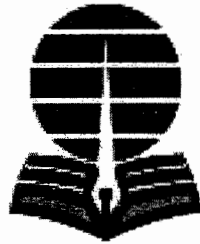


LAPORAN PENELITIAN LANJUT

BIDANG KELEMBAGAAN DAN PENGEMBANGAN SISTEM



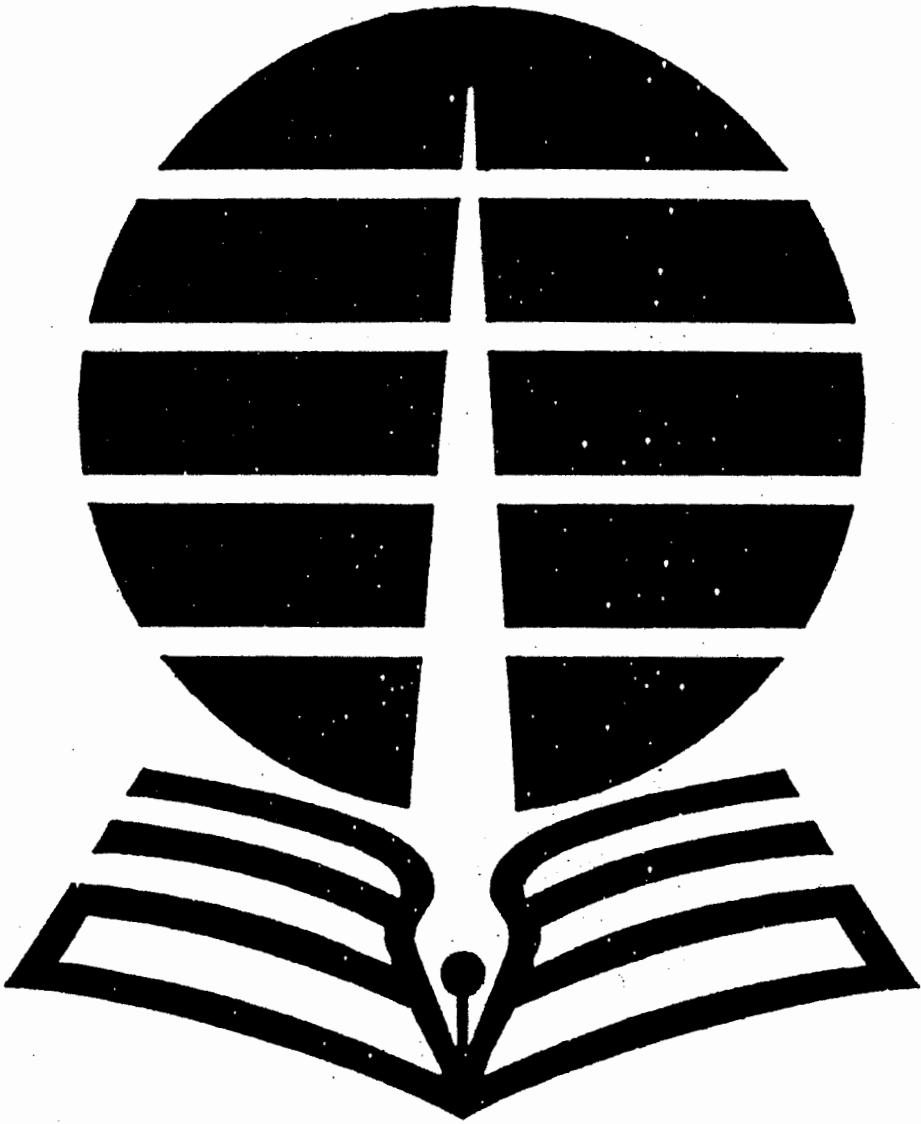
**PENGEMBANGAN MODEL SISTEM UJIAN BERDASARKAN
COMPUTERIZED ADAPTIVE TESTING SEBAGAI UPAYA
EFISIENSI PENYELENGGARAAN UJIAN
UNIVERSITAS TERBUKA**

Oleh:

**Agus Santoso
Dwi Astuti Aprijani
Unggul Utan Sufandi
Iswaya Maalik**

**LEMBAGA PENELITIAN DAN PENGABDIAN MASYARAKAT
UNIVERSITAS TERBUKA**

2010



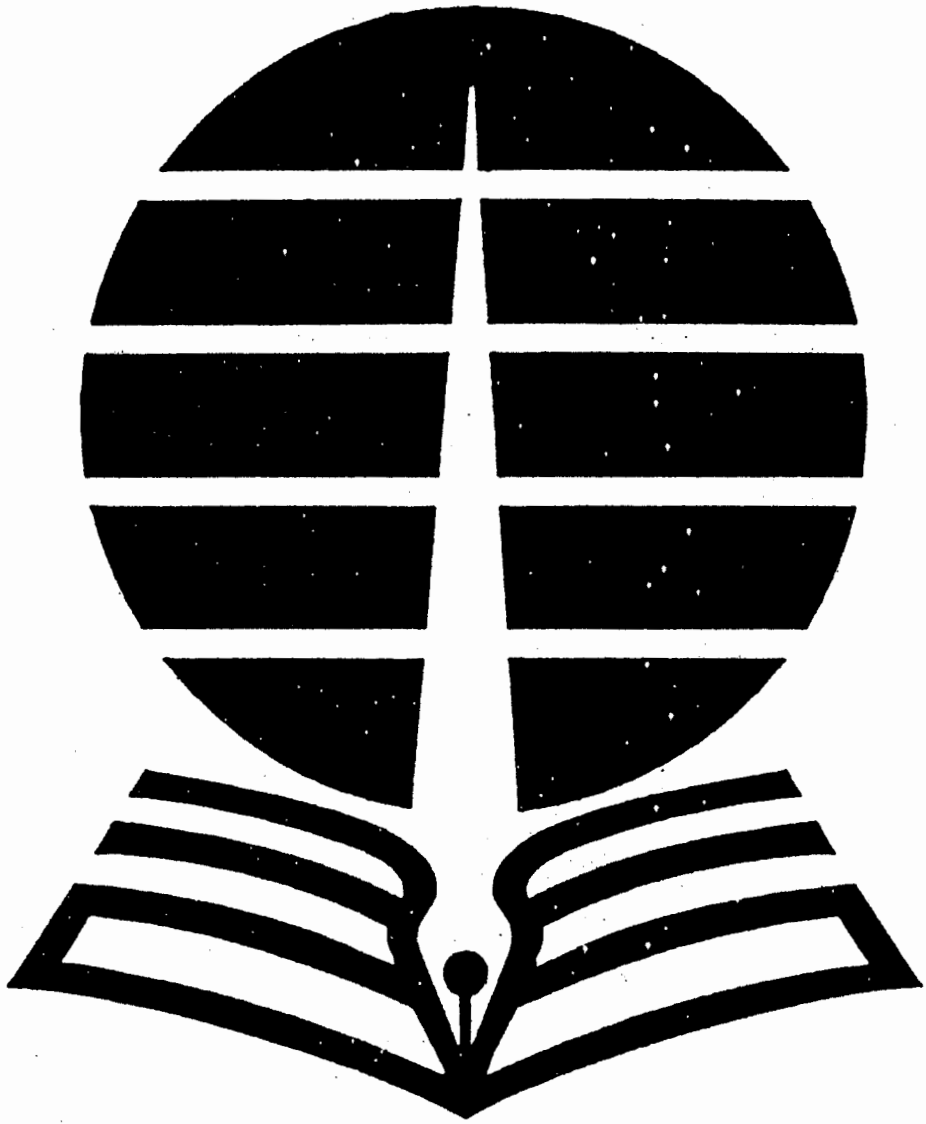
ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sistem ujian berbasis komputer yang diaplikasikan Universitas Terbuka (*Computer-Based Test* = CBT) atau Sistem Ujian Online (SUO) ke sistem *Computerized Adaptive Testing* (CAT).

Prosedur pengembangan CAT pada penelitian ini dilakukan empat tahapan. Tahap pertama adalah pemilihan matakuliah yang di-CAT-kan. Tahap kedua adalah membangun bank soal untuk keperluan CAT, tahap ketiga adalah melakukan simulasi algoritma CAT, dan tahap keempat adalah membangun program aplikasi CAT. Pada tahap pertama, bank soal untuk keperluan CAT diperoleh dari butir soal yang telah diujikan dengan *paper and pencil test* (P&P test) selama 13 semester. Kalibrasi butir soal dilakukan menggunakan model pengukuran *item respons theory*. Pada tahap simulasi, tiga desain algoritma dikembangkan yaitu desain CBT, CAT murni dan CAT yang dikendala modul atau Constrained-CAT (CCAT). Pada tahap ketiga program aplikasi CAT murni dan CCAT dikembangkan dan disesuaikan dengan aplikasi program CBT-UT.

Kalibrasi butir soal menghasilkan sebanyak 404 butir soal terpilih sebagai bank soal untuk keperluan CAT. Selanjutnya hasil simulasi menunjukkan bahwa dengan tingkat realibilitas sebesar 90% desain CAT murni hanya memerlukan panjang test rata-rata 17 butir soal untuk mengestimasi kemampuan peserta tes. Hal ini berarti CAT lebih efisien karena hanya memerlukan separoh dari jumlah butir soal pada tes konvensional dengan P&P test maupun CBT. Namun demikian, desain CCAT lebih disarankan untuk diterapkan pada sistem ujian Universitas Terbuka dibandingkan desain CAT murni karena desain ini mempertimbangkan keseimbangan isi.

Kata kunci: Computer-Based Testing, Computerized Adaptive Testing, Paper and Pencil Test.



**HALAMAN PENGESAHAN LAPORAN PENELITIAN LANJUT
BIDANG KELEMBAGAAN**

1. a. Judul Penelitian : Pengembangan Model Sistem Ujian Berdasarkan *Computerized Adaptive Testing* sebagai Upaya Efisiensi Penyelenggaraan Ujian Universitas Terbuka
b. Bidang Penelitian : Kelembagaan dan Pengembangan Sistem
c. Kategori Penelitian : Penelitian Lanjut

2. Ketua Peneliti
a. Nama Lengkap : Dr. Agus Santoso, M.Si
b. NIP : 19640217 199303 1 001
c. Pangkat dan Golongan : Penata (Gol.III/c)
d. Jabatan Fungsional : Lektor
e. Fakultas/Jurusan : FMIPA/Statistika

3. Anggota Peneliti
a. Jumlah Anggota : 3 orang
b. Nama Anggota/Unit Kerja : Dra. Dwi Astuti Aprijani, M.Kom./Puskom Unggul Utan Sufandi, S.Kom, M.Si/Puskom Iswaya Maalik, S.Kom./Puskom

4. a. Periode Penelitian : 2010
b. Lama Penelitian : 8 bulan

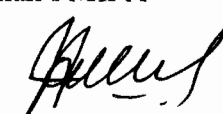
5. Biaya Penelitian : Rp. 30.000.000,- (*Tiga Puluh Juta Rupiah*)

6. Sumber Biaya : LPPM-UT

Pondok Cabe, 30 Desember 2010

Mengetahui,
Dekan FMIPA

Ketua Peneliti



Dr. Nuraini Soleiman, M.Ed.
NIP. 19540730 198601 2 001

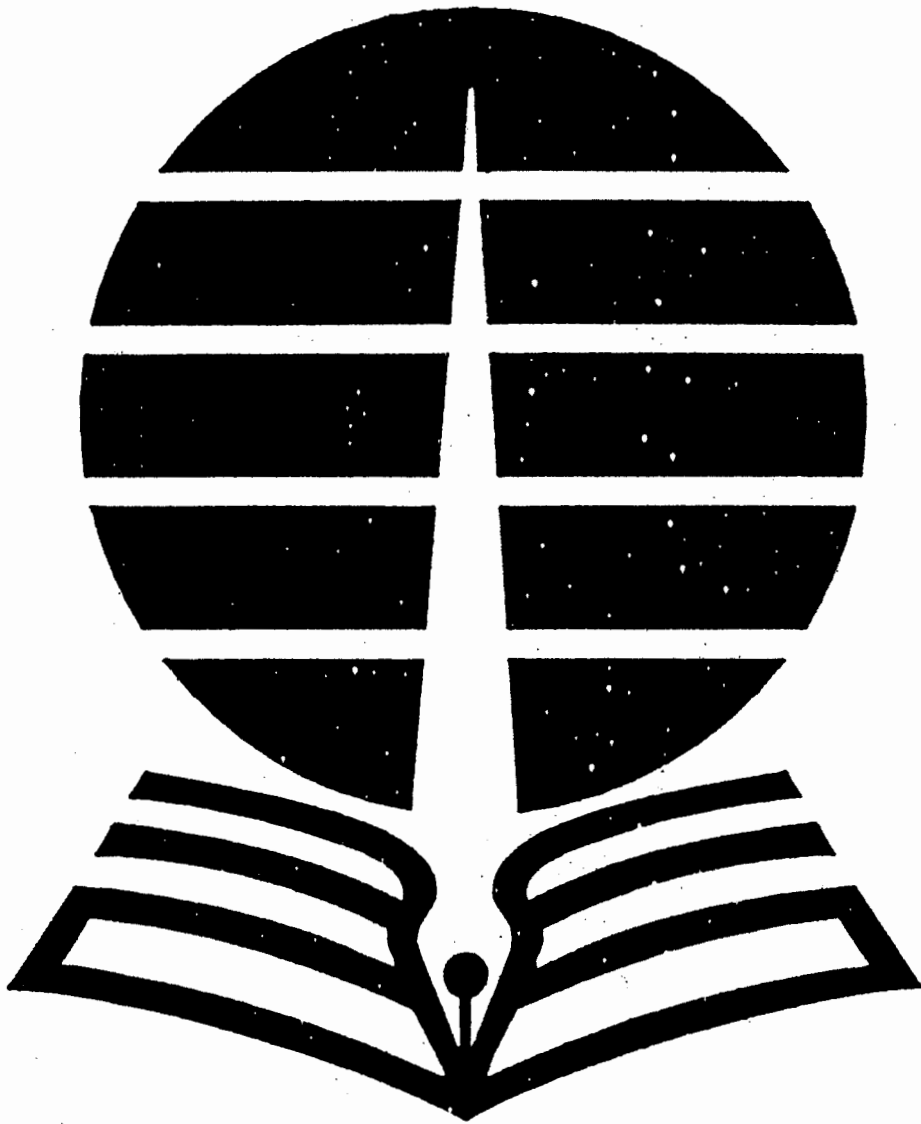

Dr. Agus Santoso, M.Si.
NIP. 19640217 199303 1 001


Menyetujui,
Kepala LPPM-UT

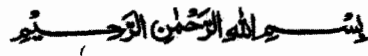

Menyetujui,
Kepala PAU-PPI-LPPM-UT


Drs. Agus Joko Purwanto, M.Si.
NIP. 19660508 199203 1 003


Dra. Trini Prastati, M.Pd.
NIP. 19660917 198601 2 001



KATA PENGANTAR



Puji syukur kami panjatkan ke hadirat Allah Swt. atas rahmat dan hidayah Nya sehingga penelitian kelembagaan ini dapat diselesaikan. Penelitian ini berjudul Pengembangan Model Sistem Ujian Berdasarkan Computerized Adaptive Testing Sebagai Upaya Efisiensi Penyelenggaraan Ujian Universitas Terbuka.

Dalam kesempatan ini, kami menyampaikan ucapan terima kasih yang tak terhingga kepada Ibu Prof. Ir. Tian Belawati, Ph.D., selaku Rektor Universitas Terbuka dan Bapak Drs. Agus Joko Purwanto, M.Si., selaku ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Terbuka yang telah memberikan kesempatan kepada kami untuk melakukan penelitian ini. Ucapan terima kasih juga kami sampaikan kepada Ibu Dr. Tri Yuni Hewindati selaku Pembantu Rektor I Universitas Terbuka dan Ibu Dr. Suciati selaku reviewer yang telah memberikan masukan dan saran perbaikan pada proposal dan laporan penelitian ini.

Kami juga mengucapkan terima kasih kepada: Ibu Irma Adnan, MA., selaku Kepala Pusat Pengujian Universitas Terbuka, dan Ibu Dr. Lia Amalia, selaku koordinator Bank Soal pada Pusat Pengujian Universitas Terbuka beserta staf, serta Bapak Ir. Edward Zubir, M.A., selaku Kepala Pusat Komputer beserta staf yang telah mengizinkan penulis untuk memperoleh informasi/data dan mengizinkan penulis mengembangkan program aplikasi *Computerized Adaptive Testing* untuk keperluan penelitian ini; Dekan Fakultas Matematika dan Ilmu

Pengetahuan Alam, Pembantu Dekan I Fakultas Matematika dan Ilmu Pengetahuan Alam, dan Ketua Jurusan Statistika dan rekan-rekan seprofesi yang telah memberikan dukungan sehingga penelitian ini dapat terselesaikan.

Akhirnya kami berharap mudah-mudahan penelitian ini memberi manfaat dan juga dapat memberikan andil dalam peningkatan kualitas pengujian di Universitas Terbuka pada khususnya dan di lembaga pendidikan lain di Indonesia pada umumnya.

Pondok Cabe, 27 Desember 2010

Tim Peneliti:

Agus Santoso (Ketua)

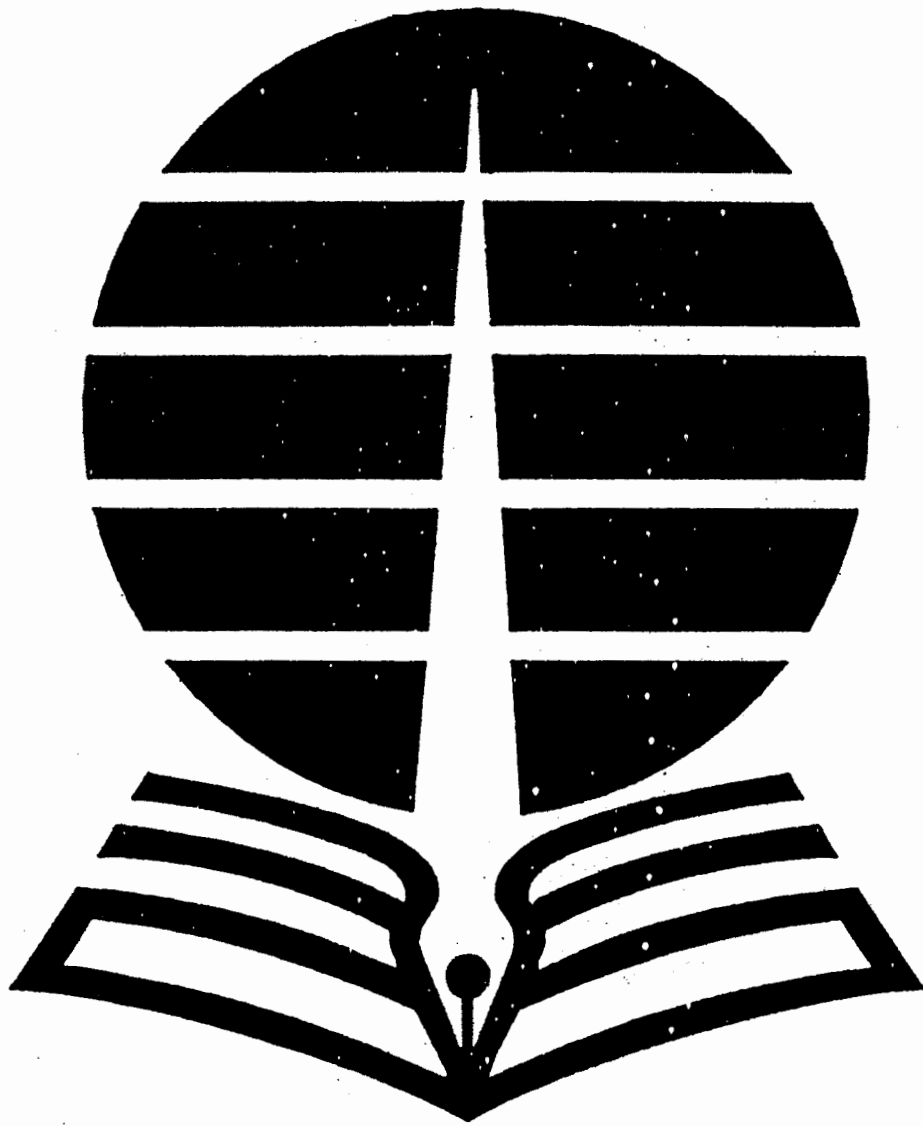
Dwi Astuti Aprijani (Anggota)

Unggul Utan Sufandi (Anggota)

Iswaya Maalik (Anggota)

DAFTAR SINGKATAN

1. UT = UNIVERSITAS TERBUKA
2. CAT = COMPUTERIZED ADAPTIVE TESTING
3. UAS = UJIAN AKHIR SEMESTER
4. IRT = ITEM RESPONSE THEORY
5. CCAT = CONSTRAINED-COMPUTERIZED ADAPTIVE TESTING
6. CBT = COMPUTER-BASED TESTING
7. P&P TEST = PAPER AND PENCIL TEST
8. MLE = MAXIMUM LIKELIHOOD ESTIMATION
9. BS = BANK SOAL
10. NIS = NOMOR INDUK SOAL



DAFTAR ISI

	Halaman
ABSTRAK	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
DAFTAR SINGKATAN	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	5
1.3. Tujuan Pengembangan	6
1.4. Pentingnya Pengembangan	6
1.5. Definisi Istilah	7
BAB II KAJIAN PUSTAKA	10
2.1. Penyelenggaraan Tes	10
1. <i>Paper and Pencil Testing</i>	10
2. <i>Computerized Testing</i>	10
3. <i>Computerized Adaptive Testing</i>	11
2.2. Komponen-Komponen CAT	13
1. Model IRT	14
2. Bank Soal	18
2.3. Pengujian Algoritma CAT	19
1. Memulai CAT (<i>Starting Point</i>)	20
2. Proses Melanjutkan (<i>Continue Process</i>)	21
3. Aturan Pemberhentian (<i>Stopping Rule</i>)	24
2.4. Algoritma CAT murni	24
2.5. Algoritma CCAT (CAT-Dikendala Modul)	25
BAB III METODOLOGI PENELITIAN	26
3.1. Tempat dan Waktu Penelitian	26
3.2. Pengembangan Model dan Prosedur Pengembangan	26
1. Tahap 1. Identifikasi Mata Kuliah	26
2. Tahap 2. Membangun Bank Soal	27
3. Tahap 3. Simulasi Algoritma CAT	27

	1. Prosedur Simulasi	28
	1.1. Desain CAT Murni	28
	1.2. Desain CCAT	29
	1.3. Desain CBT	29
	2. Kriteria Efisiensi dan Akurasi	30
	2.1. Persentase Jawaban Benar	31
	2.2. B i a s	31
	2.3. Kesalahan Baku Pengukuran	31
	4. Tahap 4. Membuat Program CAT pada sistem UAS UT	32
BAB IV	HASIL DAN PEMBAHASAN	33
	4.1. Hasil Identifikasi Mata Kuliah dan Pembentukan Bank Soal	33
	4.2. Deskripsi Bank Soal	36
	4.3. Hasil Simulasi	38
	1. Desain CAT Murni	38
	2. Desain CCAT	42
	3. Persentase Jawaban Benar	43
	4. B i a s	44
	5. Kesalahan Pengukuran	45
	4.4. Pembahasan	46
	4.5. Program Aplikasi Sistem Ujian	50
	1. Tampilan Awal Aplikasi	50
	2. Tampilan Login	51
	3. Tampilan Aplikasi CAT	52
	4. Tampilan Akhir Tes	53
BAB V	SIMPULAN DAN SARAN	54
	5.1. Simpulan	54
	5.2. Implikasi	55
	5.3. Keterbatasan Penelitian	56
	5.4. Saran	57
	5.5. Rekomendasi	58
	DAFTAR PUSTAKA	59
	LAMPIRAN-LAMPIRAN	63
	PERSONALIA PENELITIAN	86

DAFTAR TABEL

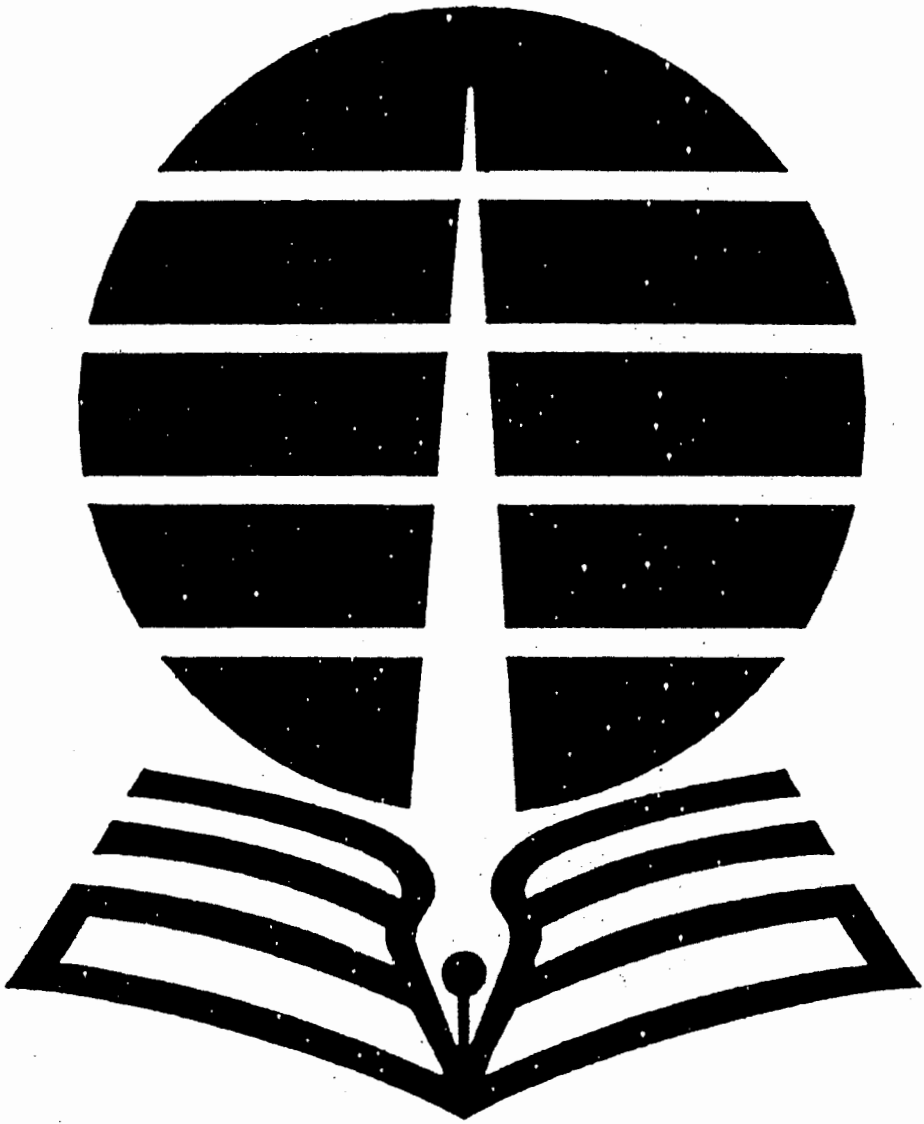
	Halaman
Tabel 1 Banyaknya Peserta Ujian Matakuliah Pengantar Statistik Sosial	33
Tabel 2 Panjang Tes Matakuliah Pengantar Statistik Sosial	34
Tabel 3 Ringkasan Statistik Tingkat Kesukaran Butir Soal pada Bank Soal	36
Tabel 4 Banyaknya dan Persentase Butir Soal per Modul pada Bank Soal dan Perangkat P&P <i>test</i>	37
Tabel 5 Nomor Induk Soal, Pola Respons, Estimasi Theta, SEM, dan Nilai Fungsi Informasi	38
Tabel 6 Banyaknya Butir Soal Yang Diperlukan pada 21 Tingkatan Theta Yang Disimulasikan	41
Tabel 7 Tingkat Kesukaran Butir Soal pada Bank Soal untuk Keperluan CAT	67

DAFTAR GAMBAR

	Halaman
Gambar 1	Proses <i>Adaptive Testing</i> 12
Gambar 2	Estimasi Kemampuan pada CAT 13
Gambar 3	Bagan Alur Pengujian Algoritma CAT 20
Gambar 4	Kesalahan Baku Estimasi Desain CCAT 42
Gambar 5	Persentase Jawaban Benar 43
Gambar 6	B i a s 44
Gambar 7	Kesalahan Pengukuran 45
Gambar 8	Kesesuaian Tingkat Kesukaran Butir Soal dengan Kemampuan 49
Gambar 9	Tampilan Awal Aplikasi 50
Gambar 10	Tampilan Login 51
Gambar 11	Tampilan Aplikasi CAT 52
Gambar 12	Tampilan Akhir Tes 53

DAFTAR LAMPIRAN

		Halaman
Lampiran 1	Contoh Output Hasil Kalibrasi Butir Soal	63
Lampiran 2	Contoh Output Kecocokan Butir (<i>item fit</i>) dengan Model IRT 1P Yang Diterapkan	65
Lampiran 3	Contoh Output Map Kedudukan Tingkat Kemampuan Peserta dengan Tingkat Kesukaran Butir Sola	66
Lampiran 4	Tabel 7. Tingkat Kesukaran Butir Soal pada Bank Soal Untuk Keperluan CAT	67
Lampiran 5	Hasil Simulasi Desain CAT Murni	79
Lampiran 6	Hasil Simulasi Desain CCAT	80
Lampiran 7	Hasil Simulasi Desain CBT	81
Lampiran 8	<i>Script</i> Program SAS	82



BAB I PENDAHULUAN

1.1. Latar Belakang Masalah

Universitas Terbuka (UT) telah menerapkan sistem ujian berbasis komputer (*Computer-Based Testing* = CBT) pada sistem Ujian Akhir Semester (UAS) sejak tahun 2006, di samping menggunakan *paper and pencil test* (P&P *test*) yang selama ini telah diselenggarakan. CBT dikembangkan berdasarkan pemanfaatan teknologi internet, dengan mempertimbangkan sarana komputer di UPBJJ-UT (Unit Program Belajar Jarak Jauh Universitas Terbuka) yang tersebar di 37 lokasi di Indonesia.

Pada P&P testing maupun CBT, setiap peserta mengerjakan sejumlah butir soal tertentu dengan jumlah butir soal adalah tetap. Penyelenggaraan tes yang memberikan sejumlah butir soal yang sama pada setiap peserta tes seperti pada sistem UAS-UT dengan CBT maupun P&P *test* kurang efisien, khususnya untuk peserta tes dengan kemampuan rendah dan tinggi (Lord, 1980: 150; Hambleton, Swaminathan, & Rogers, 1991: 145). Hal ini karena beberapa bahkan banyak butir soal yang tidak mampu memberikan informasi berguna dalam membedakan peserta tes dalam rentang kemampuan tertentu. Peserta tes yang memiliki kemampuan tinggi mendapat beberapa butir soal yang mudah, dimana mereka memiliki peluang yang kecil menjawab salah. Dengan demikian, butir soal seperti itu tidak menyediakan informasi tentang kemampuan mereka. Sebaliknya, peserta tes dengan kemampuan rendah akan mendapatkan beberapa butir soal yang sukar, dimana mereka memiliki peluang yang kecil untuk menjawab butir soal dengan

benar, akibatnya jawaban mereka yang salah hanya memberikan sedikit informasi mengenai kemampuan mereka. Pemberian sejumlah butir soal yang sama kepada peserta tes juga mengakibatkan tes tidak efektif, hal ini karena tes yang memberikan butir soal yang sama kepada semua peserta mengabaikan keragaman kemampuan peserta tes, padahal sesungguhnya kemampuan peserta tes beragam. Kondisi ini sangat mungkin terjadi pada pelaksanaan UAS-UT mengingat karakteristik mahasiswa UT yang sangat beragam.

Sejak masa ujian 2010, CBT diselenggarakan secara *online* melalui administrator di UPBJJ setelah mahasiswa melakukan pendaftaran ujian dan melakukan konfirmasi terlebih dahulu ke Pusat Pengujian UT. Selanjutnya mahasiswa mengerjakan soal ujian yang diberikan di depan komputer. Bentuk tes, komposisi soal, dan ragam soal pada perangkat tes CBT maupun P&P *test* pada sistem UAS UT sesungguhnya sama, perbedaannya hanya terletak pada media yang digunakan; pada P&P *test* menggunakan form kertas, sedangkan CBT menggunakan (*screen*) komputer.

Untuk keperluan UAS (P&P *test* maupun CBT), setiap matakuliah disyaratkan memiliki tidak kurang dari 10 set soal paralel per matakuliah untuk disimpan dalam (*data base*) bank soal UT. Butir-butir soal yang tersimpan dalam bank soal umumnya berasal dari butir soal yang telah diujikan. Butir-butir soal yang berkualitas baik dan sesuai kisi-kisi dipilih, sedangkan yang kurang baik didrop atau tidak dimasukkan dalam bank soal. Kriteria kualitas butir soal didasarkan pada pendekatan pengukuran klasik.

Penerapan pendekatan pengukuran klasik mengandung kelemahan berupa *group-dependent* dan *test-dependent* (Hambleton, Swaminathan, & Rogers, 1991: 2-5). Akibat dari kelemahan yang ada pada pengukuran klasik, maka perbandingan antara butir soal pada perangkat tes berbeda tidak dapat dilakukan. Hal ini disebabkan perbandingan antara butir soal pada perangkat tes atau beberapa tes tersebut dipengaruhi oleh kelompok yang mengerjakan tes. Karena tidak dapat melakukan perbandingan antara butir soal maka kualitas butir soal pada bank soal UT sesungguhnya sulit ditentukan.

Sebelum penyelenggaraan UAS, staf pusat pengujian bersama dengan penulis soal atau pengampu matakuliah melakukan perakitan soal. Pemilihan butir-butir soal untuk P&P *test* dipilih dari 10 set soal paralel, sedangkan untuk perangkat tes CBT dipilih secara acak hanya dari 3 set soal paralel. Perakitan butir-butir soal harus memenuhi kriteria cakupan materi dalam kisi-kisi dan sebaran tingkat kesukaran. Untuk memperoleh hasil rakitan perangkat tes yang memenuhi kriteria tersebut ternyata tidak mudah, apalagi dikaitkan dengan komposisi dan ragam soal yang diinginkan dalam satu perangkat tes.

Di Amerika, sejak tahun 1980-an mulai dikembangkan sistem pengujian menggunakan komputer berdasarkan rancangan *adaptive test* yang populer disebut *Computerized Adaptive Testing* (Weiss, 1983; Wainer, 1990; van der Linden & Glas, 2000). *Computerized Adaptive Testing* (CAT) berdasarkan *Item Response Theory*, IRT (Lord, 1980; Lord & Novick, 1968; Hambleton, Swaminathan, & Rogers, 1991). CAT versi *ASVAB=Armed Services Vocational Aptitude Battery* (Wainer et al., 1990), CAT versi *GRE = Graduated Record*

Examination (Mills, 1999), CAT untuk mengases kemampuan di bidang musik (Vispoel, 1999) merupakan tiga contoh penerapan CAT pada penyelenggaraan tes adaptif, tentu masih banyak versi CAT yang lain.

Tidak seperti pada CBT, dimana komputer hanya dimanfaatkan untuk memindahkan tes dari form kertas ke monitor komputer. Pada CAT, komputer diatur dan dimanfaatkan sedemikian hingga dapat menampilkan butir-butir soal yang sesuai dengan kemampuan individu peserta tes, karena tes disesuaikan dengan kemampuan peserta tes maka tidak ada pertanyaan yang lebih sukar ataupun lebih mudah; peserta yang berkemampuan tinggi hanya akan diberi butir-butir soal dengan tingkat kesukaran yang tinggi, sedangkan butir-butir soal yang mudah tidak perlu diberikan, begitu sebaliknya, bagi peserta yang berkemampuan rendah hanya akan diberi butir-butir soal dengan tingkat kesukaran rendah, tidak perlu diberikan butir soal yang sukar. Dengan demikian CAT lebih efisien, disamping itu CAT juga lebih reliabel, karena kesalahan pengukuran akan lebih kecil karena setiap individu hanya mendapatkan butir soal yang sesuai dengan kemampuannya.

Kelebihan lain dari CAT adalah lebih fleksibel dalam pelaksanaan, keamanan tes pun lebih terkendali, tidak memerlukan proses perakitan soal, artinya ketika *item bank* (bank soal) untuk keperluan CAT sudah terbentuk maka penerapan CAT sudah siap dimplementasikan tidak perlu ada proses perakitan sebelumnya. Masalahnya adalah bagaimana membangun suatu bank soal yang menjamin bahwa seluruh kemampuan peserta tes dapat diukur melalui butir-butir soal yang ada pada bank soal.

Tidak hanya bank soal yang dibutuhkan untuk pengembangan CAT. Menurut Wainer, et al. (1990) secara umum sistem CAT memiliki empat komponen, yaitu: bank soal, prosedur pemilihan butir soal, pendugaan tingkat kemampuan, dan aturan pemberhentian. Menurut Green, et al. (1984) dan Kingsbury & Zara (1989) pengembangan CAT memerlukan evaluasi pada enam komponen: 1) model respons butir, 2) bank soal, 3) pemilihan butir soal awal, 4) metode pendugaan kemampuan, 5) prosedur pemilihan butir soal berikutnya, dan 6) aturan pemberhentian. Masalahnya adalah bagaimana mengembangkan algoritma CAT untuk dapat diaplikasikan pada sistem UAS UT karena dari beberapa penelitian membuktikan bahwa keputusan untuk menentukan pilihan kriteria dari setiap komponen CAT memiliki konsekuensi berbeda-beda (Green, et al., 1984; Dragow & Buchanan, 1999; van der Linden, 2000).

Berdasarkan uraian di atas maka perlu adanya pengembangan sistem ujian pada UAS-UT, yaitu sistem ujian dengan menggunakan CAT. Namun demikian sebelumnya perlu dikaji berbagai komponen-komponen CAT yang akan diterapkan sesuai dengan sistem penyelenggaraan UAS UT.

1.2. Perumusan Masalah

Berdasarkan latar belakang di atas dan mengacu pada TOR (*Term of Reference*) Penelitian 2010, dimana penelitian kelembagaan harus berorientasi pada upaya efisiensi penyelenggaraan ujian UT, maka penelitian ini bertujuan untuk mengembangkan *Computer-Based Testing* ke *Computerized Adaptive Testing*. Dengan demikian masalah penelitiannya adalah bagaimana membangun

CAT yang dapat diaplikasikan pada sistem UAS UT. Secara khusus masalah penelitian ini sebagai berikut.

1. Bagaimana membangun Bank Soal untuk keperluan CAT?
2. Bagaimana membangun algoritma CAT?
3. Bagaimana mengaplikasikan CAT pada sistem aplikasi UAS UT?

1.3. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan model sistem ujian berbasis komputer yang telah diaplikasikan di UT (CBT UT) ke sistem tes adaptif berbasis komputer (CAT). Secara khusus penelitian ini bertujuan untuk:

1. Menemukan prosedur untuk membangun bank soal yang valid untuk keperluan CAT.
2. Menemukan algoritma CAT yang reliabel, valid, dan efisien untuk mengukur hasil belajar mahasiswa UT.
3. Menemukan prosedur yang praktis untuk mengaplikasikan CAT pada sistem UAS UT.

1.4. Pentingnya Pengembangan

Pengembangan model sistem ujian ini penting dilakukan karena beberapa alasan, antara lain: pertama, sistem ujian yang memberikan perangkat tes dengan jumlah butir soal yang sama pada seluruh peserta tes seperti pada *P&P testing* dan CBT UT kurang efektif dan efisien. Oleh karena itu UT perlu mengembangkan sistem ujian berbasis komputer yang bersifat adaptif (CAT). Kedua, penskoran

hasil UAS UT masih menggunakan teori klasik, hal ini sulit untuk membandingkan kemampuan antar peserta tes. Dengan mengembangkan CAT yang berbasis *item respons theory* (IRT), kemampuan antar peserta tes dapat dibandingkan. Ketiga, bank soal hasil kalibrasi menggunakan IRT dapat menjamin bahwa antar butir soalnya dapat dibandingkan, dengan demikian kualitas bank soal mudah ditentukan. Keempat, evaluasi hasil belajar mahasiswa UT sangat ditentukan oleh hasil UAS mereka. Sedikitnya 70% hasil belajar mahasiswa UT ditentukan oleh nilai UAS, sisanya 30% ditentukan oleh nilai tugas-tugas lain. Oleh karena itu perlu dikembangkan sistem pengukuran hasil belajar yang valid dan reliabel. Kelima, jumlah mahasiswa yang sangat banyak dan tersebar, maka UT perlu mengembangkan sistem penyelenggaraan ujian alternatif.

1.5. Definisi Istilah

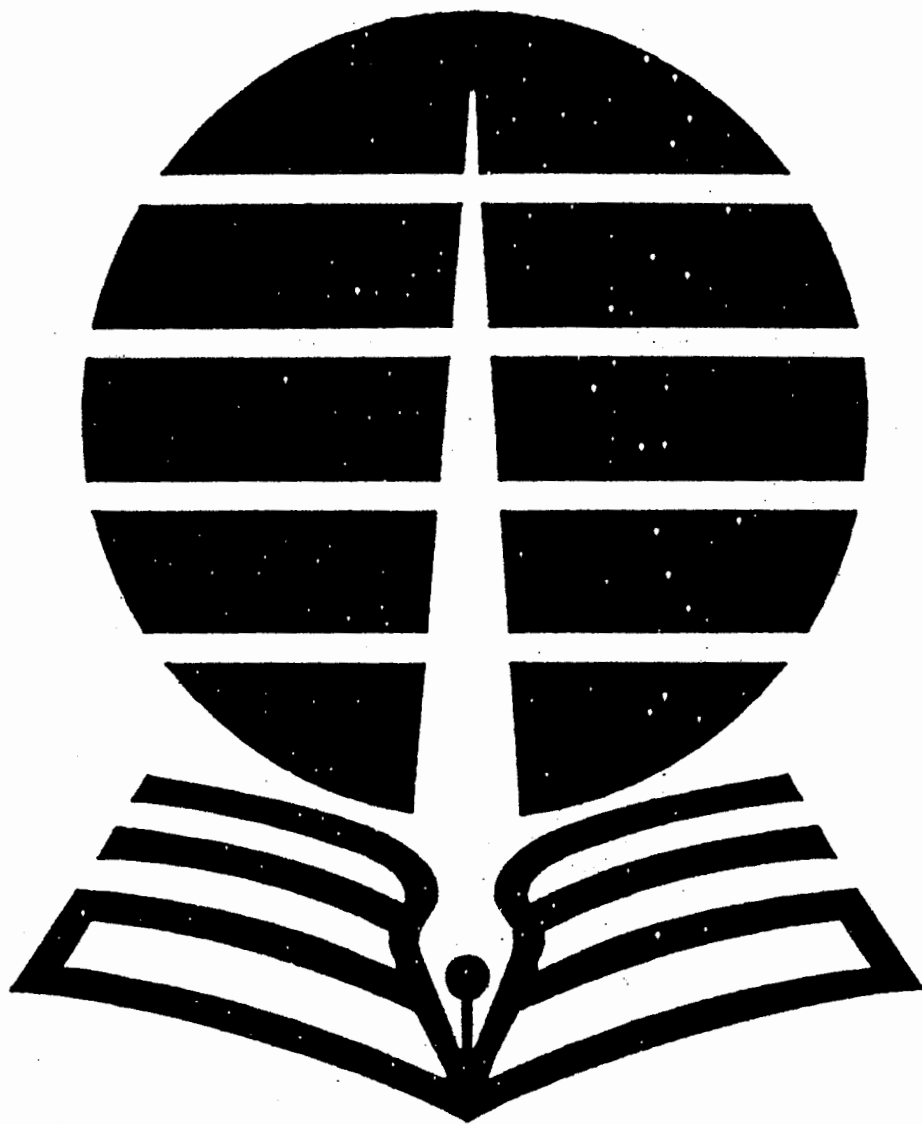
Beberapa istilah yang digunakan dalam penelitian ini, dijelaskan sebagai berikut.

1. Internet adalah singkatan dari *Interconnection Networking*, yaitu sejumlah besar *network* yang terhubung melalui protokol TCF/IP. *Network* adalah sekumpulan sistem komputer yang digandeng dan membentuk sebuah jaringan. Internet sebenarnya sebuah *network* dengan skala yang sangat besar. TCF/IP singkatan dari *Transmission Control Protocoll/Internet Protocol* dan bertugas menghubungkan jaringan komputer dan mengamati lalu lintas dalam jaringan. Protokol ini mengatur format data yang diizinkan,

penanganan kesalahan (*handling error*), lalu lintas pesan, dan standar komunikasi lainnya.

2. *Data base* adalah pangkalan data atau basis data. Basis data adalah timbunan data yang tersimpan secara elektronik dalam komputer, dipergunakan untuk berbagai keperluan melalui suatu program komputer tertentu.
3. *Online* berarti terhubung, tersambung, atau terkoneksi, sehingga sistem *online* berarti suatu sistem yang terdiri atas berbagai perangkat yang saling terhubung satu dengan lainnya.
4. Butir-butir soal paralel adalah butir-butir soal yang ditulis berdasarkan Tujuan Instruksional Khusus (TIK) yang sama. TIK tercantum dalam kisi-kisi yang merupakan representasi kompetensi matakuliah. Dengan demikian jika penulis soal atau pengampu matakuliah telah membuat atau mengembangkan sebanyak 10 butir soal untuk masing-masing TIK artinya matakuliah tersebut telah mempunyai 10 set soal paralel.
5. Pengampu matakuliah adalah staf edukatif UT yang bertanggung jawab terhadap matakuliah yang sesuai dengan kompetensi yang dimilikinya, sedangkan penulis soal adalah staf edukatif yang berasal dari UT sendiri ataupun berasal dari luar UT (*outsourcing*).
6. *Item bank* atau bank soal adalah kumpulan butir soal yang telah dikalibrasi dan telah dilengkapi dengan catatan (*record*) mengenai spesifikasi atau karakteristik butir soal.
7. Algoritma adalah urutan langkah logis untuk memecahkan suatu masalah.

8. Valid berarti absah. Bank soal yang valid adalah bank soal yang memenuhi asumsi model IRT. Asumsi model IRT adalah unidimensionalitas. Bank soal dikatakan memenuhi asumsi unidimensi jika butir-butir soal pada bank soal tersebut mengukur satu kemampuan tertentu atau mengukur secara dominan suatu kemampuan tertentu.
9. Reliabel berarti tepat. Algoritma CAT yang reliabel adalah algoritma CAT yang mampu mengukur kemampuan peserta tes dengan tepat atau algoritma yang menghasilkan kesalahan pengukuran yang kecil.
10. Efisien berarti hemat. Algoritma CAT yang efisien adalah algoritma CAT yang mampu mengukur kemampuan peserta tes hanya dengan jumlah butir soal yang lebih sedikit dibandingkan dengan jumlah butir soal pada tes konvensional (misalnya P&P *test*).
11. Praktis berarti mudah. Algoritma CAT yang valid dan praktis adalah algoritma CAT yang reliabel, efisien, dan berdasarkan pada bank soal yang valid, serta mudah diimplementasikan pada sistem aplikasi UAS UT.



BAB II

TINJAUAN PUSTAKA

2. 1. Penyelenggaraan Tes

1. *Paper and Pencil Testing*

Penyelenggaraan tes melalui *paper and pencil test (P&P test)* melibatkan penggunaan lembaran kertas. Bentuk penyelenggaraan tes ini merupakan penyelenggaraan tes konvensional karena semua penempuh tes menerima seperangkat butir soal yang sama. Kelemahan *P&P test* adalah kerahasiaan tes tidak dijamin karena dapat saja dibaca oleh orang yang tidak berwenang (Bunderson, Inouye, & Olsen, 1989). Selain itu, karena harus memberikan semua butir soal, maka diperlukan waktu penyelenggaraan yang lebih lama. Penggunaan kertas menjadi masalah tersendiri, misalnya dibutuhkan ruang untuk penyimpanan data perangkat tes.

2. *Computerized Testing*

Seiring dengan perkembangan teknologi, pengujian dengan menggunakan komputer mulai dilakukan. Awalnya, komputer hanya digunakan untuk mengotomatisasikan aktivitas pengukuran yang biasa. Tes yang semula berada di kertas dipindahkan ke dalam komputer. Penggunaan komputer seperti ini disebut *Computerized Testing* atau *Computerized-Based Testing (CBT)* dan merupakan generasi pertama penggunaan komputer untuk pengujian.

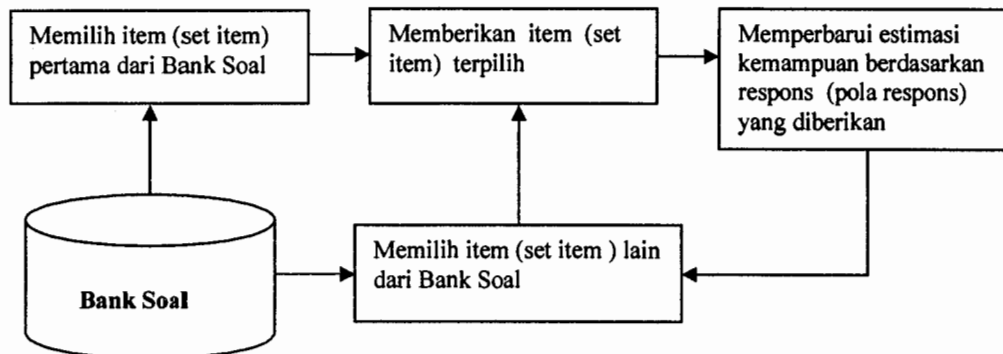
Bunderson, Inouye, & Olsen (1989 : 174) menyatakan beberapa kelebihan dari CBT, yaitu: meningkatkan standardisasi, meningkatkan keamanan tes,

meningkatkan kemampuan tampilan tes, memperkecil *error of measurement*, serta mempercepat pemberian skor dan interpretasi.

3. Computerized Adaptive Testing

Computerized Adaptive Testing (CAT) merupakan generasi kedua dari penggunaan komputer untuk pengujian (Bunderson, Inouye, & Olsen, 1989 : 383). *Adaptive* berarti bahwa butir soal yang diberikan disesuaikan dengan tingkat kemampuan setiap peserta tes atau *tailored testing* (Lord, 1980:151). Pada CAT yang berbasiskan *item response theory* (IRT), komputer tidak hanya sekedar memindahkan butir soal ke dalam komputer, tetapi komputer diatur untuk menyeleksi dan menyajikan butir soal menurut perkiraan tingkat kemampuan peserta tes. Hal ini mengakibatkan individu peserta tes yang memiliki tingkat kemampuan tinggi akan mendapatkan butir soal yang lebih sulit dibandingkan dengan individu yang memiliki tingkat kemampuan rendah. Sebaliknya individu peserta tes yang memiliki tingkat kemampuan rendah akan mendapatkan butir soal yang lebih mudah dibandingkan dengan individu peserta tes yang memiliki tingkat kemampuan tinggi. Dengan demikian CAT lebih efisien karena dapat mengestimasi kemampuan peserta tes dengan jumlah butir soal yang lebih sedikit dibandingkan P&P *test* maupun CBT tanpa mengurangi ketepatan pengukuran (Wainer, 1990:10; Hambleton, Swaminathan, & Rogers, 1991:146, Weiss & Schleisman, 1999:130).

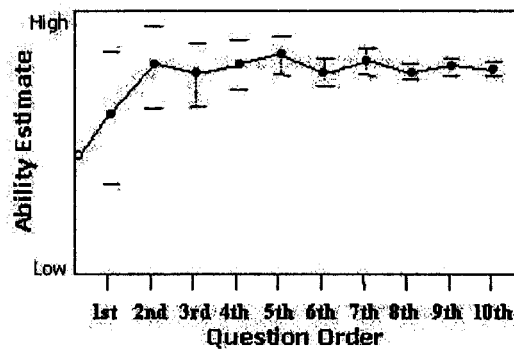
Proses *adaptive testing* secara skematik dapat dilihat pada Gambar 1.



Gambar 1. Proses *Adaptive Testing*

Berdasarkan Gambar 1, proses *adaptive testing* dimulai dengan memilih butir soal atau kelompok butir soal pertama dari bank soal. Biasanya butir soal pertama dipilih disesuaikan dengan tingkat kemampuan awal optimal dari populasi peserta tes atau dipilih dengan tingkat kesukaran sedang. Setelah butir soal atau kelompok butir soal dipilih, selanjutnya butir soal diberikan kepada peserta tes. Setelah peserta tes merespons (benar atau salah) butir soal atau kelompok butir soal pertama, kemudian tingkat kemampuan peserta diperbarui atau diestimasi kembali. Selanjutnya, berdasarkan estimasi tingkat kemampuan terbaru, butir soal atau kelompok butir soal yang lain dipilih kembali dari bank soal. Kemudian butir soal atau kelompok butir soal yang lain diberikan lagi kepada peserta tes, begitu seterusnya proses ini berlangsung dan diberhentikan setelah sebanyak butir soal yang ditentukan sudah diberikan atau setelah presisi estimasi tingkat kemampuan atau tingkat kesalahan baku pengukuran yang diinginkan telah dicapai.

Menurut Dunkel (1999), CAT adalah suatu metode penilaian secara teknologi di mana komputer menyeleksi dan menyajikan butir soal menurut perkiraan tingkat kemampuan peserta tes. Estimasi tingkat kemampuan peserta tes pada tes adaptif lebih akurat karena setiap peserta hanya diberi butir soal yang sesuai dengan kemampuannya, dengan kata lain kesalahan pengukuran (*measurement error*) akan lebih kecil. Ilustrasi estimasi tingkat kemampuan peserta tes pada CAT disajikan pada Gambar 2 berikut.



Gambar 2. Estimasi Kemampuan pada CAT (Sumber: Anonim)

Gambar 2 menunjukkan bagaimana tingkat kemampuan seorang peserta tes diestimasi lebih rendah setelah pertanyaan dijawab secara salah (pertanyaan 3, 6, 8, dan 10). Titik-titik vertikal mengindikasikan besarnya *error* dikaitkan dengan tingkat kemampuan yang diestimasi.

2.2. Komponen-Komponen CAT

Menurut Wainer (1990) secara umum sistem CAT memiliki empat komponen, yaitu: bank soal, prosedur pemilihan butir soal, pendugaan kemampuan, dan aturan pemberhentian, sedangkan dua komponen CAT yang

sering diperhatikan pada sistem CAT adalah keseimbangan konten dan kontrol butir soal yang sering muncul.

Menurut Green, et al. (1984) dan Kingsbury & Zara (1989) pengembangan CAT memerlukan evaluasi pada enam komponen berikut:

- 1) model respons butir
- 2) bank soal
- 3) pemilihan butir soal awal
- 4) metode pendugaan tingkat kemampuan
- 5) prosedur pemilihan butir soal
- 6) aturan pemberhentian.

1. Model Item Response Theory

Model *item response theory* (IRT) menggambarkan peluang menjawab butir soal secara benar berdasarkan tingkat kemampuan peserta tes dan butir soal yang diberikan. Dalam pendekatan IRT kemampuan individu atau *proficiency* (disimbolkan dengan θ) dan tingkat kesukaran butir atau *difficulty* (disimbolkan dengan b) berada pada satu dimensi yang sama (Lord, 1980:12; Hambleton, Swaminathan, & Rogers, 1991: 13; Embretson & Rouse, 2000 : 66).

Model IRT memiliki asumsi mengenai data dimana model ini diterapkan. Hambleton & Swaminathan (1985: 16) dan Hambleton, Swaminathan, & Rogers, (1991: 9) menyatakan bahwa ada tiga asumsi yang mendasari teori respons butir, yaitu unidimensi, independensi lokal, dan kecocokan spesifikasi model.

Tiga model IRT yang umum digunakan untuk butir-butir soal dengan format pilihan ganda adalah model-logistik 1 parameter (1P), 2P, dan 3P. Model

1P atau model Rasch merupakan model IRT yang paling sederhana (Hambleton & Swaminathan, 1985: 39; Hambleton, Swaminathan, & Rogers, 1991 : 12; Embretson & Reise, 2000 : 67). Pada model ini butir-butir soal diasumsikan tidak dapat dijawab benar dengan cara menebak dan mempunyai daya beda yang sama tetapi setiap butir soal mempunyai tingkat kesukaran (dinyatakan dengan b) bervariasi. Parameter b mengacu pada titik pada skala kemampuan (*ability*) dimana seorang peserta mempunyai peluang 50% menjawab butir soal dengan benar. Semakin besar b semakin sulit butir soal itu. Ketika butir-butir soal diasumsikan mempunyai parameter daya beda (dinyatakan dengan a) yang bervariasi maka model 2P lebih cocok. Nilai a tinggi menunjukkan bahwa kurva karakteristik butir (*item characteristic curve = ICC*) sangat tajam. *ICC* merupakan kurva monoton naik, semakin meningkat tingkat kemampuan peserta tes semakin besar peluang untuk menjawab benar suatu *item* (Hambleton & Swaminathan, 1985:39; Embretson & Reise, 2000 : 58). Nilai a yang rendah menunjukkan bahwa butir soal mempunyai kurva dengan kenaikan yang gradual pada skala kemampuan. Dengan demikian butir soal dengan nilai a yang tinggi lebih dapat membedakan peserta tes kedalam kelompok kemampuan yang berbeda dibandingkan dengan nilai a yang rendah. Model 3P digunakan ketika parameter tebakan semu (*pseudo-guessing*, dinyatakan dengan c) diasumsikan ada dalam model. Parameter ini merepresentasikan peluang seorang peserta tes dengan kemampuan sangat rendah menjawab butir soal dengan benar.

Model IRT dan sifat-sifatnya secara singkat dapat dijelaskan sebagai berikut. Misalkan seorang peserta tes dengan tingkat kemampuan θ menjawab

butir soal pilihan ganda i (U_i), maka peluang peserta tes menjawab butir soal dengan benar atau salah digambarkan sebagai fungsi kepadatan peluang (f.k.p) sebagai berikut:

$$P(U_i = u_i | \theta) = P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i}, \quad u_i = 0, 1, \quad -\infty < \theta < \infty, \quad (1)$$

sedangkan $u_i = 1$ jika jawaban benar, $u_i = 0$ jika jawaban salah, dan $Q(\theta) = 1 - P(\theta)$. Fungsi di atas menyatakan sebaran Bernoulli dengan parameter θ . Selanjutnya dengan menerapkan asumsi independensi lokal, f.k.p dari n butir soal dapat dihitung sebagai hasil perkalian dari f.k.p. setiap butir soal:

$$\begin{aligned} P(U_1 = u_1, U_2 = u_2, \dots, U_n = u_n | \theta) &= (P_1(\theta)^{u_1} Q_1(\theta)^{1-u_1}) \dots (P_n(\theta)^{u_n} Q_n(\theta)^{1-u_n}) \\ &= \prod_{i=1}^n P_i^{u_i} Q_i^{1-u_i} \end{aligned} \quad (2)$$

Untuk model 3P, $P_i(\theta)$ dihitung dengan

$$P(\theta) = c + (1-c) \frac{1}{1 + e^{-1,7a_i(\theta-b_i)}} \quad (3)$$

Sebagai contoh, jika $P_1=0,2$, $P_2=0,7$, $P_3=0,7$, dan pola jawaban peserta pada ketiga butir soal itu adalah $(1,1,0)$, artinya $u_1=1$, $u_2=1$, dan $u_3=0$, maka peluang peserta menjawab dengan pola jawaban seperti itu adalah $0,2 \times 0,7 \times (1-0,7)$ atau $0,042$.

Ketika spesifikasi model cocok dengan data tes maka dua sifat yang diinginkan dari IRT, yaitu sifat invariansi parameter butir dan parameter kemampuan dapat diperoleh. Kedua sifat IRT ini sangat diinginkan untuk tes adaptif yang akan memberikan butir soal yang berbeda untuk peserta tes yang berbeda pula.

Fungsi Informasi Butir dan Tes

Nilai fungsi informasi butir menggambarkan seberapa akurat suatu butir soal dapat mengestimasi tingkatan kemampuan peserta tes. Dengan menggunakan fungsi informasi, ketepatan pengukuran pada pengestimasian kemampuan peserta dapat dihitung pada setiap tingkat kemampuan.

Secara matematis, fungsi informasi butir dituliskan sebagai berikut:

$$I_i(\theta) = E \left\{ \left[\frac{\partial \ln f(U_i; \theta)}{\partial \theta} \right]^2 \right\} = \frac{[P'_i(\theta)]^2}{[P_i(\theta)][Q_i(\theta)]} \quad (4)$$

Fungsi informasi butir dinyatakan oleh Birnbaum (Hambleton, Swaminathan, & Rogers, 1991: 91) dalam persamaan berikut.

$$I_i(\theta) = \frac{2,89a_i^2(1-c_i)}{[c_i + \exp(1,7a_i(\theta - b_i))] [1 + \exp(-1,7a_i(\theta - b_i))]^2} \quad (5)$$

Persamaan (5) menunjukkan bahwa nilai informasi hanya tergantung pada parameter butir (misalnya; a , b , dan c untuk model 3P) dan tingkat kemampuan, θ . Dengan demikian untuk setiap tingkat kemampuan, θ , kontribusi informasi untuk setiap butir pada bank soal dapat dihitung.

Fungsi informasi tes merupakan jumlah dari fungsi informasi butir penyusun tes tersebut (Hambleton & Swaminathan, 1985: 94). Fungsi informasi perangkat tes secara matematis dapat dituliskan sebagai berikut.

$$I_t(\theta) = \sum_{i=1}^n I_i(\theta) \quad (6)$$

Seperti fungsi informasi butir, fungsi informasi tes menggambarkan seberapa akurat perangkat tes mengestimasi tingkat kemampuan yang berbeda.

Semakin besar informasi pada tingkat kemampuan yang diberikan semakin akurat kemampuan itu diestimasi dari perangkat tes itu.

Standard Error of Measurement (SEM)

Kesalahan baku pengukuran (*Standard Error of Measurement, SEM*) berkaitan erat dengan fungsi informasi. Fungsi informasi dengan *SEM* mempunyai hubungan yang berbanding terbalik kuadratik, semakin besar fungsi informasi maka *SEM* semakin kecil atau sebaliknya. Hubungan keduanya, menurut Hambleton, Swaminathan, & Rogers (1991 : 94) dinyatakan dengan

$$SEM(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \quad (7)$$

Semakin besar informasi pada tingkat kemampuan yang diberikan semakin akurat kemampuan itu diestimasi dari perangkat tes itu. Weiss & McBride (1984) menyatakan bahwa tes yang ideal adalah tes yang mempunyai kurva fungsi informasi tes yang tinggi dan melebar.

2. Bank Soal

CAT mengambil butir soal dari bank soal untuk diberikan kepada peserta tes. Ketersediaan butir-butir soal yang cukup dengan kualitas butir-butir soal yang baik pada bank soal sangat menentukan kualitas CAT.

Menurut Wang & Vispoel (1998), ada tiga komponen yang memberikan kontribusi terhadap kualitas bank soal, yaitu: ukuran bank soal, parameter butir soal, dan struktur isi. Ukuran bank soal minimal dipengaruhi oleh panjang tes dan ukuran peserta tes. Way (1997) menyarankan rasio 1 berbanding 6 – 8 untuk panjang tes dan banyaknya butir soal minimal yang harus ada dalam bank soal,

artinya jika panjang tes CAT dirancang sebanyak 20 butir soal maka banyaknya butir yang harus tersedia pada bank soal minimal sebanyak 120 sampai 160 butir. Stahl & Lunz (1993) menyarankan sedikitnya 500 butir untuk jumlah peserta tes sebanyak 1000 peserta, dan lebih baik lagi jika bank soal berisi 600 sampai 800 butir soal dengan sebaran yang merata pada setiap materi. Reckase (2003) menyarankan bank soal untuk keperluan CAT minimalnya 200 untuk jumlah peserta tes sebanyak 100 orang. Kemudian Flaughner (1990 : 42) menyarankan sebanyak 200 butir soal sudah cukup untuk keperluan CAT, sedangkan Bunderson, Inouye, & Olsen (1989 : 301) dan Gershon (2005: 113) menyarankan ukuran bank soal untuk keperluan CAT sedikitnya 100 butir soal.

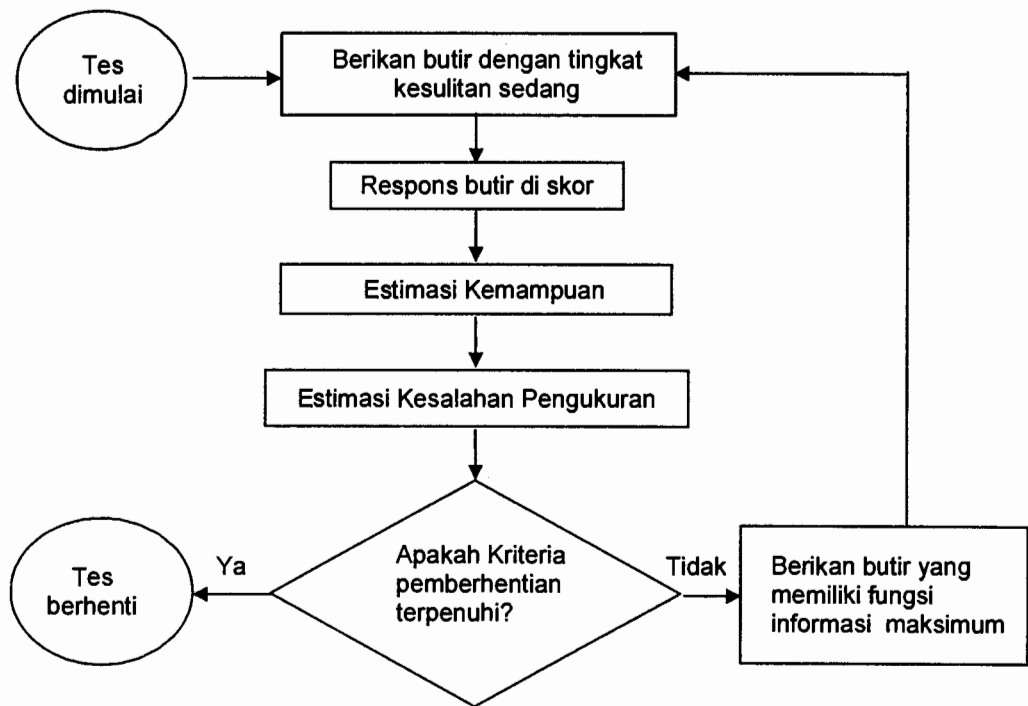
Ansley & Forsyth (1985) dan Folk & Green (1989) menyarankan bank soal untuk keperluan CAT sebaiknya berisi butir-butir dengan: daya beda bersebaran seragam antara 0,4 sampai 2,0, tingkat kesukaran bersebaran seragam antara -2,0 sampai 2,0, dan faktor *guessing* bersebaran seragam antara 0 sampai 0,3.

2. 3. Pengujian Algoritma CAT

Thissen & Mislevy (1990: 103-135) menjelaskan langkah-langkah yang lazim digunakan untuk mengembangkan algoritma CAT sebagai berikut:

- 1) Bagaimana memulai: Butir soal pertama apa yang diberikan pada peserta tes?
- 2) Bagaimana melanjutkan: Setelah ada jawaban, butir soal berikutnya yang bagaimana yang diberikan pada peserta?
- 3) Bagaimana mengakhiri: Kapan tes dihentikan?

Berikut adalah bagan alur pengujian algoritma:



Gambar 3. Bagan Alur Pengujian Algoritma CAT

1. Memulai CAT (*Starting Point*)

Ketika CAT dimulai, belum ada butir soal yang diberikan pada peserta tes, belum ada respons yang diberikan oleh peserta tes sehingga tingkat kemampuan peserta belum dapat diestimasi. Walaupun belum ada informasi mengenai kemampuan peserta sebelumnya, penyelenggaraan CAT harus dimulai. Jika tidak ada informasi awal mengenai kemampuan peserta tes, maka CAT dapat dimulai dengan memilih butir soal awal dengan tingkat kesukaran sedang (Green, et al. 1984:347; Vispoel, 1999: 156; Mills, 1999:123).

Hambleton, Zaal, & Pieters (1991: 363) menyimpulkan bahwa kemampuan awal pada tes adaptif tidak berpengaruh nyata pada estimasi kemampuan akhir jika panjang tes tidak terlalu pendek. Chen, Ankenmann, & Chang (2000) menyimpulkan bahwa setelah mencapai 10 butir soal maka estimasi kemampuan akhir peserta pada CAT tidak dipengaruhi secara signifikan oleh metode pemilihan butir awal. Lunz & Bergstrom (1994) menemukan bahwa estimasi kemampuan akhir peserta tes yang mengambil tes *American Society for Clinical Pathologies* (ASCP-CAT) tidak dipengaruhi oleh butir soal pertama mereka apakah mudah, sedang atau sukar.

2. Proses Melanjutkan (*Continue Process*)

Setelah memperoleh jawaban peserta tes terhadap butir soal yang diberikan, selanjutnya komputer menskor jawaban dengan benar atau salah, kemudian memutuskan apakah tes dilanjutkan ataukah tidak. Dua langkah untuk proses melanjutkan CAT, langkah pertama adalah mengestimasi tingkat kemampuan peserta tes; langkah kedua adalah bagaimana memilih butir soal berikutnya.

Metode Pendugaan Tingkat Kemampuan

Metode yang umum digunakan untuk mengestimasi kemampuan peserta tes adalah *Maximum Likelihood Estimation* (MLE) (Birnbaum, 1958; Baker, 1992), dan tiga metode Bayes: *Owen's Bayesian Procedure* (OWEN) (Owen, 1975), the *Expected a Posteriori Procedure* (EAP) (Bock & Aitken, 1981; Bock & Mislevy, 1982), and *Maximum a Posteriori Estimation* (MAP) (Samejima, 1969). Berikut dipaparkan secara singkat metode MLE.

Maximum Likelihood Estimation

Misalkan seorang peserta tes dengan tingkat kemampuan θ menjawab tes yang berisi n butir soal pilihan ganda dengan parameter butir soal diketahui. Peluang bersama dari peserta tes dapat dituliskan sebagai $P(U_1, U_2, \dots, U_n | \theta)$. Selanjutnya jika asumsi independensi lokal diterapkan maka fungsi kemungkinannya (*likelihood function*); $L(\theta)$, dituliskan sebagai berikut

$$L(\theta) = P(U_1 = u_1, U_2 = u_2, \dots, U_n = u_n | \theta) = \prod_{i=1}^n P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i}, \quad (8)$$
$$i = 1, 2, \dots, n, \quad -\infty < \theta < \infty$$

Tujuan MLE adalah menemukan nilai yang memaksimumkan $L(\theta)$. Nilai parameter kemampuan yang memaksimumkan fungsi kemungkinan, L disebut dengan *the maximum likelihood estimate of ability* (Hambleton, 1993). Secara matematik, hal ini sama dengan untuk menemukan nilai yang memaksimumkan nilai logaritma natural, $\ln L(\theta)$. Nilai ini dapat diperoleh dengan membuat turunan pertama dari $\ln L(\theta)$ terhadap θ sama dengan nol.

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{i=1}^n [u_i - P_i(\theta)] \frac{P_i'(\theta)}{P_i(\theta)Q_i(\theta)} = 0. \quad (9)$$

Pada praktiknya, untuk menyelesaikan sistem persamaan (9) dilakukan dengan menggunakan prosedur Newton-Raphson (Hambleton & Swaminathan, 1985:83).

Satu masalah dengan penerapan metode MLE pada tes adaptif adalah ketidakmampuan fungsi kemungkinan untuk menemukan solusi maksimum ketika peserta tes menjawab semua butir soal dengan benar atau salah.

Pemilihan Butir Soal Berikutnya

Setelah kemampuan peserta diestimasi, kemudian komputer mesti memilih butir soal berikutnya. Dua metode yang digunakan untuk memilih butir soal berikutnya, yaitu metode pemilihan dengan menggunakan informasi butir soal maksimum yang tidak dikendala dan metode informasi butir soal maksimum yang dikendala, misalnya dikendala keseimbangan konten dan atau dikendala *item exposure*.

Prosedur Informasi Maksimum

Lord (1977; 1980) menyarankan menggunakan prosedur informasi butir soal maksimum yang tidak dikendala untuk memilih butir soal berikutnya untuk diberikan pada peserta tes. Berdasarkan prosedur ini, butir soal yang mempunyai nilai fungsi informasi terbesar pada kemampuan peserta tertentu dipilih untuk diberikan pada peserta tes. Hal ini menjamin bahwa nilai fungsi informasi tes untuk setiap peserta tes adalah maksimum, artinya kesalahan baku pengukuran minimum.

Prosedur Informasi Maksimum yang Dikendala Konten

Pengembang CAT mesti memrogram komputer untuk memilih butir-butir yang menjamin bahwa semua peserta tes memperoleh butir-butir soal sesuai dengan kisi-kisi. Inilah yang disebut keseimbangan konten. Berdasarkan algoritma keseimbangan konten, butir soal berikutnya yang diberikan adalah butir soal yang mempunyai nilai fungsi informasi maksimum dan berasal dari kategori konten yang diinginkan.

3. Aturan Pemberhentian (*Stopping Rule*)

Dua metode utama yang digunakan untuk memberhentikan CAT, yaitu *equal measurement precision* dan *fixed number of items*. Kedua metode ini menghasilkan variasi kesalahan pengukuran yang berbeda. Tujuan digunakannya metode *equal measurement precision* adalah menghasilkan skor tes dengan tingkat ketepatan estimasi yang sama. Namun, panjang tes diprediksikan bervariasi dari satu peserta dengan peserta tes lainnya. Sedangkan penerapan aturan *fixed number of items* akan berakibat pada ketepatan estimasi yang tidak sama, namun demikian, kriteria ini lebih mudah diterapkan.

Terkait dengan aturan pemberhentian tes maka pada penelitian ini dikembangkan dua desain algoritma CAT, yaitu: algoritma CAT murni dan algoritma CAT yang dikendala modul disingkat *Constrained-CAT* = CCAT (Stocking & Swanson, 1993).

2. 4. Algoritma CAT Murni

Tahapan algoritma CAT murni sebagai berikut:

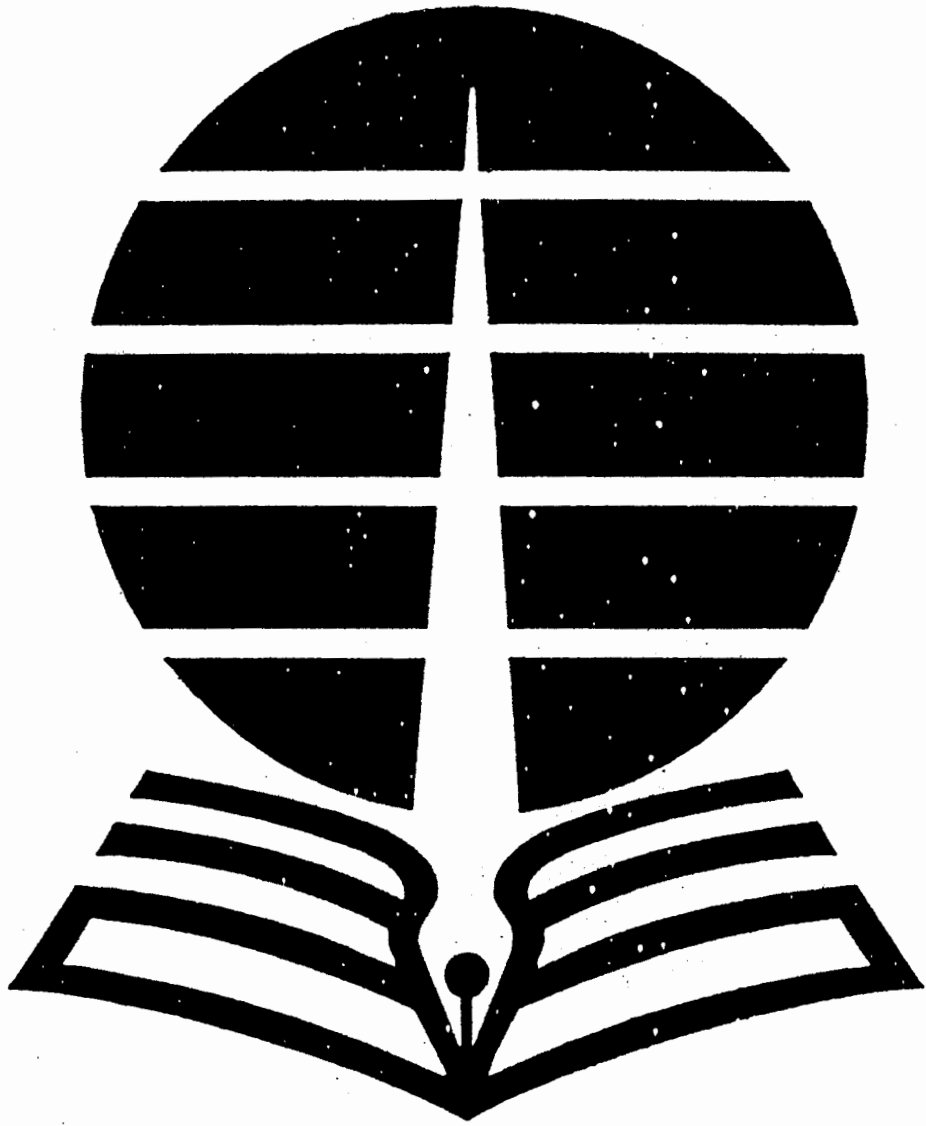
1. Dipilih butir soal awal, yaitu butir soal dengan tingkat kesukaran sedang
2. Dari pola respons peserta terhadap butir soal yang diberikan, dibentuk fungsi kemungkinan.
3. Diestimasi (sementara) tingkat kemampuan peserta
4. Dicari nilai fungsi informasi butir pada tingkat kemampuan peserta yang telah diperoleh pada tahap 3, dan dihitung pula estimasi kesalahan baku pengukurannya.

5. Dipilih lagi butir soal yang memiliki fungsi informasi tertinggi pada estimasi tingkat kemampuan peserta terbaru.
6. Ulangi langkah 2 sampai langkah 5.
7. Tes dihentikan jika estimasi kesalahan baku pengukuran sudah mencapai 0,30.
Kesalahan baku pengukuran sebesar 0,30 setara dengan reliabilitas sebesar 91% pada tes konvensional (Thissen, 1990).

2. 5. Algoritma CCAT (CAT-Dikendala Modul)

Tahapan algoritma CCAT sebagai berikut:

1. Dipilih butir soal awal, yaitu butir soal dengan tingkat kesukaran sedang dari modul 1.
2. Dari pola respons peserta terhadap butir soal yang diberikan, dibentuk fungsi kemungkinan.
3. Diestimasi (sementara) tingkat kemampuan peserta.
4. Dicari fungsi informasi butir pada estimasi tingkat kemampuan peserta yang telah diperoleh pada tahap 3, dihitung pula estimasi kesalahan baku pengukurannya.
5. Dipilih lagi butir soal yang memiliki fungsi informasi tertinggi.
6. Ulangi langkah 2 sampai 5, sampai sejumlah butir tertentu dipilih dari modul 1.
7. Ulangi langkah 2 sampai 5, untuk modul 2, modul 3, dan seterusnya sampai modul terakhir.
8. Tes dihentikan jika sejumlah butir soal yang telah ditetapkan terpenuhi.



BAB III

METODOLOGI PENELITIAN

3. 1. Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di Pusat Pengujian dan Pusat Komputer UT, Pondok Cabe, Tangerang, Banten. Penelitian ini memerlukan waktu kurang lebih delapan bulan.

3. 2. Pengembangan Model dan Prosedur Pengembangan

Jenis penelitian ini adalah penelitian pengembangan, yaitu pengembangan model pengukuran prestasi belajar mahasiswa UT. Hasil yang diharapkan dari pengembangan sistem pengujian ini adalah suatu program CAT yang dapat diimplementasikan pada sistem UAS UT.

Prosedur pengembangan dalam penelitian ini terdiri atas empat tahap: tahap pertama adalah tahap identifikasi matakuliah yang akan di-CAT-kan, tahap kedua membangun bank soal, tahap kedua simulasi algoritma CAT, dan tahap ketiga membangun program CAT pada sistem UAS UT.

1. Tahap 1. Identifikasi Matakuliah

Matakuliah yang di-CAT-kan dipilih didasarkan pada: 1) matakuliah yang pelaksanaan ujiannya dilakukan dengan P&P *test*, 2) untuk menjamin hasil estimasi parameter butir soal yang reliabel dipilih matakuliah yang diikuti oleh minimumnya 250 peserta. Setelah matakuliah ditentukan, selanjutnya diambil

data, yaitu berupa Lembar Jawaban Ujian (LJU), yang berisi respons peserta ujian terhadap perangkat tes matakuliah terpilih.

2. Tahap 2. Membangun Bank Soal

Bank Soal untuk keperluan CAT dibangun dari butir-butir soal yang telah diujikan dengan P&P *test*. Data pola respons dari 8 - 14 masa ujian dikalibrasi menggunakan pendekatan *item response theory* (IRT) menggunakan *software* QUEST (Adam & Khoo, 1996). Untuk menentukan seberapa baik model IRT *fit* dengan data tes dilakukan uji kecocokan model IRT terhadap data. Selanjutnya butir-butir soal yang *fit* dengan model IRT dipilih sebagai butir soal pada bank soal untuk keperluan CAT.

3. Tahap 3. Simulasi Algoritma CAT

Pada studi simulasi, dikembangkan dua desain algoritma CAT , yaitu desain algoritma CAT murni dan Constrained-CAT (CCAT), simulasi juga dilakukan pada desain P&P *test* (CBT). Simulasi desain CBT dilakukan untuk pembandingan terhadap desain CAT maupun CCAT. Penelitian simulasi dilakukan untuk menguji algoritma CAT yang diterapkan. Simulasi dipilih karena: 1) range tingkat kemampuan peserta dan ukuran peserta tes (*sample size*) dapat dibangkitkan sesuai yang dikehendaki, 2) rentang tingkat kesukaran butir sedang pada prosedur pemilihan butir awal dan besarnya *step size* untuk mengestimasi kemampuan awal peserta dapat dikaji, 3) simulasi dapat mengurangi biaya, karena tidak perlu menghadirkan pengawas tes dan peserta tes.

Simulasi desain CAT murni dilakukan berdasarkan bank soal empiris. Simulasi desain CAT murni dilakukan untuk mengetahui banyaknya butir soal atau panjang tes yang diperlukan untuk setiap tingkat kemampuan yang disimulasikan. Pada desain CCAT disimulasikan berdasarkan data bank soal empiris (BS-Empiris) dan diskenariokan desain CCAT, yaitu: CCAT dengan panjang tes 50% dari panjang P&P *test*.

Simulasi dilakukan menggunakan program SAS *versi* 9. Program ini dipilih karena kelebihanannya dalam mengolah data berstruktur matriks menggunakan prosedur IML (*Interactive Matrix Language*).

1. Prosedur Simulasi

1. 1. Desain CAT Murni

Prosedur simulasi untuk desain CAT murni berdasarkan pada 2100 simuli peserta tes yang disimulasikan, yang mewakili 100 simuli peserta tes untuk setiap 21 titik skala theta (θ) dari -3,0 sampai +3,0 dengan kenaikan 0,3 (-3,0, -2,7, -2,4, ..., +3,0).

Langkah simulasi untuk desain CAT murni sebagai berikut:

1. Untuk tingkat kemampuan peserta tes, θ tertentu, tes adaptif diberikan. Berdasarkan metode pemilihan butir awal (*starting point*), satu butir soal dipilih dan diberikan. Peluang peserta tes menjawab benar pada butir soal ke- i , $P_i(\theta)$ dihitung. Untuk membangkitkan jawaban atau respons dari peserta tes, nilai $P_i(\theta)$ dibandingkan dengan peubah acak x yang diambil dari sebaran uniform [0,1]. Jika x kurang dari $P_i(\theta)$ maka respons diskor 1, sebaliknya jika

x lebih dari atau sama dengan $P_i(\theta)$ maka respons diskor 0. Berdasarkan respons dan parameter butir soal selanjutnya kemampuan peserta tes, θ diestimasi. Estimasi theta dan butir soal yang diberikan dicatat untuk dianalisis lebih lanjut.

2. Berdasarkan metode pemilihan butir soal, diberikan butir soal berikutnya untuk peserta tes, θ tersebut sampai mencapai tingkat kesalahan baku pengukuran sebesar 0,30.
3. Langkah 1 dan 2 diulang untuk seluruh 2100 simuli peserta tes.
4. Banyaknya butir soal dan butir soal yang diberikan dicatat untuk dianalisis.

1. 2. Desain CCAT

Langkah-langkah simulasi desain CCAT mirip dengan CAT murni, perbedaannya hanya pada pemilihan butir soal pertama yang diberikan harus berasal dari modul 1, setelah sejumlah butir soal dari modul 1 terpenuhi butir soal berikutnya harus dari modul 2, setelah sejumlah butir soal dari modul 2 terpenuhi butir soal berikutnya harus dari modul 3, dan seterusnya sampai seluruh butir soal dari setiap modul dengan jumlah tertentu diberikan sesuai desain CCAT yang disimulasikan.

1. 3. Desain CBT

Seperti halnya pada simulasi CAT dan CCAT, simulasi CBT dilakukan pada 21 titik skala dengan 100 peserta setiap titik skala θ . Perangkat tes yang digunakan untuk simulasi CBT dilakukan dengan memilih sebanyak 35 butir soal yang berasal dari BS-Empiris atau bank soal untuk keperluan CAT. Pemilihan butir-butir soal dilakukan sedemikian hingga memenuhi kisi-kisi.

Langkah-langkah simulasi CBT sebagai berikut:

1. Pada θ tertentu, respons peserta tes untuk tingkat kemampuan θ pada n butir soal CBT diestimasi.
2. Berdasarkan respons yang diberikan dan parameter butir soal, tingkat kemampuan akhir peserta tes diestimasi dan dicatat untuk dianalisis lebih lanjut.
3. Langkah 1 dan 2 diulang untuk 2100 simulasi peserta tes.

Pada desain algoritma CAT murni maupun CCAT, selama belum diperoleh respons berpola, dilakukan pengestimasian tingkat kemampuan dengan menggunakan metode *step size* (Dodd,1990).

2. Kriteria Efisiensi dan Akurasi

Seperti telah dijelaskan sebelumnya bahwa untuk desain CAT murni berdasarkan bank soal untuk keperluan CAT yang terbentuk, simulasi dilakukan untuk mengetahui panjang tes atau banyaknya butir soal yang diperlukan pada setiap tingkatan kemampuan, di samping itu untuk mengetahui rentang tingkat kesukaran butir yang sesuai dan besarnya *step size* yang ideal untuk algoritma CAT yang akan diterapkan pada aplikasi CAT. Efisiensi desain CAT didasarkan pada banyaknya butir soal yang diperlukan sampai dapat mengestimasi kemampuan peserta tes. Keakuratan desain CCAT didasarkan pada kriteria kesalahan baku pengestimasian (*standard error of estimation* = SEE). Selanjutnya untuk membandingkan akurasi dari desain CAT, CCAT, dan CBT

berdasarkan tiga kriteria, yaitu: persentase jawaban benar, bias, dan kesalahan pengukuran.

2. 1. Persentase Jawaban Benar

Pada kemampuan peserta tes yang sesungguhnya (*true ability*), θ , dan banyaknya butir soal yang diberikan sebanyak n , rata-rata persentase jawaban benar dihitung dengan rumus:

$$P(\theta_i) = \left(\sum_{r=1}^{100} \left(\sum_{j=1}^n P_{rj} \right) / n \right) / 100 \quad (10)$$

sedangkan P_{rj} adalah respons (1 jika benar atau 0 jika salah) dari θ , pada butir soal ke- j , dan untuk simuli peserta tes ke- r .

2. 2. Bias

Bias adalah selisih antara estimasi tingkat kemampuan ($\hat{\theta}$) dengan sesungguhnya (*true ability*). Nilai ini menggambarkan kesalahan sistematik pada pengestimasian tingkat kemampuan. Secara matematis, bias dihitung dengan menggunakan rumus:

$$Bias(\theta) = \sum_{r=1}^{100} (\hat{\theta}_r - \theta) / 100 \quad (11)$$

2. 3. Kesalahan Baku Pengukuran

Kesalahan baku pengukuran (*Standard Error of Measurement* = SEM) mengindikasikan kesalahan acak pada pengestimasian tingkat kemampuan tertentu. Kesalahan ini juga menggambarkan presisi dari pengestimasian

kemampuan. SEM dihitung untuk setiap skala kemampuan, θ . Rumus untuk menghitung SEM adalah

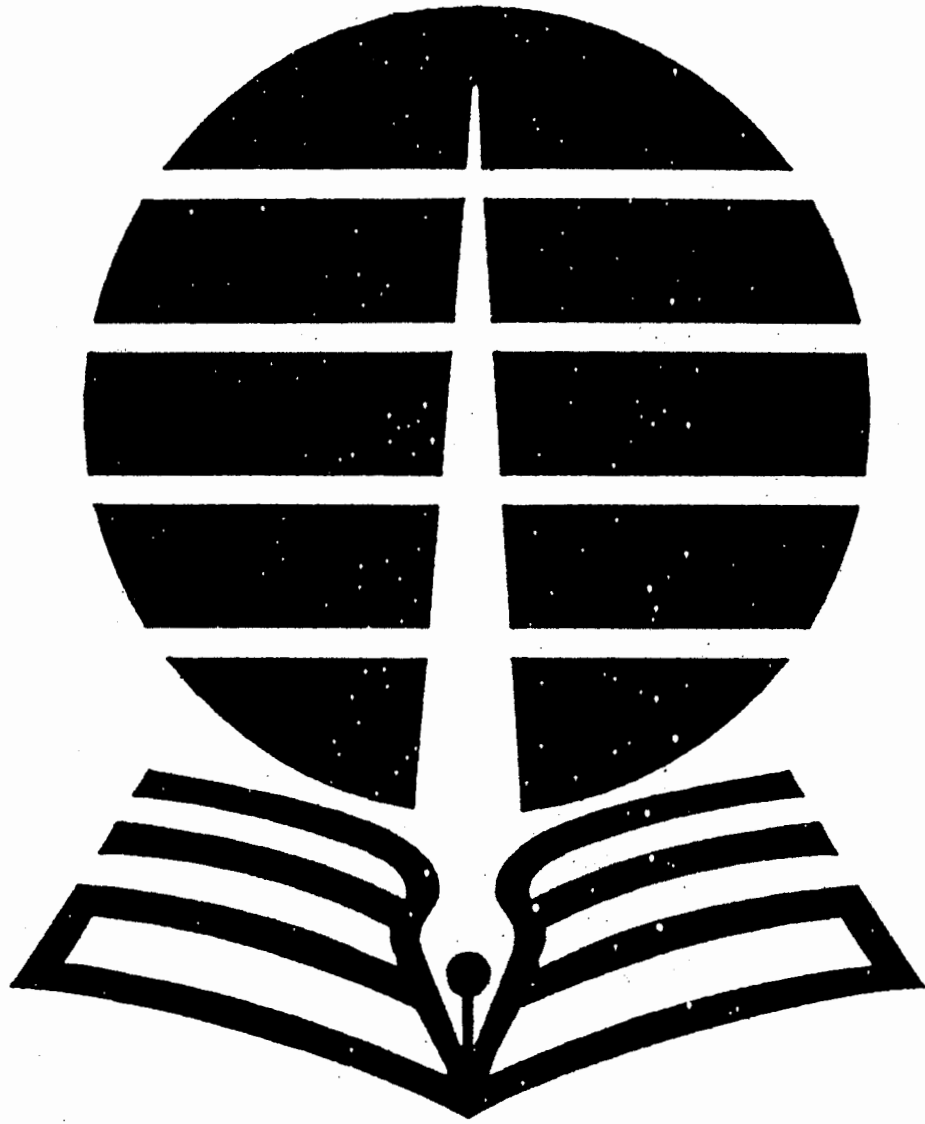
$$SEM(\theta) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \bar{\theta})^2} \quad (12)$$

dimana $\bar{\theta} = \sum_{r=1}^{100} \hat{\theta}_r / 100$ adalah rata-rata estimasi kemampuan dari 100 ulangan.

4. Tahap 4. Membuat Program CAT pada Sistem UAS UT

Berdasarkan rancangan algoritma CAT pada studi simulasi, selanjutnya dibuat suatu program (*software*) CAT. Rancangan algoritma dituliskan dalam bahasa pemrograman *Power Builder versi 10*. Tiga desain tes yang dikembangkan pada penelitian ini yaitu desain CBT, CAT, dan CCAT.

Desain CBT yang dikembangkan pada penelitian ini mirip dengan CBT UT, perbedaannya hanya pada metode penskorannya, pada CBT UT penskoran dilakukan dengan pendekatan klasik, sedangkan pada CBT yang dikembangkan pada penelitian ini, penskorannya dengan menggunakan pendekatan IRT dan bank soal untuk keperluan CBT sudah terkalibrasi berdasarkan model IRT.



BAB IV
HASIL DAN PEMBAHASAN

4. 1. Hasil Identifikasi Mata Kuliah dan Pembentukan Bank Soal

Matakuliah yang di-CAT-kan pada penelitian ini adalah matakuliah Pengantar Statistika Sosial dengan kode ISIP4215. Matakuliah ini dipilih karena mempunyai jumlah peserta tes yang cukup banyak pada setiap semesternya. Banyaknya peserta yang mengikuti ujian matakuliah Pengantar Statistika Sosial pada masa ujian 2004.1 sampai 2010.1 selengkapnya disajikan pada tabel 1 berikut.

Tabel 1. Banyaknya Peserta Ujian Matakuliah Pengantar Statitika Sosial

No	Masa Ujian	
1	2004.1	1.570
2	2004.2	2.149
3	2005.1	2.028
4	2005.2	1.553
5	2006.1	2.525
6	2006.2	1.291
7	2007.1	1.708
8	2007.2	1.956
9	2008.1	2.163
10	2008.2	2.500
11	2009.1	3.165
12	2009.2	3.243
13	2010.1	4.099

Dalam Tabel 1 terlihat bahwa banyaknya peserta ujian pada setiap semester umumnya lebih dari 1000 peserta bahkan dua masa ujian yaitu masa ujian 2009.1 (dibaca masa ujian 2009 semester 1), 2009.2 dan 2010.1 lebih dari 3000 peserta ujian. Dengan cukup besarnya jumlah peserta ujian per semesternya maka hasil kalibrasi menggunakan pendekatan pengukuran *item respons theory* lebih akurat (Hambleton, Swaminathan, & Rogers, 1991:54).

Panjang tes atau banyaknya butir soal untuk matakuliah Pengantar Statistika Sosial per semester disajikan pada tabel 2 berikut.

Tabel 2. Panjang Tes Matakuliah Pengantar Statistika Sosial per Masa Ujian

1	2004.1	46	45	
2	2004.2	47	45	
3	2005.1	48	42 (dari 45 butir)	3 butir di drop (no. 4,17, dan 32)
4	2005.2	49	32 (dari 35 butir)	3 butir di drop (no.5, 8, dan 11)
5	2006.1	51	35	
6	2006.2	52	35	
7	2007.1	53	35	
8	2007.2	54	35	
9	2008.1	55	35	
10	2008.2	11	35	
11	2009.1	13	35	
12	2009.2	16	35	
13	2010.1	20	35	

Dalam Tabel 2 terlihat bahwa untuk masa ujian 2005.1 dari panjang tes 45 butir tiga butir didrop, juga untuk masa ujian 2005.2 dari 35 butir soal ada tiga

butir soal yang didrop, sedangkan untuk masa ujian yang lain panjang tes untuk matakuliah ini umumnya adalah 35 butir soal.

Contoh hasil kalibrasi butir soal pada masa ujian 2010.1 disajikan pada Lampiran 1. Dari Lampiran 1 terlihat bahwa tingkat kesukaran butir nomor 1 adalah -0,58 ditunjukkan oleh nilai DELTA, butir soal nomor 1 ini dijawab oleh 4100 (MAXSCORE) namun yang menjawab benar sebanyak 1939 (SCORE). Butir soal nomor 2 tingkat kesukarannya adalah -0,27 dan seterusnya. Selanjutnya hasil kalibrasi juga diperoleh bahwa seluruh butir soal dari masa ujian 2010.1 adalah cocok (*fit*) dengan model *item response theory* 1 parameter. Hal ini dapat dilihat pada hasil uji fit butir soal pada Lampiran 2. Berdasarkan Lampiran 2 terlihat bahwa dengan menggunakan uji statistik Mean Square (MNSQ) dengan batas bawah 0,75 dan batas atas 1,25 (Adam & Khoo, 1996) maka seluruh butir soal pada masa ujian 2010.1 berada pada rentang batas tersebut, dengan demikian maka seluruh butir soal dari masa ujian 2010.1 dapat dipilih sebagai butir soal pada bank soal untuk keperluan CAT.

Lampiran 3 menggambarkan kedudukan kemampuan peserta tes pada masa ujian 2010.1 terhadap butir soal. Dari Lampiran 3 terlihat bahwa secara umum tingkat kesukaran butir-butir soal yang diujikan berada di atas dari tingkat kemampuan peserta tes, dengan kata lain tingkat kesukaran butir-butir soal dari perangkat tes mata kuliah Pengantar Statistik Sosial yang diujikan hanya dapat dijawab oleh peserta tes yang mempunyai kemampuan sedang sampai tinggi atau pada tingkat kemampuan -0,5 sampai 2,0 pada skala logit, hanya butir soal nomor

5, 23, dan 3 yang cocok untuk sebagian besar peserta tes yang berkemampuan rendah (-3 sampai -0,5).

4. 2. Deskripsi Bank Soal CAT

Seperti telah dijelaskan pada BAB III dan hasil analisis pendahuluan bahwa bank soal untuk keperluan CAT atau bank soal empiris dibangun dari butir-butir soal selama 13 semester (masa ujian 2004.1 sampai 2010.1). Hasil kalibrasi menghasilkan sebanyak 404 butir soal dengan rerata tingkat kesukaran adalah 0,0096 atau rerata tingkat kesukarannya adalah sedang dan standar deviasi atau simpangan baku sebesar 0,8217. Ringkasan statistik tingkat kesukaran dari 404 butir soal hasil kalibrasi disajikan pada tabel 3 berikut.

Tabel 3. Ringkasan Statistik Tingkat Kesukaran Butir Soal pada Bank Soal

Statistik	Tingkat Kesukaran Butir
Rerata	0,0096
Simpangan Baku	0,8217
Minimum	-3,14
Maksimum	2,08

Berdasarkan tabel 3 diperoleh bahwa tingkat kesukaran yang terendah adalah -3,14 sedangkan tertinggi adalah 2,08. Hal ini berarti bahwa bank soal yang terbentuk kurang menyediakan soal untuk peserta tes yang mempunyai tingkat kemampuan yang ekstrim (tinggi).

Tingkat kesukaran butir soal hasil kalibrasi dari 13 masa ujian beserta nomor induk soal (NIS), selengkapnya disajikan pada Lampiran 4.

Pada penelitian ini bank soal untuk keperluan CAT diset dalam model *item respons theory* (IRT) 3 parameter (3P), dengan menset setiap butir soal memiliki tingkat daya beda sama yaitu sebesar 1 dan faktor guessing sebesar 0,001. Hal ini dilakukan karena model IRT 3P lebih cocok pada tes pilihan ganda dibandingkan model 1P atau 2P (Hambleton, Swaminathan, & Rogers (1991).

Banyaknya butir soal per materi/modul dari Bank Soal disajikan pada Tabel 4 berikut.

Tabel 4. Banyaknya dan Persentase Butir Soal per Modul pada Bank Soal dan Perangkat P&P *test*

Materi/Modul	Bank Soal		Perangkat P&P <i>test</i>	
	Banyak Butir Soal	Persentase	Banyak Butir Soal	Persentase
Modul 1	13	3,22	1	2,86
Modul 2	51	12,62	5	14,29
Modul 3	66	16,34	5	14,29
Modul 4	44	10,89	4	11,43
Modul 5	86	21,29	8	22,86
Modul 6	35	8,66	2	5,71
Modul 7	31	7,67	3	8,57
Modul 8	35	8,66	3	8,57
Modul 9	43	10,64	4	11,43
Jumlah	404	100	35	100

Dalam Tabel 4 terlihat bahwa ketersediaan dan banyaknya butir soal serta persentase jumlah butir per materi/modul pada bank soal untuk keperluan CAT yang terbentuk umumnya mirip dengan persentase banyaknya butir pada perangkat P & P *test*.

4.3. Hasil Simulasi

1. Desain CAT Murni

Sebelum memaparkan hasil simulasi mengenai banyaknya butir soal yang diperlukan untuk mengestimasi tingkat kemampuan, berikut dipaparkan terlebih dahulu contoh hasil simulasi desain CAT murni berdasarkan bank soal empiris. Hal ini sekaligus untuk menguji proses algoritma CAT yang diterapkan.

Misalkan untuk $\theta = 0$ yang diambil secara acak. Berdasarkan hasil simulasi, peserta ini dapat diestimasi dengan butir soal sebanyak 17 butir soal. Nomor induk soal (NIS), tingkat kesukaran butir soal, pola respons untuk setiap urutan butir soal yang ditampilkan serta estimasi tingkat kemampuan (θ), kesalahan baku pengukuran ($SEM = \text{standard error of measurement}$) dan nilai fungsi informasi disajikan pada Tabel 5.

Berdasarkan Tabel 5 terlihat bahwa butir pertama yang terpilih adalah butir soal dengan NIS 37, memiliki tingkat kesukaran, $b = 0,12$, artinya ini sesuai dengan kriteria yang diterapkan pada algoritma desain CAT murni bahwa butir soal awal yang dipilih adalah butir dengan tingkat kesukaran sedang, yang dipilih secara acak pada rentang tingkat kesukaran sedang (-0,5 sampai +0,5).

Tabel 5. Nomor Induk Soal, Pola Respons, Estimasi Theta, SEM, dan Nilai Fungsi Informasi

No.Urut	1	2	3	4	5	6	...	15	16	17
N.I.S	37	359	4	264	181	200	...	340	382	301
Tk. Kesukaran	0.12	0.49	0.28	-0.14	0.17	-0.08	...	-0.08	0	-0.07
Respons	1	0	0	1	0	1	...	1	0	1
Theta	0.5	0.2961	-0.1312	0.1789	-0.0728	0.1302	...	0.0119	-0.0659	0.0043
SEM		0.8509	0.7066	0.6126	0.5472	0.4983	...	0.3111	0.3010	0.2918
Info	0.7081	0.7062	0.6219	0.6621	0.6782	0.6853	...	0.7048	0.7053	0.7060

Keterangan : Respons 1 = benar; 0 = salah

Berdasarkan Tabel 5, terlihat pula bahwa butir soal ini direspons 1, artinya dijawab benar, selanjutnya karena benar maka ditampilkan lagi butir soal dengan NIS 359. Butir soal dengan NIS 359 ini dipilih karena memiliki fungsi informasi terbesar pada theta sebesar 0,5, yaitu sebesar 0,7081. Hal ini juga telah sesuai dengan kriteria pemilihan butir soal berikutnya yang diterapkan pada algoritma CAT murni yang menggunakan kriteria *step size* sebesar 0,5. Berdasarkan kriteria *step size* ini maka ketika butir soal pertama dijawab benar maka butir soal kedua dipilih adalah butir soal yang mampu memberikan informasi maksimum bagi peserta dengan kemampuan pada tingkat 0,5, sebaliknya jika butir pertama dijawab salah maka butir soal kedua dipilih adalah butir soal yang memberikan informasi maksimum bagi peserta dengan kemampuan pada tingkatan -0,5. Pada butir soal pertama ini, kesalahan baku estimasi atau kesalahan pengukuran belum bisa ditentukan karena belum ada pola respons.

Selanjutnya ketika butir soal kedua direspons salah, maka pemilihan butir soal ketiga sudah didasarkan pada hasil pengestimasi theta. Hal ini karena metode MLE yang diterapkan pada algoritma desain tes adaptif murni akan berproses setelah respons sudah berpola (minimal ada satu benar atau satu salah). Berdasarkan metode MLE setelah menjawab butir soal nomor urut 1 benar dan nomor urut 2 salah, maka berdasarkan metode MLE kemampuan peserta ini diestimasi sebesar 0.2961, dan kesalahan pengukuran sudah dapat dihitung, yaitu sebesar 0,8509, dan karena kesalahan baku pengukuran belum mencapai 0,30 maka tes masih berlanjut.

Berdasarkan nilai fungsi informasi maksimum, maka butir soal ketiga yang dipilih adalah butir soal dengan NIS 4. Butir soal ini terpilih karena memiliki nilai fungsi informasi terbesar diantara butir-butir soal lainnya di bank soal untuk theta sebesar 0.2961. Seperti terlihat pada Tabel 2, nilai fungsi informasi butir soal ini sebesar 0.7062. Selanjutnya butir soal ketiga ini direspons, kemampuan dan kesalahan baku pengukuran diestimasi kembali, kemudian butir soal keempat dipilih, direspons, kemampuan diestimasi ulang, begitu seterusnya sampai tes dihentikan pada butir soal ke-17 karena pada butir ke-17 kesalahan baku pengukurannya telah mencapai 0,30 dengan estimasi theta sebesar 0,0043. Hal ini telah sesuai dengan kriteria pemberhentian tes yang diterapkan pada algoritma CAT murni, yaitu tes diberhentikan ketika kesalahan baku pengukuran telah mencapai 0,30.

1. 1. Banyaknya Butir Soal yang Diperlukan

Banyaknya butir soal atau panjang tes yang diperlukan untuk setiap theta yang disimulasikan berdasarkan bank soal data empiris disajikan pada Tabel 6. Dari Tabel 6 terlihat bahwa banyaknya butir yang diperlukan untuk mengestimasi tingkat kemampuan peserta tes untuk desain CAT murni berkisar antara 17 sampai 25 butir soal untuk rentang tingkat kemampuan (theta) antara -2,4 sampai 2,1. Hal ini menunjukkan bahwa dengan rancangan tes adaptif tingkat peserta tes sudah dapat diestimasi kemampuannya hanya dengan 17 sampai 25 butir soal saja untuk rentang theta antara -2,4 sampai 2,1.

Tabel 6. Banyaknya Butir Soal yang Diperlukan pada 21 Tingkatan Theta Yang Disimulasikan

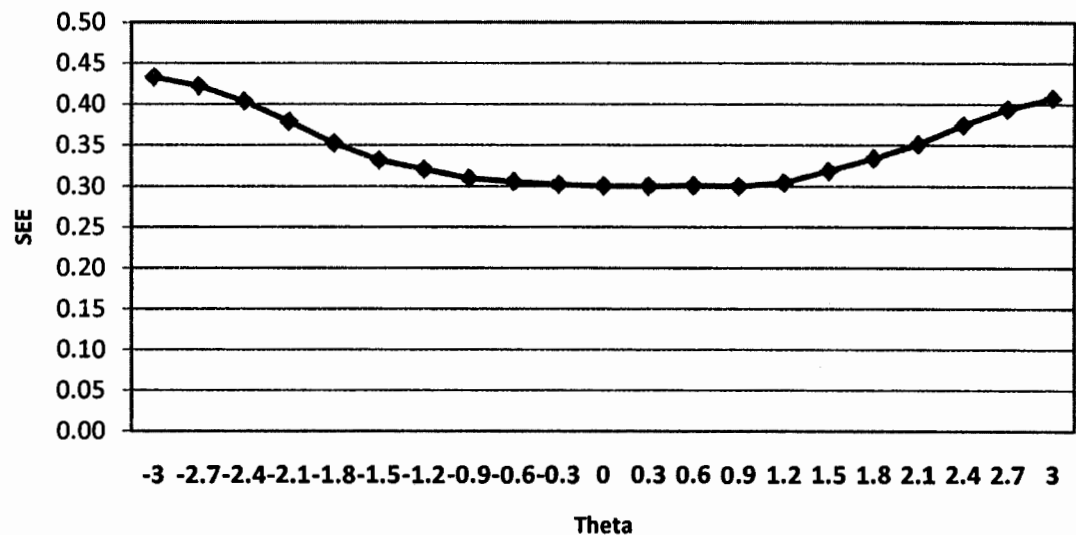
Tingkat Kemampuan (Theta)	Banyak Butir
-3	45
-2.7	36
-2.4	27
-2.1	21
-1.8	18
-1.5	18
-1.2	17
-0.9	17
-0.6	17
-0.3	17
0	17
0.3	17
0.6	17
0.9	17
1.2	17
1.5	18
1.8	20
2.1	24
2.4	39
2.7	65
3	105

Namun untuk tingkat kemampuan ekstrim (rendah maupun tinggi) dibutuhkan jumlah butir (panjang tes) yang lebih banyak, yaitu antara 35 sampai 60 butir, bahkan untuk tingkat kemampuan yang paling tinggi ($\theta=+3$) membutuhkan jumlah butir sebanyak 105 butir. Hal ini dikarenakan bank soal yang digunakan kurang menyediakan butir soal dengan tingkat kesukaran tinggi untuk peserta dengan kemampuan tinggi, seperti terlihat pada tabel 3 bahwa tingkat kesukaran tertinggi dari bank soal yang digunakan pada penelitian ini maksimum adalah sebesar 2,08.

2. Desain CCAT

Berbeda dengan desain CAT murni, pada desain CCAT didasarkan aturan pemberhentian *fixed length test*, artinya panjang tes desain CCAT adalah tetap untuk setiap theta yang disimulasikan. Dengan demikian simulasi CCAT tidak untuk membandingkan panjang tes pada setiap theta, tetapi untuk melihat tingkat keakurasian. Pada simulasi desain CCAT, keakurasian ditinjau dari kesalahan baku estimasi (*standard error of estimation* = SEE).

Gambar 4 adalah kesalahan baku estimasi untuk setiap tingkat kemampuan (theta) yang disimulasikan.



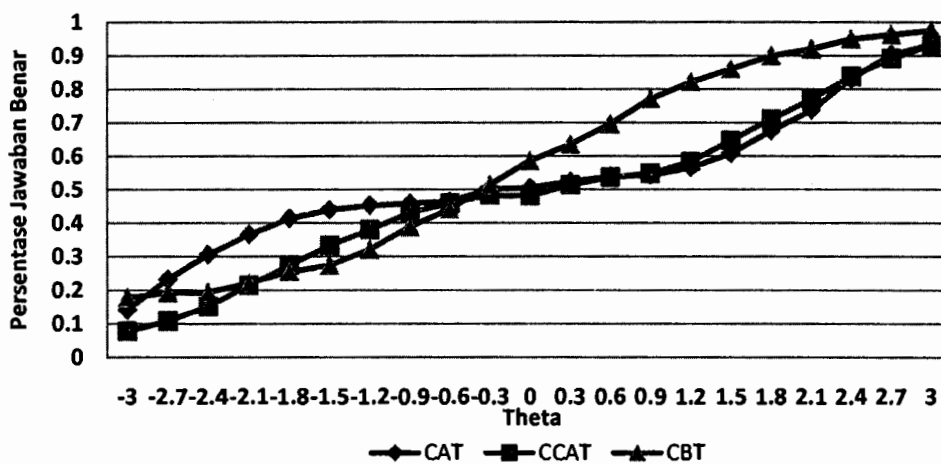
Gambar 4. Kesalahan Baku Estimasi Desain CCAT

Pada Gambar 4 terlihat bahwa desain CCAT50 umumnya memiliki kesalahan baku pengukuran cukup kecil yaitu pada rentang antara 0,30 sampai 0,43. Nilai kesalahan baku pengukuran pada rentang tersebut sebanding dengan

tingkat reliabilitas sebesar 80 sampai 90 % pada tes konvensional menggunakan *paper and pencil test* maupun CBT.

3. Persentase Jawaban Benar

Persentase jawaban benar untuk setiap tingkat kemampuan (*theta*) berdasarkan ketiga desain (CAT, CCAT, dan CBT) yang disimulasikan disajikan pada Gambar 5 berikut.



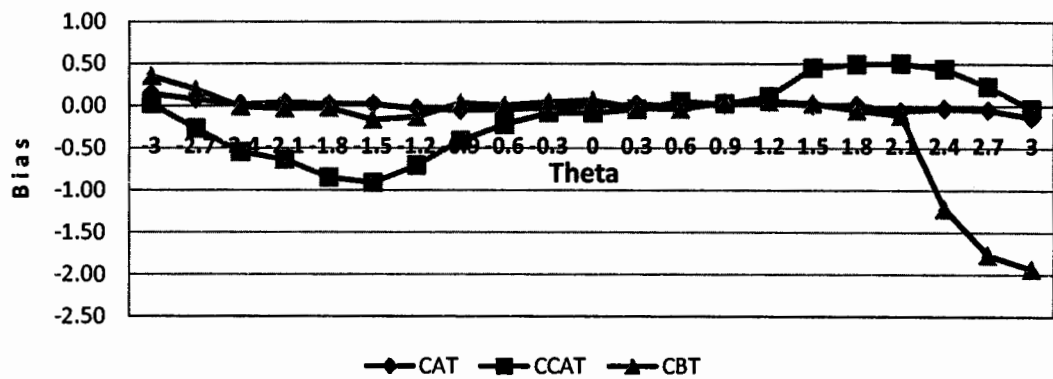
Gambar 5. Persentase Jawaban Benar

Berdasarkan Gambar 5 terlihat bahwa ketiga desain mempunyai pola persentase jawaban benar yang berbeda-beda untuk *theta* antara -3 sampai -0,9. Persentase jawaban benar desain CAT pada rentang *theta* -3 sampai -0,9 cenderung lebih besar dibandingkan desain CCAT maupun CBT. Selanjutnya pada rentang *theta* antara -0,6 dan -0,3 ketiga desain mempunyai persentase jawaban benar yang hampir sama, sedangkan pada *theta* 0,0 sampai +3 persentase jawaban benar pada desain CBT cenderung lebih besar dibandingkan desain CAT

dan CCAT. Persentase jawaban benar untuk desain CAT dan CCAT pada theta -0,9 sampai +3 memiliki pola dan persentase jawaban benar yang hampir sama.

4. Bias

Bias adalah selisih antara estimasi kemampuan (θ) dengan sesungguhnya. Bias untuk setiap kemampuan (theta) yang disimulasikan berdasarkan desain CAT, CCAT, dan CBT disajikan pada Gambar 6.

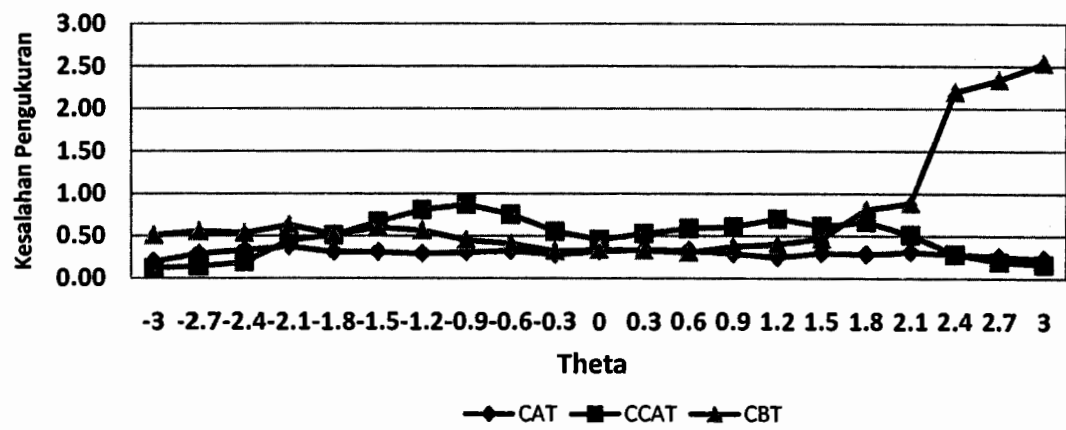


Gambar 6. Bias

Berdasarkan Gambar 6 terlihat bahwa pola bias untuk desain CAT dan CBT bervariasi sangat kecil pada rentang theta antara -3 sampai 2,1 dibandingkan desain CCAT, namun demikian bias CBT lebih besar dibandingkan desain CAT dan CCAT untuk rentang theta antara 2,4 sampai +3. Bias desain CAT umumnya cukup kecil atau hampir mendekati nol untuk setiap theta, hal ini berarti hasil estimasi theta untuk desain CAT sangat akurat, sedangkan bias desain CCAT bervariasi cukup besar untuk setiap theta yang disimulasikan.

5. Kesalahan Pengukuran

Gambar 7 menunjukkan kesalahan pengukuran (standar deviasi) desain CAT, CCAT, dan CBT. Kesalahan ini mengindikasikan kesalahan acak dalam pendugaan pada tingkat kemampuan (θ) tertentu. Kesalahan ini juga menggambarkan presisi dari pendugaan kemampuan. Kesalahan pengukuran dihitung untuk setiap θ .



Gambar 7. Kesalahan Pengukuran

Berdasarkan Gambar 7. Terlihat bahwa kesalahan pengukuran pada setiap tingkat θ untuk desain CAT umumnya lebih kecil dibandingkan desain CCAT maupun CBT, kecuali untuk θ ekstrim (rendah maupun tinggi). Kesalahan pengukuran desain CCAT lebih kecil dibandingkan desain CAT maupun CBT untuk θ ekstrim rendah yaitu pada rentang θ -3 sampai -2,4 dan pada θ ekstrim tinggi yaitu pada rentang 2,4 sampai +3. Output hasil analisis menggunakan SAS *versi* 9 mengenai bias dan kesalahan pengukuran (standard deviasi) untuk desain CAT, CCAT, dan CBT masing-masing disajikan pada

Lampiran 5, 6, dan 7. Contoh script program SAS untuk desain CAT Murni disajikan pada Lampiran 8.

4. 4. Pembahasan

Jumlah butir soal pada bank soal untuk keperluan CAT sebanyak 404 butir soal. Jumlah ini cukup ideal untuk membangun sistem ujian dengan desain CAT, namun demikian sebaran tingkat kesukaran butir soal dari bank soal yang terbentuk terlihat kurang merata untuk tingkatan kemampuan peserta tes. Tingkat kesukaran butir sedang lebih banyak dibandingkan butir soal yang mudah maupun yang sukar. Butir soal yang sukar lebih sedikit dibandingkan butir soal yang mudah. Hal ini mengakibatkan kurang tersedianya butir-butir soal yang sukar untuk peserta dengan kemampuan yang tinggi. Begitupun sebaran tingkat kesukaran dan banyaknya butir soal per materi/modul yang kurang merata, hal ini mengakibatkan terbatasnya butir soal yang akan diberikan pada desain CCAT. Hal ini selanjutnya dapat mempengaruhi hasil estimasi dari kemampuan peserta tes dan ini terbukti dengan lebih besarnya bias dan kesalahan pengukuran desain CCAT dibandingkan desain CAT.

Berdasarkan hasil simulasi CAT murni ditemukan ada variasi NIS yang digunakan untuk butir soal pertama untuk setiap peserta tes. Dengan menerapkan kriteria butir soal awal dengan tingkat kesukaran sedang pada rentang: $-0,50 < b < +0,50$, maka ada 205 NIS yang terpilih dan diberikan sebagai butir soal pertama kepada peserta tes. Diurutan kedua hanya 2 NIS yang sering dimunculkan, yaitu NIS 59 dan NIS 359. NIS 59 dimunculkan jika butir soal pertama direspons salah

dan NIS 359 dimunculkan jika butir soal pertama direspons benar. Diurutan ketiga ada 61 NIS, dengan sebanyak 8 NIS yang sering dimunculkan, yaitu NIS 180, 478, 348, 163, 240, 26, 264, dan NIS 204. Diurutan keempat ada 140 NIS yang dimunculkan, dengan sebanyak 15 NIS yang sering dimunculkan. Diurutan kelima sampai urutan ketujuh belas, variasi NIS yang dimunculkan bervariasi antara 190 sampai 300 NIS. Variasi NIS dipengaruhi oleh variasi dari pola respons pada butir soal sebelumnya. Peserta yang merespons dua butir soal sebelumnya salah dan salah akan ditampilkan butir soal ke tiga dengan NIS yang berbeda dengan peserta yang memiliki pola respons sebelumnya salah dan benar atau benar dan salah atau benar dan benar.

Dari hasil simulasi ditemukan bahwa NIS 359 dan NIS 59 yang digunakan sebagai butir kedua memiliki tingkat kesukaran 0,49 dan -0,51 (lihat Tabel 7 Lampiran 4), tingkat kesukaran ini mirip dengan nilai estimasi theta menggunakan *step size* sebesar 0,5. Hal ini menunjukkan bahwa kriteria *step size* dan fungsi informasi yang diterapkan pada algoritma CAT murni cukup efektif karena dapat memilih butir soal dengan tingkat kesukaran yang mirip dengan estimasi tingkat kemampuan (Hullin, Drasgow, & Parson, 1983: 220).

Seperti telah dijelaskan pada Bab III bahwa kriteria pengestimasian tingkat kemampuan pada penelitian ini menggunakan metode MLE dan metode *step size* (sebesar 0,5) ketika pola respons belum berpola. Dari hasil simulasi diketahui bahwa pola respons benar/salah semua, sebanyak 6 sampai 7 terjadi pada tingkat theta ekstrim (+2,7 sampai +3,0 dan -3 sampai -1,5), sedangkan untuk tingkat theta antara -1,2 sampai +2,4 metode MLE sudah dapat berproses setelah butir

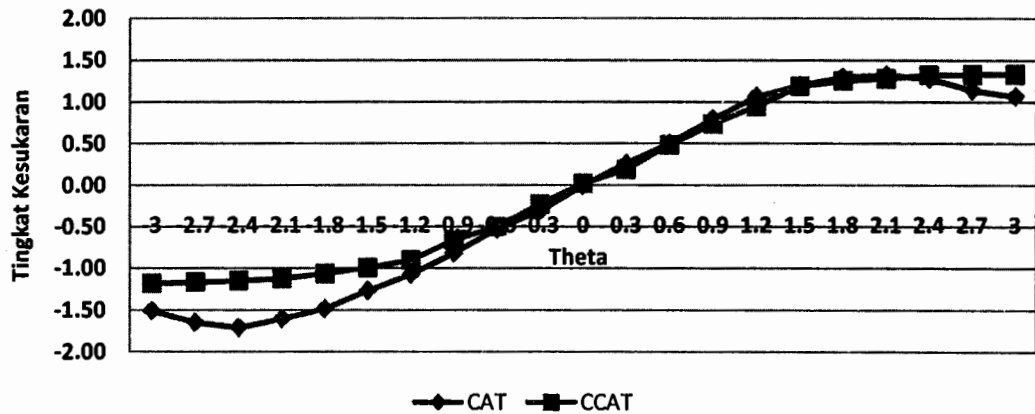
soal ke 2 atau ke 3 direspons, artinya setelah butir soal ke 2 atau ke 3 direspons pengestimasi theta sudah menggunakan metode MLE.

Berdasarkan hasil simulasi desain CAT murni memiliki pola bias dan kesalahan pengukuran yang lebih rendah dibandingkan dengan desain CCAT untuk setiap tingkat kemampuan (theta) yang disimulasikan, hal ini menunjukkan bahwa desain CAT lebih akurat dalam mengestimasi kemampuan peserta tes dibandingkan dengan desain CCAT, namun demikian dari hasil simulasi juga diperoleh bahwa pada desain CAT tidak semua modul ditampilkan untuk setiap peserta tes. Hal ini merupakan konsekuensi logis dari penerapan desain CAT murni yang menggunakan kriteria pemberian butir soal berdasarkan pada tingkatan butir soal bukan pada tingkatan materi/modul dan kriteria fungsi informasi maksimum untuk memilih butir soal berikutnya bukan pada tingkatan modul yang harus diberikan. Oleh karena itu wajar jika ada individu peserta tes pada desain CAT murni tidak memperoleh butir soal dari materi/modul tertentu.

Berbeda dengan desain CAT, desain CCAT dirancang untuk memberikan butir soal secara proporsional per materi/modul secara adatif kepada peserta tes. Dari simulasi diperoleh bahwa walaupun bias desain CCAT ini lebih besar dibandingkan desain CAT namun rancangan ini lebih cocok diterapkan pada sistem ujian Universitas Terbuka yang menuntut adanya keterwakilan butir soal dari setiap materi/modul. Ditambahkan lagi bahwa desain CCAT masih memiliki tingkat kesalahan baku pengukuran (*standard error of measurement* = SEM) yang cukup kecil. Dari gambar 4 terlihat bahwa kesalahan baku pengukuran pada tingkat kemampuan (theta) -2,1 sampai 2,7 sebesar 0,30. Nilai ini setara dengan

tingkat reliabilitas sebesar 91% pada tes konvensional menggunakan *paper and pencil test*.

Gambar 8 adalah perbandingan tingkat kesesuaian antara tingkat kesukaran butir soal dengan tingkat kemampuan peserta tes.



Gambar 8. Kesesuaian Tingkat Kesukaran Butir Soal dengan Kemampuan

Berdasarkan Gambar 8 terlihat bahwa kesesuaian antara tingkat kesukaran butir soal dengan tingkat kemampuan peserta (θ) pada desain CAT lebih baik dibandingkan dengan desain CCAT, terutama pada rentang θ antara -1,8 sampai 1,8, sedangkan untuk θ ekstrim (rendah atau tinggi) tingkat kesesuaian pada desain CAT maupun CCAT masih kurang sesuai. Hal ini disebabkan karena pada bank soal untuk keperluan CAT yang terbentuk kurang menyediakan butir soal dengan tingkat kesukaran untuk peserta tes dengan kemampuan yang ekstrim, khususnya untuk kemampuan yang tinggi.

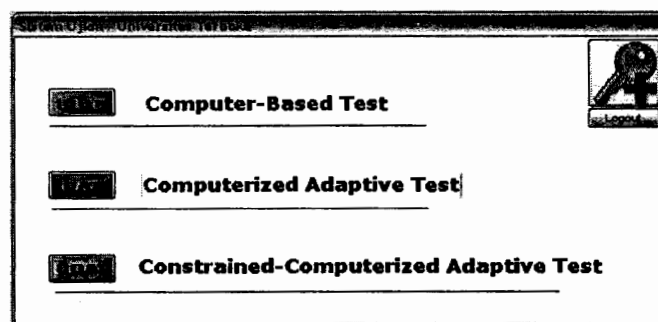
Kekurangesesuaian antara tingkat kesukaran butir soal dengan tingkat kemampuan peserta dapat terjadi karena kriteria pemilihan butir soal yang

digunakan pada penelitian ini tidak didasarkan pada tingkat kesukaran butir soal atau tidak di-macth-kan antara tingkat kesukaran butir soal dengan dengan kemampuan peserta tetapi didasarkan pada fungsi informasi butir soal, dengan kriteria ini maka butir soal yang mempunyai nilai fungsi informasi yang maksimum pada tingkatan kemampuan tertentu itulah yang diberikan kepada peserta tes. Dengan demikian, maka pola tingkat kesesuaian antara tingkat kesukaran butir dan kemampuan peserta pada kedua desain tersebut sesungguhnya sudah cukup baik atau sudah cukup efektif. Ditambahkan lagi, jika antara tingkat kesukaran butir dengan tingkat kemampuan peserta dikorelasikan maka diperoleh nilai korelasi untuk desain CAT dan CCAT masing-masing sebesar 0,97 dan 0,98. Nilai korelasi tersebut cukup tinggi, dengan demikian kedua desain ini cukup efektif dan layak diterapkan pada sistem ujian Universitas Terbuka.

4. 5. Program Aplikasi Sistem Ujian

1. Tampilan Awal Aplikasi

Aplikasi sistem ujian yang dikembangkan sebanyak tiga desain, yaitu CBT, CAT dan CCAT. Gambar 9 di bawah ini adalah tampilan awal dari sistem ujian.

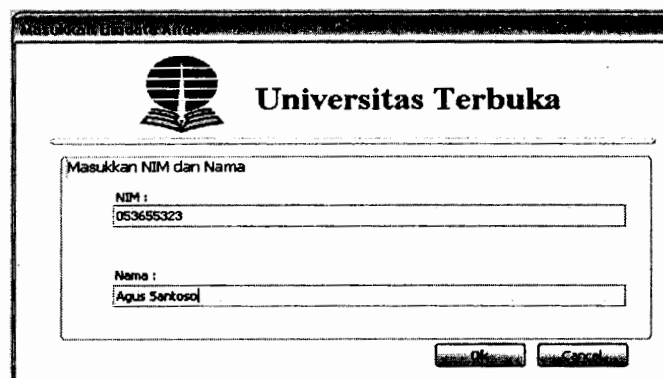


Gambar 9. Tampilan Awal Aplikasi

Seperti telah dijelaskan sebelumnya, CBT yang ada pada program aplikasi CAT ini mirip dengan CBT yang dikembangkan UT, artinya panjang tesnya sama yaitu untuk matakuliah Pengantar Statistik Sosial berjumlah 35 butir soal, banyaknya butir soal per materi/modul sama, ragam soalnya sama, dan peserta tes dapat merubah jawaban, perbedaannya hanya pada prosedur penskorannya, pada aplikasi CBT yang dikembangkan pada penelitian ini, penskoran dilakukan berdasarkan pengukuran IRT, sedangkan pada CBT UT penskoran berdasarkan pengukuran klasik.

2. Tampilan Login

Sebelum peserta tes memulai ujian dengan CAT, setiap peserta tes diwajibkan untuk login data pribadi mahasiswa (biodata) terlebih dahulu sebelum mengikuti ujian. Untuk dapat login, peserta tes harus meng-*input*-kan nomor induk mahasiswa (NIM) dan *Password*. *Password* untuk setiap peserta tes diberikan pada saat mereka mendaftarkan ujian *online* UT. Tampilan login seperti disajikan pada Gambar 10 dibawah ini.



Masukkan NIM dan Nama

Universitas Terbuka

Masukkan NIM dan Nama

NIM :
053655323

Name :
Agus Santoso

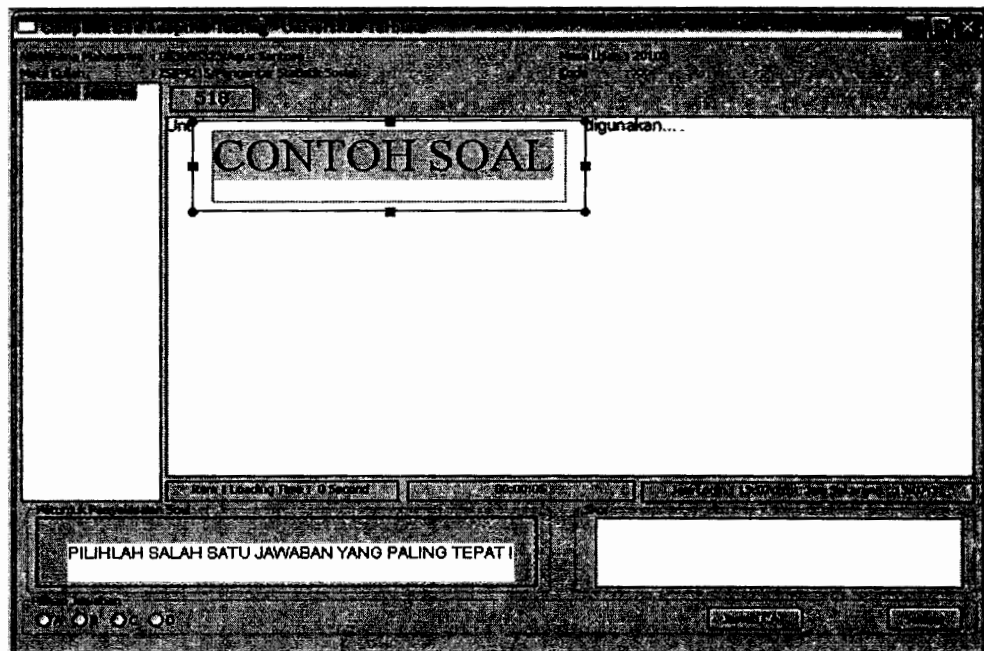
OK Cancel

Gambar 10. Tampilan Login

Ketika mahasiswa menginputkan salah pada salah satu NIM ataupun *Password* maka program tidak akan melanjutkan ke proses berikutnya. Sedangkan jika peserta tes menginput dengan benar NIM dan *Password*, maka berikutnya peserta dapat memilih sistem ujian yang dikehendaki.

3. Tampilan Aplikasi CAT

Jika peserta memilih sistem ujian CAT maka akan ditampilkan aplikasi CAT seperti pada Gambar 11.



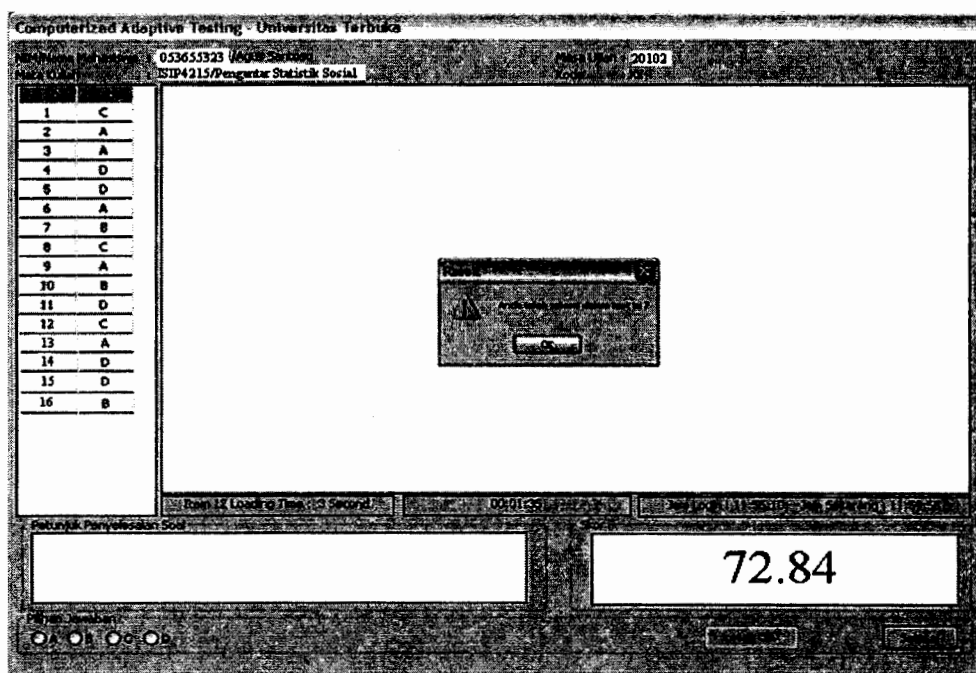
Gambar 11. Tampilan Aplikasi CAT

Pada tampilan aplikasi CAT, NIM dan nama peserta tes ditampilkan pada ujung kiri atas, beserta kode dan nama matakuliah, sedangkan pada ujung kanan atas ditampilkan masa ujian dan kode masa ujian. Pada bagian tengah ditampilkan pertanyaan/pernyataan dari butir soal yang ditampilkan (misalnya nomor induk

soal, NIS 101). Petunjuk penyelesaian soal ditampilkan pada bagian bawah sebelah kiri. Pilihan jawaban ditampilkan pada ujung bawah kiri. Urutan nomor soal dan jawaban peserta ditampilkan pada bagian tengah sebelah kiri.

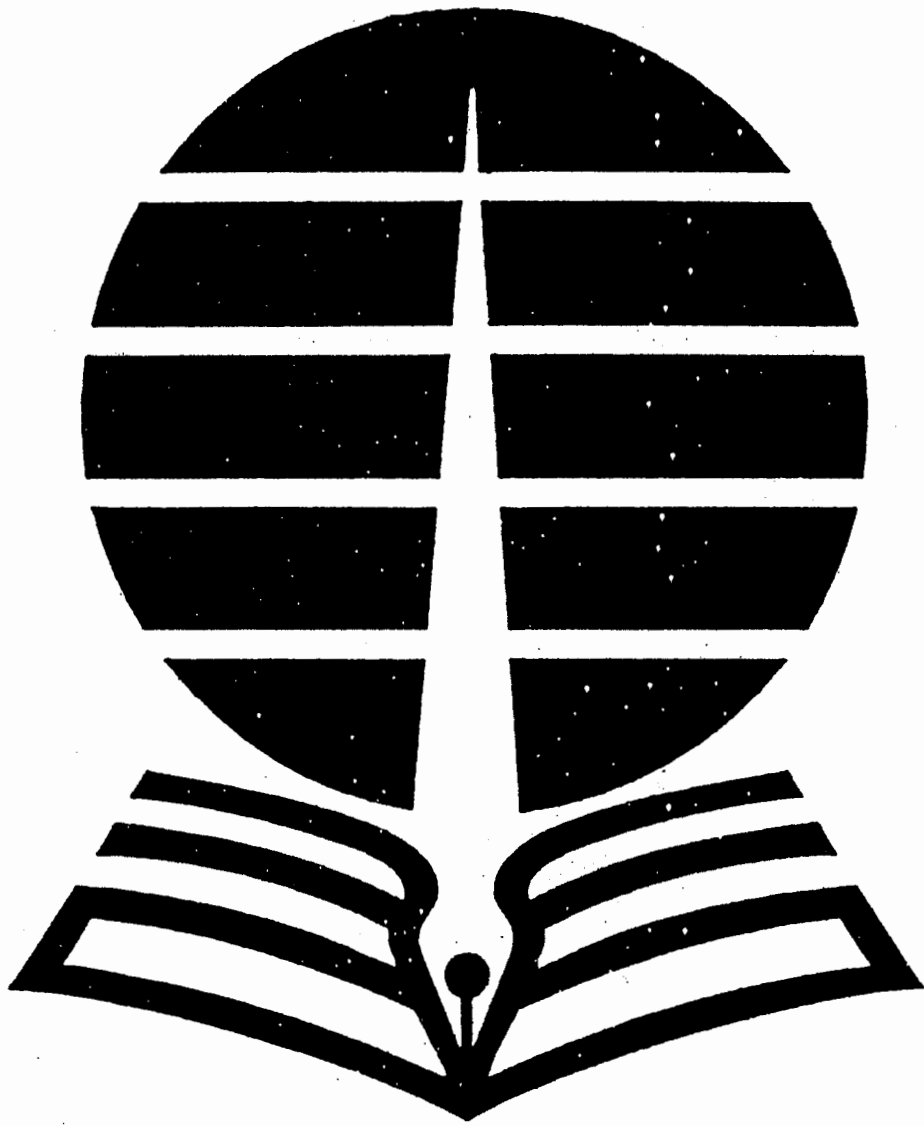
4. Tampilan Akhir Tes

Setelah tes dinyatakan berhenti maka akan ditampilkan pemberitahuan bahwa tes telah selesai seperti tampilan pada Gambar 12 berikut



Gambar 12. Tampilan Akhir Tes

Pada Gambar 12 terlihat bahwa ketika tes selesai, peserta tes diinformasikan bahwa tes telah selesai. Selanjutnya dengan mengklik tombol OK, maka skor akan ditampilkan. Untuk mengakhiri CAT, klik tombol selesai, selanjutnya administrator, mensest ulang untuk peserta tes (sesi) berikutnya.



BAB V

SIMPULAN DAN SARAN

5. 1. Simpulan

Berdasarkan hasil penelitian pendahuluan dan pembentukan Bank Soal untuk keperluan CAT serta hasil simulasi dan pembahasan pada Bab IV, dapat diambil kesimpulan sebagai berikut.

- 1) Bank soal untuk keperluan CAT sebanyak 404 butir soal dengan rerata tingkat kesukaran sedang (0,0096 pada skala logit). Tingkat kesukaran terendah -3,14 dan tertinggi 2,08.
- 2) Bank soal untuk keperluan CAT akurat untuk mengukur peserta tes dengan kemampuan rendah sampai tinggi (-2,7 sampai 2,7), namun kurang akurat dalam mengukur kemampuan yang ekstrim rendah (-3) maupun ekstrim tinggi (+3).
- 3) Banyaknya butir soal (panjang tes) yang diperlukan pada desain CAT murni, berkisar antara 17 sampai 25 butir soal untuk mengestimasi kemampuan peserta tes.
- 4) Desain CAT murni memiliki bias dan kesalahan pengukuran yang rendah dibandingkan desain CCAT.
- 5) Desain CCAT dapat dipilih sebagai desain alternatif CBT dalam sistem ujian UT karena desain ini telah mempertimbangkan keseimbangan isi.

5. 2. Implikasi

Program aplikasi desain CAT murni pada sistem ujian lebih efisien dibandingkan dengan tes konvensional menggunakan CBT maupun P&P *test*. Dengan kata lain bahwa dengan mengaplikasikan CAT, peserta tes tidak harus menempuh sebanyak 35 butir soal seperti pada pelaksanaan CBT maupun P&P *test* untuk mata kuliah Pengantar Statistik Sosial yang digunakan sebagai contoh mata kuliah yang di-CAT-kan. Dari penelitian simulasi diperoleh hasil bahwa dengan desain CAT peserta tes sudah dapat diestimasi kemampuannya hanya dengan 17 sampai 25 butir soal saja atau separoh dari jumlah butir soal yang diperlukan pada P&P test atau CBT. Hasil simulasi dengan menggunakan data bank soal empiris menunjukkan bahwa program aplikasi CAT yang dikembangkan pada penelitian ini baik digunakan untuk mengukur tingkat kemampuan peserta tes. Namun perlu disadari bahwa pada desain CAT murni ini ada kemungkinan butir soal yang diberikan kepada peserta tes tidak mewakili semua materi/modul, dengan kata lain ada butir soal dari materi/modul tertentu yang tidak terwakili sebagai butir soal CAT yang diberikan kepada peserta tes.

Program aplikasi desain CAT yang dikendala modul (CCAT) merupakan program aplikasi CAT yang baik digunakan manakala UT ingin menerapkan CAT yang mempertimbangkan keseimbangan konten/isi. Namun perlu juga disadari bahwa desain CCAT ini memiliki keterbatasan berupa ketidakadilan antar individu peserta tes. Hal ini karena desain CCAT ini tidak menggunakan kriteria *equal measurement precision* tapi menggunakan kriteria *fixed number of items*

yang berpotensi adanya ketidakadilan yang disebabkan adanya variasi tingkat kesalahan pengukuran antar individu peserta tes.

Program aplikasi desain CAT yang dikembangkan pada penelitian ini adalah sangat fleksibel, dalam arti bahwa program aplikasi CAT sangat mudah untuk disesuaikan. Setiap komponen CAT, misalnya: model IRT, bank soal, pemilihan butir soal awal, penentuan *step size*, pemilihan butir berikutnya, dan aturan pemberhentian pada program aplikasi ini mudah untuk disesuaikan. Desain CCAT yang dikembangkan pada sistem ujian mempunyai panjang tes sebanyak 50% dari panjang tes CBT atau sebanyak 18 butir soal. Pengembangan desain CCAT yang lain, misalnya CCAT dengan panjang tes 25%, 40%, 60% ataupun 75% proporsional dengan panjang test CBT mudah dilakukan, yaitu dengan mengganti kriteria banyaknya butir soal per materi/modul pada program aplikasi sistem ujian ini.

5. 3. Keterbatasan Penelitian

Pada penelitian ini, pengembangan desain CAT dirancang dengan cara dikendala, yaitu dikendala materi/modul, namun belum sampai pada pengontrolan terhadap butir soal yang sering dimunculkan. Terkait dengan keterbatasan ini maka perlu dilakukan penelitian lanjutan mengenai pengembangan sistem ujian desain CAT dengan memperhatikan keseimbangan konten dan kontrol terhadap butir soal yang sering dimunculkan.

Penelitian pengembangan model CAT ini hanya berdasarkan pada penelitian simulasi dan uji coba hanya dilakukan oleh peneliti dan beberapa sukarelawan,

sedangkan uji coba lapangan belum dilakukan. Dengan demikian faktor-faktor yang berkaitan dengan manusia, seperti: ketidaktahuan menggunakan komputer, kemampuan konsentrasi, yang diduga mempengaruhi skor peserta tes diabaikan.

5. 4. Saran

Berdasarkan kesimpulan, implikasi, dan keterbatasan penelitian yang telah dijelaskan, dapat disarankan hal-hal sebagai berikut.

1) Kepada Universitas Terbuka

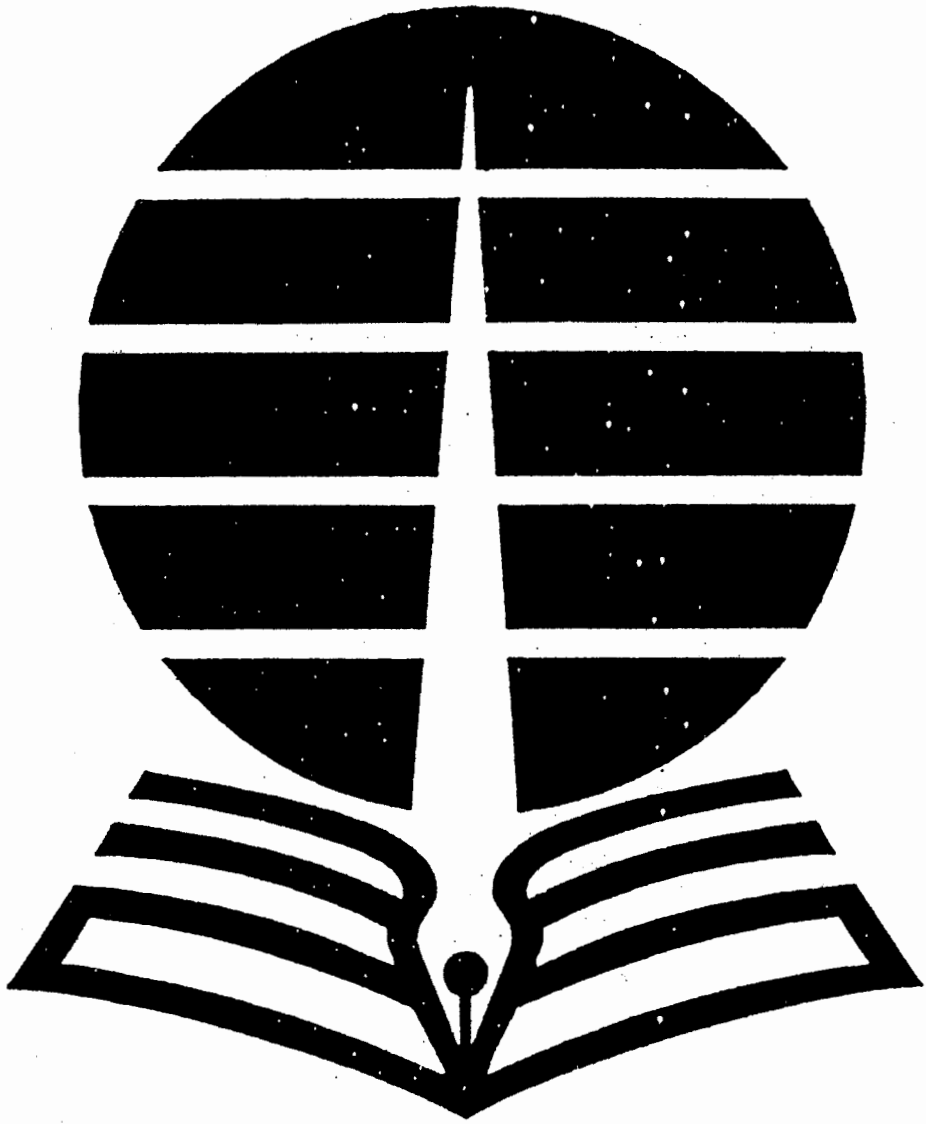
- a) Mengingat bank soal yang terbentuk untuk keperluan CAT kurang menyediakan butir soal dengan tingkat kesukaran tinggi, akibatnya peserta tes berkemampuan tinggi kurang akurat bila diukur melalui CAT ini dibandingkan dengan peserta tes berkemampuan rendah dan sedang. Oleh karena itu, perlu ada penambahan butir soal yang sukar pada bank soal.
- b) Untuk desain CAT yang mempertimbangkan keseimbangan isi, UT dapat menggunakan sistem ujian CCAT yang hanya memerlukan separoh dari banyaknya butir soal CBT (*P&P test*). Namun perlu disadari bahwa pada desain ini ada kesalahan pengukuran yang berbeda untuk setiap peserta tes.
- c) Sistem ujian CAT merupakan sesuatu yang baru pertama kali dikembangkan dan belum diujicobakan. Oleh karena itu sistem CAT ini perlu diujicobakan pada peserta tes (mahasiswa).

2) Kepada pengembang sistem ujian

- a) Penelitian tentang CAT di Indonesia masih sangat jarang, oleh karena itu perlu terus dilakukan penelitian tentang CAT. Perlu dilakukan penelitian lanjutan mengenai pengembangan CAT yang tidak hanya memperhatikan keseimbangan konten tetapi sekaligus mengontrol butir soal yang sering dimunculkan.
- b) Pada penelitian ini, desain CAT dikembangkan berdasarkan model IRT dikotomous. Oleh karena itu perlu dilakukan penelitian desain CAT untuk data berjenis politomous. Dengan demikian pengembangan dalam domain CAT berbasis pengukuran modern atau *item respons theory* lebih beragam.

5. 5. Rekomendasi

CAT lebih efisien dibanding CBT, sehingga direkomendasikan kepada UT untuk mengaplikasikan CAT ini sebagai sistem ujian alternatif pada penyelenggaraan UAS UT. Desain CAT yang dikendala modul (CCAT) lebih direkomendasikan kepada UT dibandingkan desain CAT murni, karena desain ini telah mempertimbangkan keseimbangan isi. Selanjutnya, karena CAT merupakan sesuatu yang baru pertama kali dikembangkan maka penerapannya sebaiknya dilakukan pada tingkatan UPBJJ terlebih dulu.



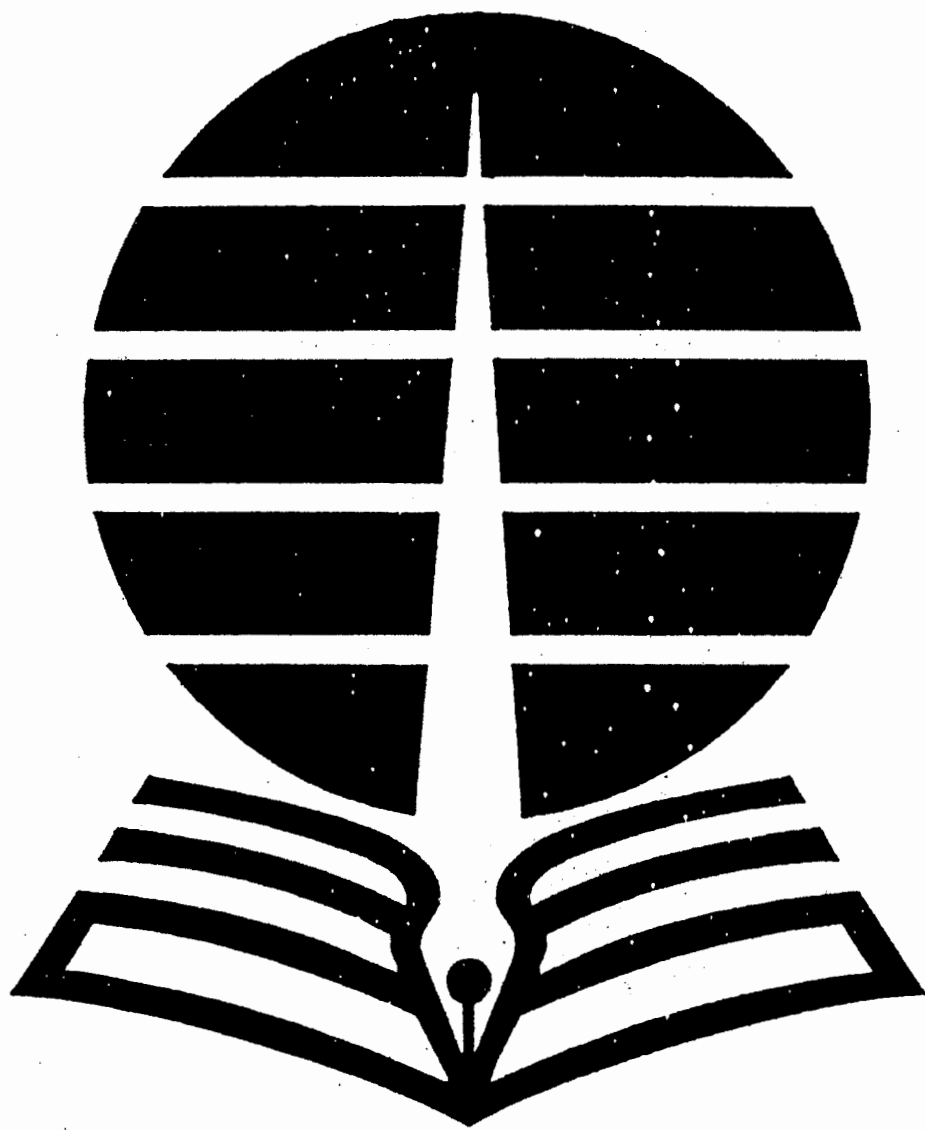
DAFTAR PUSTAKA

- Adam , R.J. & Khoo S. T. (1996). Quest. *The interactive test analysis system* [Computer software]. Camberwell, Victoria: Australian Council for Educational Research.
- Anonim (2006). *Adaptive testing*. Diambil pada tanggal 18 Juni 2006, dari http://www.windowsgalore.com/cert/adaptive_testing/.
- Ansley, T.N., & Forsyth, R.A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two dimensional data. *Applied Psychological Measurement*, 1, 37 – 48.
- Baker, F.B. (1992). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker, Inc.
- Birnbaum, A. (1958). *On the estimation of mental ability*. (Series Report 15, Project No. 7755 – 23). Randolph Air Force Base, TX:USAF School of Aviation Medicine.
- Bock, R.D. & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 4, 443 – 459.
- Bock, R.D., & Mislevy, R.J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 4, 431 – 444.
- Bunderson, C.V., Inouye, D.K., & Olsen, J.B. (1989). The four generations of computerized educational measurement. Dalam R. L. Linn (Eds.), *Educational Measurement* (3rd ed., pp. 367–407). New York: American Council on Education & Macmillan Publishing Company.
- Chen, S.Y., Ankenmann, R.D., & Chang, H.-H. (2000). A comparison of item selection rules at the early stages of computerized adaptive testing. *Applied Psychological Measurement*, 3, 241 – 255.
- Dodd, B.G. (1990). The effect of item selection procedure and stepsize on computerized adaptive attitude measurement using the rating scale model. *Applied Psychological Measurement*, 4, 355 – 366.
- Dragow, F., & Olson-Buchanan, J.B. (1999). *Innovations in computerized assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Dunkel, P.A. (1999). *Considerations in developing and using computer adaptive tests to asses second language proficiency*. Diambil pada tanggal 16 juni 2006, dari <http://www.cal.org/resources/digest/cat.html>.
- Embretson, S.E. & Rouse, S.P. (2000). *Item response theory for psychologist*. London: Lawrence Erlbaum Associates, Inc.

- Flaugher, R. (1990). Item pool. Dalam H. Wainer (Eds.), *Computerized Adaptive Testing: A Primer* (2nd ed., pp. 41–63). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Folk, V.G., & Green, B.F. (1989). Adaptive estimation when the unidimensionality assumption of IRT violated. *Applied Psychological Measurement*, 4, 373 – 389.
- Gershon, R.C. (2005). Computer adaptive testing. *Journal of Applied Measurement*, 1, 190 – 127.
- Green, B.F., Bock, R.D., Humphyers, L.G., et al. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 4, 347 – 360.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*, Boston, MA: Kluwer Academic Publishers.
- Hambleton, R.K., Zaal, J.N., & Pieters, J.M.P. (1991). Computerized adaptive testing: Theory, applications, and standards. Dalam R. K. Hambleton & J. N. Zaal (Eds.), *Advances in Educational and Psychological Testing: Theory and Applications* (pp. 341–366). Boston, MA: Kluwer Academic Publishers.
- Kingsbury, G.G., & Zara, A.R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, 4, 359 – 375.
- Lord, F.M., (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ : Lawrence Erlbaum Associates.
- Lord, F.M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lord, F.M. (1977). A broad-range tailored test of verbal ability. *Applied Psychological Measurement*, 1 , 95–100.
- Lunz, M.E., & Bergstrom, B.A. (1994). An empirical study of computer adaptive test administration formats. *Journal of Educational Measurement*, 3, 251 – 263.
- Mills, C.N. (1999). Development and introduction of a computer adaptive graduate record examinations general test. Dalam F. Drasgow & J. B. Olson-Buchanan (Eds), *Innovations in Computerized Assessment* (pp. 117–136). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Owen, R.J. (1975). A bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, 70, 351 – 356.

- Reckase, M.D. (April, 2003). *Item pool design for computerized adaptive test*. Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.
- Stahl, J.A., & Lunz, M.E. (April, 1993). *Assesing the extent of overlap of items among computer adaptive tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta.
- Stocking, M.L., & Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement*, 3, 277 – 292.
- Thissen, D. (1990). Reliability and measurement precision. Dalam H. Wainer (Eds.), *Computerized Adaptive Testing: A Primer* (2nd ed., pp. 161–186). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Thissen, D., & Mislevy, R.J. (1990). Testing algorithms. Dalam H. Wainer (Eds.), *Computerized Adaptive Testing: A Primer* (2nd ed., pp. 103–135). Hillsdale, NJ: Lawrence Erlbaum Associates.
- van der Linden, W.J. & Glas, C.A.W. (2000). *Computerized adaptive testing: Theory and practice*. The Netherlands: Kluwer Academic Publishers.
- van der Linden, W.J. (2000). Optimal assembly of test with item sets. *Applied Psychological Measurement*, 3, 225 – 240.
- Vispoel, W.P. (1999). Creating computerized adaptive test of music aptitude : Problem, solusions, and future directions. Dalam F. Drasgow, & J. B. Olson-Buchanan (Eds.), *Innovations in Computerized Assessment* (pp. 151 – 176). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Wainer, H., et al. (1990). *Computerized adaptive testing: A primer* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wainer, H. (1990). Introduction and history. Dalam H. Wainer (Eds.), *Computerized Adaptive Testing: A Primer* (2nd ed., pp. 1–21). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wainer, H. & Mislevy, R.J. (1990). Item response theory, item calibration and proficiency estimation. Dalam H. Wainer (Eds.), *Computerized Adaptive Testing: A Primer* (2nd ed., pp. 65–102). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wang, T., & Vispoel, W.P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 2, 109–136.
- Way, W.D. (Maret,1997). *Protecting the integrity of computerized testing item pools*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

- Weiss, D.J. (1983). *New horizons in testing: Latent trait test theory and computerized adaptive testing*. New York: Academic Press.
- Weiss, D.J. & Schleisman, J.L. (1999). Adaptive testing. Dalam G. N. Masters & J. P. Keeves (Eds.), *Advances in Measurement in Educational Research and Assessment* (pp. 129–137). Pergamon, NY: Elsevier Science Ltd.
- Weiss, D.J., & McBride, J. R. (1984). Bias and information of Bayesian adaptive testing. *Applied Psychological Measurement*, 3, 273-285.
- Weiss, D.J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 2, 70 - 84.



LAMPIRAN 1

QUEST: The Interactive Test Analysis System

 Item Estimates (Category Deltas) In input Order 28/ 7/2010 20:44
 all on all (N = 4102 L = 35 Probability Level=0.50)

ITEM NAME	SCORE	MAXSCR	DELTA	INFT	OUTFT	INFT	OUTFT
			1	MNSQ	MNSQ	t	t
1 item 1	1939	4100	-0.58	0.95	0.94	-6.2	-3.2
			.03				
2 item 2	1637	4094	-0.27	1.07	1.08	6.6	4.0
			.03				
3 item 3	2686	4089	-1.37	0.99	1.00	-0.7	0.1
			.03				
4 item 4	2231	4091	-0.88	0.95	0.94	-6.1	-3.4
			.03				
5 item 5	3217	4096	-2.04	0.98	0.97	-1.1	-1.0
			.04				
6 item 6	1157	4061	0.27	1.07	1.10	4.2	3.7
			.04				
7 item 7	1761	4088	-0.40	1.00	1.00	-0.4	-1.0
			.03				
8 item 8	256	4084	2.08	1.04	1.20	0.7	2.7
			.07				
9 item 9	1283	4079	0.12	1.05	1.08	3.7	3.1
			.04				
10 item 10	2258	4093	-0.91	0.90	0.89	-11.6	-6.2
			.03				
11 item 11	1523	4096	-0.15	0.94	0.93	-5.9	-3.2
			.03				
12 item 12	1231	4070	0.18	1.03	1.03	1.9	1.3
			.04				
13 item 13	1198	4059	0.21	1.05	1.07	3.0	2.9
			.04				
14 item 14	836	4094	0.72	1.01	1.06	0.5	1.8
			.04				
15 item 15	1688	4092	-0.32	0.96	0.96	-4.3	-2.2
			.03				
16 item 16	1224	4088	0.19	0.95	0.93	-3.5	-2.9
			.04				

17	item 17		1321	4056	0.06	1.08	1.11	6.0	4.4
					.04				
18	item 18		680	4093	0.98	1.00	1.05	0.2	1.4
					.04				
19	item 19		2104	4090	-0.76	1.01	1.02	1.9	1.1
					.03				
20	item 20		1380	4092	0.01	0.89	0.87	-8.4	6.1
					.03				
21	item 21		2024	4096	-0.67	0.92	0.92	-9.7	-4.9
					.03				
22	item 22		944	4089	0.56	0.98	1.01	-0.9	0.2
					.04				
23	item 23		2967	4092	-1.71	0.94	0.90	-3.9	-3.9
					.04				
24	item 24		1387	4089	0.00	0.90	0.90	-7.6	-4.9
					.03				
25	item 25		914	4092	0.61	0.90	0.88	-4.8	4.1
					.04				
26	item 26		994	4056	0.48	1.06	1.10	3.2	3.3
					.04				
27	item 27		2294	4096	-0.95	0.97	0.96	-3.8	-2.0
					.03				
28	item 28		1502	4079	-0.13	1.05	1.07	4.7	3.0
					.03				
29	item 29		600	4081	1.13	1.04	1.10	1.2	2.5
					.05				
30	item 30		1183	4084	0.24	1.11	1.16	6.8	6.0
					.04				
31	item 31		1078	4082	0.37	1.03	1.04	1.6	1.4
					.04				
32	item 32		787	4072	0.79	1.03	1.08	1.4	2.4
					.04				
33	item 33		909	4068	0.60	1.04	1.07	2.1	2.3
					.04				
34	item 34		849	4060	0.69	1.05	1.10	2.1	3.2
					.04				
35	item 35		739	4041	0.86	1.04	1.09	1.6	2.5
					.04				

	Mean				0.00	1.00	1.02	-0.7	0.1
	SD				0.84	0.06	0.09	4.7	3.3
=====									

LAMPIRAN 2

QUEST: The Interactive Test Analysis System

Item Fit

28/ 7/2010 20:44

all on all (N = 4102 L = 35 Probability Level=0.50)

INFIT

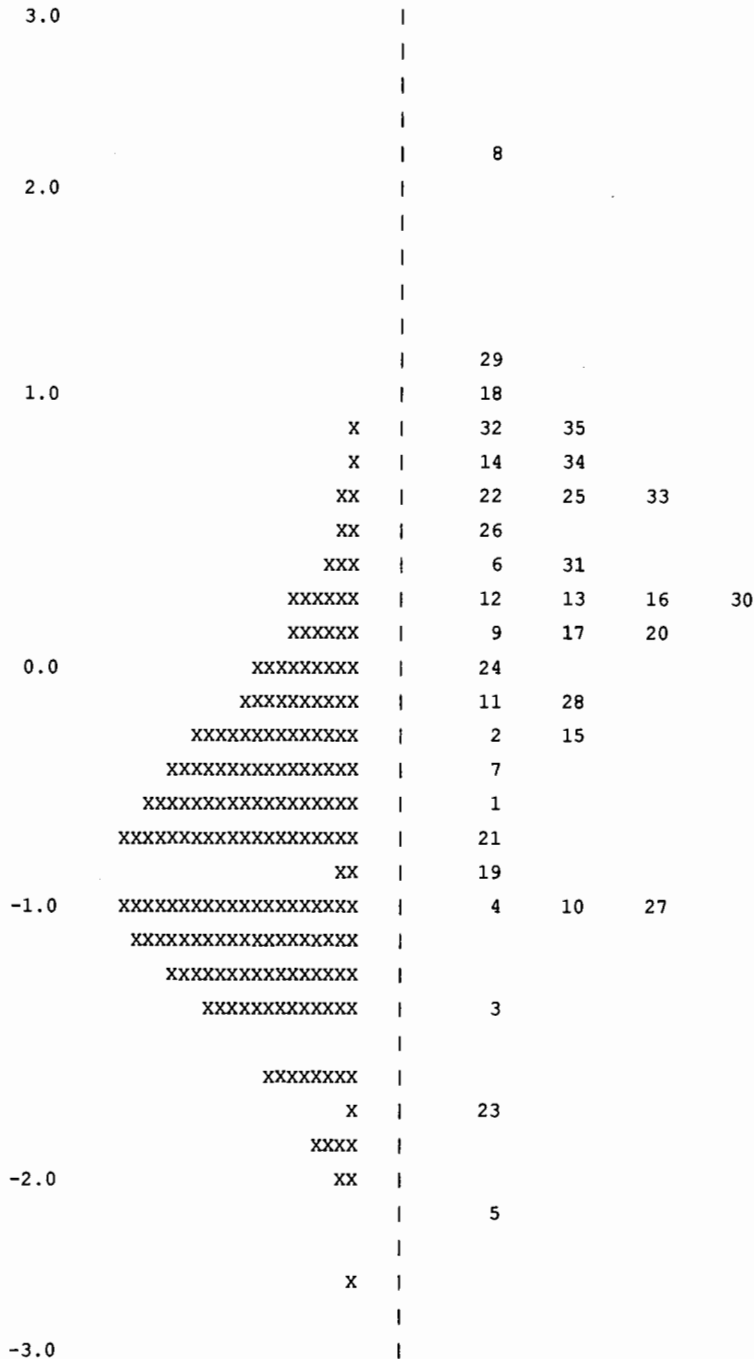
MNSQ	0.63	0.71	0.83	1.00	1.20	1.40	1.60
item 1				1			
item 2				2			
item 3				3			
item 4				4			
item 5				5			
item 6				6			
item 7				7			
item 8				8			
item 9				9			
item 10				10			
item 11				11			
item 12				12			
item 13				13			
item 14				14			
item 15				15			
item 16				16			
item 17				17			
item 18				18			
item 19				19			
item 20				20			
item 21				21			
item 22				22			
item 23				23			
item 24				24			
item 25				25			
item 26				26			
item 27				27			
item 28				28			
item 29				29			
item 30				30			
item 31				31			
item 32				32			
item 33				33			
item 34				34			
item 35				35			

LAMPIRAN 3

Case Estimates In input Order

28/ 7/2010 20:44

all on all (N = 4102 L = 35 Probability Level=0.50)



Each X represents 21 students

LAMPIRAN 4**Tabel 7. Tingkat Kesukaran Butir Soal pada Bank Soal untuk Keperluan CAT**

Item	1	2	3
MU.041	1	-0.99	1
	2	-0.94	2
	3	-1.06	3
	4	0.28	4
	5	-0.12	5
	6	-0.26	6
	7	0.27	7
	8	-0.53	8
	9	0.95	9
	10	-0.28	10
	11	-0.82	11
	12	0.59	12
	13	-0.24	13
	14	1.05	14
	15	0.07	15
	16	0.05	16
	17	-1.09	17
	18	0.5	18
	19	-0.37	19
	20	0.35	20
	21	-0.75	21
	22	0.35	22
	23	0.77	23
	24	-1.19	24
	25	-1.06	25
	26	0.04	26
	27	-1.68	27
	28	0.65	30
	29	-0.12	28
	30	0.51	29
	31	-0.25	31
	32	0.45	32
	33	0.29	33
	34	0.4	34
	35	-0.32	35
	36	0.54	36
	37	0.12	37
	38	0.19	38

	39	-0.06	39
	40	0.27	40
	41	0.35	41
	42	0.82	42
	43	0.6	43
	44	0.62	44
	45	1.06	45
MU.042	1	-1	46
	2	-0.65	55
	3	0.41	64
	4	-0.43	317
	5	-0.47	73
	6	-0.38	82
	7	-1.16	91
	8	-0.41	318
	9	-0.12	100
	10	0.12	319
	11	-1.27	109
	12	1.5	118
	13	-1.27	320
	14	-1.18	127
	15	-0.41	321
	16	0.06	136
	17	0.15	322
	18	0.77	145
	19	-0.32	323
	20	0.18	154
	21	-0.07	163
	22	-0.02	324
	23	1.01	172
	24	-0.73	325
	25	0.17	181
	26	-0.75	326
	27	-1.09	190
	28	1.39	218
	29	-0.07	199
	30	0.1	208
	31	0.16	227
	32	-0.38	236
	33	0.66	245
	34	0.38	254
	35	0.35	263

	36	1.07	272
	37	1.2	327
	38	0.08	281
	39	1.35	328
	40	0.6	290
	41	0.21	329
	42	-0.16	299
	43	-0.3	308
	44	0.37	330
	45	0.34	331
MU.051	1	-2.63	47
no.4,17,&32	2	0.14	56
diGratisikan	3	0.43	65
	5	-3.07	83
	6	-0.54	92
	7	-0.04	101
	8	-1.82	110
	9	0.51	119
	10	-0.72	128
	11	-0.92	137
	12	0.12	146
	13	0.26	155
	14	0.8	164
	15	1.93	173
	16	-0.43	344
	18	1.34	191
	19	0.43	219
	20	-0.08	200
	21	0.85	209
	22	0.13	228
	23	0.68	237
	24	-0.55	246
	25	0.31	255
	26	-0.14	264
	27	0.07	273
	28	-1.11	342
	29	1.45	282
	30	0.38	345
	31	0.41	291
	33	-0.07	300
	34	0.69	309
	35	0.36	346

	36	0.25	332
	37	0.08	333
	38	-0.46	334
	39	-0.29	335
	40	0.73	336
	41	-0.09	337
	42	0.26	338
	43	0.51	339
	44	-0.08	340
	45	-0.05	348
MU.052	1	-1.54	48
no.5,8,&11	2	-2.19	66
diGratiskan	3	-0.35	349
	4	0.9	75
	6	-0.58	93
	7	-0.38	102
	9	-0.04	120
	10	-0.61	129
	12	0.9	353
	13	-0.12	147
	14	0.57	356
	15	-0.21	156
	16	0.92	174
	17	0.21	358
	18	0.72	183
	19	0.32	192
	20	-0.57	165
	21	1.13	220
	22	-0.4	201
	23	0.48	210
	24	0.36	229
	25	-0.4	247
	26	-0.38	265
	27	-0.66	274
	28	-0.62	360
	29	1.75	283
	30	-0.3	292
	31	-0.36	362
	32	-0.08	301
	33	0.26	310
	34	0.34	363
	35	0.93	365

MU.061	1	-0.97	60
	2	0.05	72
	3	0.62	350
	4	0.31	81
	5	-3.14	85
	6	0.07	94
	7	-2.75	104
	8	0.53	113
	9	0.23	123
	10	-1.13	131
	11	1.32	142
	12	-0.73	355
	13	0.14	148
	14	-0.55	374
	15	-0.2	159
	16	-1.01	180
	17	0.49	359
	18	0.3	184
	19	0.72	195
	20	0.17	166
	21	0.48	224
	22	0.38	202
	23	0.68	212
	24	0.02	240
	25	1.25	248
	26	0.46	266
	27	-0.98	276
	28	-0.54	387
	29	0.23	289
	30	0.81	296
	31	0.6	389
	32	1	304
	33	0.32	316
	34	0.45	398
	35	0.38	403
MU.062	1	0.34	53
	2	-1	67
	3	-0.43	351
	4	-1.05	76
	5	-1.39	86
	6	-0.37	95
	7	-0.43	106

	8	0.15	114
	9	-0.86	125
	10	-0.39	133
	11	-0.46	144
	12	-0.19	354
	13	-0.22	153
	14	0.21	378
	15	-0.31	157
	16	-1.04	177
	17	1.02	383
	18	0.54	188
	19	1.95	193
	20	-1.26	167
	21	0.25	225
	22	0.31	205
	23	1.14	213
	24	0.58	244
	25	0.01	249
	26	-0.03	267
	27	-0.31	275
	28	-0.37	388
	29	-0.04	288
	30	1.24	297
	31	0.27	390
	32	0.9	302
	33	0.7	313
	34	-0.48	397
	35	1.02	402
MU.071	1	-0.51	59
	2	0.12	71
	3	0	367
	4	0.51	78
	5	-0.39	90
	6	0.66	96
	7	0.18	105
	8	-1.9	115
	9	-0.86	126
	10	-0.2	135
	11	-0.51	143
	12	-1.67	373
	13	0.13	149
	14	-0.39	375

Area	Year	Value	Total
	15	0.52	160
	16	-1.71	175
	17	0	382
	18	-0.4	185
	19	1.54	196
	20	0.4	168
	21	1.11	226
	22	-0.1	204
	23	-0.06	214
	24	0.41	241
	25	-0.21	250
	26	0.17	271
	27	0.27	280
	28	0.66	361
	29	-0.03	285
	30	0.5	298
	31	0.09	392
	32	0.45	305
	33	0.87	311
	34	-0.11	395
	35	0.45	404
MU.072	1	-1.29	60
	2	-0.03	68
	3	-0.55	352
	4	-0.55	77
	5	-1.62	88
	6	-0.47	93
	7	-0.46	107
	8	-2.06	115
	9	-0.11	124
	10	-1.28	134
	11	0.6	139
	12	-1.6	373
	13	-1.23	150
	14	0.16	376
	15	1.04	161
	16	0.04	178
	17	-0.06	380
	18	1.28	186
	19	2.01	194
	20	0.92	169
	21	0.35	221

Year	Revenue	Change	Total
	22	-0.84	207
	23	0.4	211
	24	1.41	239
	25	0.6	252
	26	-0.21	268
	27	-1.6	277
	28	-0.36	388
	29	1.17	284
	30	0.69	40
	31	1.35	391
	32	0.47	307
	33	0.27	308
	34	0.7	398
	35	0.86	400
MU.081	1	-0.1	1
	2	-0.98	69
	3	-0.04	368
	4	2.01	79
	5	-0.28	84
	6	-1.01	97
	7	1.84	104
	8	1.11	110
	9	0.18	121
	10	-0.04	130
	11	0.28	140
	12	-0.14	370
	13	1.23	151
	14	0.07	357
	15	-0.42	158
	16	0.67	177
	17	0.09	379
	18	0.58	187
	19	0	197
	20	0.49	170
	21	-0.35	224
	22	-1.26	201
	23	-1.01	215
	24	-0.47	243
	25	0.04	251
	26	-0.7	265
	27	0.37	275
	28	0.18	387

	29	0.15	287
	30	-0.52	293
	31	-0.41	394
	32	-0.9	303
	33	-0.22	314
	34	-0.62	399
	35	0.19	401
MU.082	1	-0.35	59
	2	0.45	70
	3	0.01	369
	4	-1.38	76
	5	-0.67	87
	6	-0.02	97
	7	-1.12	103
	8	-0.04	112
	9	-1.76	122
	10	-0.89	132
	11	1.57	138
	12	-0.11	371
	13	-0.64	152
	14	0.02	378
	15	0.82	161
	16	0.19	178
	17	0.66	381
	18	0.45	188
	19	-1.01	197
	20	-0.05	171
	21	1.38	220
	22	-1.16	203
	23	1.32	216
	24	0.32	238
	25	-0.06	253
	26	-0.26	269
	27	-0.77	276
	28	-0.31	360
	29	1.03	286
	30	1.38	294
	31	0.04	390
	32	0.45	305
	33	0.67	315
	34	-0.62	396
	35	0.45	366

MU.091	1	-2.28	49
	2	0.92	72
	3	-0.44	352
	4	-0.56	80
	5	-0.85	89
	6	-0.41	95
	7	0.53	108
	8	-2.82	116
	9	0.16	124
	10	-0.23	133
	11	-0.04	141
	12	-0.8	372
	13	0.23	149
	14	0.21	376
	15	0.46	160
	16	-0.19	176
	17	0.07	358
	18	0.84	189
	19	0.56	192
	20	0.03	166
	21	0.29	225
	22	-0.28	206
	23	0.77	210
	24	0.3	233
	25	0.41	259
	26	0.33	271
	27	0.59	278
	28	-0.7	385
	29	0.29	288
	30	0.85	295
	31	0.44	393
	32	0.66	304
	33	0.38	316
	34	0.04	364
	35	0.27	403
MU.092	1	-0.44	59
	2	-0.21	68
	3	-0.63	349
	4	1.03	75
	5	-0.57	87
	6	-0.02	97
	7	-0.33	105

	8	-0.32	114
	9	-0.9	125
	10	-1	134
	11	0.52	139
	12	0.44	353
	13	-0.01	147
	14	-0.51	323
	15	-1.21	162
	16	0.52	179
	17	0.03	359
	18	-0.21	184
	19	1.18	193
	20	-1.29	167
	21	0.39	222
	22	0.33	202
	23	0.41	217
	24	0.29	227
	25	0.24	261
	26	-0.22	270
	27	-0.16	279
	28	-0.39	387
	29	0.26	285
	30	0.6	296
	31	0.76	347
	32	0.58	306
	33	0.46	313
	34	0.17	330
	35	0.21	346
MU.101	1	-0.58	426
	2	-0.27	428
	3	-1.37	433
	4	-0.88	434
	5	-2.04	438
	6	0.27	442
	7	-0.4	445
	8	2.08	446
	9	0.12	451
	10	-0.91	452
	11	-0.15	456
	12	0.18	459
	13	0.21	461
	14	0.72	464

	15	-0.32	469
	16	0.19	470
	17	0.06	475
	18	0.98	478
	19	-0.76	481
	20	0.01	483
	21	-0.67	489
	22	0.56	491
	23	-1.71	536
	24	0	495
	25	0.61	497
	26	0.48	502
	27	-0.95	504
	28	-0.13	510
	29	1.13	517
	30	0.24	518
	31	0.37	523
	32	0.79	526
	33	0.6	527
	34	0.69	532
	35	0.86	534

LAMPIRAN 5

Hasil Simulasi Desain CAT Murni

The SAS System

14:00 wednesday, December 5, 2007 1

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
thetam3_0	100	-2.8536733	0.1948234	-3.0000000	-2.2197695
thetam2_7	100	-2.6208768	0.2917720	-3.0000000	-1.7486051
thetam2_4	100	-2.3810072	0.3369183	-3.0000000	-1.4336147
thetam2_1	100	-2.0595397	0.3784340	-2.8732305	-0.6705668
thetam1_8	100	-1.7780024	0.3128238	-2.4105722	-0.7953015
thetam1_5	100	-1.4754327	0.3131974	-2.2712765	-0.7912728
thetam1_2	100	-1.2309343	0.2920642	-1.8139337	-0.4731324
thetam0_9	100	-0.9518559	0.3097779	-1.5314091	-0.1830679
thetam0_6	100	-0.6446735	0.3274867	-1.5680153	0.0486519
thetam0_3	100	-0.3211490	0.2800860	-1.0825685	0.3696288
theta0	100	-0.000704501	0.3221073	-0.7034803	0.8949555
thetap0_3	100	0.3201998	0.3376539	-0.4449530	1.1596238
thetap0_6	100	0.6010776	0.3269633	-0.2354903	1.4565462
thetap0_9	100	0.9173407	0.2931282	0.1533050	1.5897214
thetap1_2	100	1.2465660	0.2482301	0.5438799	1.7203754
thetap1_5	100	1.5068425	0.2982919	0.5478199	2.3464588
thetap1_8	100	1.8109632	0.2852444	1.1279253	2.4087520
thetap2_1	100	2.0411857	0.3127111	1.1747379	2.8702003
thetap2_4	100	2.3705358	0.2838142	1.4999318	3.0000000
thetap2_7	100	2.6504184	0.2586723	1.9384294	3.0000000
thetap3_0	100	2.8643732	0.2314083	2.1255028	3.0000000

The SAS System

14:00 wednesday, December 5, 2007 2

Obs	theta	mean	std	bias
1	-3.0	-2.85367	0.19482	0.14633
2	-2.7	-2.62088	0.29177	0.07912
3	-2.4	-2.38101	0.33692	0.01899
4	-2.1	-2.05954	0.37843	0.04046
5	-1.8	-1.77800	0.31282	0.02200
6	-1.5	-1.47543	0.31320	0.02457
7	-1.2	-1.23093	0.29206	-0.03093
8	-0.9	-0.95186	0.30978	-0.05186
9	-0.6	-0.64467	0.32749	-0.04467
10	-0.3	-0.32115	0.28009	-0.02115
11	-0.0	-0.00070	0.32211	-0.00070
12	0.3	0.32020	0.33765	0.02020
13	0.6	0.60108	0.32696	0.00108
14	0.9	0.91734	0.29313	0.01734
15	1.2	1.24657	0.24823	0.04657
16	1.5	1.50684	0.29829	0.00684
17	1.8	1.81096	0.28524	0.01096
18	2.1	2.04119	0.31271	-0.05881
19	2.4	2.37054	0.28381	-0.02946
20	2.7	2.65042	0.25867	-0.04958
21	3.0	2.86437	0.23141	-0.13563

LAMPIRAN 6

Hasil Simulasi Desain CCAT

The SAS System 11:23 wednesday, December 5, 2007 1

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
thetam3_0	100	-2.9782516	0.1224753	-3.0000000	-1.9214721
thetam2_7	100	-2.9697773	0.1406474	-3.0000000	-2.0711354
thetam2_4	100	-2.9430432	0.1912681	-3.0000000	-1.9736811
thetam2_1	100	-2.7337601	0.4482943	-3.0000000	-1.4278298
thetam1_8	100	-2.6465899	0.5105649	-3.0000000	-0.7630173
thetam1_5	100	-2.4060170	0.6699788	-3.0000000	-0.7944084
thetam1_2	100	-1.9040200	0.8091526	-3.0000000	0.0467364
thetam0_9	100	-1.3114052	0.8727042	-3.0000000	0.4928958
thetam0_6	100	-0.8187608	0.7546599	-3.0000000	0.8995047
thetam0_3	100	-0.3800536	0.5536857	-2.8212750	0.6470656
theta0	100	-0.0852235	0.4592660	-1.2102237	0.8880128
thetap0_3	100	0.2612923	0.5274280	-1.5750836	1.5375170
thetap0_6	100	0.6510913	0.5894749	-0.9462390	3.0000000
thetap0_9	100	0.9271749	0.6099904	-0.9278657	3.0000000
thetap1_2	100	1.3142789	0.7034221	-0.2118653	3.0000000
thetap1_5	100	1.9507776	0.6168259	0.5271643	3.0000000
thetap1_8	100	2.2957262	0.6668996	0.7333852	3.0000000
thetap2_1	100	2.6042620	0.5129877	1.0017059	3.0000000
thetap2_4	100	2.8394647	0.2878707	1.8873060	3.0000000
thetap2_7	100	2.9329930	0.1960190	1.8051197	3.0000000
thetap3_0	100	2.9656704	0.1636172	1.9428901	3.0000000

The SAS System 11:23 wednesday, December 5, 2007 2

obs	theta	mean	std	bias
1	-3.0	-2.97825	0.12248	0.02175
2	-2.7	-2.96978	0.14065	-0.26978
3	-2.4	-2.94304	0.19127	-0.54304
4	-2.1	-2.73376	0.44829	-0.63376
5	-1.8	-2.64659	0.51056	-0.84659
6	-1.5	-2.40602	0.66998	-0.90602
7	-1.2	-1.90402	0.80915	-0.70402
8	-0.9	-1.31141	0.87270	-0.41141
9	-0.6	-0.81876	0.75466	-0.21876
10	-0.3	-0.38005	0.55369	-0.08005
11	-0.0	-0.08522	0.45927	-0.08522
12	0.3	0.26129	0.52743	-0.03871
13	0.6	0.65109	0.58947	0.05109
14	0.9	0.92717	0.60999	0.02717
15	1.2	1.31428	0.70342	0.11428
16	1.5	1.95078	0.61683	0.45078
17	1.8	2.29573	0.66690	0.49573
18	2.1	2.60426	0.51299	0.50426
19	2.4	2.83946	0.28787	0.43946
20	2.7	2.93299	0.19602	0.23299
21	3.0	2.96567	0.16362	-0.03433

LAMPIRAN 7

Hasil Simulasi Desain CBT

The SAS System 05:41 Thursday, December 5, 2007 1

The MEANS Procedure

variable	N	Mean	Std Dev	Minimum	Maximum
thetam_3.0	100	-2.6451295	0.5113724	-3.0000000	-0.9271099
thetam_2.7	100	-2.5042522	0.5590794	-3.0000000	-0.7106790
thetam_2.4	100	-2.4037898	0.5377786	-3.0000000	-1.2119948
thetam_2.1	100	-2.1280853	0.6326948	-3.0000000	-0.8405687
thetam_1.8	100	-1.8216335	0.5195930	-3.0000000	-0.5666390
thetam_1.5	100	-1.6618341	0.6028682	-3.0000000	-0.5486375
thetam_1.2	100	-1.3272520	0.5664306	-3.0000000	-0.3880463
thetam_0.9	100	-0.8638547	0.4548820	-2.2299264	0.3851738
thetam_0.6	100	-0.5935969	0.4113691	-2.3177119	0.3859739
thetam_0.3	100	-0.2531045	0.3212992	-1.2450924	0.5625522
theta_0.0	100	0.0659212	0.3395855	-0.7462727	1.0242066
thetap_0.3	100	0.2781742	0.3335244	-0.7401053	1.0753463
thetap_0.6	100	0.5651520	0.3071492	-0.1193880	1.4463984
thetap_0.9	100	0.9438338	0.3777062	-0.1324226	1.9479872
thetap_1.2	100	1.2514982	0.4032511	0.4117947	2.8589188
thetap_1.5	100	1.5309150	0.4700351	0.3773560	2.8873072
thetap_1.8	100	1.7506166	0.8141506	-3.0000000	2.8873072
thetap_2.1	100	1.9898928	0.8920098	-3.0000000	2.8873072
thetap_2.4	100	1.1812315	2.2011237	-3.0000000	2.8873072
thetap_2.7	100	0.9349964	2.4875966	-3.0000000	2.8873072
Thetap_3.0	100	0.5737247	2.7625822	-3.0000000	2.8873072

The SAS System 05:41 Thursday, December 5, 2007 3

Obs	theta	mean	std	bias
1	-3.0	-2.64513	0.51137	0.35487
2	-2.7	-2.50425	0.55908	0.19575
3	-2.4	-2.40379	0.53778	-0.00379
4	-2.1	-2.12809	0.63269	-0.02809
5	-1.8	-1.82163	0.51959	-0.02163
6	-1.5	-1.66183	0.60287	-0.16183
7	-1.2	-1.32725	0.56643	-0.12725
8	-0.9	-0.86385	0.45488	0.03615
9	-0.6	-0.59360	0.41137	0.00640
10	-0.3	-0.25310	0.32130	0.04690
11	-0.0	0.06592	0.33959	0.06592
12	0.3	0.27817	0.33352	-0.02183
13	0.6	0.56515	0.30715	-0.03485
14	0.9	0.94383	0.37771	0.04383
15	1.2	1.25150	0.40325	0.05150
16	1.5	1.53092	0.47004	0.03092
17	1.8	1.75062	0.81415	-0.04938
18	2.1	1.98989	0.89201	-0.11011
19	2.4	1.18123	2.20112	-1.21877
20	2.7	0.93500	2.34199	-1.76513
21	3.0	0.57372	2.53778	-1.92991

LAMPIRAN 8

Script Program SAS (Desain CAT Murni)

```
**** Impor data karakteristik soal ****/
PROC IMPORT OUT= WORK.Dataolah
            DATAFILE= "F:\DATA\BS_ISIP.xls"
            DBMS=EXCEL REPLACE;

RUN;

**** Pilih data untuk soal awal ****/
data dataawal;
set dataolah;
if b >= -0.5 & b <= 0.5;

**** Masuk algoritma CAT murni ****/
proc iml;

**** Subroutine fungsi peluang ****/
start peluang(data,theta);
  a = data[,4];
  b = data[,5];
  c = data[,6];
  P = c + (1 - c)/(1+exp(-1.7#a#(theta-b)));
  return(p);
finish peluang;

**** Subroutine fungsi inform ****/
start inform(data,theta);
  a = data[,4];
  b = data[,5];
  c = data[,6];
  tmp1 = 2.89#a##2#(1-c);
  tmp2 = c+exp(1.7#a#(theta-b));
  tmp3 = (1+exp(-1.7#a#(theta-b)))##2;
  inform = tmp1/(tmp2#tmp3);
  return(inform);
finish inform;

/** Subroutine membuang baris soal yang sudah ditampilkan **/
start hapus(modul,nis);
  n = nrow(modul);
  if n = 1 then modul = modul;
  else do;
    i = 0;
    do until(tmp=nis);
      i = i + 1;
      tmp = modul[i,2];
    end;
    if i = 1 then modul = modul[2:n,];
    else if i = n then modul = modul[1:(n-1),];
    else modul = modul[1:(i-1),]//modul[(i+1):n,];
  end;
  return(modul);
finish hapus;
```



```

/**** Subroutine dugaan ML ****/
start ML(jawaban,a,b,c,e,maxit);
  r = jawaban[,+];
  n = ncol(jawaban);
  if r = 0 then thetaduga0 = log(1/(n-1));
  else if r = n then thetaduga0 = log((n-1)/1);
  else thetaduga0 = log(r/(n-r));
  p = c + (1 - c)/(1+exp(-1.7#a#(thetaduga0-b)));
  iterasi = 1;
  do until (selisih<e | iterasi>maxit);
    tmp1 = a#(jawaban-p)#(p-c)/(p#(1-c));
    turunan1 = 1.7*tmp1[,+];
    tmp2 = a##2#((1-p)/p)#((p-c)/(1-c))##2;
    turunan2 = -1.7##2#tmp2[,+];
    thetaduga = thetaduga0 - turunan1/turunan2;
    if thetaduga > 3 then thetaduga = 3;
    if thetaduga < -3 then thetaduga = -3;
    selisih = abs(thetaduga-thetaduga0);
    thetaduga0 = thetaduga;
    p = c + (1 - c)/(1+exp(-1.7#a#(thetaduga0-b)));
    iterasi = iterasi+1;
  end;
  return(thetaduga);
finish ML;

ulangan = 100;
jumsoal = nrow(allmodul);
nis = j(21*ulangan,jumsoal,.);
modul = j(21*ulangan,jumsoal,.);
jawab = j(21*ulangan,jumsoal,.);
nsoal = j(21*ulangan,1,.);
informasi = j(21*ulangan,jumsoal,.);
msem = j(21*ulangan,jumsoal,.);
thetadugaml = j(ulangan,21,.);
ma = j(21*ulangan,jumsoal,.);
mb = j(21*ulangan,jumsoal,.);
mc = j(21*ulangan,jumsoal,.);
n1 = nrow(dataawal);
e = 0.0001;
maxit = 200;
it = 1;
it_theta = 1;
do theta = -3 to 3 by 0.3;
  do i = 1 to ulangan;

/***** SOAL PERTAMA *****/
  soal = 1;
  tmodul = allmodul;
  if n1 = 1 then nourut_soall = 1; else nourut_soall =
ceil(ranuni(1)*(n1-1));
  soall = dataawal[nourut_soall,];
  soaltampil = soall;
  jawaban1 = ranuni(1) <= peluang(soall,theta);
  nis[it,1] = soall[,2];
  modul[it,1] = 1;
  ma[it,1] = soall[,4];

```

```

    mb[it,1] = soall[,5];
    mc[it,1] = soall[,6];
    jawab[it,1] = jawaban1;
    tmodul = hapus(tmodul,nis[it,1]);
    if jawaban1 = 1 then thetaduga = 0.5;
    else thetaduga = -0.5;

/*****      ITERASI HINGGA JAWABAN RESPONDEN BERPOLA      *****/
do until (jawab[it,soal] <> jawab[it,soal-1]);
    soal = soal + 1;
    info = inform(tmodul,thetaduga);
    max_info = max(info);
    n_info = nrow(tmodul);
    j = 0;
    do until(tmp=max_info);
        j = j + 1;
        tmp = info[j,];
    end;
    soalt = tmodul[j,];
    soaltampil = soaltampil//soalt;
    nis[it,soal] = tmodul[j,2];
    modul[it,soal] = tmodul[j,3];
    ma[it,soal] = tmodul[j,4];
    mb[it,soal] = tmodul[j,5];
    mc[it,soal] = tmodul[j,6];
    jawabant = ranuni(1) <= peluang(soalt,theta);
    jawab[it,soal] = jawabant;
    if jawabant = 1 then thetaduga = thetaduga + 0.5;
    else thetaduga = thetaduga - 0.5;
    if thetaduga > 3 then thetaduga = 3;
    if thetaduga < -3 then thetaduga = -3;
    tmodul = hapus(tmodul,nis[it,soal]);
    end;
vjawaban = jawab[it,1:soal];
a = ma[it,1:soal];
b = mb[it,1:soal];
c = mc[it,1:soal];
thetaduga = ml(vjawaban,a,b,c,e,maxit);
info = inform(soaltampil,thetaduga);
informasi[it,1:soal] = info`;
sem = 1/sqrt(sum(informasi[it,1:soal]));
msem[it,soal] = sem;
if sem > 0.3 then do;

/*ITERASI HINGGA SEM < 0.3 DIAWALI PENDUGAAN THETA DGN METODE ML*/
do until(sem < 0.3 | soal = jumsoal);
    soal = soal + 1;
    info = inform(tmodul,thetaduga);
    max_info = max(info);
    n_info = nrow(tmodul);
    j = 0;
    do until(tmp=max_info);
        j = j + 1;
        tmp = info[j,];
    end;
    soalt = tmodul[j,];

```

```

        soaltampil = soaltampil//soalt;
        nis[it,soal] = tmodul[j,2];
        modul[it,soal] = tmodul[j,3];
        ma[it,soal] = tmodul[j,4];
        mb[it,soal] = tmodul[j,5];
        mc[it,soal] = tmodul[j,6];
        jawabant = ranuni(1) <= peluang(soalt,theta);
        jawab[it,soal] = jawabant;
        tmodul = hapus(tmodul,nis[it,soal]);
        vjawaban = jawab[it,1:soal];
        a = ma[it,1:soal];
        b = mb[it,1:soal];
        c = mc[it,1:soal];
        thetaduga = ml(vjawaban,a,b,c,e,maxit);
        info = inform(soalt,thetaduga);
        informasi[it,soal] = info;
        sem = 1/sqrt(sum(informasi[it,1:soal]));
        msem[it,soal] = sem;
    end;
end;

/*ITERASI KONVERGEN, DILANJUTKAN PENYIMPANAN NILAI DUGAAN THETA */
        nsoal[it,] = soal;
        thetadugaml[i,it_theta] = thetaduga;
        it = it + 1;
    end;
    it_theta = it_theta + 1;
end;
nama_soal = 'soal1' : 'soal404';
maxsoal = max(nsoal);
nis = nis[,1:maxsoal];
modul = modul[,1:maxsoal];
jawab = jawab[,1:maxsoal];
msem = msem[,1:maxsoal];
informasi = informasi[,1:maxsoal];
create work.nis from nis[colname=nama_soal];
append from nis;
create work.modul from modul[colname=nama_soal];
append from modul;
create work.jawab from jawab[colname=nama_soal];
append from jawab;
create work.msem from msem[colname=nama_soal];
append from msem;
create work.informasi from informasi[colname=nama_soal];
append from informasi;
create work.thetadugaml from thetadugaml[colname=nama_soal];
append from thetadugaml;
create work.nsoal from nsoal;
append from nsoal;
quit;

PROC EXPORT DATA= WORK.nis
        OUTFILE= "F:\DATA\CATM.xls"
        DBMS=EXCEL REPLACE;
        SHEET="nis";
RUN;

```

PERSONALIA PENELITIAN

1. Ketua Peneliti

- Nama Lengkap : Agus Santoso
- Jenis Kelamin : Laki-laki
- NIP : 19640217 19930 1 001
- Bidang Ilmu : Pengukuran Pendidikan
- Pangkat/Golongan : Penata (III/c)
- Jabatan Fungsional : Lektor
- Fakultas/Jurusan : FMIPA/Statistika
- Waktu Penelitian : 8 jam/minggu

2. Anggota Peneliti

- Nama Lengkap : Dwi Astuti Aprijani
- Jenis Kelamin : Perempuan
- NIP : 19670415 199802 2 001
- Bidang Ilmu : Matematika Terapan
- Pangkat/Golongan : Penata (III/c)
- Jabatan Fungsional : Lektor
- Fakultas/Jurusan : FMIPA/Matematika
- Waktu Penelitian : 5 jam/minggu

3. Anggota Peneliti

- Nama Lengkap : Unggul Utan Sufandi
- Jenis Kelamin : Laki-laki
- NIP : 19710911 199903 1 002
- Bidang Ilmu : Ilmu Komputer
- Pangkat/Golongan : Penata Tk. I (III/b)
- Jabatan Struktural : Penanggung Jawab Pengolahan Data PUSKOM
- Waktu Penelitian : 5 jam/minggu

4. Anggota Peneliti

- Nama Lengkap : Iswaya Maalik
- Jenis Kelamin : Laki-laki
- NIP : 19780926 200312 1 003
- Bidang Ilmu : Sistem Informasi
- Pangkat/Golongan : Penata Tk. I (III/b)
- Jabatan Fungsional : Staf ICT PUSKOM
- Waktu Penelitian : 3 jam/minggu