

## MEMBANDINGKAN IRT MODEL 3PL DENGAN IRT MODEL 4PL UNTUK PENILAIAN MENGGUNAKAN COMPUTERIZED ADAPTIVE TEST

Aslam Fatkhudin<sup>1)</sup>, M. Fikri Hidayatullah<sup>2)</sup>

<sup>1)</sup>Manajemen Informatika, Politeknik Muhammadiyah Pekalongan  
email: Fatkhudin@gmail.com

<sup>2)</sup>Manajemen Informatika, Politeknik Muhammadiyah Pekalongan  
email: muhammadfikri.uad@gmail.com

### Abstract

*One of the computer-based testing is the Computerized Adaptive Test (CAT), which is a computer-based testing system where the items were given to the participants adapted to test the ability of the participants. Assessment methods are usually applied in CAT is Item Response Theory (IRT). IRT models are most commonly used today is the model 3 Parameter Logistic (3PL), which is about the discrimination, difficulty and guessing. However 3PL IRT models have not provided information more objectively test the ability of participants. The opinion of the test participants were tested items were also to be considered. In this study using CAT in combination with IRT model of 4PL.*

*In this research, the development of CAT which uses about 4 parameters, namely the discrimination, difficulty, guessing and questionnaires. The questions used were about UAS 1 English subjects. Samples were taken from 40 students answer with the best value of the total 172 students spread across 6 classes to measure the parameter estimation problem. Further testing using CAT application 4PL IRT models compared to CAT 3PL IRT models.*

*From research done shows that the CAT application combined with IRT models 4PL can measure the ability of the test taker shorter or faster and also opportunities participants correctly answered the test items was done tend to be better than the 3PL IRT models.*

**Keywords :** Ability, CAT, IRT, 3PL, 4PL, Probability, Test

### 1. PENDAHULUAN

*Item Response Theory (IRT) adalah teori yang digunakan dalam pengukuran pendidikan (Baker, 2001). IRT merupakan model matematika yang memperhitungkan kemungkinan orang memberikan jawaban dengan benar untuk setiap item (Huang, Lin, & Cheng, 2009). Menurut teori ini, jika dua tes berbeda yang berisi pertanyaan berbeda diterapkan pada orang yang sama, diperkirakan tingkat kemampuan akan sama. Skor yang diperoleh pada akhir dari tes ini adalah bukan nilai ujian tetapi estimasi kemampuan. Tingkat kemampuan ini dikenal sebagai  $\theta$  ( $\theta$ ) yang memiliki nilai variasi antara +3 sampai -3. Pada skala  $\theta$ , nilai 0 merupakan tingkat kemampuan rata-rata, nilai negatif merupakan tingkat kemampuan yang lebih rendah dari nilai rata-rata dan positif merupakan tingkat kemampuan lebih tinggi dari rata-rata (Huang, 2009). Fungsi distribusi kemungkinan memberikan jawaban yang tepat*

*untuk pertanyaan oleh peserta didik disebut Item Characteristic Curve (ICC) (Huang, 2009). ICC menunjukkan kemungkinan adanya ujian memberikan jawaban yang tepat pada tingkat kemampuan ( $\theta$ ). Ada berbagai model yang digunakan dalam IRT. Di antara model IRT yang ada, Model Tiga Parameter Logistik (3PL) adalah salah satu model yang paling umum digunakan, yang merupakan bentuk yang lebih umum dari model Dua Parameter Logistik (2PL) dan Model Satu Parameter Logistik (1PL) (Huang, 2009). Ketiga parameter tersebut adalah daya beda soal, tingkat kesulitan soal dan tebakan. Namun IRT model 3PL dianggap belum memberikan informasi kemampuan peserta tes lebih objektif. Pendapat peserta tes terhadap butir soal yang diujikan juga perlu dipertimbangkan. Untuk itu diperlukan IRT model 4PL dengan kuisioner sebagai parameter ke-empatnya dari peserta tes dalam*

memberi penilaiannya terhadap butir soal yang diujikan.

Kemajuan terbaru dalam teknologi komputer dan teori pengujian telah mempercepat perubahan format ujian dari konvensional (ujian menggunakan pensil dan kertas) ke pengujian berbasis komputer. Ada 2 macam pengujian berbasis komputer, yaitu CBT (*Computerized Based Test*) dan CAT (*Computerized Adaptive Test*). CBT adalah jenis pengujian berbasis komputer yang menyajikan beberapa soal yang berurutan (*linier*) kemudian peserta ujian menjawab kesemua soal tersebut. Nilai tergantung pada jumlah soal yang dijawab dengan benar. Sama seperti tes konvensional, hanya saja CBT disajikan dalam format digital (Triantafillou dkk, 2008). Keunggulan CBT adalah tidak perlu menggandakan kertas-kertas soal dan lembar jawaban untuk dibagikan ke peserta tes, tidak memerlukan tim khusus untuk mengoreksi soal karena sistem yang akan langsung mengoreksi dan mengkalkulasi jumlah soal yang benar dan salah serta hasil tes dapat diketahui saat itu juga dengan cepat sesaat setelah peserta selesai mengikuti tes (Bartram, 2001).

Berbeda dengan CAT, dimana butir soal yang diberikan kepada peserta tes disesuaikan dengan kemampuan peserta tes (Ozyurt dkk, 2013). CAT ini tidak hanya efisien dalam waktu pengujian dan jumlah item pengujian, tetapi juga dapat memungkinkan diagnosis yang lebih baik pada tingkat yang lebih tinggi, misalnya CAT dapat digunakan dilingkungan perguruan tinggi maupun dilingkungan kerja, untuk mengukur kemampuan mahasiswa atau pegawai. CAT dianggap mampu menggambarkan kemampuan akademis seseorang (Horward, 1990). Keunggulan lain dari CAT adalah peserta ujian dapat mengikuti tes kapanpun dan dimanapun dia siap (Lilley & Barker, 2003). Metode penilaian yang biasanya diterapkan dalam CAT adalah *Item Response Theory* (IRT) atau Teori Respon Butir (Baker, 2001).

Pada penelitian ini kami mencoba menggunakan CAT yang dikombinasikan dengan IRT model 3PL dan 4PL sebagai

pembandingnya sehingga diharapkan dapat mengukur kemampuan peserta tes lebih akurat.

## 2. KAJIAN LITERATUR

Ada sejumlah studi di mana CAT dikombinasikan dengan IRT telah diterapkan. Seperti Implementasi Modul Penilaian Individu Dalam Sistem E-Learning. Penelitian ini berkaitan dengan proses desain dan pengembangan CAT yang diintegrasikan dengan UZWEBMAT (*Turkish abbreviation of Adaptive and INtelligent WEB based MAThematics teaching-learning system*). UZWEBMAT adalah sistem pakar pembelajaran elektronik (*e-learning*) yang mendukung kemampuan dan kecerdasan individu. Hal ini ditujukan untuk mengajar tingkat probabilitas subjek. Pertama, sebuah bank soal dibuat dalam penelitian ini. Bank soal ini diberikan kepada peserta didik dalam bentuk lembaran kerja. Tes ini dianalisis sesuai dengan IRT (*Item Response Theory*) atau Teori Respon Butir. Model-Data kemampuan beradaptasi diuji pada akhir analisis. Disimpulkan bahwa data dapat disesuaikan dengan IRT model 3PL. Pada akhir operasi ini, bank soal akhir yang berisi 752 pertanyaan secara total diperoleh. Peserta didik diajarkan dengan UZWEBMAT untuk menyelesaikan semua mata pelajaran dengan mengambil bagian terakhir dari tes yang berisi 30 pertanyaan/soal secara acak. Item/soal dalam tes ini secara dinamis dipilih dari bank soal sesuai dengan penilaian adaptif. Dari hasil tersebut diperoleh perkiraan tingkat kemampuan peserta ujian serta informasi tingkat kesulitan tiap pertanyaan untuk selanjutnya disampaikan kepada peserta ujian berikutnya. Tingkat kemampuan peserta ujian itu kembali diperkirakan dan diperbarui sesuai dengan jawaban dari setiap pertanyaan. Berkas modul ini, peserta didik secara individual diajarkan dengan UZWEBMAT disediakan dengan penilaian adaptif sesuai dengan kualifikasi dan kapasitas mereka sendiri. Berbeda dengan tes klasik, penilaian individu menjadi kenyataan melalui penilaian adaptif yang memberikan kesempatan untuk

mengetahui peserta didik sesuai dengan tingkat kemampuan mereka (Ozyurt H, dkk, 2012).

Penelitian lain yang mengkombinasikan antara CAT dengan IRT misalnya menganalisa multi faktor DIF (*Differential Item Functioning*) dalam memberikan ujian nasional mata pelajaran matematika sekolah dasar di Brasil. Hasil analisis data menunjukkan bahwa soal berdasarkan faktor jenis sekolah lebih memuaskan dan konsisten dalam mendistribusikan kemampuan peserta didik dibandingkan dengan soal yang berdasarkan faktor geografis (Goncalves dkk, 2013).

### 3. METODE PENELITIAN

#### *Bahan Penelitian*

Sumber bahan yang dipergunakan dalam penelitian ini ada 3 (tiga), yaitu soal yang didapat dari guru pengampu mata pelajaran, jawaban dari siswa/peserta ujian dan kuisisioner penelaahan soal dari siswa/peserta ujian. Tujuan adanya kuisisioner penelaahan soal dari siswa/peserta ujian supaya penilaian lebih objektif karena bisa jadi soal yang diberikan oleh guru pengampu mata pelajaran, menyimpang atau tidak sesuai dengan materi soal yang telah diajarkannya.

Soal yang didapat dari guru pengampu kemudian *diinputkan* ke dalam sistem untuk selanjutnya menjadi data bank soal. *Penginput* soal bisa dari *administrator* sistem atau guru yang bersangkutan, dalam hal ini berarti harus terdapat data guru yang didapat dari pihak sekolah yang sudah *diinputkan* ke dalam sistem oleh *admin*. Agar siswa/peserta ujian bisa mulai mengerjakan soal, siswa harus *login* terlebih dahulu ke dalam sistem. Karenanya perlu *diinputkan* juga data siswa oleh *admin*. Untuk mendapatkan nilai parameter *a*, *b* dan *c* digunakan program ITEMAN, sedangkan nilai *d* sebagai parameter ke-empat dari metode IRT didapatkan dari Kuisisioner yang selanjutnya melekat pada masing-masing butir dari bank soal yang sudah *diinputkan* ke dalam sistem.

#### *Alat Penelitian*

Alat penelitian yang berkaitan dengan hardware yang dipergunakan adalah laptop sebagai alat utama penelitian sekaligus servernya dengan spesifikasi processor core-i3 2,4 GHz; RAM DDR3 dengan kapasitas 4 Gbyte; HardDisk 500 Gbyte; VGA onboard, Monitor dengan ukuran 14 inchi; DVD+RW, LAN; WLAN serta Bluetooth.

Sedangkan siswa/peserta ujian dalam mengerjakan soal dengan menggunakan komputer yang ada di laboratorium komputer sekolah mereka. Spesifikasi komputernya adalah processor dual core 1,75 GHz; RAM DDR2 dengan kapasitas 2 Gbyte, HardDisk 120 Gbyte, VGA onboard, LED Monitor dengan ukuran 15 inch serta DVD+RW. Karena sistem yang dibuat berbasis *localhost*, maka dibutuhkan alat tambahan berupa *switch* dan kabel UTP yang sudah diinstalasi konektor RJ-45 untuk menghubungkan semua komputer dalam laboratorium tersebut.

#### *Prosedur Penelitian*

Ada 2 (dua) prosedur atau tahapan utama yang kami gunakan pada penelitian ini, yaitu tahapan persiapan dan tahapan perancangan.

#### *Tahap Persiapan*

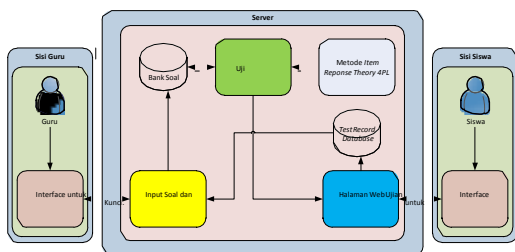
Pada tahap ini kami mencari dan mempelajari literatur pustaka dari sejumlah jurnal atau penelitian serta buku-buku referensi yang terkait dengan metode yang digunakan, yaitu *Item Response Theory* dan *Computerized Adaptive Test*. Selanjutnya dilakukan observasi untuk mengambil data dari tempat penelitian di SMA Muhammadiyah 1 Pekajangan. Data yang dibutuhkan adalah data soal beserta kunci jawaban dan kisi-kisinya, data guru dan data siswa serta kuisisioner butir soal yang didapat dari siswa.

#### *Tahap Perancangan*

Pada tahap ini dibuat arsitektur sistem informasi dan *flowchart* programnya untuk kemudian dirancang DFD dan tabel.

Arsitektur sistem terdiri dari dua bagian, yaitu sisi guru dan sisi siswa (Gambar 1). Di

sisi guru, dengan menggunakan desktop atau laptop, guru mengakses halaman untuk guru. Di halaman atau *interface* tersebut, guru bisa menginputkan soal beserta kunci jawabannya. Setelah soal tersebut dikerjakan oleh peserta ujian/siswa, sistem akan melaksanakan sebuah penilaian ujian akhir semester untuk uji adaptif dari soal. Pada sisi siswa, dengan menggunakan desktop atau laptop seorang siswa mengerjakan ujian adaptif. Untuk mengembangkan sistem, diperlukan sebuah bank soal. Sistem akan menerapkan metode IRT model 4 parameter logistik untuk memperkirakan kemampuan siswa setelah ia menjawab soal tes.

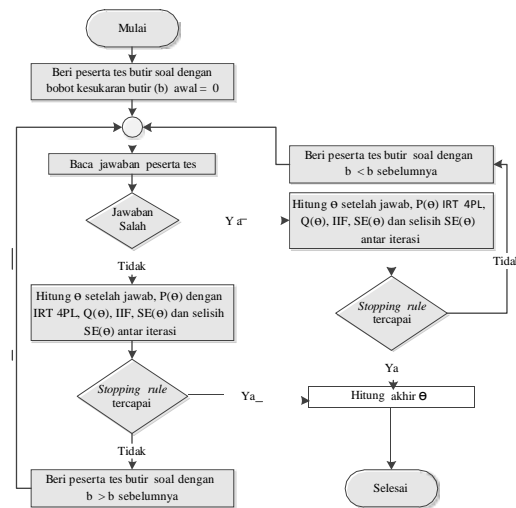


Gambar 1. Arsitektur CAT Dikombinasikan Dengan IRT Model 4PL

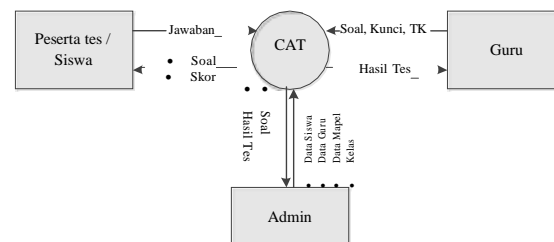
Gambaran *flowchart* programnya adalah peserta tes diberi soal dengan nilai parameter *difficulty* dari soal awal sedang ( $b$  awal = 0) karena dianggap semua peserta tes mempunyai tingkat kemampuan ( $\theta$ ) sedang ( $\theta$  awal = 0). Jika soal awal tersebut dapat dijawab dengan benar, diberikan soal berikutnya yang lebih sulit, tetapi jika soal awal tersebut dijawab salah maka berikutnya diberi soal yang lebih mudah. Proses tersebut dilakukan berulang-ulang sampai *stopping rule* tercapai. Kemudian tingkat kemampuan akhir ( $\theta$ ) peserta tes dan probabilitas yang menggunakan IRT model 4PL dihitung. Berikutnya menghitung nilai total tes. Terakhir, sistem menampilkan informasi berupa diagram dan keterangan dari hasil perhitungan tersebut. Gambar 2 menunjukkan *flowchart* program *Computerized Adaptive Test* yang dikombinasikan dengan IRT model 4PL.

Sedangkan *Data Flow Diagram* (DFD)-nya adalah siswa dan guru memberikan datanya ke dalam sistem. Data diterima oleh

Administrator untuk dikonfirmasi kemudian dimasukkan ke dalam sistem sebagai data yang *valid*. Guru menginputkan soal ke dalam sistem. Soal yang sudah dimasukkan ke sistem kemudian akan dikerjakan oleh siswa. Siswa dan guru otomatis akan menerima skor dan hasil tes dari sistem. Alur tersebut bisa digambarkan dengan Diagram Konteks sebagai ditunjukkan pada gambar 3.



Gambar 2. Flowchart CAT Dikombinasikan Dengan IRT Model 4PL



Gambar 3. Konteks Diagram

Dari DFD di atas, dihasilkan beberapa tabel sebagai berikut:

Tabel 1. Struktur Tabel Peserta

No.	Nama Field	Tipe dan Panjang Field
1	Kd_peserta	Int (10)
2	Nama_peserta	Varchar (50)
3	Nomor_peserta	Varchar (40)
4	Password	Varchar (10)
5	Kelas	Varchar (10)

Tabel 2. Struktur Tabel Kelas

No.	Nama Field	Tipe dan Panjang Field
1	Kd_kelas	Int (10)
2	Kelas	Varchar (50)

Tabel 3. Struktur Tabel Mapel

No.	Nama Field	Tipe dan Panjang Field
1	Kd_mapel	Int (10)
2	Mapel	Varchar (100)

Tabel 4. Struktur Tabel Guru

No.	Nama Field	Tipe dan Panjang Field
1	Kd_Guru	Int (5)
2	Kode	Varchar (30)
3	Nama	Varchar (50)
4	Mapel	Varchar (50)
5	Password	Varchar (50)

Tabel 5. Struktur Tabel Soal

No.	Nama Field	Tipe dan Panjang Field
1	Kd_Soal	Int (10)
2	Isi_Soal	Text
3	Kunci_Soal	char (1)
4	Tingkat_kesulitan	float
5	Pengecoh	Char (1)

Tabel 6. Struktur Tabel Daftar Tes

No.	Nama Field	Tipe dan Panjang Field
1	Kd_judul_tes	Int (10)
2	Judul	Varchar (50)

Tabel 7. Struktur Tabel Tes

No.	Nama Field	Tipe dan Panjang Field
1	Kd_Tes	Int (10)
2	Kd_judul_tes	Int (10)
3	Waktu_test	Timestamp
4	Kd_peserta	Int (10)

Tabel 8. Struktur Tabel Detail Tes

No.	Nama Field	Tipe dan Panjang Field
1	Kd_DetailTes	Int (20)
2	Kd_tes	Int (10)
3	Kd_Soal	Int (10)
4	b	float
5	Jawaban	Char (1)
6	Skor_Jawaban	Int (4)
7	Teta_awal	float
8	Teta_akhir	float
9	d	float
10	P	float
11	Q	float
12	I	float
13	SE	float
14	Selisih_SE	float

#### 4. HASIL DAN PEMBAHASAN

##### Hasil

Penelitian ini mengambil subjek siswa/siswi peserta Ujian Akhir Semester (UAS) 1 mata pelajaran Bahasa Inggris di SMA Muhammadiyah 1 Pekajangan Pekalongan Tahun Ajaran 2015/2016. Objek yang dipilih adalah butir soal objektif pilihan ganda dan respons butir peserta UAS 1

Bahasa Inggris. Data penelitian berupa butir soal pilihan ganda sejumlah 45 butir dan respons butir peserta UAS 1 Bahasa Inggris yang dikumpulkan dengan teknik dokumentasi dan kuisioner. Populasi penelitian ini adalah peserta UAS 1 Bahasa Inggris meliputi 6 kelas atau 172 peserta. Dari 172 peserta tersebut, sebagai sampel diambil 40 peserta yang memilih jawaban benar dengan rentang lebih dari 28 soal ( $\pm 23,26\%$ ). Berdasarkan sampel yang diperoleh, selanjutnya dilakukan penskoran dan analisa butir soal dengan metode *Item Response Theory* (IRT) model 4 Parameter Logistik (PL) dan kemudian dibuatkan sistem *Computerized Adaptive Test* (CAT).

*Hasil Analisis Butir Soal dengan metode IRT model 4 PL.*

Langkah pertama adalah membuat tabel masukan yang berisi kunci jawaban dan pengecoh dari masing-masing butir soal serta respon butir atau jawaban dari masing-masing peserta. Dari tabel masukan kemudian dibuat tabel penskoran. Jika jawaban sesuai kunci jawaban (benar), maka akan bernilai 1, sebaliknya jika jawaban tidak sesuai kunci jawaban (salah), maka akan bernilai 0. Dari tabel penskoran akan didapatkan skor/nilai dengan menjumlah semua jawaban benar untuk masing-masing peserta ujian kemudian dikalikan 100 dan dibagi jumlah soal yaitu 45 butir soal. Dari proses pemasukan dan penskoran, didapatkan hasil sebaran 9 (sembilan) kelompok nilai seperti yang ditampilkan pada tabel 9.

Selanjutnya tabel penskoran tersebut diurutkan secara *descending* atau dari nilai terbesar ke nilai terkecil. Dari tabel pengurutan bisa diketahui ranking, nilai terbesar yaitu 82 dan nilai terkecilnya yaitu 62. Nilai tersebut dipakai untuk menghitung *proportion correct* (proporsi jawaban benar peserta ujian) dan nilai estimasi kemampuan peserta ujian ( $\theta$ ). Nilai  $\theta$  ini nantinya akan digunakan untuk menghitung *probabilitas* peserta ujian ( $P \theta$ ).

Langkah berikutnya setelah tabel pengurutan terbentuk adalah membaginya

dalam dua kelompok, yaitu 50% kelompok atas (kelompok nilai besar) dan 50% kelompok bawah (kelompok nilai kecil). Tujuannya untuk mendapatkan nilai daya beda ( $a$ ), nilai tingkat kesulitan ( $b$ ) dan nilai pengecoh ( $c$ ). Tabel 10. menunjukkan keterangan kualitas soal hasil dari ketiga nilai parameter logistik tersebut.

Daya beda berfungsi untuk menghitung tingkat validitas dan menggambarkan tingkat kemampuan soal dalam membedakan antar peserta didik yang sudah memahami materi yang diujikan dengan peserta didik yang belum/tidak memahami materi yang diujikan. Adapun klasifikasinya adalah seperti berikut ini (Crocker dan Algina, 1986) :

1. 0,40 - 1,00 soal diterima baik (baik sekali)
2. 0,30 - 0,39 soal diterima tetapi perlu diperbaiki (baik)
3. 0,20 - 0,29 soal diperbaiki
4. 0,19 - 0,00 soal tidak dipakai/dibuang

Tingkat kesulitan soal berfungsi untuk mengetahui peluang menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Indeks tingkat kesulitan ini pada umumnya dinyatakan dalam bentuk proporsi yang besarnya berkisar 0,00 - 1,00 (Aiken, 1994). Semakin besar indeks tingkat kesulitan yang diperoleh dari hasil hitungan, berarti semakin mudah soal itu. Suatu soal memiliki TK= 0,00 artinya bahwa tidak ada siswa yang menjawab benar dan bila memiliki TK=1,00 artinya bahwa siswa menjawab benar. Perhitungan indeks tingkat kesulitan ini dilakukan untuk setiap nomor soal.

Tabel 9. Tabel hasil sebaran kelompok nilai

NO.	Kelompok Nilai	Jumlah Siswa
1	82	2
2	78	1
3	76	2
4	71	4
5	67	1
6	69	3
7	67	3
8	64	5

9	62	9
<b>Jumlah</b>		<b>30</b>

Tabel 10. Tabel keterangan kualitas soal

No.	Parameter	Keterangan	Jumlah
1	Daya Beda	Soal baik sekali	2
		Soal baik	1
		Soal diperbaiki	13
		Soal dibuang	29
		<b>Jumlah</b>	<b>45</b>
2	Tingkat Kesulitan	Soal sukar	6
		Soal sedang	14
		Soal mudah	25
		<b>Jumlah</b>	<b>45</b>
3	Pengecoh	Fungsi	35
		Tidak Fungsi	10
		<b>Jumlah</b>	<b>45</b>

Adapun klasifikasinya adalah seperti berikut ini :

1. 0,00 - 0,30 soal tergolong sukar
2. 0,31 - 0,70 soal tergolong sedang
3. 0,71 - 1,00 soal tergolong mudah

Faktor pengecoh berfungsi untuk mengetahui berfungsi tidaknya jawaban yang tersedia. Butir soal yang baik, pengecohnya akan dipilih secara merata oleh siswa/siswi yang menjawab salah (5%).

Sedangkan parameter ke-4 (parameter *carelessness*) atau nilai *d* diambilkan dari kuisisioner peserta ujian untuk menilai sendiri butir soal. Hal ini dimaksudkan agar penilaian terhadap butir soal lebih objektif. Setelah semua nilai parameter diketahui, berikutnya memasukkan nilai-nilai tersebut dalam program CAT.

#### Computerized Adaptive Test

Dengan mengambil sebuah contoh seorang siswa yang mengerjakan *computerized*

*adaptive test* didapatkan hasil seperti tertera pada tabel 11. Detail Hasil Tes.

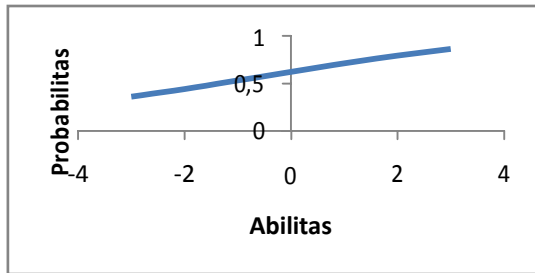
Tabel 11 menggambarkan jumlah soal yang dijawab oleh peserta tes. Dari total jumlah soal 45 butir, dengan pola jawaban yang diinputkan, peserta tes hanya menjawab 10 butir soal secara acak seperti yang ditunjukkan pada kolom i. Kolom i merupakan kode soal. Kolom b adalah bobot atau nilai tingkat kesukaran butir soal. Pada kolom skor menampilkan nilai dari jawaban peserta tes.

Tabel 11. Tabel detail hasil tes (bagian 1)

$\theta$					
i	b	Skor	$\theta$ Awal	Stl Jwb	d
138	0	1	0	0	1,00
131	0,07	0	0	0	0,67
105	-	1	0	-	-
	0,02				0,33
144	0,18	1	-0,02	0,02	0,33
113	0,2	1	0,18	0,2	0,33
140	0,38	0	0,2	0,2	0,67
141	-	0	0,2	0,2	1,00
	0,11				-
130	-	1	0,2	-	1,00
	0,16			0,16	-
132	0,41	0	-0,16	0,16	0,67
	-				-
119	-	1	-0,16	-	-
	0,39			0,39	0,33

Tabel 12. Detail hasil tes (bagian 2)

Pi( $\theta$ )	Q( $\theta$ )	Selisih
SE		
0,50	0,50	2,00
0,69	0,31	0,53
0,17	0,83	0,18
0,17	0,83	0,13
0,17	0,83	0,10
0,72	0,28	0,10
0,37	0,63	0,09
0,50	0,50	0,07
0,82	0,18	0,04
0,17	0,83	0,03



Gambar 4. Grafik *Probabilitas* IRT model 4PL

Kolom skor ini berkorelasi dengan kolom  $\theta$  setelah menjawab. Jika peserta tes menjawab dengan benar, maka akan bernilai 1 dan nilai pada kolom  $\theta$  setelah menjawab akan berubah positif yang menunjukkan bahwa kemampuan peserta tes meningkat, sebaliknya jika jawaban peserta tes salah, maka akan bernilai 0 dan nilai pada kolom  $\theta$  setelah menjawab akan berubah negatif, menunjukkan kemampuan peserta tes menurun. Nilai 0 merupakan tingkat kemampuan rata-rata. Kolom  $d$  adalah nilai kuisioner yang diberikan oleh peserta tes di tiap butir soal yang dikerjakan. Kolom  $P_i(\theta)$  menunjukkan nilai *probabilitas*/peluang peserta tes menjawab dengan benar. Sedangkan  $Q(\theta)$  menunjukkan nilai *probabilitas*/peluang peserta tes menjawab salah. Kolom selisih SE menunjukkan nilai mutlak selisih kesalahan baku pengukuran. Nilai pada kolom SE inilah yang menjadi acuan penghentian penyajian soal. Bila selisih SE sudah 0,01 maka tes akan berhenti. Di kolom selisih SE tabel 9 di atas nilainya yang terakhir belum 0,01 tetapi penyajian soal tetap berhenti di soal ke-10, itu terjadi karena kemungkinan bila penyajian soal dilanjutkan, selisih SE berikutnya akan bernilai dibawah 0,01 yaitu 0,00.

Dari data pada tabel 11 dapat diperoleh grafik hubungan *probabilitas* dengan tingkat kemampuan setiap kali peserta ujian menjawab soal yang diberikan oleh sistem. Dengan mengambil contoh soal terakhir yang dikerjakan oleh peserta tes yaitu soal ke-10 didapatkan grafik probabilitas yang naik seiring perubahan nilai kemampuan peserta ujian, seperti ditampilkan pada Gambar 4 di atas. Artinya peluang peserta tes untuk

menjawab benar butir soal meningkat, *ekuivalen* dengan kemampuan peserta tes.

## 2.2. Pembahasan

Membandingkan CAT sebelumnya yang menggunakan metode IRT model 3PL tanpa menggunakan nilai  $d$  (nilai kuisioner dari peserta tes di tiap butir soal) sebagai parameter ke empatnya, dengan CAT pada penelitian ini yang menggunakan metode IRT model 4PL (menggunakan nilai  $d$ ), menggunakan data dari hasil penelitian seperti pada tabel 11 dan gambar 4 grafik *probabilitas*, diperoleh hasil seperti ditampilkan pada tabel 13.

Tabel 13 menunjukkan peserta tes menjawab soal sebanyak 13 butir secara acak dari sejumlah butir soal yang ada yaitu 45 butir. Ditunjukkan pada nilai terakhir di kolom selisih SE (kesalahan baku pengukuran) yaitu 0,01 yang menjadi batas acuan untuk menghentikan penyajian soal.

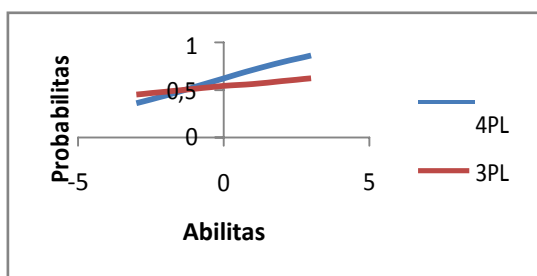
Tabel 13. Tabel detail hasil tes CAT yang menggunakan metodol IRT model 3PL (bagian 1)

i	b	Skor	$\theta$	
			Awal	Stl Jwb
138	0	1	0	0
131	0,07	0	0	0
105	-	1	0	-
	0,02			0,02
144	0,18	1	-0,02	0,18
113	0,2	1	0,18	0,2
140	0,38	0	0,2	0,2
141	-	0	0,2	0,2
	0,11			
130	-			-
	0,16	1	0,2	0,16
132	0,41	0	-0,16	0,16
119	-			-
	0,39	1	-0,16	0,39
115	0,8	1	-0,39	0,8
137	1,4	0	0,8	0,8
110	-			
	0,72	0	0,8	0,8



Tabel 14. Tabel detail hasil tes CAT yang menggunakan metodel IRT model 3PL (bagian 2)

$P_i(\theta)$	$Q(\theta)$	Selisih SE
0,50	0,50	2,00
0,53	0,47	0,58
0,50	0,50	0,26
0,50	0,50	0,15
0,50	0,50	0,11
0,58	0,42	0,08
0,37	0,63	0,06
0,50	0,50	0,05
0,72	0,28	0,03
0,50	0,50	0,04
0,50	0,50	0,03
0,73	0,27	0,02
0,07	0,93	0,01



Gambar 5. Grafik perbandingan probabilitas IRT model 3PL dengan IRT model 4PL

Berbeda dengan CAT yang menggunakan metode IRT model 4PL seperti ditunjukkan pada tabel 11 di atas, dimana peserta tes hanya menjawab soal sebanyak 10 butir saja. Hal ini menjelaskan kelebihan dan kontribusi dari penelitian ini. Aplikasi CAT dengan menggunakan metode IRT model 4PL dapat mengukur kemampuan peserta tes lebih singkat atau cepat.

Apabila data di tabel 13 dibuat grafik, dengan mengambil contoh soal terakhir yang dikerjakan oleh peserta tes yaitu soal ke-13 kemudian dibandingkan dengan grafik probabilitas IRT model 4PL yang terdapat pada gambar 5 di atas, akan diperoleh grafik perbandingan seperti gambar 6 di atas. Pada grafik ditunjukkan bahwa peluang (probabilitas) peserta tes menjawab benar pada IRT model 4PL lebih baik dibandingkan

peluang (probabilitas) peserta tes menjawab benar pada IRT model 3PL. Grafik tersebut juga menjelaskan makin banyak parameter logistik yang digunakan, maka kurva probabilitasnya akan semakin meningkat.

Sekali lagi ini menjelaskan kelebihan dari sistem pada penelitian, yaitu *probabilitas/peluang* peserta tes menjawab benar cenderung naik.

## 5. SIMPULAN

Simpulan lah melakukan pengujian terhadap metode IRT model 4PL yang dikombinasikan dengan *Computerized Adaptive Tes* (CAT) maka didapatkan kesimpulan bahwa estimasi terhadap parameter soal dapat dilakukan secara otomatis oleh sistem. Dengan menggunakan algoritma *Computerized Adaptive Test* (CAT) dengan *Item Response Theory* model 4PL, pembuat soal dapat mengestimasi parameter soal tanpa harus bergantung kepada para ahli.

Parameter soal yang diperoleh dari estimasi parameter secara otomatis tersebut dapat digunakan untuk *adaptive online test*.

Dari data hasil penelitian yang terdapat pada pembahasan, menunjukkan bahwa aplikasi CAT yang menggunakan metode IRT model 4PL dapat mengukur tingkat kemampuan dan *probabilitas/peluang* peserta tes menjawab benar butir soal dengan dengan lebih baik.

Kelebihan dari aplikasi CAT yang menggunakan metode IRT model 4PL adalah dapat mengukur kemampuan peserta tes lebih singkat atau cepat dan juga peluang peserta tes menjawab benar butir soal yang dikerjakan cenderung lebih baik dibandingkan dengan aplikasi CAT yang menggunakan metode IRT model 4PL.

Semakin banyak parameter logistik yang digunakan, maka kurva probabilitasnya akan semakin meningkat.

## 6. REFERENSI

- Aiken, Lewis R., 1994, *Psychological Testing and Assessment*, (Eight Edition), Boston : Allyn and Bacon.
- Baker F., 2001, *The basics of item response theory*. ERIC clearinghouse on assessment

- and evaluation. College Park, MD: University of Maryland.
- Bartram, Dave SHL Group plc, Thames Ditton, Surrey, UK dan Hambleton, Ronald K, 2001, *Computer-Based Testing and the Internet*, USA : University of Massachusetts at Amherst.
- Crocker, L. & Algina, J., 1986, *Introduction to Classical and Modern Test Theory*. New York : Holt, Rinehart and Winston, Inc.
- Goncalves F.B., Gamerman D., Soares T.M., 2013, *Simultaneous multifactor DIF analysis and detection in Item Response Theory*, Computational Statistics and Data Analysis 59, 144 – 160.
- Horward W., 1990, *Computerized adaptive testing: A primer*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Huang Y.M., Lin Y.T., & Cheng S.C., 2009, *An adaptive testing system for supporting versatile educational assessment*, Journal of Computers & Education 52, 53–67.
- Lilley M., & Barker T., 2003, *An evaluation of a computer adaptive test in a UK university context*. 7th Computer assisted assessment conference, 8th and 9th July, 2003, Loughborough.
- Ozyurt H., Ozyurt O., Baki A., & Guven B., 2012, *Integrating computerized adaptive testing into UZEWEBMAT : Implementation of individualized assessment module in an e-learning system*, Journal Expert System with Application, 39, 9837 – 9847.
- Ozyurt H., Ozyurt O., & Baki A., 2013, *Design and development of an innovative individualized adaptive and intelligent e-learning system for teaching–learning of probability unit: Details of UZWEBMAT*, Journal Expert System with Application, 40, 2914 – 2940.
- Triantafillou E., Georgiadou E., & Economides A.A., 2008, *The design and evaluation of a computerized adaptive test on mobile devices*, Journal of Computers & Education 50, 1319–1330.