

PENERAPAN *COMPUTERIZED ADAPTIVE TEST* PADA TES ONLINE MENGGUNAKAN ALGORITMA TEORI RESPON BUTIR MODEL 3 PL

Wahyuni¹⁾, Kusrini²⁾

^{1), 2)} Magister Teknik Informatika Universitas Amikom Yogyakarta
Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta, 55281
Email : wahyuni@wicida.ac.id¹⁾, kusrini@amikom.ac.id²⁾

Abstrak

Sistem ini dikembangkan menjadi sebuah sistem online dimana sistem ini dapat berjalan di internet atau jaringan lokal. Pada dasarnya sistem ini hampir sama dengan sistem CBT (Computerized Based Testing), namun sistem CAT (Computerized Adaptive Test) dapat merandom dan memberikan soal sesuai dengan kemampuan pesertanya. Jika peserta tes tersebut memiliki kemampuan rendah, maka sistem akan memberikan soal yang cenderung mudah. Sebaliknya, jika peserta tes memiliki kemampuan tinggi, maka sistem akan memberikan soal yang cenderung sulit.

Sistem ini menggunakan algoritma teori respon butir model 3 PL dimana parameter yang digunakan adalah tingkat kesukaran butir soal (b), daya beda butir soal (a) dan tebakan semu (c). Sistem CAT ini akan memberikan soal secara random berdasarkan respon dari peserta terhadap soal yang dikerjakan sebelumnya, maka tiap peserta tes cenderung tidak akan mendapatkan urutan soal yang sama dan jumlah soal yang sama dari sistem. Pemberhentian pemberian butir soal kepada peserta tes akan terjadi jika tingkat ketelitian estimasi kemampuan telah tercapai yaitu selisih $SE \leq 0.01$ atau jika butir soal telah habis.

Hasil penelitian menunjukkan bahwa sistem ini dapat menampung bank soal yang biasa digunakan untuk keperluan tes dalam jumlah banyak. Sistem ini juga dapat merandom soal sesuai dengan kemampuan peserta dan dapat mempersingkat waktu pengerjaan tes, karena soal yang diberikan kepada para peserta rata-rata hanya sebanyak 19 butir soal. Jika estimasi kemampuan peserta sudah tercapai, maka sistem secara otomatis akan menghitung skor akhir.

Kata kunci: computerized adaptive test, teori respon butir, CAT.

1. Pendahuluan

Computerized Adaptive Test (CAT) adalah sistem pengujian berbantuan komputer yang lebih maju dibanding Computerized Base Testing[1]. Tidak seperti CBT, komputer hanya dimanfaatkan untuk memindahkan tes dari form kertas ke monitor komputer, pada CAT komputer diatur dan dimanfaatkan sedemikian rupa

hingga dapat memilih dan menampilkan butir-butir soal yang sesuai dengan kemampuan individu peserta tes. Karena tes disesuaikan dengan kemampuan peserta tes, maka tak ada pertanyaan yang lebih sukar ataupun lebih mudah. Peserta yang berkemampuan tinggi hanya akan diberikan butir soal dengan tingkat kesukaran tinggi, begitupun sebaliknya, bagi peserta yang berkemampuan rendah hanya akan diberi butir-butir soal dengan tingkat kesukaran rendah dan tidak perlu diberikan butir soal yang sukar. Karena CAT hanya menyajikan soal yang tingkat kesukarannya sesuai dengan kemampuan peserta tes, maka jumlah butir soal yang disajikan oleh CAT bisa lebih sedikit dibandingkan dengan soal yang disajikan dalam bentuk Paper Pencil Test (PPT) dengan ketelitian pengukuran yang sama bahkan lebih baik. Dengan demikian, penggunaan CAT dapat mengurangi jumlah waktu dan biaya untuk pengadministrasian tes dan penyusunan butir-butir soal dalam bank soal.

Teori respon butir atau *Item Response Theory (IRT)* menggambarkan peluang menjawab butir soal secara benar berdasarkan tingkat kemampuan peserta tes dan butir soal yang diberikan[1]. Ada tiga model logistik dalam teori respons butir, yaitu model satu parameter logistik (1PL), model dua parameter logistik (2PL) dan model tiga parameter logistik (3PL). Perbedaan dari ketiga model ini terletak pada banyaknya parameter yang digunakan dalam menggambarkan karakteristik butir soal dalam model yang digunakan. Parameter-parameter yang digunakan tersebut adalah indeks daya beda butir (disimbolkan dengan a), indeks kesukaran (disimbolkan dengan b) dan indeks tebakan semu (disimbolkan dengan c). Sedangkan kemampuan peserta tes dilambangkan dengan θ .

Melalui sistem yang dibuat ini diharapkan dapat lebih membantu para penguji untuk mengukur kemampuan masing-masing individu peserta tes dengan tidak hanya melihat dari nilai yang dicapai seperti yang dilakukan pada PPT, namun juga dapat menganalisis titik kelemahan peserta tes melalui butir-butir soal yang diberikan oleh sistem. Selain itu sistem ini juga diharapkan dapat meminimalisir waktu pelaksanaan tes, dan meminimalisir waktu dan biaya pengadministrasian dalam mempersiapkan ujian atau tes tersebut.

Sebuah penelitian telah membandingkan IRT model tiga parameter logistic (3PL) dengan IRT model empat parameter logistik (4PL) untuk penilaian menggunakan *computerized adaptive test*[2]. Penelitian tersebut menggunakan algoritma teori respon butir model 3PL dan 4PL. Algoritma teori respon butir model 3PL adalah model yang paling umum digunakan, namun Fatkhudin & Hidayatullah mencoba menambahkan parameter ke empat yaitu *carelessness* sehingga menjadi model 4PL. Hasil dari penelitian tersebut adalah algoritma model 4PL dapat mengukur kemampuan peserta tes lebih singkat dan cepat, dan peneliti menegaskan peluang peserta tes menjawab benar cenderung naik. Namun pada penelitian ini, tidak dijelaskan apakah hasil pengukuran atau estimasi kemampuan peserta menjadi lebih baik atau tidak. Penelitian selanjutnya adalah mengembangkan *computerized adaptive test* menggunakan algoritma *fuzzy* dengan metode Tsukamoto berbasis web[3]. Pada penelitian ini peneliti mencoba menggantikan algoritma teori respon butir dengan algoritma *fuzzy* menggunakan metode Tsukamoto dalam pengambilan keputusan untuk pemberian butir soal. Hasil dari penelitian ini mengatakan bahwa algoritma *fuzzy* dengan metode Tsukamoto dapat diterapkan dalam pengambilan keputusan untuk memilih soal pada sistem CAT, namun tidak dijelaskan apakah hasil tersebut dapat berjalan lebih baik dari pada penerapan algoritma teori respon butir yang umum digunakan atau tidak. Selain itu sebelumnya telah dikembangkan sistem *computerized adaptive test* menggunakan teori respon butir namun hanya menggunakan model 1PL[4]. Dimana tentunya akan lebih baik menggunakan model 3PL sehingga didapatkan hasil yang lebih akurat. Pada penelitian ini akan diterapkan sistem *computerized adaptive test* pada tes online yang akan menggunakan algoritma teori respon butir model 3PL dalam perhitungan untuk mengestimasi kemampuan peserta tes. Peneliti mengobservasi soal TOEFL yang dimiliki oleh Balai Bahasa STMIK Widya Cipta Dharma. Balai Bahasa STMIK Widya Cipta Dharma memiliki tiga level soal TOEFL dimana soal tersebut dimuat dalam buku *Longman Introductory Course for the TOEFL Test: The Paper Test*. Alasan peneliti menggunakan soal TOEFL yang dimiliki oleh Balai Bahasa Wicida Samarinda dikarenakan soal-soal tersebut sudah diuji coba kepada responden dan pelevelan terhadap soal-soal tersebut sudah ada, sehingga soal-soal tersebut dapat dimasukkan atau dijadikan butir soal pada sistem CAT ini. Pada penelitian ini, peneliti hanya mengambil soal bagian *Structure and Written* dengan alasan hasil penilaian dari bagian tersebut sudah mencerminkan sebagian besar kemampuan berbahasa Inggris para peserta tes.

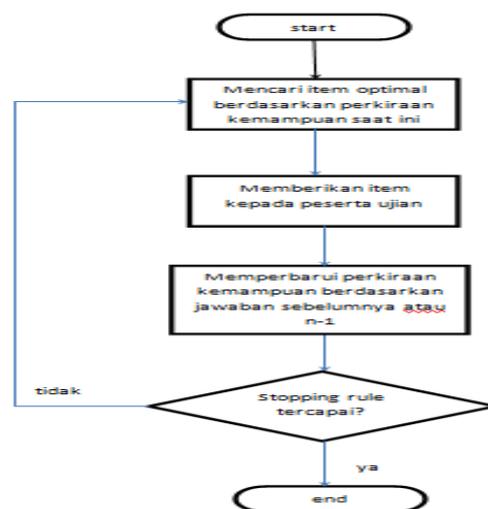
2. Pembahasan

Prinsip *Computerized Adaptive Test* (CAT) adalah sebagai berikut [5]:

- Membangun bank soal, bank soal umumnya menggunakan teori respon butir.

- Prosedur pemilihan item awal (*starting rule*), biasanya CAT dapat dimulai dengan memilih butir soal dengan tingkat kesukaran sedang.
- Prosedur pemilihan item selama tes berjalan.
- Prosedur untuk mengakhiri tes (*stopping rule*), biasanya tes CAT berhenti apabila item pada bank soal telah habis dan tingkat estimasi kemampuan peserta telah tercapai.
- Mengestimasi kemampuan siswa.

Berdasarkan cara kerja CAT di atas, maka dapat digambarkan cara kerja dari CAT seperti pada Gambar 1 di bawah ini:



Gambar 1. Cara Kerja CAT

Prinsip kerja dari algoritma teori respon butir adalah sebagai berikut[1]. Mula-mula peserta tes diberi soal dengan tingkat kesulitan awal yang sedang (b awal=0) karena dianggap mempunyai tingkat kemampuan awalnya (θ awal) = sedang = 0. Peserta tes diberi kesempatan untuk menjawab soal dengan alokasi waktu tertentu.

Jika soal dengan tingkat kesulitan sedang tersebut dijawab dengan benar, peserta diberi soal baru yang lebih sulit, jika dijawab salah maka peserta diberi soal yang lebih mudah. Kemudian kemampuan (θ) setelah menjawab soal baru $P(\theta)$, $Q(\theta)$, $I_i(\theta)$, $SE(\theta)$ dan harga mutlak selisih kesalahan baku antar penyajian soal dihitung.

Rumus yang digunakan untuk menghitung kemampuan (θ), probabilitas menjawab benar berdasarkan kemampuan ($P(\theta)$), probabilitas menjawab salah ($Q(\theta)$), fungsi informasi butir ($I_i(\theta)$), dan kesalahan baku ($SE(\theta)$) adalah sebagai berikut:

$$\theta = b_i + \frac{1}{Da_i} \ln (0.5 (1 + \sqrt{(1 + 8c_i)})) \quad (1)$$

$$P_i(\theta) = c_i + \frac{(1-c_i)e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad (2)$$

$$Q_i(\theta) = 1 - P_i(\theta) \quad (3)$$

$$I_i(\theta) = P_i(\theta) Q_i(\theta) \quad (4)$$

$$SE(\theta) = \frac{1}{\sum_{i=1}^N I_i(\theta)} \quad (5)$$

Keterangan:

- θ = tingkat kemampuan peserta tes.
- $P_i(\theta)$ = probabilitas peserta tes yang memiliki kemampuan θ dapat menjawab butir i dengan benar.
- a_i = indeks daya pembeda butir ke- i .
- b_i = indeks kesukaran butir ke- i .
- c_i = indeks tebakan semu butir ke- i .
- e = bilangan natural yang nilainya mendekati 2,718.
- D = faktor penskalaan yang harganya 1,7.
- $Q_i(\theta)$ = probabilitas menjawab salah.
- $I_i(\theta)$ = fungsi informasi butir ke- i .
- $SE(\theta)$ = kesalahan baku.

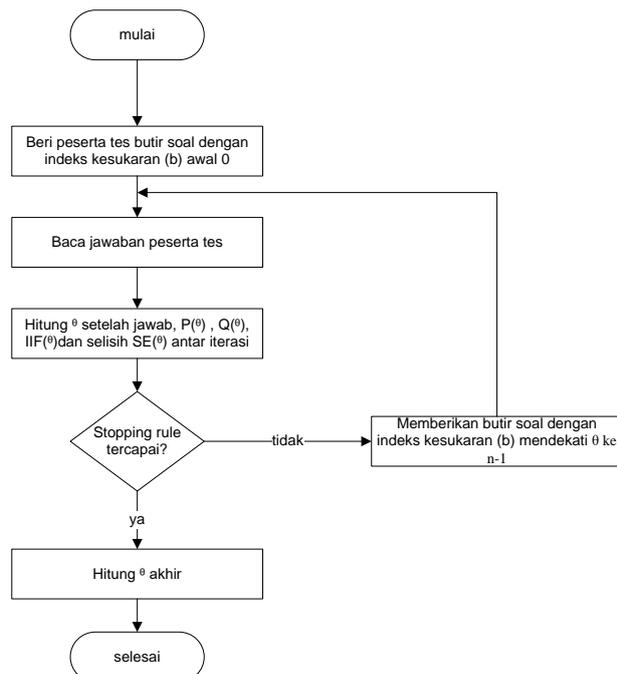
Proses tersebut dilakukan berulang-ulang sampai *stopping rule* tercapai, kemudian kemampuan akhir (θ) peserta tes dihitung.

Besarnya kemampuan setelah menjawab butir pertanyaan pada baris pertama dihitung dengan mempertimbangkan benar salahnya jawaban atau skor jawaban terhadap butir soal yang disajikan program dengan logika sebagai berikut:

Jika Skor Jawaban = 1, Maka θ setelah jawab dihitung (6)

Jika Skor Jawaban = 0, Maka θ setelah jawab = θ awal (7)

Adapun flowchart algoritma teori respon butir yang digunakan adalah sebagai berikut:



Gambar 2. Flowchart Algoritma CAT dengan Teori Respons Butir

Besarnya $P(\theta)$ dan $Q(\theta)$ dalam tabel 1 saling berkebalikan dan nilai maksimum salah satu dari keduanya adalah 1. Agar dapat memperoleh harga yang berkebalikan sesuai harapan, maka benar salahnya jawaban atau skor jawaban dijadikan patokan untuk menentukan besarnya $P(\theta)$. Logika yang digunakan untuk menghitung $P(\theta)$ adalah sebagai berikut:

Jika skor jawaban = 1, maka $P(\theta)$ dihitung sesuai rumus 1PL, 2PL atau 3PL. (8)

Jika skor jawaban = 0, maka $P(\theta) = 1 - P(\theta)$ yang dihitung sesuai rumus 1PL, 2PL atau 3 PL. (9)

Algoritma yang biasa digunakan dalam CAT adalah algoritma teori respon butir termasuk dalam memilih butir soal selama tes berjalan. Strategi yang dapat digunakan untuk memberikan item berikutnya adalah *maximum information* dan *minimum expected posterior standard deviation*, yang disebut juga *Bayesian estimation*[6]. Pada *maximum information* dilakukan dengan memilih item pada setiap tahap yang memiliki nilai b (indeks kesukaran) mendekati perkiraan θ (estimasi kemampuan peserta) pada saat itu.

Adapun range untuk masing-masing parameter dan estimasi kemampuan peserta tes sudah ditentukan[7]. Tingkat kesulitan soal (b) dikatakan baik jika nilai berkisar antara -2 sampai +2. Daya beda (a), indeks daya pembeda butir tidak bersifat negatif sehingga range daya beda adalah 0 - 1. Tebakan semu (c) atau kebetulan menjawab benar memiliki range 0 - 0.25 jika pilihan soal pilihan ganda terdapat empat buah pilihan. Tingkat kemampuan (θ) yang digunakan adalah -3 sampai +3.

Berikut adalah tabel range tabel berdasarkan tingkat kesukaran dan daya beda.

Tabel 1. Tabel Range Indeks Kesukaran

Range	Keterangan
$-2 \leq b < -0.5$	Mudah
$-0.5 \leq b < 0.5$	Sedang
$0.5 \leq b < 2$	Sukar

Tabel 2. Tabel Range Daya Beda

Range	Keterangan
$1 \geq a \geq 0.7$	Sangat Baik
$0.7 > a \geq 0.4$	Baik
$0.4 > a \geq 0.2$	Sedang
$0.2 > a \geq 0$	Kurang

Pada penelitian ini indeks kesukaran sudah ditentukan berdasarkan level soal yang dimiliki oleh Balai Bahasa STMIK Widya Cipta Dharma, sehingga peneliti hanya perlu mencari daya beda dari masing-masing soal tersebut. Setelah dicari daya beda dari ketiga level

tersebut maka didapatkan jumlah soal yang dapat dimasukkan ke dalam bank soal CAT ini yaitu sebanyak 15 soal mudah, 41 soal sedang, dan 29 soal sukar.

Halaman awal dari sistem CAT yang sudah dibuat dapat dilihat pada gambar 3.



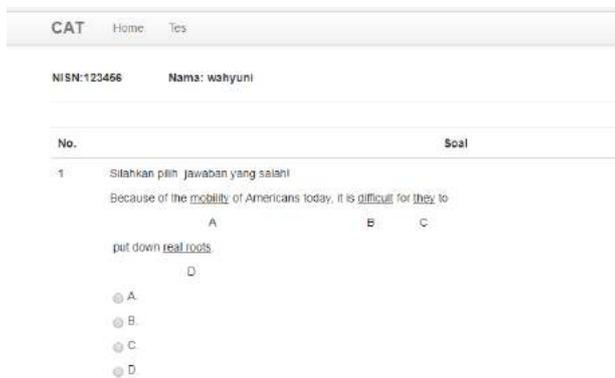
Gambar 3. Halaman Awal Sistem CAT

Pada halaman awal peserta tes diminta untuk menginputkan id peserta dan password yang sebelumnya sudah didaftarkan oleh admin dari sistem tersebut.



Gambar 4. Halaman Daftar Tes

Setelah peserta berhasil login, maka peserta diminta untuk memilih tes yang akan diikuti pada halaman daftar tes. Pada halaman tersebut sudah tersedia pilihan tes yang bisa diikuti oleh peserta.



Gambar 5. Halaman Tes

Setelah peserta memilih tes yang akan diikuti, maka peserta akan masuk ke halaman tes dimana sudah tersedia soal yang harus di jawab oleh peserta tes satu persatu.



Gambar 6. Halaman Hasil Tes Peserta

Setelah pemberian soal selesai dan peserta tes telah menyelesaikan tes tersebut, maka akan muncul halaman hasil tes peserta. Pada halaman tersebut kita bisa melihat jumlah soal yang dikerjakan oleh peserta tes. Berapa jumlah jawaban yang benar dan salah, durasi pengerjaan tes dan nilai yang didapat oleh peserta. Karena nilai estimasi kemampuan peserta (θ) antara -3 sampai + 3, maka nilai tersebut perlu dikonversikan dengan skala terendah 0 dan tertinggi 100 lebih mudah diterima oleh banyak pihak. Untuk mendapatkan nilai akhir tersebut, maka digunakan rumus sebagai berikut:

$$Skor(100) = 50 + \frac{50}{3} \theta \quad (10)$$

No.	b	a	c	Skor	θ Awal	θ SE	Jwb	P(θ)	Q(θ)	IIF	SE(θ)	Selisih SE
1	0	0.91	0	1	0	0	0	0.5	0.5	0.26	2	2
2	0.1	0.78	0.1	0	0	0	0	0.48	0.52	0.26	1.414	0.596
3	0	0.69	0.08	1	0	0.054	0.549	0.451	0.248	1.156	0.256	
4	0.1	1	0.06	1	0.064	0.16	0.564	0.446	0.247	1.003	0.153	
5	0.2	1	0.09	0	0.16	0.16	0.47	0.53	0.248	0.997	0.196	
6	0.1	0.82	0.08	0	0.16	0.16	0.441	0.559	0.247	0.919	0.078	
7	0	0.46	0.19	0	0.16	0.16	0.358	0.615	0.237	0.761	0.658	
8	0.1	0.78	0.05	1	0.16	-0.06	0.536	0.462	0.248	0.711	0.05	
9	0	0.73	0.19	0	-0.06	-0.06	0.42	0.59	0.244	0.671	0.04	
10	0.1	0.82	0	0	-0.06	-0.06	0.456	0.514	0.26	0.936	0.035	
11	0.2	0.55	0.1	1	-0.06	-0.149	0.501	0.439	0.248	0.607	0.029	
12	-0.1	0.55	0.07	1	-0.149	-0.062	0.543	0.457	0.248	0.961	0.026	
13	0.1	0.78	0.1	1	-0.062	0.172	0.571	0.409	0.245	0.666	0.023	
14	0.2	1	0.08	1	0.172	0.277	0.57	0.43	0.245	0.536	0.02	
15	0.3	0.36	0.15	0	0.277	0.277	0.426	0.573	0.245	0.52	0.016	
16	0.2	0.27	0	0	0.277	0.277	0.401	0.509	0.26	0.603	0.017	
17	0.1	0.75	0	0	0.277	0.277	0.445	0.555	0.247	0.496	0.015	
18	-0.3	0.45	0.11	0	0.277	0.277	0.348	0.652	0.227	0.475	0.013	
19	-0.41	0.73	0.12	0	0.277	0.277	0.263	0.737	0.184	0.465	0.01	
θ Akhir					0.277							
Skor												54.62

Gambar 7. Halaman Detail Hasil Tes

Untuk melihat detail hasil tes dari peserta, maka kita bisa melihatnya melalui halaman admin. Pada gambar 7 di atas, dapat kita lihat bahwa sistem akan memberikan soal pertama yang tingkat kesukarannya (b) sedang yaitu 0. Setelah menjawab respon dari peserta tes, maka sistem akan menghitung (θ) setelah menjawab soal baru, P (θ), Q (θ), I_i(θ), SE(θ) dan harga mutlak selisih kesalahan baku antar penyajian soal. Jika respon ppeserta tes untuk soal tersebut benar, maka diberikan nilai 1. Jika salah maka

diberikan nilai 0. Jika nilai respon peserta = 1, maka (θ) setelah menjawab soal baru akan dihitung sesuai dengan persamaan 1, tetapi jika respon jawaban peserta bernilai 0, maka (θ) setelah menjawab soal baru = (θ) sebelum menjawab soal baru. Pada gambar 7, dapat dilihat bahwa peserta tes mendapatkan soal sebanyak 19 buah. Banyaknya soal yang diberikan oleh sistem bergantung pada respon dari peserta itu sendiri. Sistem akan berhenti memberikan soal atau tercapainya *stopping ruel* adalah jika selisih kesalahan baku sudah mencapai 0.01 atau kurang dari 0.01 atau jika soal yang diberikan sudah habis.

3. Kesimpulan

Dari hasil uji coba di atas, dapat dikatakan bahwa sistem CAT menggunakan algoritma teori respon butir 3PL ini benar-benar dapat mempersingkat waktu pengerjaan tanpa mengurangi keakuratan perhitungan hasil akhir. Dapat kita lihat bahwa peserta tes hanya mendapatkan 19 buah soal dari sistem dimana 8 soal dijawab benar dan 11 soal di jawab salah dengan hasil akhir yang didapat adalah 55 dimana hasil tersebut tidak jauh berbeda jika kita melakukan tes tersebut secara manual. Dengan durasi waktu yang dibutuhkan peserta untuk menyelesaikan tes tersebut hanya 9 menit 24 detik, hal itu sudah membuktikan bahwa sistem CAT memang mempunyai kelebihan yaitu dapat meminimalisir waktu pengerjaan suatu tes tanpa mengabaikan hasil akhir yang didapat.

Daftar Pustaka

- [1] S. Hadi, "Pengembangan *Computerized Adaptive Test* Berbasis Web," Yogyakarta: Aswaja Pressindo, 2013.
- [2] A. Fatkhudin, M.F. Hidayatullah, "Membandingkan IRT Model 3PL Dengan IRT Model 4PL Untuk Penilaian Menggunakan *Computerized Adaptive Test*," SAINTEKBU: Jurnal Sains dan Teknologi, Volume 9 No. 1 November 2016.
- [3] L. M. Anggrenita, I. Muslim, M.R.A. Saf, "Pengembangan *Computerized Adaptive Test (CAT)* Menggunakan Algoritma Fuzzy Dengan Metode Tsukamoto Berbasis Web", Jurnal Aksara Komputer Terapan Politeknik Caltex Riau, Vol. 5, No. 1, Tahun 2016.
- [4] L. Erawan, S. Santosa, "*Computer-Adaptive Test* Dengan Pendekatan *Item Response Theory* Satu Parameter," Jurnal Teknologi Informasi, Volume 5 Nomor 2, Oktober 2009, ISSN 1414-9999.
- [5] Winarno, "Pengembangan *Computerized Adaptive Testing (CAT)* Menggunakan Metode Pohon Segitiga Keputusan," Jurnal Penelitian dan Evaluasi Pendidikan Tahun 16, Nomor 2, 2012.
- [6] H.C. Bagus, "Administrasi Ujian Nasional (UN) Dengan Menggunakan Model *Computerized Adaptive Testing (CAT)*," Jurnal Pendidikan dan Kebudayaan, Vol. 18, Nomor 1, Maret 2012.