (0.7014)	6.0
BUTIR19	-0.662 0.133*	0.584 0.132*	1.134 0.286*	0.504 0.114*	0.000 0.000*	4.3 (0.6310)	6.0
BUTIR20	0.121 0.125*	0.771 0.146*	-0.157 0.163*	0.611 0.116*	0.000 0.000*	1.0 (0.9874)	6.0
BUTIR21	0.203 0.134*	0.813 0.150*	-0.250 0.156*	0.631 0.116*	0.000 0.000*	9.4 (0.2236)	7.0
BUTIR22	0.711 0.176*	1.052 0.205*	-0.675 0.143*	0.725 0.142*	0.000 0.000*	7.7 (0.1747)	5.0
BUTIR23	0.268 0.135*	0.893 0.169*	-0.300 0.149*	0.666 0.126*	0.000 0.000*	0.8 (0.9973)	7.0
BUTIR24	0.156 0.131*	0.886 0.171*	-0.176 0.147*	0.663 0.128*	0.000 0.000*	2.5 (0.8738)	6.0
BUTIR25	0.167 0.138*	1.027 0.192*	-0.163 0.135*	0.717 0.134*	0.000 0.000*	0.8 (0.9920)	6.0
BUTIR26	0.335 0.150*	1.115 0.201*	-0.300 0.130*	0.744 0.134*	0.000 0.000*	5.5 (0.3620)	5.0



BUTIR27	0.018	0.934	-0.019	0.682	0.000	10.4	5.0
	0.134*	0.160*	0.144*	0.117*	0.000*	(0.0642)	
BUTIR28	-0.505	0.957	0.528	0.691	0.000	3.4	5.0
	0.156*	0.218*	0.146*	0.158*	0.000*	(0.6407)	
BUTIR29	0.294	0.879	-0.335	0.660	0.000	4.4	6.0
	0.135*	0.162*	0.151*	0.122*	0.000*	(0.6263)	
BUTIR30	0.427	0.842	-0.508	0.644	0.000	5.1	7.0
	0.137*	0.156*	0.164*	0.119*	0.000*	(0.6445)	

\* STANDARD ERROR

LARGEST CHANGE = 0.000781

159.4 178.0  
(0.8378)

### Output Analisis untuk model 3P

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
BUTIR01	-1.651 0.617*	0.904 0.386*	1.826 0.508*	0.671 0.286*	0.217 0.055*	12.2 (0.0325)	5.0
BUTIR02	-1.231 0.431*	1.133 0.415*	1.086 0.213*	0.750 0.275*	0.123 0.047*	9.0 (0.1091)	5.0
BUTIR03	-0.153 0.247*	0.665 0.194*	0.231 0.343*	0.554 0.161*	0.247 0.087*	10.2 (0.2513)	8.0
BUTIR04	0.010 0.210*	0.704 0.190*	-0.015 0.300*	0.576 0.155*	0.207 0.080*	16.5 (0.0056)	5.0
BUTIR05	0.444 0.215*	0.790 0.212*	-0.562 0.329*	0.620 0.166*	0.255 0.094*	10.2 (0.1185)	6.0
BUTIR06	0.479 0.212*	0.713 0.191*	-0.672 0.364*	0.581 0.156*	0.265 0.097*	6.0 (0.4184)	6.0
BUTIR07	-1.156 0.418*	0.727 0.273*	1.590 0.461*	0.588 0.221*	0.206 0.063*	19.9 (0.0057)	7.0
BUTIR08	-0.018	0.684	0.027	0.564	0.231	9.5	7.0

	0.226*	0.183*	0.328*	0.151*	0.086*	(0.2213)		
BUTIR09	0.366	0.792	-0.462	0.621	0.226	6.7	6.0	
	0.206*	0.196*	0.309*	0.154*	0.087*	(0.3487)		
BUTIR10	0.411	0.608	-0.676	0.520	0.254	10.2	5.0	
	0.201*	0.161*	0.401*	0.137*	0.095*	(0.0699)		
BUTIR11	0.289	1.616	-0.179	0.850	0.282	4.1	5.0	
	0.303*	0.552*	0.212*	0.291*	0.082*	(0.5395)		
BUTIR12	0.538	1.409	-0.382	0.816	0.208	2.3	5.0	
	0.243*	0.389*	0.204*	0.225*	0.077*	(0.8072)		
BUTIR13	0.001	1.651	0.000	0.855	0.216	5.1	5.0	
	0.296*	0.523*	0.179*	0.271*	0.071*	(0.4092)		
BUTIR14	-0.267	1.166	0.229	0.759	0.162	3.2	6.0	
	0.260*	0.292*	0.191*	0.190*	0.061*	(0.7897)		
BUTIR15	0.147	1.338	-0.110	0.801	0.233	2.2	6.0	
	0.268*	0.384*	0.216*	0.230*	0.078*	(0.8999)		
BUTIR16	-0.008	1.010	0.008	0.711	0.182	5.0	6.0	
	0.230*	0.241*	0.227*	0.169*	0.070*	(0.5393)		
BUTIR17	0.074	1.283	-0.058	0.789	0.240	6.2	6.0	
	0.283*	0.311*	0.228*	0.191*	0.077*	(0.4037)		

BUTIR18	-0.093 0.292*	1.273 0.338*	0.073 0.218*	0.786 0.209*	0.225 0.073*	3.7 (0.8141)	7.0
BUTIR19	-1.391 0.476*	0.999 0.403*	1.393 0.316*	0.707 0.285*	0.159 0.050*	4.2 (0.5242)	5.0
BUTIR20	-0.277 0.275*	1.035 0.288*	0.268 0.224*	0.719 0.200*	0.197 0.071*	5.7 (0.5717)	7.0
BUTIR21	-0.587 0.472*	1.801 0.635*	0.326 0.180*	0.874 0.308*	0.251 0.063*	3.0 (0.8867)	7.0
BUTIR22	0.387 0.255*	1.444 0.401*	-0.268 0.205*	0.822 0.228*	0.224 0.077*	5.1 (0.5299)	6.0
BUTIR23	-0.190 0.298*	1.259 0.370*	0.151 0.210*	0.783 0.230*	0.216 0.072*	2.5 (0.9270)	7.0
BUTIR24	-0.327 0.320*	1.339 0.417*	0.244 0.193*	0.801 0.249*	0.202 0.068*	7.6 (0.2671)	6.0
BUTIR25	-0.194 0.272*	1.282 0.333*	0.151 0.190*	0.788 0.205*	0.169 0.063*	5.5 (0.4853)	6.0
BUTIR26	-0.026 0.270*	1.480 0.388*	0.017 0.180*	0.829 0.217*	0.176 0.064*	3.6 (0.4559)	4.0

BUTIR27	-0.299	1.173	0.255	0.761	0.149	4.7	5.0
	0.266*	0.296*	0.188*	0.192*	0.057*	(0.4512)	
BUTIR28	-0.849	1.162	0.731	0.758	0.116	6.9	5.0
	0.327*	0.368*	0.167*	0.240*	0.047*	(0.2273)	
BUTIR29	-0.234	1.382	0.169	0.810	0.233	4.1	7.0
	0.322*	0.374*	0.206*	0.219*	0.071*	(0.7664)	
BUTIR30	0.073	1.122	-0.065	0.747	0.216	4.1	5.0
	0.256*	0.283*	0.237*	0.188*	0.077*	(0.5377)	

\* STANDARD ERROR

88 LARGEST CHANGE = 0.000794 199.1 176.0  
(0.1120)

Berdasarkan output program *BILOG MG* di atas diperoleh temuan berikut.

Tabel 13. Kecocokan Butir Tes Paket I

No	IP		2P		3P	
	$\alpha$	Kategori	$\alpha$	Kategori	A	Kategori
1	0,0654	Cocok	0,0090	Tidak cocok	0,0325	Tidak cocok
2	0,6720	Cocok	0,9423	Cocok	0,1091	Cocok
3	0,3956	Cocok	0,3608	Cocok	0,2513	Cocok
4	0,0496	Tidak cocok	0,1377	Cocok	0,0056	Tidak cocok
5	0,5755	Cocok	0,5848	Cocok	0,1185	Cocok
6	0,6091	Cocok	0,7081	Cocok	0,4184	Cocok
7	0,2244	Cocok	0,0998	Cocok	0,0057	Tidak cocok
8	0,4026	Cocok	0,7294	Cocok	0,2213	Cocok
9	0,8012	Cocok	0,7793	Cocok	0,3487	Cocok
10	0,3219	Cocok	0,3488	Cocok	0,0699	Cocok
11	0,0799	Cocok	0,7959	Cocok	0,5395	Cocok
12	0,2055	Cocok	0,6806	Cocok	0,5395	Cocok
13	0,1042	Cocok	0,3492	Cocok	0,4092	Cocok
14	0,5848	Cocok	0,7261	Cocok	0,7897	Cocok
15	0,2379	Cocok	0,5611	Cocok	0,8999	Cocok
16	0,6689	Cocok	0,4834	Cocok	0,5393	Cocok
17	0,1000	Cocok	0,5574	Cocok	0,4037	Cocok
18	0,4827	Cocok	0,7014	Cocok	0,8141	Cocok
19	0,7996	Cocok	0,6310	Cocok	0,5242	Cocok
20	0,4424	Cocok	0,9874	Cocok	0,5717	Cocok
21	0,1669	Cocok	0,2236	Cocok	0,8867	Cocok
22	0,7144	Cocok	0,1747	Cocok	0,5299	Cocok
23	0,9726	Cocok	0,9973	Cocok	0,9270	Cocok
24	0,9589	Cocok	0,8738	Cocok	0,2671	Cocok
25	0,4639	Cocok	0,9920	Cocok	0,4853	Cocok
26	0,2652	Cocok	0,3620	Cocok	0,4559	Cocok
27	0,2362	Cocok	0,0642	Cocok	0,4512	Cocok
28	0,4502	Cocok	0,6407	Cocok	0,2273	Cocok
29	0,3254	Cocok	0,6263	Cocok	0,7664	Cocok
30	0,6395	Cocok	0,6445	Cocok	0,5377	Cocok

Tabel 13 menunjukkan bahwa pada model 1 parameter ada 1 butir tidak cocok, pada model 2 parameter ada 1 butir tidak cocok, dan pada model 3 parameter ada 3 butir tidak cocok. Ternyata model paling sedikit butir tidak cocok terjadi pada model 1 parameter dan 2 parameter, yaitu masing–masing ada 1 butir tidak cocok. Dalam hal ini diputuskan untuk menggunakan model 2 parameter, dengan pertimbangan bahwa melalui model 2 parameter penulis dapat menggali parameter lebih banyak dibandingkan model 1 parameter. Pada model 2 parameter dapat dianalisis parameter tingkat kesulitan butir dan daya beda, sedangkan model 1 parameter hanya tingkat kesukaran butir yang dianalisis.

### 3. Parameter Butir Tes Paket I

#### a. Tingkat kesukaran butir

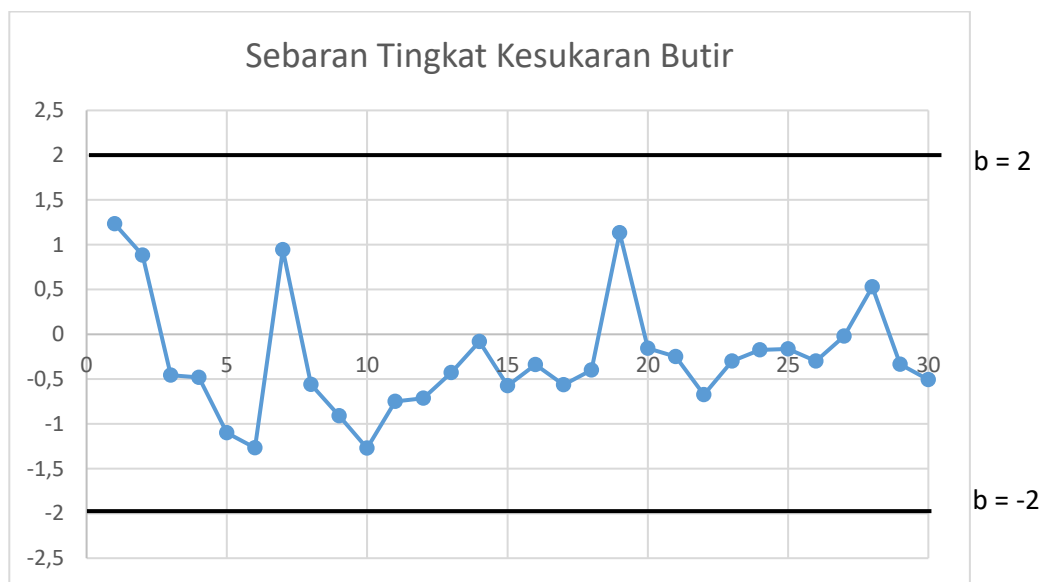
Tingkat kesukaran butir dapat ditemukan dalam output *BILOG MG* pada nilai *threshold* di *output fase 2*. Berdasarkan hasil analisis data diperoleh output tingkat kesukaran yang disajikan pada Tabel 14.

Tabel 14. *Output* Tingkat Kesukaran Butir Paket I

Butir	Tingkat Kesukaran	Butir	Tingkat Kesukaran	Butir	Tingkat Kesukaran
1	1,233	11	-0,750	21	-0,250
2	0,883	12	-0,713	22	-0,675
3	-0,456	13	-0,428	23	-0,300
4	-0,482	14	-0,080	24	-0,176
5	-1,101	15	-0,575	25	-0,163
6	-1,266	16	-0,337	26	-0,300
7	0,944	17	-0,564	27	-0,019
8	-0,559	18	-0,400	28	0,528
9	-0,909	19	1,134	29	-0,335
10	-1,272	20	-0,157	30	-0,508

Tabel 14 menunjukkan butir-butir pada tes kompetensi guru matematika Paket I memiliki tingkat kesukaran bervariasi. Butir dengan tingkat kesukaran terendah adalah butir 10 ( $b = -1,272$ ), dan butir dengan tingkat kesukaran tertinggi adalah butir 1 ( $b = 1,233$ ). Hal ini berarti kemampuan minimal pada skala untuk dapat menjawab benar dengan peluang 50% dari butir-butir di atas paling kecil  $-1,272$  dan paling tinggi  $1,233$ .

Secara grafis, sebaran tingkat kesukaran butir Paket I disajikan pada Gambar 13.



Gambar 13. Sebaran Tingkat Kesukaran Butir Tes Paket I

Gambar 13 menunjukkan bahwa tingkat kesukaran seluruh butir berada di antara  $b = -2$  dan  $b = 2$ . Hal ini berarti seluruh butir tersebut termasuk dalam kategori baik dilihat dari tingkat kesukaran butirnya.



b. Daya beda butir

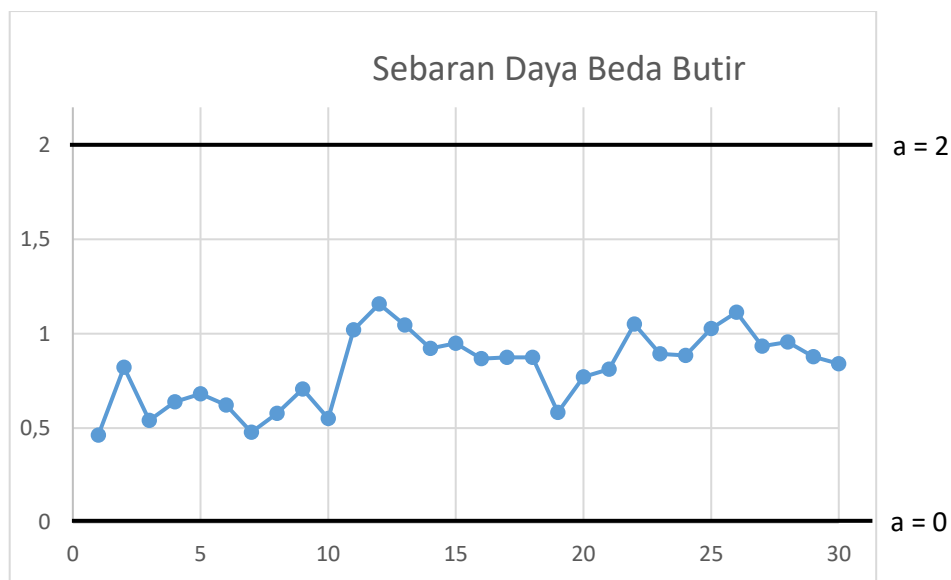
Daya beda butir dapat ditemukan dalam output *BILOG MG* pada nilai *slope* di *output fase 2*. Berdasarkan hasil analisis data diperoleh *output* daya beda butir yang disajikan pada Tabel 15.

Tabel 15. *Output* Daya Beda Butir Paket I

Butir	Slope (a)	Butir	Slope (a)	Butir	Slope (a)
1	0,464	11	1,021	21	0,813
2	0,822	12	1,158	22	1,052
3	0,541	13	1,046	23	0,893
4	0,639	14	0,923	24	0,886
5	0,682	15	0,949	25	1,027
6	0,622	16	0,868	26	1,115
7	0,479	17	0,876	27	0,934
8	0,578	18	0,875	28	0,957
9	0,707	19	0,584	29	0,879
10	0,552	20	0,771	30	0,842

Tabel 15 menunjukkan bahwa butir-butir pada tes kompetensi guru Paket I memiliki daya beda yang bervariasi. Butir dengan daya beda terendah adalah butir 10 ( $b = -1,272$ ), dan butir dengan daya beda tertinggi adalah butir 12 ( $b = 1,158$ ).

Secara grafis, sebaran daya beda butir Paket I disajikan pada Gambar 14. Gambar 14 menunjukkan daya beda seluruh butir berada di antara  $a = 0$  dan  $a = 2$ . Hal ini berarti seluruh butir tersebut termasuk dalam kategori baik dilihat dari daya beda butirnya.

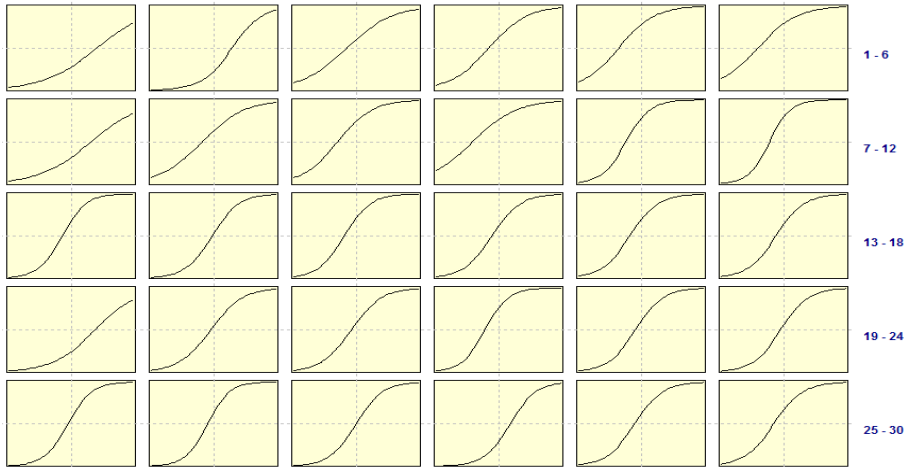


Gambar 14. Sebaran Daya Beda Butir Tes Paket I

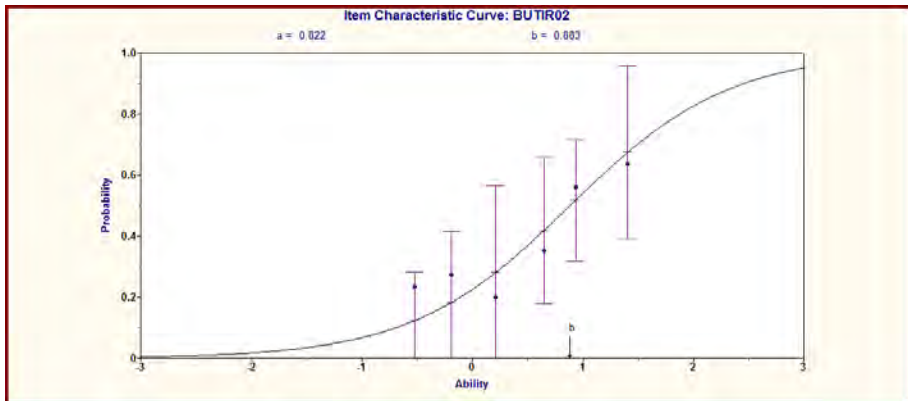
c. Kurva karakteristik

Kurva karakteristik butir soal tes kompetensi guru matematika Paket I yang dijelaskan dalam buku ini merupakan bentuk kurva karakteristik butir soal berdasarkan model logistik 2 parameter. Butir soal yang ideal adalah butir soal dengan kurva karakteristik berbentuk huruf S yang cukup landai. Kurva karakteristik diperoleh dari output *IRT Graphis*, disajikan pada Gambar 15.

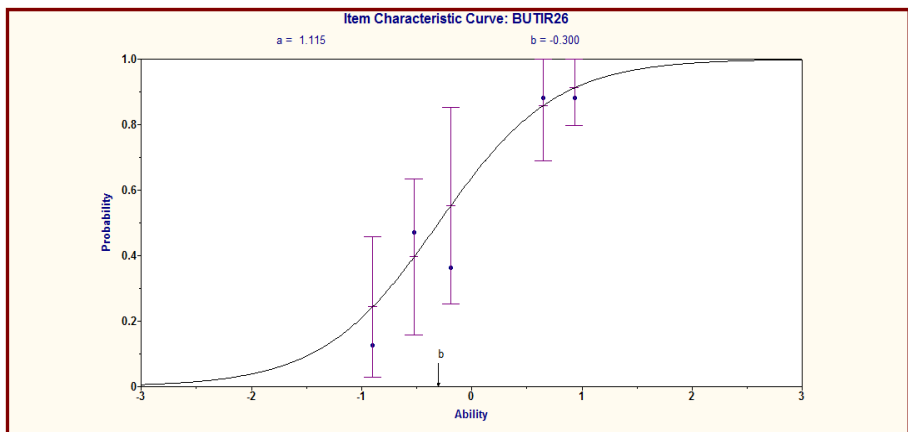
Grafik 15 menunjukkan bahwa secara umum seluruh butir menampilkan kurva karakteristik yang baik. Untuk mendapatkan gambaran lebih jauh tentang butir, berikut disajikan Gambar 16 dan Gambar 17 yang menunjukkan grafik dari 2 butir soal, yaitu butir nomor 2 dan 26.



Gambar 15. Kurva Karakteristik Butir (Butir 1 – 30)



Gambar 16. Kurva Karakteristik Butir Nomor 2



Gambar 17. Kurva Karakteristik Butir Nomor 26

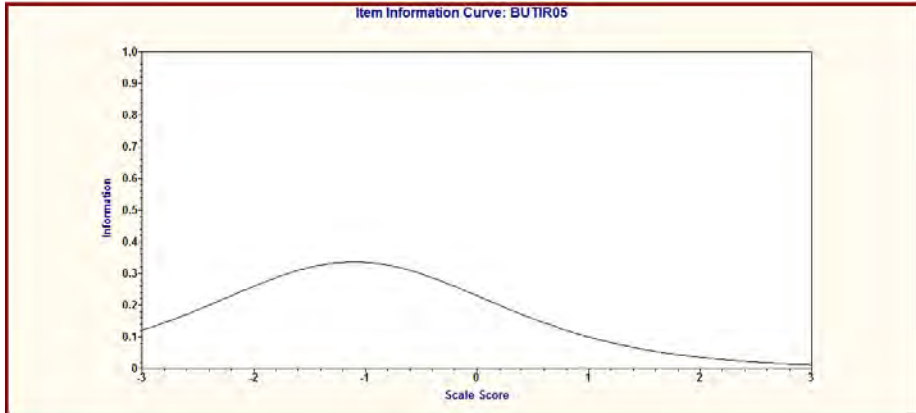
Gambar 16 menunjukkan butir nomor 2 memiliki tingkat kesukaran 0,883 ( $b = 0,883$ ) dan daya beda 0,822 ( $a = 0,822$ ). Hal ini menunjukkan bahwa diperlukan kemampuan 0,883 pada skala untuk dapat menjawab benar butir nomor 2 dengan peluang 50%. Selain itu,  $a = 0,822$  menunjukkan butir nomor 2 memiliki daya beda butir sebesar 0,822, termasuk dalam batas kategori baik.

Gambar 17 menunjukkan bahwa butir nomor 26 memiliki tingkat kesukaran  $-0,300$  ( $b = -0,300$ ) dan daya beda 1,115 ( $a = 1,115$ ). Hal ini menunjukkan bahwa diperlukan kemampuan  $-0,300$  pada skala untuk dapat menjawab benar butir nomor 26 dengan peluang 50%. Selain itu,  $a = 1,115$  menunjukkan bahwa butir nomor 26 memiliki daya beda butir sebesar 1,115, termasuk dalam batas kategori baik.

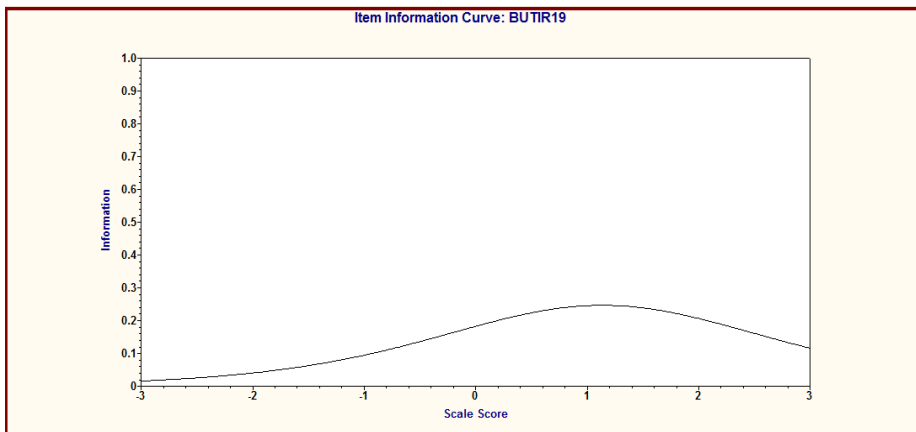
d. Fungsi informasi

Fungsi informasi butir diperoleh dari output *IRT Graphis* pada program *BILOG MG*. Fungsi informasi disajikan dalam koordinat kartesius, dimana sumbu X menunjukkan level abilitas peserta tes, dan sumbu Y menunjukkan besarnya fungsi informasi. Sebagai gambaran disajikan fungsi informasi dari 2 butir, yaitu butir nomor 5 dan butir nomor 19. Kedua fungsi informasi disajikan pada Gambar 18 dan Gambar 19.

Gambar 18 menunjukkan bahwa butir nomor 2 akan memberikan banyak informasi apabila diberikan pada peserta didik dengan level abilitas rendah. Gambar 19 menunjukkan bahwa butir nomor 19 akan memberikan banyak informasi apabila diberikan pada peserta didik dengan level abilitas tinggi.



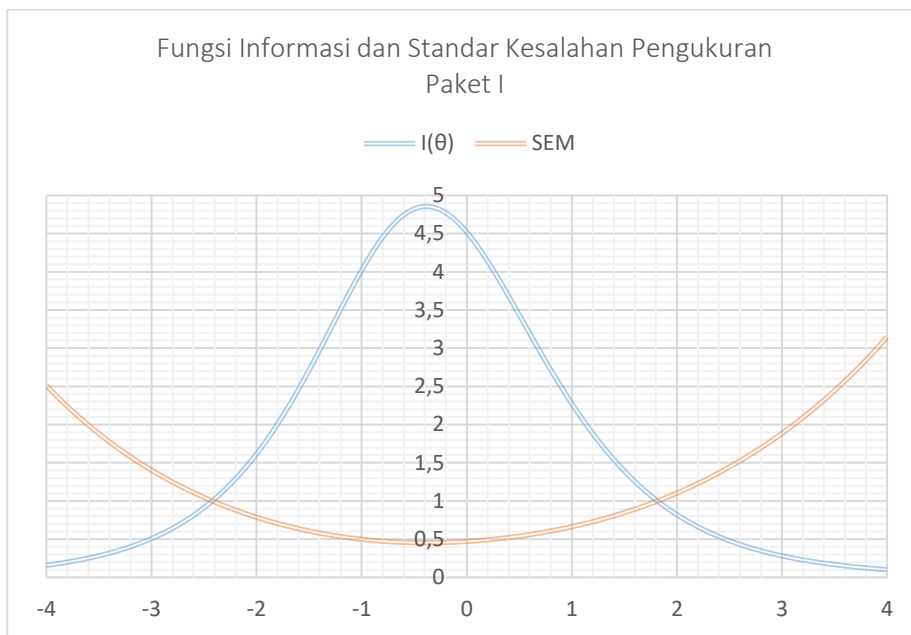
Gambar 18. Fungsi Informasi Butir Nomor 5



Gambar 19. Fungsi Informasi Butir Nomor 19

Nilai parameter butir yang diperoleh dalam analisis ini merupakan hasil estimasi sehingga kebenarannya bersifat probabilitas dan tidak terlepas dengan kesalahan pengukuran (*Standard Error Measurement/SEM*). Fungsi informasi dengan SEM mempunyai hubungan berbanding terbalik, yaitu semakin besar fungsi informasi maka SEM semakin kecil dan sebaliknya semakin kecil fungsi informasi maka akan semakin besar nilai SEM. Grafik fungsi informasi dan SEM disajikan pada Gambar 20. Garis biru pada Gambar 20 merupakan grafik dari fungsi informasi, dan garis merah merupakan

grafik dari SEM. Semakin besar nilai fungsi informasi maka semakin kecil nilai SEM, dan sebaliknya.



Gambar 20. Kurva Fungsi Informasi dan SEM Butir Soal Paket I

#### 4. Parameter Kemampuan

Parameter kemampuan dapat ditentukan dengan memanfaatkan perpotongan kurva fungsi informasi dan SEM, seperti tersaji pada Gambar 20. Gambar 20 menunjukkan bahwa nilai fungsi informasi tes maksimal tercapai pada kemampuan siswa sebesar  $-0,4$ . Gambar 20 juga menunjukkan bahwa fungsi informasi dan SEM berpotongan sebelah kiri pada  $-2,4$  dan sebelah kanan  $1,8$ . Hal ini berarti bahwa tes akan memberikan informasi paling optimal apabila diberikan pada peserta tes dengan level abilitas dari  $-2,4$  sampai  $1,8$ . Apabila tes ini diberikan pada peserta dengan abilitas di bawah  $-2,4$  atau di atas  $1,8$  maka kesalahan pengukuran lebih tinggi daripada fungsi informasi yang bisa diperoleh dari tes tersebut.

## C. Terapan Teori Respon Butir untuk Analisis Tes Kompetensi Guru Matematika Paket II

### 1. Uji Asumsi

#### a. Asumsi unidimensi

Asumsi unidimensi tes kompetensi guru matematika Paket II diuji dengan menggunakan analisis faktor, yaitu dengan melihat nilai eigen pada matriks varians kovarians inter – butir. Analisis faktor untuk menguji asumsi unidimensi tes kompetensi guru matematika Paket II ini dilakukan dengan bantuan program pengolah data SPSS.

Uji asumsi dengan analisis faktor ini diawali dengan melakukan analisis kecukupan sampel melalui uji *KMO* dan *Bartlett*. Hasil analisis disajikan pada Tabel 16.

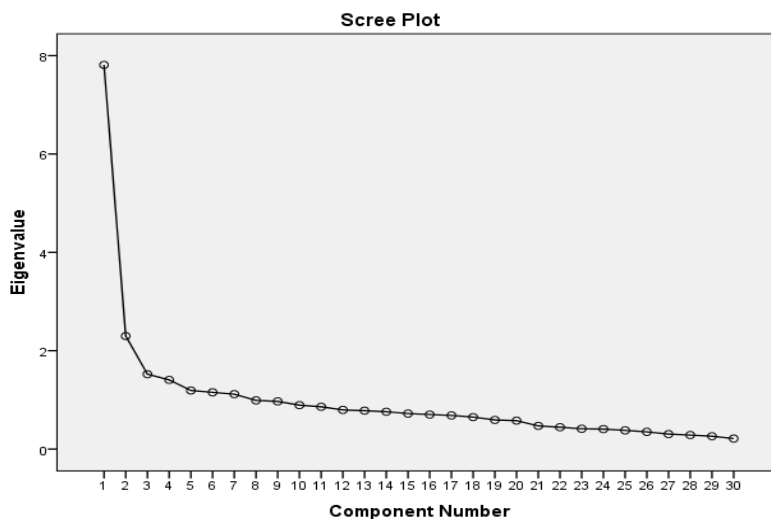
Tabel 16. Hasil Uji *KMO* dan *Bartlett* Paket II

<b><i>KMO and Bartlett's Test</i></b>		
<i>Kaiser–Meyer–Olkin Measure of Sampling Adequacy.</i>	0,842	
<i>Bartlett's Test of Sphericity</i>	<i>Approx. Chi–Square</i>	1436,496
	<i>df</i>	435
	<i>Sig.</i>	0,000

Berdasarkan output uji *KMO* dan *Bartlett* pada Tabel 16 diketahui bahwa nilai *KMO* 0,823 dan nilai *Khi*–kuadrat pada uji *Bartlett* sebesar 1436,496 dengan derajat kebebasan 435 dan nilai sig. 0,00 (kurang dari 0,01). Hasil ini menunjukkan bahwa sampel sebesar 152 yang digunakan dalam proses ini cukup.

Setelah diperoleh hasil uji kecukupan sampel, langkah berikutnya adalah melakukan analisis faktor untuk memperoleh output *scree–plot*

dan nilai eigen. Berdasarkan analisis yang dilakukan dihasilkan *scree-plot* yang disajikan pada Gambar 21.



Gambar 21. *Output Scree-Plot* Tes Paket II

Gambar 21 menunjukkan bahwa terdapat terdapat satu curaman dominan, kemudian agak melandai mulai nilai eigen kedua, dan berikutnya semakin melandai mulai nilai eigen ketiga. Hal ini menunjukkan bahwa tes kompetensi guru matematika Paket II mengukur 1 faktor dominan.

Untuk menguatkan analisis terhadap *scree-plot* di atas, penulis melakukan analisis terhadap nilai eigen. *Output* nilai eigen disajikan pada Tabel 17. Tabel 17 menunjukkan bahwa nilai eigen komponen 1 cukup besar dengan persentase 26,037% terhadap total. Menurut Wells & Purwono (2008; Retnawati, 2015: 144), persentase varians lebih besar dari 20% dikatakan bahwa perangkat mengukur dimensi utama. Dengan demikian dapat dikatakan bahwa asumsi unidimensi terpenuhi. Hal ini selaras dengan hasil analisis *scree-plot* di atas.



Tabel 17. *Output* Nilai Eigen Paket II

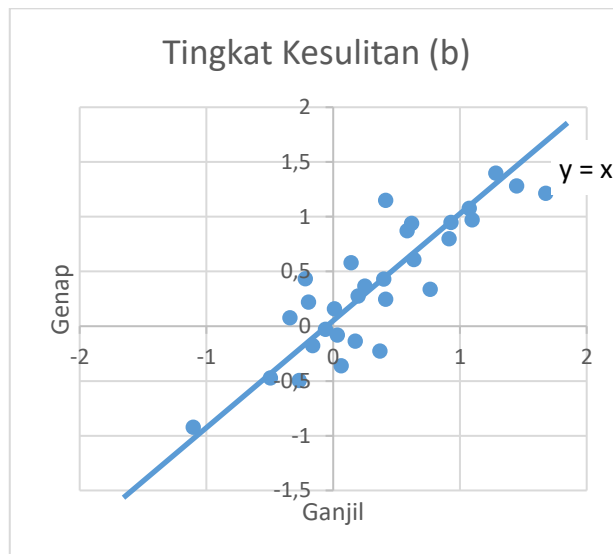
Komponen	Nilai Eigen	% varians	Kumulatif % Varians
1	7,811	26,037	26,037
2	2,298	7,659	33,696
3	1,522	5,075	38,771
4	1,407	4,689	43,461
5	1,191	3,971	47,432
6	1,154	3,845	51,277
7	1,116	3,720	54,997
8	0,990	3,300	58,297
9	0,969	3,230	61,528
10	0,893	2,975	64,503
Dst sampai 30			

b. Asumsi independensi lokal

Asumsi independensi lokal tes kompetensi guru matematika Paket II tidak diuji secara tersendiri tetapi didasarkan pada hasil uji unidimensi yang telah dilakukan. Hal ini didasarkan pendapat Hambleton, Swaminathan, dan Rogers (1991: 11) yang menyatakan bahwa asumsi independensi lokal butir terpenuhi bila bersifat unidimensi. Hal senada dinyatakan Retnawati (2015: 7) bahwa asumsi independensi lokal otomatis terbukti setelah dibuktikan dengan unidimensionalitas data respon terhadap tes. Oleh karena itu, dengan telah terbuktinya asumsi unidimensi tes kompetensi guru matematika Paket II di atas, disimpulkan bahwa asumsi independensi lokal juga terpenuhi.

c. Asumsi invariansi parameter

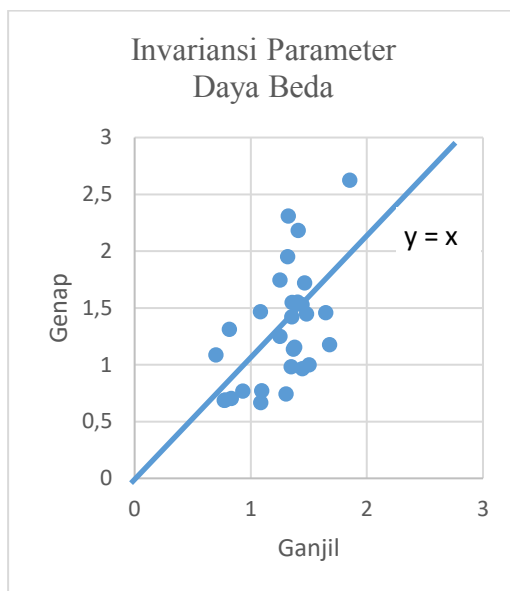
Ada dua parameter yang diuji asumsinya pada tes Paket II, parameter butir dan parameter kemampuan. Parameter butir meliputi tingkat kesulitan, daya beda, dan *pseudo guessing*. Uji asumsi invariansi parameter tingkat kesukaran butir dibuktikan dengan terlebih dahulu melakukan estimasi parameter tingkat kesukaran butir pada peserta kelompok ganjil dan genap, yang hasilnya kemudian disajikan dalam diagram pencar (*scatter plot*) pada Gambar 22.



Gambar 22. Invariansi Paramter Tingkat Kesukaran Butir Tes Paket II

Gambar 22 menunjukkan bahwa masing–masing titik berada relatif dekat dengan garis  $y = x$ . Hal ini menunjukkan bahwa tidak terjadi variasi parameter hasil estimasi pada peserta kelompok genap dan ganjil. Dengan demikian dapat dikatakan bahwa invariansi tingkat kesukaran butir terpenuhi.

Uji asumsi invariansi parameter daya beda butir juga dibuktikan dengan terlebih dahulu melakukan estimasi parameter daya beda butir pada peserta kelompok ganjil dan genap, yang hasilnya kemudian disajikan dalam sebuah gambar diagram pencar (*scatter plot*). Diagram pencar hasil estimasi parameter daya beda butir pada peserta kelompok ganjil dan genap disajikan pada Gambar 23.

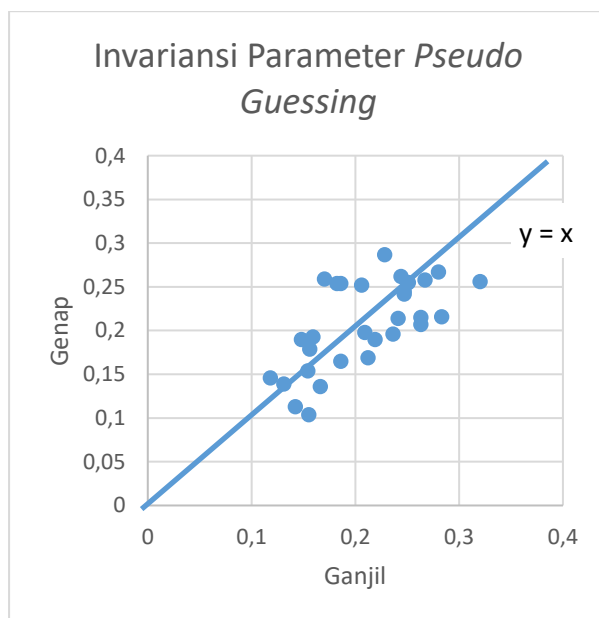


Gambar 23. Invariansi Paramter Daya Bada Butir Paket II

Gambar 23 menunjukkan bahwa masing–masing titik berada relatif dekat dengan garis  $y = x$ . Hal ini menunjukkan bahwa tidak terjadi variasi parameter hasil estimasi pada peserta kelompok genap dan ganjil. Dengan demikian dapat dikatakan bahwa invariansi daya beda butir terpenuhi.

Uji asumsi invariansi parameter butir berikutnya adalah *pseudo-guesing*. Asumsi ini juga dibuktikan dengan terlebih dahulu melakukan estimasi parameter *pseudo-guesing* pada peserta

kelompok ganjil dan genap, yang hasilnya kemudian disajikan dalam sebuah gambar diagram pencar (*scatter plot*). Diagram pencar hasil estimasi parameter *pseudo-guessing* pada peserta kelompok ganjil dan genap disajikan pada Gambar 24.

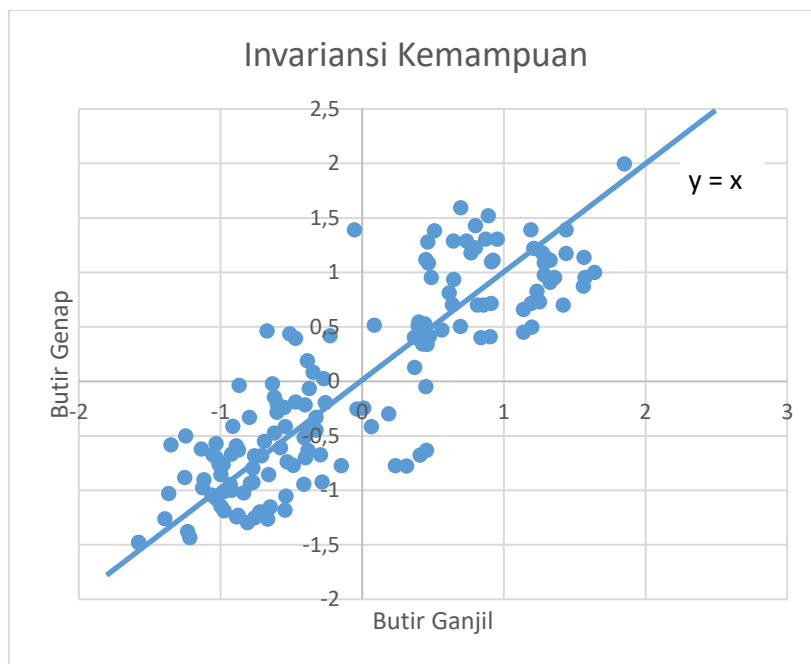


Gambar 24. Invariansi Paramter *Pseudo-Guesing* Tes Paket II

Gambar 24 menunjukkan bahwa masing–masing titik berada relatif dekat dengan garis  $y = x$ . Hal ini menunjukkan bahwa tidak terjadi variasi parameter hasil estimasi pada peserta kelompok genap dan ganjil. Dengan demikian dapat dikatakan bahwa invariansi *pseudo-guessing* butir terpenuhi.

Uji asumsi parameter berikutnya dilakukan terhadap parameter kemampuan. Uji asumsi invariansi parameter kemampuan dibuktikan dengan terlebih dahulu melakukan estimasi parameter kemampuan pada butir ganjil dan genap, yang hasilnya kemudian disajikan dalam

sebuah gambar diagram pencar (*scatter plot*). Diagram pencar hasil estimasi parameter kemampuan pada butir ganjil dan genap disajikan pada Gambar 25.



Gambar 25. Invariansi Parameter Kemampuan Mengerjakan Butir Ganjil dan Genap

Gambar 25 menunjukkan bahwa masing–masing titik berada relatif dekat dengan garis  $y = x$ . Hal ini menunjukkan bahwa tidak terjadi variasi parameter hasil estimasi pada butir genap dan ganjil. Dengan demikian dapat dikatakan bahwa invariansi parameter kemampuan terpenuhi.

Berdasarkan uji asumsi di atas dapat disimpulkan bahwa asumsi–asumsi untuk diterapkannya teori respons butir, yaitu asumsi parameter butir dan asumsi parameter kemampuan pada tes kompetensi guru matematika Paket II, dapat terpenuhi.

## 2. Uji Kecocokan Model

Ada 3 model yang akan diuji kecocokan modelnya pada tes kompetensi guru matematika SMA Paket II, yaitu model 1P, 2P, dan 3P. Pada proses ini, uji kecocokan model dilakukan melalui kecocokan secara statistik yang didasarkan pada nilai probabilitas (signifikansi, *sig*). Butir dikatakan tidak cocok dengan model apabila nilai  $sig < \alpha$  ( $\alpha = 0,05$ ). Model logistik yang selanjutnya digunakan untuk analisis data adalah model yang memiliki butir cocok paling banyak.

Uji kecocokan model pada contoh ini dianalisis dengan menggunakan bantuan program *BILOG MG*. Analisis dilakukan untuk model 1P, 2P, dan 3P, yang hasilnya kemudian dijadikan dasar untuk menentukan model mana yang paling cocok. Berikut disajikan output analisis dengan program *BILOG MG output fase 2* dari tes Pakat 1, baik untuk model 1P, 2P, maupun 3P.

## Output Analisis untuk model 1P

ITEM	INTERCEPT	SLOPE	THRESHOLD	LOADING	ASYMPTOTE	CHISQ	DF
	S.E.	S.E.	S.E.	S.E.	S.E.	(PROB)	
BUTIR01	1.073 0.128*	0.774 0.026*	-1.388 0.165*	0.612 0.021*	0.000 0.000*	7.8 (0.1689)	5.0
BUTIR02	-0.382 0.113*	0.774 0.026*	0.493 0.146*	0.612 0.021*	0.000 0.000*	12.7 (0.0809)	7.0
BUTIR03	0.000 0.105*	0.774 0.026*	0.000 0.136*	0.612 0.021*	0.000 0.000*	12.5 (0.1316)	8.0
BUTIR04	0.484 0.109*	0.774 0.026*	-0.626 0.140*	0.612 0.021*	0.000 0.000*	9.5 (0.1454)	6.0
BUTIR05	0.484 0.107*	0.774 0.026*	-0.626 0.138*	0.612 0.021*	0.000 0.000*	10.6 (0.1022)	6.0
BUTIR06	-0.125 0.107*	0.774 0.026*	0.161 0.138*	0.612 0.021*	0.000 0.000*	22.6 (0.0020)	7.0
BUTIR07	0.145 0.109*	0.774 0.026*	-0.188 0.141*	0.612 0.021*	0.000 0.000*	22.1 (0.0047)	8.0
BUTIR08	-0.805	0.774	1.041	0.612	0.000	7.0	6.0

	0.119*	0.026*	0.154*	0.021*	0.000*	(0.3239)	
BUTIR09	0.734	0.774	-0.949	0.612	0.000	11.9	5.0
	0.113*	0.026*	0.147*	0.021*	0.000*	(0.0356)	
BUTIR10	-0.104	0.774	0.134	0.612	0.000	18.4	8.0
	0.105*	0.026*	0.136*	0.021*	0.000*	(0.0184)	
BUTIR11	0.506	0.774	-0.654	0.612	0.000	1.2	6.0
	0.115*	0.026*	0.149*	0.021*	0.000*	(0.9744)	
BUTIR12	0.550	0.774	-0.711	0.612	0.000	7.4	5.0
	0.121*	0.026*	0.156*	0.021*	0.000*	(0.1952)	
BUTIR13	0.528	0.774	-0.683	0.612	0.000	4.5	5.0
	0.120*	0.026*	0.155*	0.021*	0.000*	(0.4778)	
BUTIR14	0.376	0.774	-0.486	0.612	0.000	11.5	6.0
	0.116*	0.026*	0.150*	0.021*	0.000*	(0.0743)	
BUTIR15	-0.381	0.774	0.493	0.612	0.000	9.9	7.0
	0.126*	0.026*	0.162*	0.021*	0.000*	(0.1931)	
BUTIR16	0.832	0.774	-1.075	0.612	0.000	2.5	4.0
	0.126*	0.026*	0.163*	0.021*	0.000*	(0.6511)	
BUTIR17	0.104	0.774	-0.134	0.612	0.000	15.1	6.0
	0.117*	0.026*	0.151*	0.021*	0.000*	(0.0192)	



BUTIR18	0.021 0.118*	0.774 0.026*	-0.027 0.152*	0.612 0.021*	0.000 0.000*	23.3 (0.0007)	6.0
BUTIR19	-0.883 0.124*	0.774 0.026*	1.142 0.160*	0.612 0.021*	0.000 0.000*	4.6 (0.4608)	5.0
BUTIR20	-0.754 0.117*	0.774 0.026*	0.975 0.152*	0.612 0.021*	0.000 0.000*	5.2 (0.6384)	7.0
BUTIR21	0.376 0.114*	0.774 0.026*	-0.486 0.147*	0.612 0.021*	0.000 0.000*	7.8 (0.2546)	6.0
BUTIR22	0.355 0.122*	0.774 0.026*	-0.459 0.158*	0.612 0.021*	0.000 0.000*	7.9 (0.2459)	6.0
BUTIR23	-0.562 0.124*	0.774 0.026*	0.726 0.160*	0.612 0.021*	0.000 0.000*	6.4 (0.3766)	6.0
BUTIR24	0.334 0.117*	0.774 0.026*	-0.431 0.151*	0.612 0.021*	0.000 0.000*	6.5 (0.2597)	5.0
BUTIR25	0.355 0.118*	0.774 0.026*	-0.459 0.153*	0.612 0.021*	0.000 0.000*	2.4 (0.7857)	5.0
BUTIR26	-0.104 0.123*	0.774 0.026*	0.134 0.159*	0.612 0.021*	0.000 0.000*	8.4 (0.3984)	8.0

BUTIR27	-0.779	0.774	1.008	0.612	0.000	10.1	7.0
	0.129*	0.026*	0.167*	0.021*	0.000*	(0.1847)	
BUTIR28	-0.729	0.774	0.943	0.612	0.000	3.2	6.0
	0.128*	0.026*	0.165*	0.021*	0.000*	(0.7896)	
BUTIR29	-0.295	0.774	0.381	0.612	0.000	5.7	8.0
	0.117*	0.026*	0.151*	0.021*	0.000*	(0.6774)	
BUTIR30	0.145	0.774	-0.188	0.612	0.000	5.2	7.0
	0.116*	0.026*	0.150*	0.021*	0.000*	(0.6318)	

\* STANDARD ERROR

LARGEST CHANGE = 0.000712

283.9 187.0  
(0.0000)

## Output Analisis untuk model 2P

108

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
BUTIR01	0.973 0.133*	0.587 0.127*	-1.657 0.360*	0.506 0.110*	0.000 0.000*	3.8 (0.5726)	5.0
BUTIR02	-0.357 0.113*	0.668 0.127*	0.535 0.178*	0.556 0.106*	0.000 0.000*	14.7 (0.0228)	6.0
BUTIR03	-0.003 0.103*	0.563 0.109*	0.005 0.183*	0.491 0.095*	0.000 0.000*	15.2 (0.0335)	7.0
BUTIR04	0.434 0.113*	0.591 0.129*	-0.734 0.207*	0.509 0.111*	0.000 0.000*	15.7 (0.0077)	5.0
BUTIR05	0.414 0.107*	0.510 0.105*	-0.811 0.246*	0.454 0.094*	0.000 0.000*	2.8 (0.7370)	5.0
BUTIR06	-0.114 0.106*	0.601 0.117*	0.190 0.175*	0.515 0.100*	0.000 0.000*	23.3 (0.0007)	6.0
BUTIR07	0.135 0.107*	0.651 0.117*	-0.207 0.169*	0.546 0.098*	0.000 0.000*	7.6 (0.1803)	5.0
BUTIR08	-0.740	0.626	1.181	0.531	0.000	3.2	5.0

	0.129*	0.132*	0.249*	0.112*	0.000*	(0.6686)	
BUTIR09	0.656	0.575	-1.140	0.499	0.000	11.4	5.0
	0.121*	0.136*	0.271*	0.118*	0.000*	(0.0439)	
BUTIR10	-0.094	0.550	0.171	0.482	0.000	9.7	6.0
	0.104*	0.108*	0.188*	0.095*	0.000*	(0.1366)	
BUTIR11	0.505	0.760	-0.665	0.605	0.000	2.0	5.0
	0.120*	0.138*	0.166*	0.110*	0.000*	(0.8555)	
BUTIR12	0.667	1.058	-0.631	0.727	0.000	4.1	3.0
	0.186*	0.263*	0.119*	0.181*	0.000*	(0.2549)	
BUTIR13	0.639	1.049	-0.609	0.724	0.000	0.9	3.0
	0.186*	0.238*	0.118*	0.164*	0.000*	(0.8272)	
BUTIR14	0.442	0.978	-0.452	0.699	0.000	7.0	4.0
	0.148*	0.189*	0.123*	0.135*	0.000*	(0.1333)	
BUTIR15	-0.450	1.119	0.402	0.745	0.000	1.7	4.0
	0.144*	0.171*	0.120*	0.114*	0.000*	(0.7861)	
BUTIR16	0.993	1.075	-0.924	0.732	0.000	0.3	3.0
	0.236*	0.303*	0.143*	0.206*	0.000*	(0.9604)	
BUTIR17	0.140	0.997	-0.141	0.706	0.000	4.0	5.0
	0.139*	0.172*	0.130*	0.122*	0.000*	(0.5506)	

BUTIR18	0.038 0.121*	0.923 0.131*	-0.041 0.131*	0.678 0.096*	0.000 0.000*	9.3 (0.0974)	5.0
BUTIR19	-0.827 0.134*	0.658 0.129*	1.257 0.247*	0.550 0.108*	0.000 0.000*	0.8 (0.9921)	6.0
BUTIR20	-0.680 0.121*	0.591 0.113*	1.150 0.255*	0.509 0.098*	0.000 0.000*	8.3 (0.2137)	6.0
BUTIR21	0.422 0.142*	0.916 0.186*	-0.461 0.128*	0.675 0.137*	0.000 0.000*	6.0 (0.1959)	4.0
BUTIR22	0.466 0.141*	1.124 0.192*	-0.415 0.114*	0.747 0.128*	0.000 0.000*	3.5 (0.4833)	4.0
BUTIR23	-0.619 0.140*	0.956 0.153*	0.647 0.142*	0.691 0.110*	0.000 0.000*	1.6 (0.9072)	5.0
BUTIR24	0.402 0.145*	1.003 0.187*	-0.401 0.121*	0.708 0.132*	0.000 0.000*	2.8 (0.6000)	4.0
BUTIR25	0.448 0.160*	1.070 0.222*	-0.419 0.115*	0.731 0.151*	0.000 0.000*	0.7 (0.9476)	4.0
BUTIR26	-0.099 0.126*	1.037 0.158*	0.096 0.122*	0.720 0.110*	0.000 0.000*	3.8 (0.5745)	5.0

BUTIR27	-0.855	0.937	0.913	0.684	0.000	1.7	5.0
	0.154*	0.170*	0.163*	0.124*	0.000*	(0.8841)	
BUTIR28	-0.785	0.907	0.866	0.672	0.000	3.4	6.0
	0.144*	0.148*	0.161*	0.110*	0.000*	(0.7582)	
BUTIR29	-0.295	0.816	0.362	0.632	0.000	2.4	6.0
	0.118*	0.137*	0.147*	0.106*	0.000*	(0.8807)	
BUTIR30	0.163	0.856	-0.191	0.650	0.000	6.1	5.0
	0.118*	0.134*	0.136*	0.102*	0.000*	(0.2957)	

\* STANDARD ERROR

LARGEST CHANGE = 0.000565 177.8 147.0  
(0.0424)

### Output Analisis untuk model 3P

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
BUTIR01	0.732 0.186*	0.618 0.137*	-1.184 0.397*	0.526 0.117*	0.248 0.095*	7.6 (0.1068)	4.0
BUTIR02	-0.900 0.360*	1.000 0.312*	0.900 0.206*	0.707 0.220*	0.164 0.057*	11.9 (0.0644)	6.0
BUTIR03	-0.409 0.260*	0.753 0.201*	0.544 0.274*	0.601 0.161*	0.203 0.074*	13.6 (0.0595)	7.0
BUTIR04	-0.005 0.251*	0.778 0.214*	0.007 0.321*	0.614 0.169*	0.280 0.093*	5.4 (0.3727)	5.0
BUTIR05	0.050 0.223*	0.631 0.156*	-0.079 0.362*	0.534 0.132*	0.259 0.093*	6.8 (0.4506)	7.0
BUTIR06	-0.508 0.261*	0.773 0.209*	0.657 0.253*	0.611 0.165*	0.181 0.067*	18.1 (0.0060)	6.0
BUTIR07	-0.185 0.224*	0.802 0.187*	0.231 0.250*	0.626 0.146*	0.186 0.071*	10.8 (0.0565)	5.0
BUTIR08	-1.330	0.956	1.391	0.691	0.129	2.9	4.0

	0.426*	0.333*	0.257*	0.240*	0.046*	(0.5687)	
BUTIR09	0.294	0.657	-0.448	0.549	0.281	4.5	5.0
	0.213*	0.171*	0.366*	0.143*	0.100*	(0.4857)	
BUTIR10	-0.477	0.725	0.658	0.587	0.185	12.9	6.0
	0.255*	0.196*	0.269*	0.158*	0.069*	(0.0444)	
BUTIR11	0.184	0.909	-0.203	0.673	0.222	5.3	5.0
	0.218*	0.199*	0.259*	0.148*	0.084*	(0.3836)	
BUTIR12	-0.058	1.683	0.035	0.860	0.313	0.5	5.0
	0.348*	0.750*	0.201*	0.383*	0.081*	(0.9905)	
BUTIR13	0.115	1.827	-0.063	0.877	0.274	6.7	4.0
	0.319*	0.685*	0.181*	0.329*	0.078*	(0.1507)	
BUTIR14	-0.407	2.031	0.200	0.897	0.293	8.6	6.0
	0.466*	0.593*	0.191*	0.262*	0.066*	(0.1956)	
BUTIR15	-0.897	1.504	0.597	0.833	0.103	3.4	5.0
	0.369*	0.368*	0.139*	0.204*	0.039*	(0.6371)	
BUTIR16	0.492	1.567	-0.314	0.843	0.329	2.8	5.0
	0.313*	0.697*	0.217*	0.375*	0.094*	(0.7346)	
BUTIR17	-1.093	2.640	0.414	0.935	0.240	6.1	6.0
	0.728*	0.963*	0.155*	0.341*	0.050*	(0.4171)	



BUTIR18	-0.340 0.274*	1.185 0.241*	0.287 0.193*	0.764 0.156*	0.158 0.057*	10.3 (0.1142)	6.0
BUTIR19	-1.637 0.583*	1.187 0.470*	1.379 0.211*	0.765 0.303*	0.128 0.039*	2.7 (0.7409)	5.0
BUTIR20	-1.631 0.622*	1.239 0.500*	1.316 0.197*	0.778 0.314*	0.154 0.042*	1.3 (0.9697)	6.0
BUTIR21	-0.678 0.601*	2.197 0.830*	0.309 0.188*	0.910 0.344*	0.325 0.063*	1.7 (0.9439)	6.0
BUTIR22	0.092 0.244*	1.371 0.290*	-0.067 0.183*	0.808 0.171*	0.187 0.069*	5.2 (0.3900)	5.0
BUTIR23	-1.217 0.475*	1.476 0.459*	0.824 0.133*	0.828 0.257*	0.109 0.038*	3.6 (0.6073)	5.0
BUTIR24	-0.230 0.358*	1.768 0.461*	0.130 0.183*	0.870 0.227*	0.250 0.068*	1.3 (0.9300)	5.0
BUTIR25	-0.208 0.362*	1.815 0.640*	0.114 0.176*	0.876 0.309*	0.255 0.070*	0.8 (0.9744)	5.0
BUTIR26	-0.549 0.309*	1.411 0.331*	0.389 0.160*	0.816 0.192*	0.142 0.051*	6.0 (0.3077)	5.0

BUTIR27	-1.756	1.699	1.033	0.862	0.107	10.9	5.0
	0.716*	0.686*	0.126*	0.348*	0.034*	(0.0542)	
BUTIR28	-1.545	1.578	0.979	0.845	0.103	3.9	5.0
	0.596*	0.573*	0.125*	0.306*	0.034*	(0.5585)	
BUTIR29	-0.908	1.264	0.718	0.784	0.164	7.3	7.0
	0.405*	0.391*	0.167*	0.242*	0.052*	(0.3966)	
BUTIR30	-0.264	1.221	0.216	0.774	0.190	4.3	5.0
	0.278*	0.285*	0.197*	0.181*	0.065*	(0.5102)	

\* STANDARD ERROR

LARGEST CHANGE = 0.207516 187.2 161.0  
(0.0771)

Berdasarkan output program *BILOG MG* di atas diperoleh temuan berikut

Tabel 18. Kecocokan Butir Tes Kompetensi Guru Paket II

No	IP		2P		3P	
	$\alpha$	Kategori	$\alpha$	Kategori	$\alpha$	Kategori
1	0,1689	Cocok	0,5726	Cocok	0,1068	Cocok
2	0,0809	Cocok	0,0228	Tidak Cocok	0,0644	Cocok
3	0,1316	Cocok	0,0335	Tidak Cocok	0,0595	Cocok
4	0,1454	Cocok	0,0077	Tidak Cocok	0,3727	Cocok
5	0,1022	Cocok	0,7370	Cocok	0,4506	Cocok
6	0,0020	Tidak Cocok	0,0007	Tidak Cocok	0,0060	Tidak Cocok
7	0,0047	Tidak Cocok	0,1803	Cocok	0,0565	Cocok
8	0,3239	Cocok	0,6686	Cocok	0,5687	Cocok
9	0,0356	Tidak Cocok	0,0439	Tidak Cocok	0,4857	Cocok
10	0,0184	Tidak Cocok	0,1366	Cocok	0,0444	Tidak Cocok
11	0,9744	Cocok	0,8555	Cocok	0,3836	Cocok
12	0,1952	Cocok	0,2549	Cocok	0,9905	Cocok
13	0,4778	Cocok	0,8272	Cocok	0,1507	Cocok
14	0,0743	Cocok	0,1333	Cocok	0,1956	Cocok
15	0,1931	Cocok	0,7861	Cocok	0,6371	Cocok
16	0,6511	Cocok	0,9604	Cocok	0,7346	Cocok
17	0,0192	Tidak Cocok	0,5506	Cocok	0,4171	Cocok
18	0,0007	Tidak Cocok	0,0974	Cocok	0,1142	Cocok
19	0,4608	Cocok	0,9921	Cocok	0,7409	Cocok
20	0,6384	Cocok	0,2137	Cocok	0,9697	Cocok
21	0,2546	Cocok	0,1959	Cocok	0,9439	Cocok
22	0,2459	Cocok	0,4833	Cocok	0,3900	Cocok
23	0,3766	Cocok	0,9072	Cocok	0,6073	Cocok
24	0,2597	Cocok	0,6000	Cocok	0,9300	Cocok
25	0,7857	Cocok	0,9476	Cocok	0,9744	Cocok
26	0,3984	Cocok	0,5745	Cocok	0,3077	Cocok
27	0,1847	Cocok	0,8841	Cocok	0,0542	Cocok
28	0,7896	Cocok	0,7582	Cocok	0,5585	Cocok
29	0,6774	Cocok	0,8807	Cocok	0,3966	Cocok
30	0,6318	Cocok	0,2957	Cocok	0,5102	Cocok

Tabel 18 menunjukkan bahwa pada model 1 parameter ada 6 butir tidak cocok, pada model 2 parameter ada 5 butir tidak cocok, dan pada model 3 parameter ada 2 butir tidak cocok. Ternyata model paling sedikit butir tidak cocok terjadi pada model 3 parameter. Dengan demikian disimpulkan bahwa model paling cocok adalah model 3 parameter.

### 3. Parameter Butir Tes

#### a. Tingkat kesukaran butir

Tingkat kesukaran butir dapat ditemukan dalam *output BILOG MG* pada nilai *threshold* di *output fase 2*. Berdasarkan hasil analisis data diperoleh output tingkat kesukaran yang disajikan pada Tabel 19.

Tabel 19. *Output* Tingkat Kesukaran Butir Paket II

Butir	Tingkat Kesukaran	Butir	Tingkat Kesukaran	Butir	Tingkat Kesukaran
1	-1,184	11	-0,203	21	0,309
2	0,900	12	0,035	22	-0,067
3	0,544	13	-0,063	23	0,824
4	0,007	14	0,200	24	0,130
5	-0,079	15	0,597	25	0,114
6	0,657	16	-0,314	26	0,389
7	0,231	17	0,414	27	1,033
8	1,391	18	0,287	28	0,979
9	-0,448	19	1,379	29	0,718
10	0,658	20	1,316	30	0,216

Tabel 19 menunjukkan bahwa butir-butir pada tes kompetensi guru Paket I memiliki tingkat kesukaran yang bervariasi. Butir dengan tingkat kesukaran terendah adalah butir 1 ( $b = -1,184$ ), dan butir

dengan tingkat kesukaran tertinggi adalah butir 8 ( $b = 1,391$ ). Hal ini berarti kemampuan minimal pada skala untuk menjawab benar dengan peluang 50% dari butir-butir di atas paling kecil  $-1,184$  dan paling tinggi  $1,391$ .

Secara grafis, sebaran tingkat kesukaran butir Paket II disajikan pada Gambar 26. Gambar 26 menunjukkan bahwa tingkat kesukaran seluruh butir berada di antara  $b = -2$  dan  $b = 2$ . Hal ini berarti seluruh butir tersebut termasuk dalam kategori baik dilihat dari tingkat kesukaran butirnya.



Gambar 26. Sebaran Tingkat Kesukaran Butir Tes Paket II

b. Daya beda butir

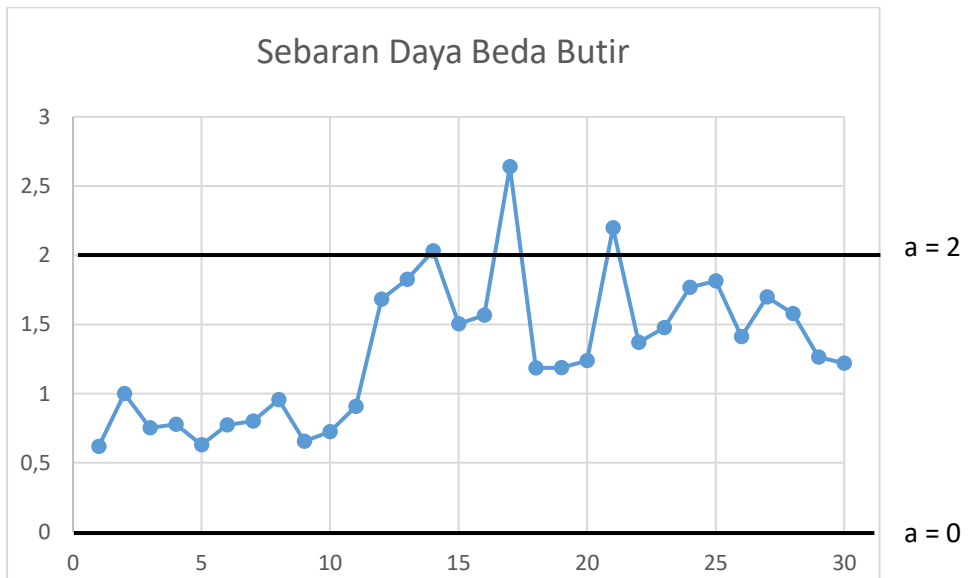
Daya beda butir dapat ditemukan dalam output BILOG MG pada nilai *slope* di *output fase 2*. Berdasarkan hasil analisis data diperoleh *output* daya beda yang disajikan pada Tabel 20.

Tabel 20. *Output* Daya Bada Butir Paket II

Butir	<i>Slope</i> ( <i>a</i> )	Butir	<i>Slope</i> ( <i>a</i> )	Butir	<i>Slope</i> ( <i>a</i> )
1	0,618	11	0,909	21	2,197
2	1,000	12	1,683	22	1,371
3	0,753	13	1,827	23	1,476
4	0,778	14	2,031	24	1,768
5	0,631	15	1,504	25	1,815
6	0,773	16	1,567	26	1,411
7	0,802	17	2,640	27	1,699
8	0,956	18	1,185	28	1,578
9	0,657	19	1,187	29	1,264
10	0,725	20	1,239	30	1,221

Tabel 20 menunjukkan bahwa butir-butir pada tes kompetensi guru Paket II memiliki daya beda yang bervariasi. Butir dengan daya beda terendah adalah butir 1 ( $a = 0,618$ ), dan butir dengan daya beda tertinggi adalah butir 17 ( $a = 2,640$ ).

Secara grafis, sebaran daya butir Paket II disajikan pada Gambar 27.



Gambar 27. Sebaran Daya Bada Butir Paket II

Gambar 27. menunjukkan bahwa daya beda 27 butir berada di antara  $a = 0$  dan  $a = 2$ , dan 3 butir daya bedanya di atas 2. Hal ini berarti dari 30 butir soal yang dikembangkan, 27 butir memiliki daya beda butir dalam kategori baik, dan 3 butir memiliki daya beda dalam kategori kurang baik.

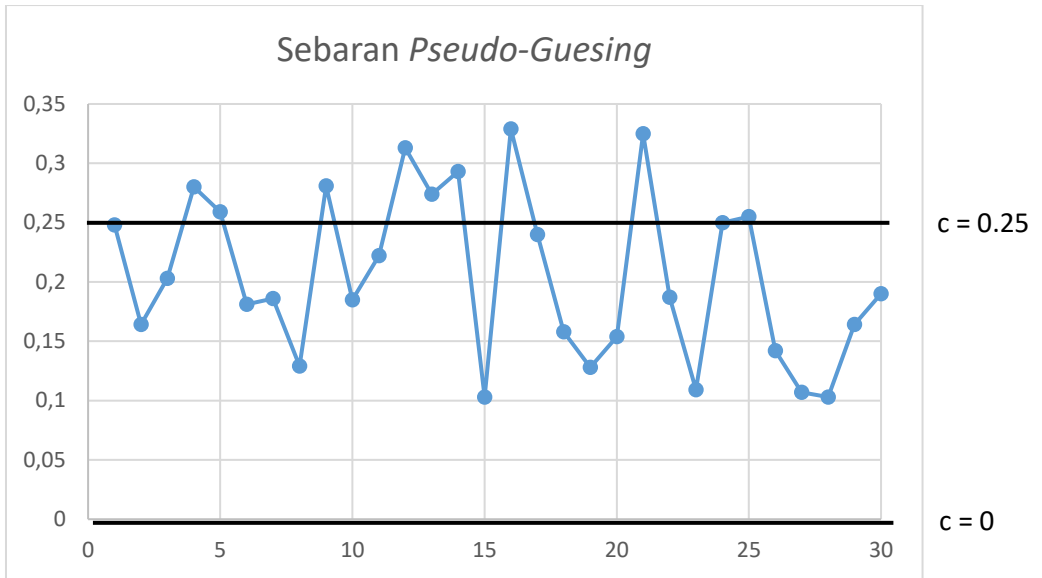
c. *Pseudo Guesing*

Nilai tebak semu dapat ditemukan dalam output *BILOG MG* pada nilai *asymptote* di *output fase 2*. Berdasarkan hasil analisis data diperoleh *output pseudo-guesing* yang disajikan pada Tabel 21.

Tabel 21. *Output Pseudo-Guesing* Butir Paket II

Butir	<i>Asymptote</i> (a)	Butir	<i>Asymptote</i> (a)	Butir	<i>Asymptote</i> (a)
1	0,248	11	0,222	21	0,325
2	0,164	12	0,313	22	0,187
3	0,203	13	0,274	23	0,109
4	0,280	14	0,293	24	0,250
5	0,259	15	0,103	25	0,255
6	0,181	16	0,329	26	0,142
7	0,186	17	0,240	27	0,107
8	0,129	18	0,158	28	0,103
9	0,281	19	0,128	29	0,164
10	0,185	20	0,154	30	0,190

Tabel 21 menunjukkan bahwa butir-butir pada tes kompetensi guru Paket II memiliki *pseudo-guesing* yang bervariasi. Butir dengan *pseudo-guesing* terendah adalah butir nomor 15 ( $c = 0,103$ ) dan butir nomor 28 ( $c = 0,103$ ), dan butir dengan *pseudo-guesing* tertinggi adalah butir nomor 16 ( $c = 3,290$ ). Secara grafis, sebaran *pseudo-guesing* Paket II disajikan pada Gambar 28.



Gambar 28. Sebaran *Pseudo-Guesing* Butir Paket II

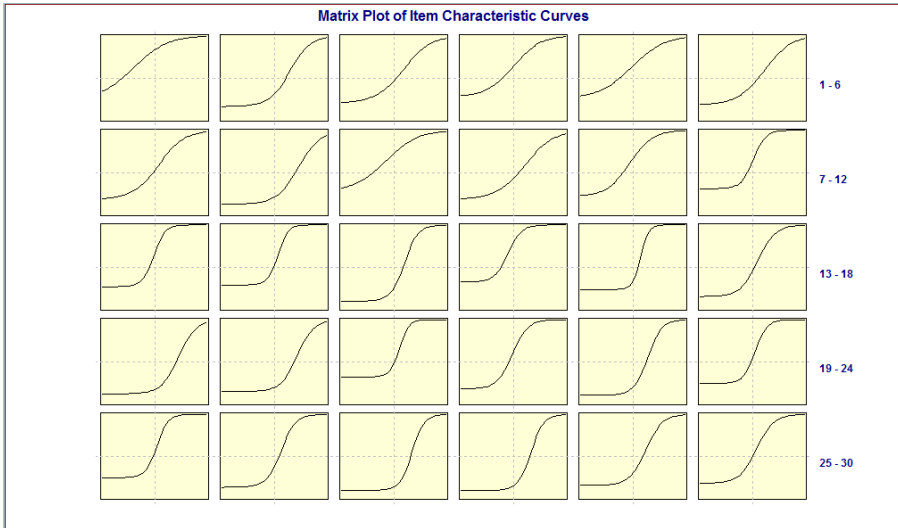
Gambar 28 menunjukkan bahwa 21 butir memiliki *pseudo-guesing* di antara  $c = 0$  dan  $c = 0,25$ , dan 9 butir memiliki *pseudo-guesing* di atas 0,25. Hal ini berarti dari 30 butir soal yang dikembangkan, 21 butir memiliki *pseudo-guesing* butir dalam kategori baik, dan 9 butir memiliki *pseudo-guesing* dalam kategori kurang baik.

d. Kurva karakteristik

Kurva karakteristik butir soal tes kompetensi guru matematika Paket II yang dijelaskan dalam buku ini merupakan bentuk kurva karakteristik butir soal berdasarkan model logistik 3 parameter. Butir soal yang ideal adalah butir soal dengan kurva karakteristik berbentuk huruf S yang cukup landai.

Kurva karakteristik diperoleh dari output *IRT Graphis*, disajikan pada Gambar 29.





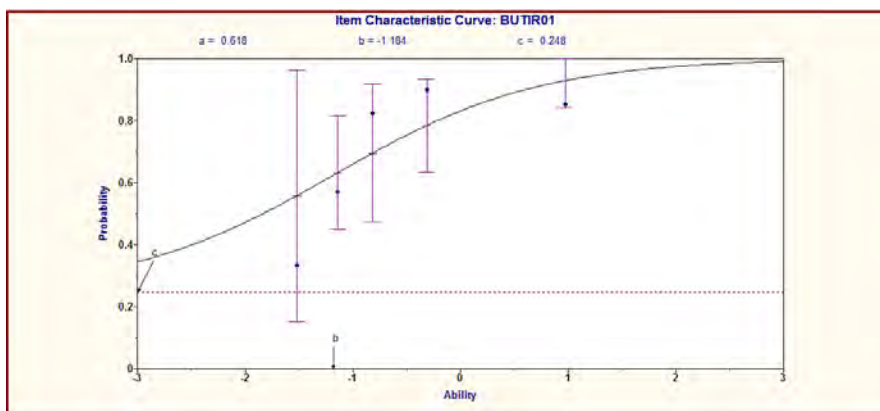
Gambar 29. Kurva Karakteristik Butir (Butir 1 – 30)

Gambar 29 menunjukkan bahwa kurva karakteristik yang terbentuk memiliki kemiringan yang bervariasi. Variasi kemiringan kurva ini menunjukkan daya beda butir. Butir soal nomor 17 memiliki kemiringan paling curam, menunjukkan bahwa butir tersebut memiliki daya beda paling tinggi. Butir soal nomor 1 memiliki kemiringan kurva paling landai, menunjukkan bahwa butir soal nomor 1 memiliki daya beda butir paling rendah.

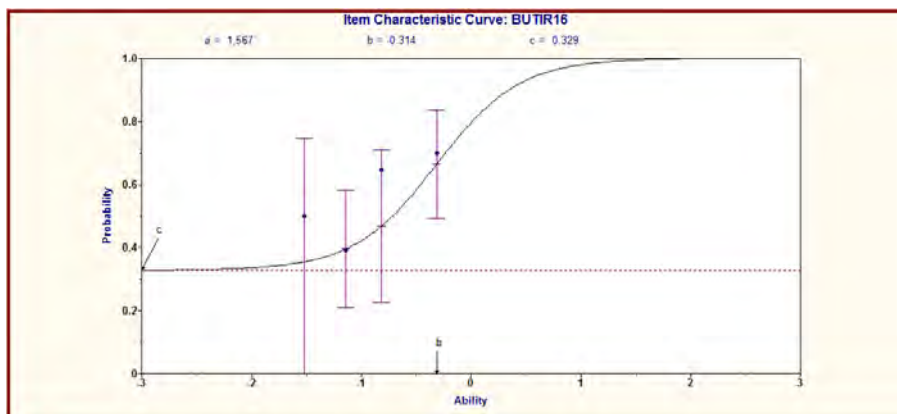
Gambar 29 juga menunjukkan bahwa butir-butir soal yang dikembangkan tidak ada yang melewati titik  $(0,0)$  dan memotong sumbu Y di atas  $y = 0$ . Hal ini berarti bahwa seluruh butir memuat *pseudo-guessing*. Adanya parameter *pseudo-guessing* ini menunjukkan bahwa probabilitas peserta dengan kemampuan rendah menjawab benar pada butir yang memiliki indeks kesukaran tidak sesuai dengan kemampuan peserta tersebut. Gambar 29 juga menunjukkan bahwa butir yang memotong sumbu Y paling tinggi butir nomor 16, dan paling rendah butir nomor 15 dan 28. Hal ini berarti bahwa butir nomor 16

merupakan butir dengan probabilitas terbesar peserta dengan kemampuan rendah menjawab benar pada suatu butir yang memiliki indeks kesukaran tidak sesuai kemampuan peserta tersebut. Butir soal nomor 15 dan 18 adalah butir soal dengan probabilitas terkecil peserta dengan kemampuan rendah menjawab benar pada butir yang memiliki indeks kesukaran tidak sesuai dengan kemampuan peserta tersebut.

Untuk mendapatkan gambaran lebih jauh tentang butir, Gambar 30 dan Gambar 31 menyajikan kurva karakteristik dari 2 butir soal, yaitu butir nomor 1 dan 16.



Gambar 30. Kurva Karakteristik Butir Nomor 1



Gambar 31. Kurva Karakteristik Butir Nomor 16

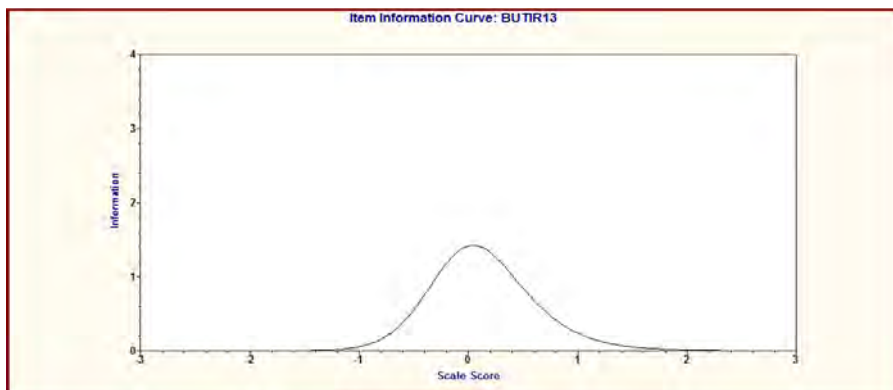
Gambar 30 menunjukkan bahwa butir nomor 1 memiliki tingkat kesukaran daya beda 0,16. Kurva karakteristik butir soal nomor 1 memiliki kemiringan landai, sesuai dengan daya beda butir yang cukup kecil. Pada kurva karakteristik tersebut juga disajikan data *pseudo-guessing* (c) adalah 0,280, terlihat juga pada grafik dimana kurva karakteristik butir soal nomor 1 tidak memotong sumbu tegak pada titik 0. Pada grafik kurva karakteristik butir soal nomor 1 juga disajikan data tentang tingkat kesukaran butir, yaitu  $b = -1,184$ , yang menunjukkan bahwa diperlukan kemampuan  $-1,184$  pada skala untuk dapat menjawab benar butir nomor 1 dengan peluang 50%.

Gambar 31 menunjukkan bahwa butir nomor 16 memiliki tingkat kesukaran daya beda 1,567. Kurva karakteristik butir soal nomor 16 memiliki kemiringan yang curam, sesuai dengan daya beda butir yang besar. Pada kurva karakteristik tersebut juga disajikan data *pseudo-guessing* (c) butir nomor 16 adalah 0,329, terlihat juga pada grafik dimana kurva karakteristik butir soal nomor 16 tidak memotong sumbu tegak pada titik 0. Pada grafik kurva karakteristik butir soal nomor 16 juga disajikan data tentang tingkat kesukaran butir, yaitu  $b = -0,314$ , yang menunjukkan bahwa diperlukan kemampuan  $-0,314$  pada skala untuk dapat menjawab benar butir 16 dengan peluang 50%.

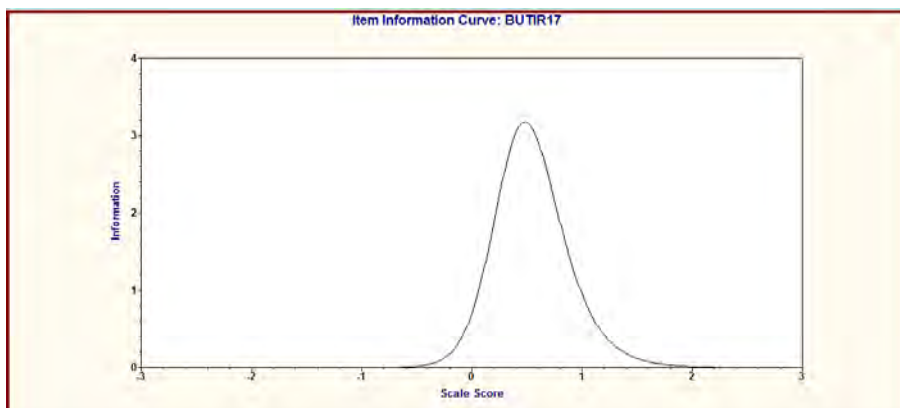
e. Fungsi informasi

Fungsi informasi butir diperoleh dari output *IRT Graphis* pada program *BLOG MG*. Fungsi informasi disajikan dalam koordinat kartesius, dimana sumbu X menunjukkan level abilitas peserta tes, dan sumbu Y menunjukkan besarnya fungsi informasi. Semakin tinggi puncak dari fungsi informasi, semakin tinggi pula informasi yang bisa diberikan

oleh suatu model. Sebagai gambaran disajikan fungsi informasi dari 2 butir, yaitu butir nomor 5 dan butir nomor 17. Kedua fungsi informasi disajikan pada Gambar 32 dan Gambar 33.



Gambar 32. Fungsi Informasi Butir Nomor 13

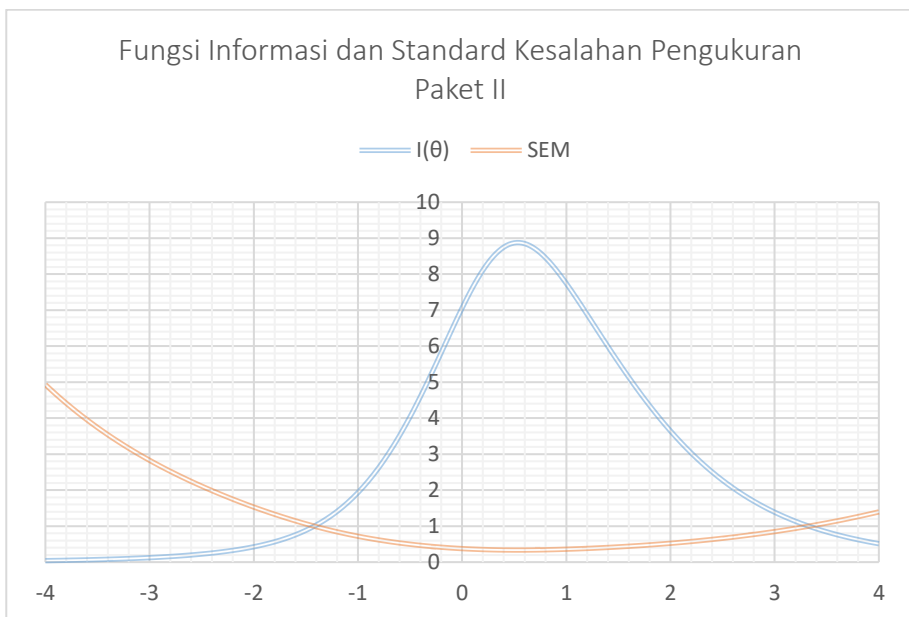


Gambar 33. Fungsi Informasi Butir Nomor 17

Gambar 32 menunjukkan bahwa butir nomor 13 akan memberikan banyak informasi apabila diberikan pada peserta didik dengan level abilitas sedang. Gambar 33 menunjukkan bahwa butir nomor 17 akan memberikan banyak informasi apabila diberikan pada peserta didik dengan level abilitas lebih tinggi. Puncak kurva fungsi informasi butir nomor 17 lebih tinggi dibandingkan nomor 13, berarti bahwa informasi

yang dapat diberikan oleh butir nomor 17 lebih banyak dibandingkan informasi yang diberikan butir nomor 13.

Seperti telah dijelaskan sebelumnya bahwa fungsi informasi memiliki hubungan berbanding terbalik dengan *SEM*, yaitu semakin besar fungsi informasi maka *SEM* semakin kecil dan sebaliknya semakin kecil fungsi informasi maka akan semakin besar nilai *SEM*. Grafik fungsi informasi dan *SEM* disajikan pada Gambar 34. Garis biru pada Gambar 34 merupakan grafik fungsi informasi dan garis merah merupakan grafik dari *SEM*. Semakin besar nilai fungsi informasi maka semakin kecil nilai *SEM*, dan sebaliknya.



Gambar 34. Kurva Fungsi Informasi dan SEM Butir Soal Paket II

#### 4. Parameter Kemampuan

Parameter kemampuan dapat ditentukan dengan memanfaatkan perpotongan kurva fungsi informasi dan *SEM*, seperti tersaji pada

Gambar 34. Gambar 34 menunjukkan bahwa nilai fungsi informasi tes maksimal tercapai pada kemampuan siswa sebesar 0.5. Gambar 42 menunjukkan bahwa fungsi informasi dan *SEM* berpotongan sebelah kiri pada  $-1,4$  dan sebelah kanan  $3,3$ . Hal ini menunjukkan bahwa tes akan memberikan informasi paling optimal apabila diberikan pada peserta tes dengan level abilitas dari  $-1,4$  sampai  $3,3$ . Apabila tes ini diberikan pada peserta dengan abilitas di bawah  $-1,4$  atau di atas  $3,3$  maka kesalahan pengukuran lebih tinggi daripada fungsi informasi yang bisa diperoleh dari tes tersebut.

Berdasarkan hasil analisis di atas secara umum dapat diperoleh bahwa tes kompetensi guru matematika SMA Paket I cocok untuk model 2P. Hasil analisis menggunakan model 2P menemukan bahwa tes kompetensi guru matematika SMA Paket I memiliki tingkat kesukaran butir terendah adalah  $-1,272$  (butir 10), dan butir dengan tingkat kesukaran tertinggi  $1,233$  (butir 1). Hal ini berarti bahwa semua butir tes Paket I termasuk kategori baik karena berada di interval  $-2$  dan  $2$ .

Analisis terhadap daya beda butir ditemukan bahwa tes kompetensi guru Paket I memiliki daya beda yang bervariasi. Butir dengan daya beda terendah adalah butir 10 ( $b = -1,272$ ), dan butir dengan daya beda tertinggi adalah butir 12 ( $b = 1,158$ ). Hal ini menunjukkan bahwa semua butir termasuk dalam kategori baik dilihat dari daya beda butirnya karena memiliki daya beda di antara  $a = 0$  dan  $a = 2$ .

Selain diketahui tingkat kesukaran dan daya beda, berdasarkan analisis terhadap fungsi informasi dan standar kesalahan pengukuran ditemukan bahwa tes akan memberikan informasi terbaik pada peserta tes dengan abilitas  $-0.4$ . Hal ini diketahui dari nilai fungsi informasi tes yang mencapai

maksimal pada kemampuanabilitas $-0,4$ . Fungsi informasi dan *SEM* berpotongan sebelah kiri pada  $-2,4$  dan sebelah kanan  $1,8$ . Hal ini menunjukkan bahwa tes akan memberikan informasi paling optimal apabila diberikan pada peserta tes dengan level abilitas dari  $-2,4$  sampai  $1,8$ .

Berdasarkan analisis terhadap tes Paket II ditemukan bahwa tes Paket II cocok untuk model 3P. Hasil analisis menggunakan model 3P menemukan bahwa tes kompetensi guru matematika SMA Paket II memiliki tingkat kesukaran butir terendah adalah  $-1,184$  (butir 1), dan butir dengan tingkat kesukaran tertinggi  $1,391$  (butir 8). Hal ini berarti bahwa semua butir tes Paket I termasuk kategori baik karena berada di interval  $-2$  dan  $2$ . Kemampuan minimal pada skala untuk dapat menjawab benar dengan peluang 50% dari tes Paket I paling kecil  $-1,184$  dan paling tinggi  $1,391$ .

Analisis terhadap daya beda butir ditemukan bahwa tes kompetensi guru Paket II memiliki daya beda yang bervariasi. Butir dengan daya beda terendah adalah butir 1 ( $b = -0,618$ ), dan butir dengan daya beda tertinggi adalah butir 17 ( $b = 2,640$ ). Hal ini menunjukkan bahwa semua butir termasuk dalam kategori baik dilihat dari daya beda butirnya karena memiliki daya beda di antara  $a = 0$  dan  $a = 2$ .

Analisis tes Paket II juga dilakukan terhadap parameter *pseudo-guessing*. Hasil analisis yang telah dilakukan menunjukkan bahwa butir-butir pada tes kompetensi guru Paket II memiliki *pseudo-guessing* yang bervariasi. Butir dengan *pseudo-guessing* terendah adalah butir nomor 15 ( $c = 0,103$ ) dan butir nomor 28 ( $c = 0,103$ ), dan butir dengan *pseudo-guessing* tertinggi adalah butir nomor 16 ( $c = 3,290$ ). Dilihat dari sebaran nilai *pseudo-guessing* ditemukan bahwa *pseudo-guessing* 21 butir berada di antara  $c = 0$  dan  $c = 0,25$ , dan 9 butir *pseudo-guessing* nya di atas  $0,25$ . Hal ini berarti dari 30 butir

soal yang dikembangkan, 21 butir memiliki *pseudo-guessing* butir dalam kategori baik, dan 9 butir memiliki *pseudo-guessing* dalam kategori kurang baik.

Analisis juga dilakukan terhadap fungsi informasi dan *SEM*. Fungsi informasi dan *SEM* mempunyai hubungan berbanding terbalik, yaitu semakin besar informasi item untuk tingkat kemampuan yang diberikan, semakin sedikit kesalahan yang terlibat dalam memperkirakan kemampuan peserta tes, dan sebaliknya (Reid, dkk., 2007: 180). Berdasarkan analisis fungsi informasi dan *SEM* ditemukan bahwa fungsi informasi dan *SEM* berpotongan sebelah kiri pada  $-1,4$  dan sebelah kanan  $3,3$ . Hal ini menunjukkan bahwa tes akan memberikan paling informasi apabila diberikan pada peserta tes dengan level abilitas dari  $-1,4$  sampai  $3,3$ . Berdasarkan analisis yang telah dilakukan juga ditemukan bahwa nilai fungsi informasi tes maksimal tercapai pada kemampuan siswa sebesar  $0,5$ .



## DAFTAR PUSTAKA

- Asamoah, D. & Ocansey, M.K. (2019). Item discrimination and distractor analysis: a technical report on thirty multiple choice core mathematics achievement test items. *International Journal of Research and Scientific Innovation (IJRSI)*, VI(IX), 23 – 33. Retrieved from [https://www.academia.edu/40383048/Item\\_Discrimination\\_and\\_Distractor\\_Analysis\\_A\\_Technical\\_Report\\_on\\_Thirty\\_Multiple\\_Choice\\_Core\\_Mathematics\\_Achievement\\_Test\\_Items](https://www.academia.edu/40383048/Item_Discrimination_and_Distractor_Analysis_A_Technical_Report_on_Thirty_Multiple_Choice_Core_Mathematics_Achievement_Test_Items)
- Ashraf, Z.A. & Jaseem, K. (2020). Classical and Modern Methods in Item Analysis of Test Tools. *International Journal of Research and Review*, 7(5), 397 – 403. Retrieved form [www.ijrrjournal.com](http://www.ijrrjournal.com)
- Bichi, A.A., Embong, R., Mamat, M., & Maiwada, D.A. (2015). Comparison of Classical Test Theory and Item Response Theory:A Review of Empirical Studies. *Australian Journal of Basic and Applied Sciences*, 9(7), 549–556. Retrieved from: [https://www.academia.edu/12523294/Comparison\\_of\\_Classical\\_Test\\_Theory\\_and\\_Item\\_Response\\_Theory\\_A\\_Review\\_of\\_Empirical\\_Studies](https://www.academia.edu/12523294/Comparison_of_Classical_Test_Theory_and_Item_Response_Theory_A_Review_of_Empirical_Studies)
- Bichi, A.A. (2016). Classical test theory: An introduction to linear modeling approach to test and item analysis. *International Journal for Social Studies*, 2(9), 27 – 33. Retrieved from: <https://edupediapublications.org/journals>
- Bichi, A.A. & Talib, R. (2018). Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development. *International Journal of Evaluation and Research in Education (IJERE)*, 7(2), 142 – 151. <https://doi.org/10.11591/ijere.v7.i2.pp142-151>
- Bulut, O. (2015). Applying Item Response Theory Models to Entrance Examination for Graduate Studies: Practical Issues and Insights. *Journal of Measurement and Evaluation in Education and Psychology*, 6(2); 313–330. <https://doi.org/10.21031/epod.17523>
- Champlain, A. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education* 2010; 44: 109–117. <https://doi:10.1111/j.1365-2923.2009.03425>

- Crocker, L. & Algina, G. (2008). *Introduction to Classical and Modern Test Theory*. Ohio: Cengage Learning
- Erguven, M. (2013). Two approaches to psychometric process: Classical test theory and item response theory. *Journal of Education*, 2(2), 23 – 30. Retrieved from: <https://jeps.ibsu.edu.ge/jms/index.php/je/article/view/84>
- Glas, C. (2008). Item response theory in educational assessment and evaluation. *Mesure et évaluation en éducation*, 31(2), 19–34. <https://doi.org/10.7202/1025005ar>.
- Hambleton, R.K. & Jones, R.W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *ITEMS – Instructional Topics in Educational Measurement*, 12(3), 38 – 47. <https://doi.org/10.1111/j.1745-3992.1993.tb00543.x>
- Hambleton, R.K & Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications Inc.
- Hambleton, R.K. & van der Linden, W.J. (1982). Advances in Item Response Theory and Applications: An Introduction. *Applied Psychological Measurement*, 6(4), 373–378. <https://doi.org/10.1177/014662168200600401>
- Han, K.T. (2012). Fixing the c Parameter in the Three-Parameter Logistic Model. *Practical Assessment, Research & Evaluation*, 17(1), 1 – 24. Retrieved from: <http://pareonline.net/getvn.asp?v=17&n=1>
- Harvey, R.J. & Hammer, A.L. (1999). *Item response theory*. *The Counseling Psychologist*, 27(3), 353–383. <https://doi.org/10.1177/0011000099273004>
- Kahrahman, N. (2014). An Explanatory Item Response Theory Approach for a Computer-Based Case Simulation Test. *Eurasian Journal of Educational Research*, Issue 54, 117–134. Retrieved from <http://ejer.com.tr/en/archives/2014-issue-54/an-explanatory-item-response-theory-approach-for-a-computer-based-case-simulation-test>
- Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1–11.

Retrieved

from

<https://www.researchgate.net/publication/228299950>

- Mahmud, J. (2017). Item response theory: A basic concept. *Educational Research and Reviews*, 12(5), 258–266, <https://doi/10.5897/ERR2017.3147>
- Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan (Edisi 2)*. Yogyakarta: Parama Publishing.
- Mardapi, D., Hadi, S. & Retnawati, H. (2015). Menentukan kriteria ketuntasan minimal berbasis peserta didik. *Jurnal Penelitian dan Evaluasi Pendidikan*, 19(1), 38–45. Retrieved from <http://journal.uny.ac.id/index.php/jpep>
- Musa, A., Shaheen, S., & Ahmed, A. (2018). Distractor analysis of multiple choice questions: A descriptive study of physiology examinations at the Faculty of Medicine, University of Khartoum. *Khartoum Medical Journal*, 11(1), 1444 – 1453. Retrieved from <http://kmjuofk.com/ojs/index.php/kmjhome/issue/view/4>
- Novick, M.R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3, 1–18. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- Ogunsakin, I. B., & Shogbesan, Y. O. (2018). Item Response Theory (IRT): A modern statistical theory for solving measurement problem in 21st century. *International Journal of Scientific Research in Education*, 11(3B), 627–635. Retrieved from <http://www.ij sre.com/past-issues.html>
- Reid, C.A. dkk. (2007). Modern psychometric methodology: applications of item response theory. *Rehabilitation Counseling Bulletin*, 50(3), 177–188. <https://doi.org/10.1177/00343552070500030501>
- Retnawati, H. (2015). *Teori respons butir dan penerapannya untuk peneliti, praktisi pengukuran dan pengujian, dan mahasiswa pascasarjana*. Yogyakarta: Nuha Medika.
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian, panduan peneliti, mahasiswa, dan psikometrian*. Yogyakarta: Parama Publishing.

- Retnawati, H. (2017). *Validitas, reliabilitas & karakteristik butir (panduan untuk peneliti, mahasiswa, dan psikometrian)*. Yogyakarta: Parama Publishing.
- Stage, C. (2003). Classical test theory or item response theory: The Swedish experience. Paper written for, and published in Spanish by Centro de Estudios Públicos, Santiago, Chile. Available at [www.cepchile.cl](http://www.cepchile.cl)
- Thissen, D. & Wainer, H. (2001). *Test scoring*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers
- Wang Z. & Osterlind S.J. (2013). *Classical test theory*. In: Teo T. (eds) *Handbook of Quantitative Methods for Educational Research*. Rotterdam: SensePublishers
- Zeiba, A. (2017). The item information function in one and two-parameter logistic models – a comparison and use in the analysis of the results of school tests. *Didactics of Mathematics*, 10(14), 87 – 96. [Http://doi.org/10.15611/dm.2013.10.08](http://doi.org/10.15611/dm.2013.10.08)

## **SEKILAS PENULIS**



*Sumaryanta, lahir di Sleman, tahun 1975. Saat ini penulis tinggal di Sleman, Yogyakarta.*

*Penulis menyelesaikan pendidikan sekolah dasar di SD Negeri Mlesen tahun 1987, dilanjutkan ke SMP N 2 Sleman, lulus tahun 1990. Tahun 1990 penulis melanjutkan studi ke SMA Negeri 1 Sleman, diselesaikan tahun 1993.*

*Pendidikan S1 ditempuh di Prodi Pendidikan Matematika FPMIPA IKIP N Yogyakarta, tahun 1993 – 1998. Pada tahun 2002, penulis melanjutkan studi S2 pada Prodi Penelitian dan Evaluasi Pendidikan Program Pascasarjana Universitas Negeri Yogyakarta, lulus tahun 2004. Pada tahun 2017, penulis melanjutkan studi S3 pada prodi dan kampus yang sama, yaitu Prodi Penelitian dan Evaluasi Pendidikan Program Pascasarjana Universitas Negeri Yogyakarta, lulus tahun 2020. Penulis memulai karier sebagai guru matematika di SMA N 1 Nglipar Gunungkidul Yogyakarta tahun 2000. Pada tahun 2010 penulis mutasi kerja ke PPPPTK Matematika, Kementerian Pendidikan dan Kebudayaan, berlokasi di Yogyakarta. Saat ini, penulis masih aktif bekerja di PPPPTK Matematika sebagai widyaiswara bidang pendidikan matematika.*

## SEKILAS ISI BUKU

Buku ini memuat penjelasan konseptual tentang teori tes klasik dan teori respon butir disertai dengan contoh praktis penerapannya. Ada 3 bab dalam buku ini, yaitu:

1. Bab I. Teori Tes

Memuat penjelasan konseptual tentang teori tes klasik dan teori respon butir.

2. Bab II. Terapan Teori Tes Klasik

Memuat penjelasan tentang: kerangka analisis dengan teori tes klasik dan contoh terapan teori tes klasik. Ada 2 (dua) contoh yang dibahas, masing-masing dijelaskan tentang analisis data dengan menggunakan aplikasi (ITEMAN), analisis butir dengan teori tes klasik (tingkat kesukaran, daya beda, dan keberfungsian distraktor), dan tindak lanjut hasil analisis yang dilakukan.

3. Bab III. Terapan Teori Tes Respon Butir

Memuat penjelasan tentang: kerangka analisis dengan teori respon butir dan contoh terapan teori respon butir. Ada 2 (dua) contoh yang dibahas, masing-masing dijelaskan tentang analisis data dengan aplikasi (BILOG MG), analisis butir dengan teori respon butir (uji asumsi, uji kecocokan model, penentuan parameter butir dan kemampuan).

Semoga kehadiran buku ini dapat bermanfaat bagi para pembaca sekaligus menginspirasi pemanfaatannya untuk pengembangan tes.

Penerbit



CONFIDENT

*Unggul dan Luhur*

ISBN : 978-623-6834-42-8



9 786236 834428