

PENGUJIAN HASIL BELAJAR DAN PENILAIAN PENDIDIKAN BERBANTUAN KOMPUTER

Djemari Mardapi, Haryanto, dan Samsul Hadi

FT Universitas Negeri Yogyakarta

email: haryanto.ftuny@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengetahui kinerja sistem pengujian hasil belajar berbantuan komputer dalam hal proses pemilihan butir-butir tes yang tepat bagi siswa serta menilai kemampuan hasil proses pelaksanaan program pembelajaran siswa dengan bantuan komputer. Jenis penelitiannya *Research and Development (R&D)*, yang terbagi atas pengembangan program pengujian dan penilaian berbantuan komputer serta pengujian kinerja program dalam proses pengujian kemampuan siswa SMA di DIY dan penilaian pelaksanaan program pembelajaran. Teknik pengambilan data dilakukan dengan observasi, dokumentasi, angket, dan pengujian dengan teknik analisis deskriptif kuantitatif dan evaluatif. Hasil penelitian meliputi *pertama*, dalam pengelolaan tes, program komputer mampu melakukan pengadministrasian *bank* soal, pengemasan butir-butir tes secara otomatis berdasar algoritma yang diberikan, pengemasan jumlah butir tes sesuai dengan kemampuan siswa, pengacakan letak jawaban benar pada alternatif pilihan jawaban dari masing-masing butir tes, dan penyimpanan rekaman hasil tes, baik secara individu maupun kelompok. *Kedua*, program omputer mampu memberikan penilaian terhadap kemampuan siswa, baik dalam pengujian maupun pelaksanaan proses pembelajaran secara otomatis.

Kata kunci: pengujian hasil belajar, penilaian pendidikan

TESTING LEARNING OUTCOMES AND ASSESSMENT COMPUTER AIDED EDUCATION

Abstract

This study broadly aims to determine the performance of the test system computer-aided and to assess students' learning outcomes in computer-assisted program implementation. This research was conducted with the approach of Research and Development (R & D), which is divided into the development of the test program and computer-assisted assessmen, and testing the performance of the program. Techniques of data retrieval is done by observation, documentation, questionnaires, and testing. Data analysis technique used is quantitative descriptive and evaluative. The results obtained include first, in the management of the test, a computer program capable of performing the administration of a question bank, packaging the test items automatically based algorithms are given, packaging of the test items in accordance with students' abilities, randomization of the true answer to the alternative choice of answers from each test item, and keeping records of test results either individually or in groups; and second, the computer program is capable of performing the testing and evaluation of the implementation learning process performed automatically by a computer program.

Keywords: testing the learning outcomes, educational assessment

PENDAHULUAN

Indikasi sekolah berkualitas pada umumnya dilihat dari banyaknya prestasi yang diraih. Misalnya, tingginya persentase kelulusan siswa dari ujian nasional, tingginya pencapaian nilai rata-rata hasil ujian nasional, persentase lulusan sekolah yang melanjutkan ke jenjang pendidikan berikutnya, dan persentase lulusan sekolah menengah kejuruan yang masuk ke dunia usaha dan dunia industri, persentase tingkat *drop out* atau siswa putus sekolah. Untuk itu, sekolah melalui berbagai upaya, berusaha meningkatkan kualitas pendidikan, misal melalui pendidikan dan latihan komputer untuk guru-guru, penggunaan media dan proses pembelajaran berbasis TI (Baumgartner & Jackson, 1995; Chee & Wong, 2003). Permasalahan yang perlu dikaji adalah seberapa besar penggunaan TI dapat meningkatkan kualitas hasil pendidikan?

Target keberhasilan peningkatan kualitas pendidikan, khususnya dalam pembelajaran melalui penilaian berbasis TI, juga telah mulai diupayakan (Chee & Wong, 2003). Penggunaan TI untuk keperluan tes menurut Chee & Wong (2003) dan Towndrow & Vallenge (2004), ditujukan untuk meningkatkan efektivitas dan efisiensi pelaksanaan dan penyelenggaraan tes. Era TI melalui *Computerized Based Testing* (CBT), komputer diprogram untuk mengolah urutan tampilan butir-butir soal ujian dan menerima pilihan jawaban dari peserta (Chee & Wong, 2003; Towndrow & Vallenge, 2004). Namun sejauh ini, mengenai bagaimanakah mengembangkan butir tes berbasis TI yang tepat merupakan permasalahan yang menarik untuk diteliti.

Berdasar observasi terhadap program yang digunakan oleh lembaga/instansi penyelenggara pengujian berbasis komputer, terdapat beberapa hal yang perlu untuk dikaji dan diteliti lebih lanjut. Proses acak pengambilan butir test dari basis data *bank* soal dalam pengujian berbasis komputer, masih terjadi ketidaksesuaian antara tingkat

kesulitan soal dengan kemampuan peserta tes. Di sisi lain, juga banyak dijumpai, bahwa pengambilan butir tes dari basis data *bank* soal dilakukan secara sekuensial. Hal itu menyebabkan urutan butir tes yang tampil dari waktu ke waktu tidak berubah, sehingga mudah untuk dihafalkan. Mekanisme pengambilan butir tes dari basis data *bank* soal secara berurutan, menurut teori pengukuran akan mengurangi validitas hasil tes.

Perkembangan terkini, seiring dengan usaha pembuatan bank soal dan peningkatan kualitas tes, teknologi komputer sangat membantu untuk keperluan tersebut (Hambleton, Swaminathan, & Rogers, 1991). Peningkatan kualitas tes dimaksudkan untuk menyesuaikan tingkat kesulitan butir tes dengan kemampuan peserta tes untuk tes dengan tujuan untuk mengetahui kemampuan siswa pada mata pelajaran tertentu. Pemilihan butir tes dengan menggunakan teknologi komputer disebut dengan istilah "*Computerized Adaptive Test/CAT*". Disebut *computerized* (terkomputerisasi) karena dalam pelaksanaan pengujian tidak lagi menggunakan "kertas dan pensil". *Adaptive* berarti tampilan butir-butir soal tes didasarkan pada hasil analisis dengan algoritma dalam sebuah program, agar butir-butir tes yang diberikan kepada peserta sesuai dengan kemampuannya. Kesesuaian butir soal didasarkan pada respons jawaban yang diberikan peserta tes.

Berbagai analisis statistik dengan berdasarkan teori pengukuran digunakan untuk mengestimasi kemampuan peserta tes, seperti teori tes klasik dan teori respons butir. Pada teori respons butir, estimasi kemampuan peserta tes dilakukan dengan Maximum Likelihood dan statistik Bayesian. Proses analisis dimulai dari butir tes yang pertama hingga butir tes yang terakhir. Analisis penalaran yang demikian akan memunculkan persepsi terbaru yang dapat mengubah persepsi yang telah ada, sehingga tindakan yang dipilih berdasarkan informasi yang lebih akurat.

Perkembangan di bidang komputasi dewasa ini maju dengan pesat. Berbagai teori evolusi komputasi untuk mendukung kebutuhan manusia terus-menerus diupayakan. Salah satu perkembangan tersebut untuk kepentingan tes adalah komputasi yang didasarkan dengan algoritma "*Artificial Intelligence (AI)*" atau kecerdasan buatan (Luger, 2005). Teori ini menggambarkan mengenai upaya membuat mesin komputer agar dapat belajar sehingga dapat bekerja dengan kecerdasannya seperti yang dimiliki manusia. Terkait dengan permasalahan *testing* hasil pembelajaran, komputer diprogram melalui pemberian *knowledge* untuk mengetahui kemampuan peserta tes, didukung dengan basis data *bank* soal, dapat dimunculkan butir tes yang tepat dengan kemampuan peserta tes. Analisis yang dilakukan komputer melalui program *artificial intelligence* adalah berupa sekumpulan *knowledge* berbasis penalaran monotonik. Uraian di atas memberi gambaran bahwa pendekatan AI untuk keperluan *testing* diduga memiliki prospek yang baik sehingga perlu dikaji dan diteliti lebih lanjut.

Berdasar uraian di atas, terdapat beberapa permasalahan yang perlu diteliti seperti berikut: 1) Keefektifan dan efisiensi penggunaan TI untuk mendukung pencapaian kualitas pendidikan belum banyak dikaji khususnya untuk mengungkap penilaian hasil-hasil pembelajaran dan pendidikan; 2) Proses acak dalam pengambilan butir tes dari basis data *bank* soal pada pengujian berbasis komputer belum mampu memilih tingkat kesulitan butir tes yang sesuai dengan kemampuan siswa; 3) Mekanisme pengambilan butir tes secara berurutan dari basis data *bank* soal dalam pengujian berbasis komputer belum mampu menjamin ketepatan hasil tes, 4) Belum dilakukan perbaikan teknik analisis pemilihan butir tes pada penalaran non-monotonik agar mampu mendeteksi persepsi terbaru/terakhir yang telah stabil, 5) Belum dilakukan upaya pengurangan beban komputasi pada perlakuan yang berulang-

ulang dalam teknik analisis pemilihan butir tes dengan menggunakan penalaran non-monotonik, 6) Belum ada kajian model sistem inferensi dengan pendekatan teknik AI yang tepat untuk digunakan dalam pemilihan butir-butir tes dalam program CAT, 7) Belum diketahui ketepatan program berbantuan komputer dengan algoritma AI untuk mendeskripsikan kemampuan siswa dan penilaian proses program pembelajaran/pendidikan, 8) Belum diketahui hasil kerja program berbantuan komputer dengan algoritma AI dalam memilih butir-butir tes yang tepat/sesuai dengan kemampuan siswa dan penilaian proses program pembelajaran/pendidikan.

Fokus penelitian ini adalah pengujian hasil belajar dan penilaian pembelajaran/pendidikan yang dilakukan melalui program terkomputerisasi dengan menggunakan algoritma AI. Untuk itu, maka penelitian ini dibatasi pada: 1) Ketepatan program komputer dengan algoritma AI dalam memilih butir-butir tes yang tepat dengan kemampuan siswa, 2) Kinerja program komputer dengan algoritma AI dalam mendeskripsikan kemampuan siswa dan penilaian pelaksanaan program pembelajaran/pendidikan.

Berdasar identifikasi dan batasan masalah di atas maka masalah penelitian ini dapat dirumuskan sebagai berikut: 1) Bagaimanakah hasil kerja program sistem pengujian hasil belajar dan penilaian pembelajaran/pendidikan berbantuan komputer dengan algoritma *Artificial Intelligence (AI)* dalam memilih butir-butir tes yang sesuai dengan kemampuan siswa?, 2) Bagaimanakah kinerja program pengujian dan penilaian berbantuan komputer dalam mendeskripsikan kemampuan siswa dan penilaian pelaksanaan program pembelajaran/pendidikan?

Tujuan penelitian ini adalah untuk menghasilkan: 1) sistem pengujian hasil belajar dan penilaian pendidikan melalui program berbantuan komputer dengan

algoritma AI untuk memilih butir-butir tes yang sesuai dengan kemampuan siswa dan melakukan penilaian terhadap proses pelaksanaan pembelajaran/pendidikan, 2) program pengujian hasil belajar dan penilaian pendidikan berbantuan komputer dengan algoritma AI yang teruji dengan baik untuk mendeskripsikan kemampuan siswa.

Manfaat utama penelitian ini adalah diperolehnya sistem ujian berbantuan program komputer yang lebih efektif dan efisien. Efektif karena sistem ujian tetap menggunakan kisi-kisi soal ujian berdasarkan kurikulum yang berlaku, sedang efisien lama waktu pengujian akan lebih singkat dengan informasi yang lebih akurat.

METODE

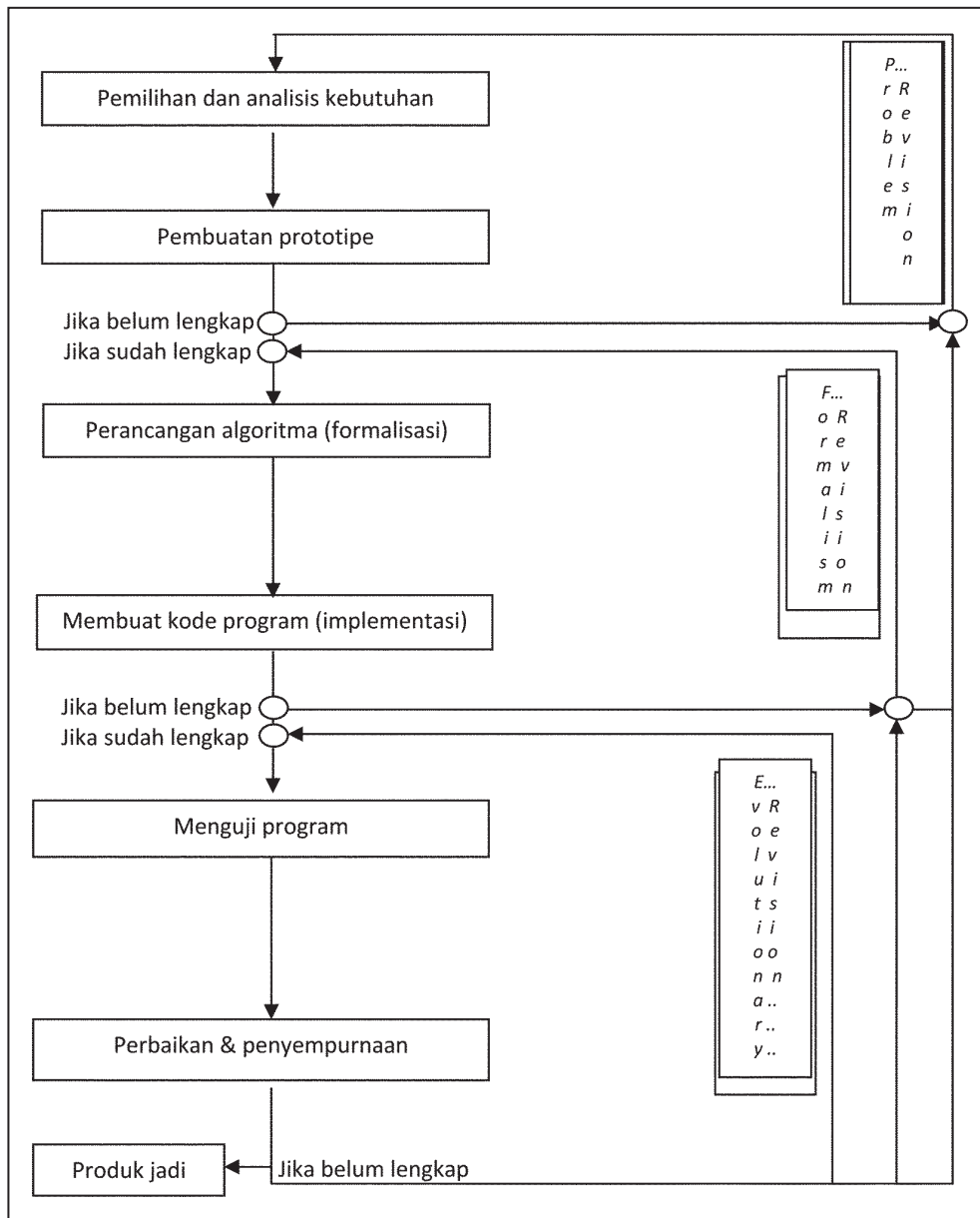
Penelitian ini menggunakan pendekatan penelitian *Research and Development*. Terdapat dua tahap dalam proses pelaksanaan, yaitu: tahap pengembangan produk dan tahap implementasi produk. Pada tahap pengembangan produk, proses yang dilakukan adalah mengembangkan perangkat butir tes mata pelajaran matematik SMA dan mengembangkan perangkat lunak model CAT dengan algoritma cerdas. Gambar 1 adalah model yang digunakan untuk mengembangkan produk CAT dengan algoritma cerdas, mengacu model Pressman (1997) dan Rolston (1988).

Pada Gambar 1, langkah (1) Pemilihan dan analisis kebutuhan sistem (perangkat lunak model CAT dengan algoritma cerdas), adalah: pengumpulan informasi yang berfungsi untuk *need assessment* sebagai dasar menyusun model. Berdasar informasi yang terkumpul, selanjutnya dibuat prototipe perangkat lunak. Pada langkah ini, proses akan kembali ke pemilihan dan analisis kebutuhan, jika dalam pembuatan prototipe masih terdapat kekurangan informasi. Langkah pertama ini disebut dengan *problem revision*. Pada langkah ini dilakukan secara terus menerus dalam rangka memperoleh permasalahan yang

representatif. Proses akan berlanjut ke langkah berikutnya bila ruang lingkup permasalahan yang akan diselesaikan telah terpenuhi. Langkah (2) Perancangan algoritma, yaitu membuat sistematika urutan kerja program perangkat lunak yang didasarkan pada langkah (1).

Proses selanjutnya adalah menerjemahkan algoritma tersebut ke dalam kode program. Pada langkah ini, proses akan kembali ke perancangan algoritma, jika dalam pembuatan kode program terdapat algoritma yang kurang sesuai. Langkah ini disebut dengan *formalism revision*. Proses juga dapat berulang kembali menuju langkah pertama (*problem revision*) jika terdapat algoritma yang kurang sesuai disebabkan adanya informasi yang belum lengkap pada langkah (1). Proses akan berlanjut ke langkah berikutnya, bila target yang dikehendaki telah tercapai, yaitu dihasilkan sebuah program yang mampu digunakan untuk menyelesaikan masalah.

Langkah (3) adalah pengujian perangkat lunak (*debugging*) untuk menemukan kesalahan yang mungkin terjadi. Dalam hal ini terdapat tiga kemungkinan kesalahan yang dapat terjadi yaitu: *syntax error* (kesalahan kalimat), *run time error* (kesalahan saat dijalankan), dan *logic error* (kesalahan hasil ditinjau dari penalaran logika). Berdasarkan kesalahan-kesalahan tersebut yang ditemukan saat pengujian, dijadikan sebagai dasar proses perbaikan dan penyempurnaan program. Pada langkah ketiga ini, proses akan kembali ke pengujian program, jika masih terjadi kesalahan yang menyebabkan program belum berfungsi seperti yang diharapkan. Langkah ketiga ini disebut dengan *evolutionary revision*. Proses juga dapat berulang kembali menuju langkah kedua (*formalism revision*) jika terjadi kesalahan program disebabkan oleh adanya algoritma dan penulisan kode program yang belum sesuai pada langkah (2). Atau bahkan Proses juga dapat berulang kembali menuju langkah pertama (*problem revision*) jika adanya kesalahan algoritma dan



Gambar 1. Pengembangan Model CAT dengan Algoritma Cerdas

kode program yang kurang sesuai disebabkan adanya informasi yang belum lengkap pada langkah (1).

Langkah-langkah tersebut, merupakan proses siklus hidup yang digunakan untuk pengembangan perangkat lunak model CAT dengan algoritma cerdas. Siklus hidup pengembangan program tersebut mengharuskan proses berulang, jika sampai pada langkah tertentu masih ditemukan

adanya kesalahan. Proses akan berulang pada bagian yang mengandung kesalahan. Dalam pengembangan program juga dimungkinkan proses berulang dimulai dari langkah awal. Siklus tersebut dilakukan terus menerus, hingga diperoleh perangkat lunak yang secara operasional dapat berfungsi sesuai dengan tujuan yang telah ditentukan.

Tahap kedua dari penelitian ini adalah menerapkan produk dalam situasi yang

sesungguhnya di kelas. Untuk menerapkan produk mengikuti langkah-langkah yang dikemukakan oleh Borg & Gall (1983), termasuk di dalamnya prosedur untuk pengembangan produk melengkapi model Pressman (1997) dan Rolston (1988) yang telah diurai di atas, yaitu: langkah (1) *Research and information collecting*, yakni melakukan *review literature*, observasi terhadap kelas yang akan digunakan untuk implementasi, dan mempersiapkan pelaksanaannya. Langkah (2) *Planning*, meliputi: mendefinisikan *skill* yang diamati, menentukan urutan-urutan tujuan yang hendak dicapai, dan menguji kemudahan dalam pelaksanaannya. Langkah (3) *Develop preliminary form of product*, dalam hal ini adalah mempersiapkan materi instruksional yang akan diberikan (memilih, mengorganisir, dan mengemas materi), mempersiapkan buku pegangan yang digunakan untuk pembelajaran, dan mempersiapkan peralatan evaluasi yang akan digunakan untuk mengukur keberhasilan pencapaian tujuan. Langkah (4) *Preliminary field testing*, adalah menggunakan produk dalam situasi sebenarnya di kelas. Produk diterapkan di kelas dengan melibatkan 12 peserta didik.

Proses yang dilakukan dalam langkah ini adalah *interview* mengenai penggunaan produk, observasi terhadap hasil dan dampak penggunaan produk, mengumpulkan data menggunakan berbagai alat, dan menganalisis data dan hasil pengamatan. Pada langkah ini merupakan pengujian produk pada sisi pengguna, yaitu: pengguna pertama/ pengajar (*first user*) dengan *alpha testing*, dan pengguna akhir/peserta didik (*end user*) dengan *beta testing*. Pengujian tersebut berupa verifikasi dan validasi produk. Langkah (5) *Main product revision*, merupakan proses perbaikan terhadap produk berdasar informasi hasil analisis data pada perlakuan langkah keempat. Dalam hal ini produk disempurnakan berdasarkan masukan-masukan yang diperoleh.

Selanjutnya langkah (6) *Main field testing*, yakni menggunakan produk hasil perbaikan di dalam suasana pembelajaran di kelas. Produk diterapkan di kelas dengan melibatkan 64 peserta didik. Dalam hal ini dilakukan kembali *interview*, observasi, pengumpulan data, dan analisis terhadap data yang terkumpul. Langkah (7) *Operational product revision*, adalah melakukan kembali proses perbaikan dan penyempurnaan produk dengan mengacu pada masukan-masukan hasil analisis data yang terkumpul. Langkah (8) *Operational field testing*, adalah menggunakan kembali hasil produk yang telah mengalami perbaikan. Produk diterapkan di kelas dengan melibatkan 40 hingga 200 peserta didik. Untuk melihat hasil dan dampak program, kembali dilakukan *interview*, observasi, pengumpulan data, dan analisis data. Langkah (9) *Final product revision*, merupakan perbaikan produk terakhir agar diperoleh produk yang lebih sempurna. Langkah (10) *Disemination and implementation*, yakni melaporkan produk terakhir yang telah sempurna dan telah teruji melalui beberapa tahapan serta melaporkan hasil temuan yang diperoleh dan siap untuk diimplementasikan pada skala pemakaian yang lebih luas.

Penelitian ini dilakukan di Laboratorium Komputer PEP Pascasarjana UNY, untuk proses pengembangan/*development* produk program CAT. Siswa sekolah menengah sebagai tempat untuk implementasi/*research* produk program CAT dalam situasi yang sebenarnya. Waktu penelitian, dilakukan pada bulan April 2011 sampai dengan Juni 2011 untuk pembuatan program CAT. Bulan Juli sampai dengan September 2011 dilakukan uji implementasi atau pengujian program CAT tahap pertama.

Subjek yang digunakan dalam penelitian untuk uji coba meliputi guru mata pelajaran yang diujinasionalkan khususnya guru matematika untuk proses penggalian informasi dan identifikasi kebutuhan sistem yang diperlukan dalam program CAT dan

menilai produk program CAT. Di samping itu, juga untuk memperoleh informasi mengenai pemilihan materi tes untuk uji coba. Siswa kelas XI, untuk uji fungsionalitas dan unjuk kinerja produk program CAT. Mata pelajaran yang digunakan adalah matematika.

Data yang diperlukan meliputi data kualitatif berupa fungsi-fungsi yang diperlukan program, calon pengguna program, standar minimal perangkat keras yang digunakan program, dan hasil uji internal terhadap fungsionalitas program. Data berikut adalah data kuantitatif berupa tanggapan guru terhadap produk program CAT dan hasil unjuk kerja program dalam proses pengujian hasil belajar siswa.

Penelitian meliputi dua kegiatan, yaitu: (a) pengembangan produk model CAT dengan algoritma cerdas dan (b) implementasi hasil produk dalam situasi nyata di kelas. Teknik pengumpulan data meliputi: Untuk pengembangan produk, menggunakan: (1) observasi, mengenai: ketepatan instruksi (*syntax error*), ketepatan proses (*run time error*), ketepatan hasil (*logic error*), dan verifikasi dan validasi produk (*white/black box testing*). (2) kuesioner dan wawancara, mengenai: kelengkapan dan ketepatan fungsi perangkat lunak. (3) dokumentasi, mengenai: materi, bentuk dan model tes. Untuk implementasi produk, menggunakan: (1) observasi, mengenai: kebenaran, ketepatan, fungsionalitas, dan kemampuan produk dalam menguji kemampuan siswa dan menilai proses pembelajaran, dan (2) dokumentasi, mengenai: data nilai prestasi belajar siswa.

Teknik analisis data yang digunakan dalam rangka menjawab rumusan masalah pada pendahuluan yang kemudian dirinci dalam pertanyaan penelitian yang diajukan dalam penelitian ini adalah teknik analisis deskriptif kuantitatif dan evaluatif. Pada tahap pertama, penelitian ini menguji kelayakan produk *software* yang digunakan untuk mengevaluasi kemampuan siswa peserta uji melalui pengukuran terkomputerisasi dengan

algoritma cerdas. Teknik analisis deskriptif evaluatif dilakukan untuk menentukan kelayakan, kemampuan dan efektivitas kinerja produk dalam fungsinya mengukur kemampuan siswa peserta uji.

Tahap kedua, fokus penelitian adalah pada penerapan program pengujian adaptif terkomputerisasi dengan algoritma cerdas untuk proses pengujian dan penilaian kemampuan peserta didik sebagai hasil proses pembelajaran/pendidikan. Teknik analisis deskriptif kuantitatif dilakukan untuk mengetahui gambaran tingkat kemampuan siswa peserta uji yang diukur menggunakan tes terkomputerisasi dengan algoritma cerdas. Dalam penelitian ini, kriteria kemampuan dideskripsikan ke dalam: (1) kemampuan siswa tidak lulus atau belum tuntas/di bawah batas ketuntasan belajar, (2) kemampuan siswa lulus baik, yaitu tuntas belajar/di atas atau sama dengan batas ketuntasan belajar, dan (3) kemampuan siswa lulus sangat baik, tuntas sangat baik. Hasil pengujian hasil belajar ini berupa skor yang menunjukkan tingkat kemampuan peserta uji.

HASIL PENELITIAN DAN PEMBAHASAN

Proses yang terjadi di dalam program CAT dimulai dari peserta tes melakukan login, mengerjakan tes, hingga memperoleh hasil tes. Adapun skenarionya sebagai berikut. *Pertama*, memilih 2 butir soal dengan nilai b dalam interval $-0.5 < b < +0.5$ (*default*), catat semua informasi tentang identitas butir soal, termasuk kunci dan parameter a , b , dan c . *Kedua*, menampilkan kedua butir soal tersebut satu per satu. Ambil respon dari *testee* (berupa ABCDE), selanjutnya *confirm* respon tersebut dengan kunci untuk memperoleh skor. Skor=1 jika benar, skor=0 jika salah. *Ketiga*, memilih soal berikutnya berdasarkan pola skor. (Kemungkinan skor yang diperoleh adalah 00, 01, 10, atau 11); jika SKOR = 00 → ambil soal dengan b awal dikurangi 0.2; jika skor 01 atau 10 → ambil soal lain dalam rentang seperti pada 2 butir

pertama; jika skor 11 → ambil soal dengan b awal ditambah 0.2. *Keempat*, menampilkan soal yang terpilih, ambil respon dari testee, selanjutnya *confirm* dengan kunci untuk memperoleh skor. *Kelima*, kemungkinan skor yang diperoleh adalah 000, 001, 010, 100, ..., dan seterusnya, atau 111. *Keenam*, menghitung THETA (3 skor) dengan rumus MLE, dan hitung SEM. *Ketujuh*, emilih soal berikutnya berdasarkan jawaban butir ke-3; jika SALAH (skor=0) → ambil soal dengan b terakhir dikurangi 0.20; jika BENAR (skor=1) → ambil soal dengan b terakhir ditambah 0.20. *Kedelapan*, menampilkan soal yang terpilih, ambil respon dari testee, selanjutnya *confirm* dengan kunci untuk memperoleh skor. *Kesembilan*, menghitung nilai THETA dari 4 skor terakhir. Demikian juga SEM. *Kesepuluh*, LOOP (mengulang); mengulangi Langkah f-i dan AKHIRI TES jika: (1) jumlah butir soal sudah mencapai

15 atau (2) nilai SEM lebih kecil atau sama dengan 0.33; mencatat transaksi, nilai THETA terakhir, dan nilai SEM terakhir, khusus untuk CONTENT BASED CAT, ditambahkan syarat PEMERATAAN SK secara proporsional.

Penetapan kriteria dilakukan sebelum pelaksanaan ujian. Ada dua pilihan dalam penetapan kriteria, yaitu: (1) menggunakan DEFAULT atau (2) merubah nilai DEFAULT. Jika tidak ada perubahan nilai, maka pertanyaan berikut dijawab No (Gambar 2).

Metode estimasi theta yang digunakan dalam studi ini adalah *Maximum Likelihood Estimation* (MLE). Contoh dapat dilihat pada Gambar 3. Gambar 4 adalah contoh estimasi theta.

Fungsi Informasi Butir (*Item Information Function=IIF*) dapat dilihat pada Gambar 5. *Test Information Function & Standard Error Measurement* (TIF & SEM) diperoleh dengan menjumlahkan seluruh IIF

QUESTION:
DO YOU WANT TO CANGE
ANY CRITERION?

YES → Edit/Entry
the value

NO → Use the value
Of DEFAULT

Component	DEFAULT	ALTERNATIVE (example)	REMARK
Initial THETA	0.00 ± 0.50	-1.00 ± 0.50	Misal ada info ttg perkiraan kemampuan testee 1SD di bawah rata-2 → -1.00 ± 0.50
Content Based	No	Yes	Yes= SK/KD dijadikan pertimbangan dalam NEXT ITEM → <i>content balancing</i>
Scale Transformation	UMEAN=1 USCALE=0	UMEAN=55 USCALE=15	Skala asli dengan interval -3 s/d +3 diubah menjadi 10 s/d 100
Time Constraint	No	Yes	Yes= maximum duration setiap butir soal di-AKTIF-kan
Max Item	15	20	Stopping rules-1
Min SEM	0.30	0.40	Stopping rules-2
D Value	1.7	1.0	Konstanta untuk fungsi logistik. Dengan D=1.7, kurva akan mendekati OGIVE
Next_Value	0.2	0.1	Penambahan/pengurangan nilai b dalam pemilihan NEXT ITEM

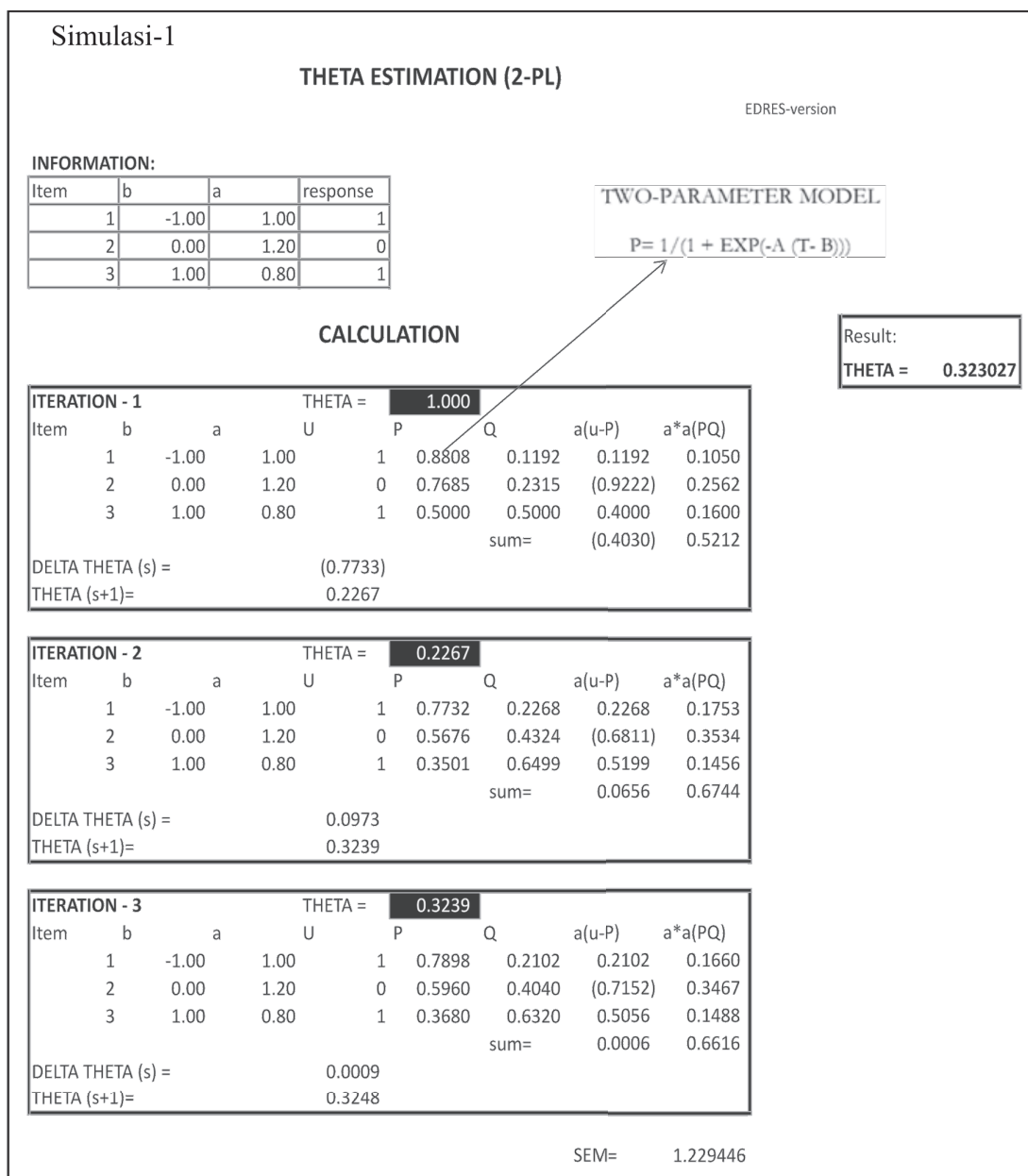
Gambar 2. Penetapan Kriteria Butir

dalam satu tes. *Standard Error Measurement* (SEM) = $1/(\text{SQRT}(\text{TIF}))$. Contoh dapat dilihat pada Gambar 6.

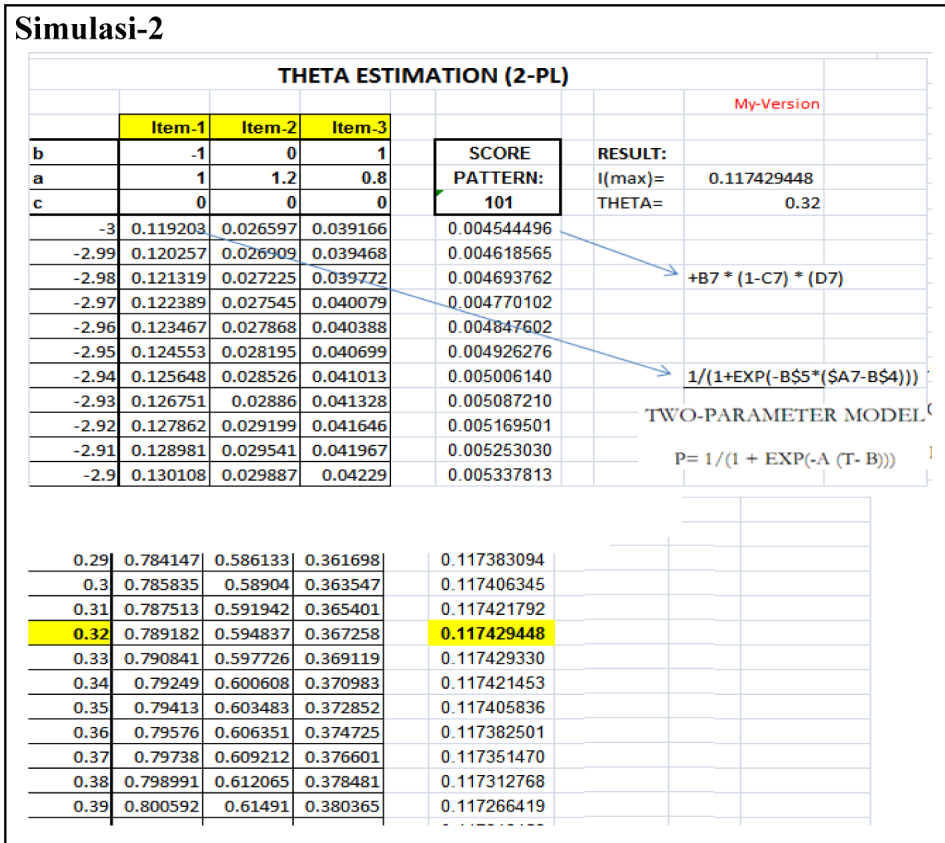
User Interface untuk *Testee*. Setelah proses *login*, peserta tes (*testee*) menjawab pertanyaan atau soal dengan tampilan pada Gambar 7. Demikian seterusnya, dan tes akan berakhir ketika kriteria STOPPING RULES terpenuhi.

Setiap akhir pelaksanaan tes dengan CAT, untuk setiap peserta diperoleh laporan seperti dalam Gambar 8. Informasi tersebut dapat digunakan untuk menelusuri dan mengevaluasi apakah algoritma valid secara psikometris dan matematis.

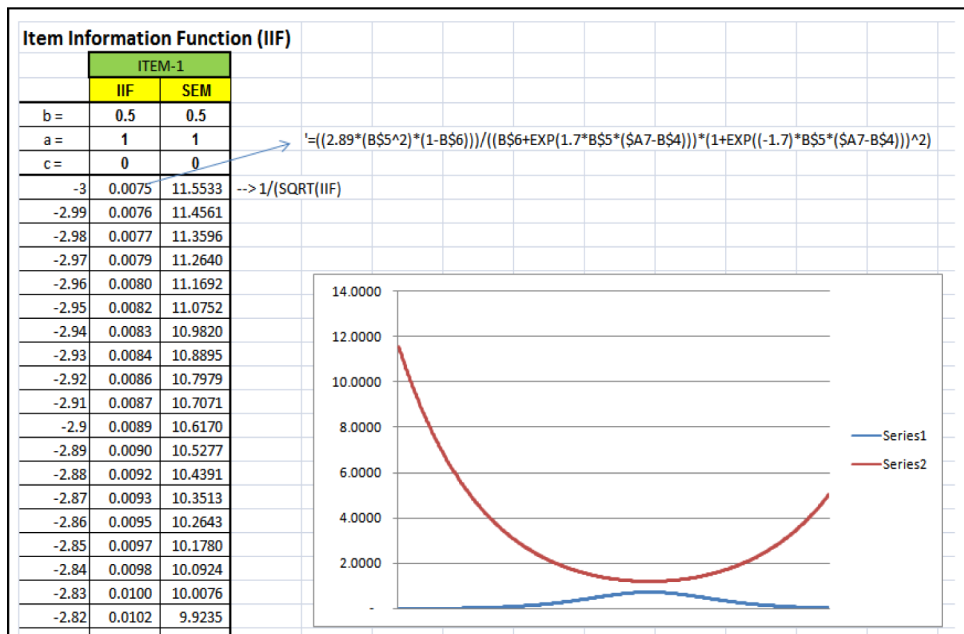
Secara umum algoritma CAT (*initial/ the first item, estimation & next item*), dan



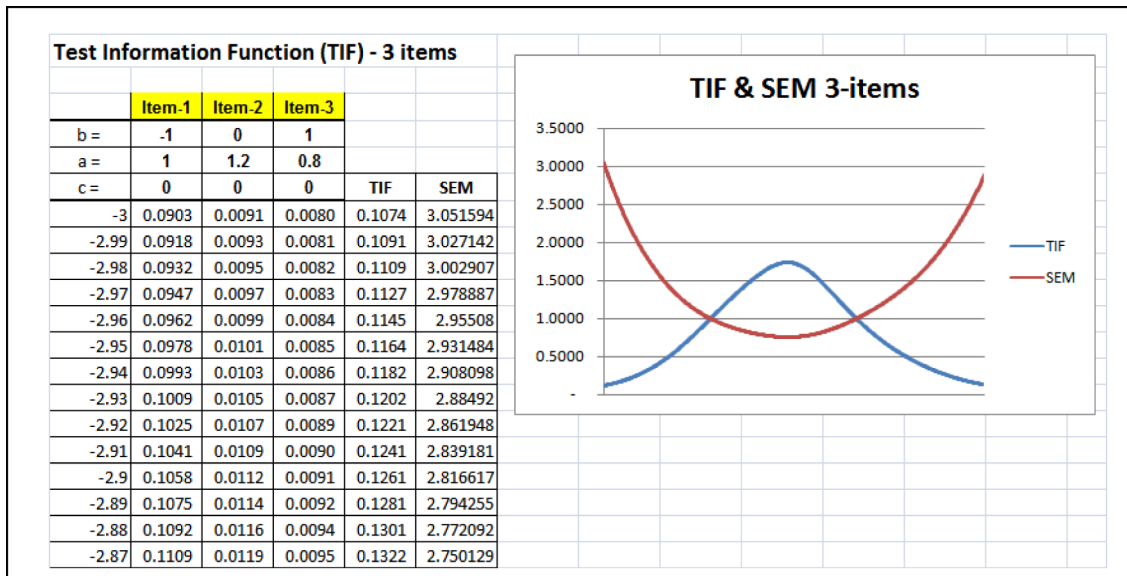
Gambar 3. Estimasi Kemampuan dengan Maximum Likelihood



Gambar 4. Estimasi Theta



Gambar 5. Fungsi Informasi Butir



Gambar 6. Fungsi Tes Informasi

CAT #1

Suku kedua dari barisan geometri adalah $\frac{3}{2}$. Jika suku kelima barisan tersebut adalah $\frac{3}{16}$, maka suku pertamanya adalah

A. 12
 B. 6
 C. 3
 D. 2
 E. 0,5

Soal no 2 dari 15

Gambar 7. Antarmuka Pemakai

stopping rules) sudah berjalan sesuai dengan rancangan. Namun demikian, masih ada sedikit masalah dengan perhitungan atau kalkulasi sebagai berikut. *Pertama*, hasil estimasi theta dengan MLE adalah BENAR dengan membandingkan hasil perhitungan EXCEL, bahkan lebih akurat karena menggunakan 4 digit desimal. *Kedua*, hasil perhitungan IIF sudah BENAR, tetapi ketika

menghitung TIF yang merupakan akumulasi dari IIF, BENAR untuk butir pertama sampai dengan ketiga, tetapi pada butir ke-4 tampak ada KESALAHAN. Untuk itu perlu dicek dan dievaluasi kembali. *Ketiga*, disamping itu, tampilan pada RAPORT perlu diurutkan berdasarkan nomor urut butir soal yang tampil, serta perlu ditambahkan field SUMBER untuk menelusuri butir soal yang dipilih.

No	A	B	C	RES	KEY	Theta	IIF	TIF	SEM
3	1	0.06	0	1	C	2.2555	0.0225	0.76	1.1471
2	1	0.06	0	0	C	-2.7878	0.7206	0.7403	1.1622
1	1	-0.11	0	0	A	-2.7878	0.0197	0.0197	7.1247
4	1	0.07	0	1	A	-2.3562	0.066	1.52	0.8111
5	1	0.08	0	0	A	1.7393	0.0453	3.04	0.5735
6	1	0.09	0	1	C	-0.8963	0.1533	6.08	0.4056
7	1	-0.2	0	0	B	-0.2141	0.3835	12.16	0.2868
8	1	-0.3	0	0	B	-0.5368	0.7224	24.32	0.2028

Gambar 8. Hasil Tes Berbantuan Komputer

SIMPULAN

Model CAT dengan algoritma cerdas mampu memilih butir-butir tes dengan tingkat kesulitan yang sesuai dengan benar-salah respons jawaban siswa. Di samping itu, juga mampu mengatur banyak butir tes yang diberikan kepada siswa sesuai dengan tingkat kemampuannya. Sistem inferensi model CAT dengan algoritma penalaran dalam memilih butir-butir tes adalah mampu mengatur bahwa (a) Siswa dengan kemampuan tinggi memperoleh butir-butir tes dengan tingkat kesulitan tinggi. (b) Siswa dengan kemampuan sedang memperoleh butir-butir tes dengan tingkat kesulitan sedang, dan (c) Siswa dengan kemampuan rendah memperoleh butir-butir tes dengan tingkat kesulitan rendah. Ketepatan sistem inferensi dalam memilih butir-butir tes yang sesuai dengan respons jawaban siswa diwujudkan pada nilai. Kemampuan model CAT dengan algoritma penalaran dalam mendeskripsikan kemampuan siswa berdasar butir-butir tes yang diterima dan nilai-nilai estimasi hasil

analisis. Hasil temuan penelitian ini, adalah: (1) model CAT dengan algoritma penalaran digunakan untuk membentuk sistem inferensi. (2) Sistem inferensi model CAT, dengan algoritma penalaran: (a) Mampu mengelola basis data *bank* soal yang digunakan untuk proses pengujian terkomputerisasi. (b) Mampu memilih secara adaptif butir-butir tes yang sesuai dengan kemampuan siswa berdasar pada respons jawaban yang diberikan dalam proses pengujian terkomputerisasi.

Penelitian yang dilakukan adalah untuk mengetahui kemampuan kognitif, pada mata pelajaran Matematika. Berdasar silabi, untuk setiap pokok bahasan, kemampuan kognitif yang diungkap meliputi tiga tingkatan, yaitu: pengetahuan, pemahaman, dan aplikasi. Dengan demikian, tiga tingkatan kemampuan kognitif yang selanjutnya belum dapat diungkap dan hal ini menjadi keterbatasan penelitian ini. Program CAT yang dikembangkan digunakan untuk mengetahui kemampuan kognitif. Oleh karena itu, untuk mengungkap kemampuan afektif

dan psikomotorik/keterampilan menjadi keterbatasan penelitian ini.

Jenis dan model tes yang biasa dikembangkan dalam proses pengujian memiliki variasi yang banyak. Antara lain: tes dengan jawaban pendek, tes dengan jawaban panjang, tes menjodohkan, tes melengkapi, tes mengurutkan, tes benar-salah, tes sebab akibat, dan tes pilihan ganda. Untuk itu, keterbatasan penelitian ini adalah belum mampu menangani semua jenis dan model tes tersebut. Jenis tes yang dikembangkan dalam penelitian ini terbatas hanya untuk tes pilihan ganda.

Program CAT dapat digunakan untuk butir tes modern, bila pada bagian instruksi untuk perhitungan matematis parameter butir tes diubah dan disesuaikan dengan parameter butir tes modern. Untuk itu, estimasi kemampuan (θ) dan $p(\theta)$ ke p untuk menggambarkan kemampuan yang sebenarnya, keakuratannya masih perlu ditingkatkan dalam penelitian ini.

Proporsi perbandingan jumlah butir yang memenuhi klasifikasi (rentang) tingkat kesulitan dan daya beda juga berpengaruh terhadap kinerja program. Dalam penelitian ini, basis data *bank* soal yang digunakan terbatas sejumlah 20 butir untuk masing-masing tingkat kesulitan. Jumlah basis data *bank* soal yang baik tentunya yang memenuhi batas kecukupan, yakni jika soal yang diujikan 20 butir maka jumlah minimal basis data *bank* soal adalah 60 butir dan masing-masing tingkat kesulitan minimal memiliki 20 butir tes.

Jumlah basis pengetahuan yang digunakan untuk membangun sistem inferensi dalam penelitian ini adalah empat *rule*. Jumlah basis pengetahuan berpengaruh terhadap ketepatan hasil inferensi. Belum ada aturan untuk jumlah basis pengetahuan yang baik. Menurut peneliti, jumlah basis pengetahuan sebaiknya yang memenuhi kecukupan untuk membangun sistem inferensi. Dalam penelitian ini, basis pengetahuan yang

dibangun dibatasi pada empat *rule*, karena telah memenuhi kemungkinan-kemungkinan kombinasi klasifikasi tingkat kesulitan butir yang dibentuk.

Pengujian berbasis komputer dengan model CAT dapat dilakukan bila fasilitas komputer mencukupi untuk sejumlah peserta (satu komputer satu peserta), dan peserta telah memiliki literasi tentang komputer. Berdasar hasil penelitian, hal ini menjadi keterbatasan yang cukup besar. Literasi komputer yang dimaksud adalah: siswa mampu bekerja dengan komputer yang meliputi: pengetahuan mematikan dan menyalakan komputer sistem *windows*, pengetahuan sistem *login*, pengetahuan model navigasi.

Keterbatasan lain adalah dalam penggunaan model CAT ini, bahwa peserta tidak diijinkan ke butir tes awal untuk melakukan perubahan. Butir tes yang sudah dikerjakan tidak dapat diubah dan tidak dapat dilihat lagi. Karena hal itu akan berpengaruh pada prediksi kemampuan yang telah dilakukan. Peserta harus telah yakin terhadap jawaban yang dipilih pada alternatif pilihan jawaban. Untuk itu, model CAT menyediakan fasilitas *pass* untuk melompat/*skip* ke butir tes yang selanjutnya. Butir tes yang di *skip* tidak akan dihitung dalam penentuan skor.

Program CAT menggunakan algoritma penalaran merupakan salah satu alternatif upaya membangun perangkat lunak untuk keperluan pengujian. Mengingat dasar pemilihan butir tes dilakukan dengan cara menganalisisnya terlebih dahulu, maka sangat dimungkinkan hasil analisis tidak sama dengan tipe butir tes di dalam basis data *bank* soal. Untuk itu sangat disarankan agar tipe-tipe butir tes diperbanyak jumlahnya.

Klasifikasi tingkat kesulitan butir tes dalam penelitian ini dibagi dalam tiga kelompok, yaitu: tingkat kesulitan tinggi, tingkat kesulitan sedang, dan tingkat kesulitan mudah. Oleh sebab itu, jumlah butir-butir tes yang memenuhi syarat untuk dijadikan basis

data *bank* soal jumlahnya harus memenuhi ketiga kelompok tersebut. Jika banyak soal (N) yang akan diujikan adalah 20 butir tes, maka dalam basis data *bank* soal minimal adalah butir sulit dan butir mudah masing-masing 20 butir tes dan 40 butir tes sedang, sehingga total 80 butir tes. Dengan rincian 20 butir tes tingkat kesulitan tinggi, 40 butir tes tingkat kesulitan sedang, dan 20 butir tes tingkat kesulitan rendah. Syarat butir tes yang bagus minimal memiliki parameter tingkat kesulitan butir dengan kisaran nilai -3 sampai 3 untuk tes modern dan daya beda butir 0,0 hingga 0,4 serta tebakan 0,00 sampai 0,2 (jika ada 45 opsi jawaban per butir).

Program CAT menggunakan algoritma penalaran merupakan model pengujian terkomputerisasi, sehingga dalam pelaksanaannya harus dilakukan di dalam laboratorium komputer. Dengan demikian disarankan agar sebelum menyelenggarakan pengujian dengan program CAT ini, sekolah harus memiliki minimal sebuah laboratorium komputer yang memenuhi standard minimal (spesifikasi standard komputer minimal ada di lampiran). Di samping itu, siswa yang akan diuji kemampuannya menggunakan program CAT ini, seyogyanya dilatih terlebih dahulu cara mengoperasikan dan bekerja dengan komputer. Hal itu dimaksudkan agar dalam mengerjakan ujian terkomputerisasi dengan program CAT, siswa telah terbiasa (*familier*) menggunakan komputer. Perangkat keras dengan memori 256 Mb dan kecepatan di bawah 1 GHz, juga dapat mempengaruhi kinerja program, sehingga disarankan dalam menggunakan program CAT agar memenuhi standard minimal perangkat keras.

Perlu dilakukan penelitian lebih lanjut dengan memperhatikan pembagian klasifikasi tingkat kesulitan butir, daya beda butir yang lebih sempit, dan basis pengetahuan

(*knowledge/rule base*) yang memadai, agar respons butir yang dimunculkan program CAT menjadi lebih halus. Sempitnya range klasifikasi tingkat kesulitan butir tes dan daya beda butir akan membawa dampak pada *rule-rule base* yang dibuat, sehingga diharapkan pemilihan terhadap butir-butir tes yang dimunculkan menjadi lebih tepat.

DAFTAR PUSTAKA

- Baumgartner, T.A., & Jackson, A. S. 1995. *Measurement for Evaluation in Physical Education and Exercise Science* (5thed.). New York: WCB Brown & Benchmark Publishers.
- Borg, W.R., & Gall, M. D. 1983. *Educational Research: an Introduction* (4thed.). New York: Longman Inc.
- Chee, T.S., & Wong, A.F.L. 2003. *Teaching and Learning with Technology*". Singapore: Prentice Hall.
- Hambleton, R.K., Swaminathan, H., & Rogers, H. 1991. *Fundamentals of Item Response Theory*. New Delhi: Sage Pub. Inc.
- Luger, G. F. 2005. *Artificial Intelligence, Structure and Strategies for Complex Problem Solving*(5thed). New York: Addison Wesley.
- Pressman, R. S. 1997. *Software Engineering: a Practitioner's Approach*. New York: McGraw Hills Book, Inc.
- Rolston, D. W. 1988. *Principles of Artificial Intellogence and Expert Systems Development*. Singapore: McGraw Hills Book, Co.
- Towndrow, P. A., & Vallence, M. 2004. *Using IT in The Language Classroom: a Guide for Teachers and Students in Asia* (3rded.). Singapore: Longman Pearson Education South Asia Pte. Ltd.