

PENGEMBANGAN *COMPUTERIZED ADAPTIVE TESTING* (CAT) DENGAN ALGORITMA LOGIKA *FUZZY*

Haryanto

Jurusan Elektro FT UNY
Jl. Colombo No.1 Yogyakarta 55281
haryanto.ftuny@gmail.com

Abstrak

Penelitian ini bertujuan menghasilkan: (1) model sistem inferensi dalam mengambil keputusan untuk memilih butir-butir tes yang tepat bagi siswa, (2) perangkat lunak *Computerized Adaptive Testing* (CAT) dengan algoritma logika *fuzzy* dalam mendeskripsikan kemampuan siswa. Penelitian dengan pendekatan *Research and Development* (R & D), terdiri atas dua bagian yaitu: (1) pengembangan program CAT, (2) pengujian program CAT pada siswa SMA Negeri 6 Yogyakarta kelas XII sebagai sampel, pada mata pelajaran Matematika dengan pokok bahasan Notasi sigma, barisan dan deret. Data dikumpulkan melalui observasi, dokumentasi, dan angket dan dianalisis secara deskriptif kuantitatif. Hasil penelitian menunjukkan program CAT: (1) mudah digunakan, tampilan interaktif, memiliki sistem keamanan, mudah diakses, dan mengacu standar kekinian, (2) dapat mengenali tiga macam pengguna saat proses *login*, (3) memiliki tiga macam basis data, (4) memiliki sembilan menu utama, (5) menggunakan model sistem inferensi algoritma logika *fuzzy*, (6) model tampilan program CAT, (7) dapat bekerja sesuai dengan fungsi dan jenis pengguna, dan (8) mampu mengelola: (a) *bank* soal, (b) mengemas butir-butir tes secara otomatis sesuai dengan kemampuan siswa, (c) pilihan jawaban dimunculkan secara acak, dan (d) menyimpan rekaman hasil tes secara individu maupun bersama-sama.

Kata kunci: *program CAT, bank soal*

DEVELOPING A COMPUTERIZED-ADAPTIVE TESTING WITH FUZZY LOGIC ALGORITHM

Haryanto

Electro Department Engineering Faculty UNY
Jl. Colombo No.1 Yogyakarta 55281
haryanto.ftuny@gmail.com

Abstract

The development of the *Computerized-Adaptive Testing* (CAT) aimed to investigate: 1) the performance of inference system in making decision to select the appropriate test items for the students, and 2) the performance of CAT program using *fuzzy* logic algorithm in describing the competence of students. This research and development (R&D) consists of two parts: (1) developing the CAT, (2) tested testing the CAT on mathematics at the topics of Sigma notation, sequence and series. Data were collected through observation, documentation and questionnaire and analyzed using quantitative descriptive technique. The result showed that: (1) the CAT user friendly, interactive, secure, accessible, and current and (2) able to recognize three types of users while logging-in three types of data bases, that nine options on the main menu it is able to work properly according to the function and the type of users, and administrered the bank and organize test items automatically, based on students' competence, randomize the answer in the options, and record the result of the test simultaneously and individually.

Keywords: *CAT program*

Pendahuluan

Perkembangan dan kemajuan Ilmu Pengetahuan Teknologi dan Seni (IPTEKS) khususnya bidang Teknologi Informasi (TI), dewasa ini menunjukkan peningkatan yang cukup signifikan bagi kepentingan umat manusia. Dalam dunia pendidikan yang dapat dilihat antara lain: masuknya mata pelajaran TIK dalam kurikulum, proses pembelajaran dan media pendidikan berbasis TI, hingga proses evaluasi hasil belajar berbantuan komputer (PP No: 22, 2005; Baumgartner & Jackson, 1995; Bates & Poole, 2003; Chee & Wong, 2003). Hal itu menunjukkan komitmen sekolah untuk meningkatkan kualitas pendidikan melalui hadirnya TI.

Target 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 & Vallence (2004), ditujukan untuk efektivitas dan efisiensi pelaksanaan dan penyelenggaraan tes. Era TI melalui *Computerized Base Test* (CBT), komputer diprogram untuk mengolah urutan tampilan butir-butir soal ujian dan menerima respons jawaban dari peserta (Chee & Wong, 2003; Towndrow & Vallence, 2004).

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 *bank* soal dalam pengujian berbasis komputer, memunculkan problematika ketidaksesuaian antara tingkat kesulitan soal dengan kemampuan peserta tes. Di sisi lain, pengambilan butir tes dari *bank* soal secara berurut yang juga banyak dijumpai, menurut teori pengukuran mengurangi validitas hasil tes, karena butir-butir tes yang muncul mudah untuk dihafalkan.

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 kesesuaian butir tes terhadap kemampuan peserta tes atau sering disebut dengan istilah "*Computerized Adaptive Test/CAT*". Disebut *computerized* (terkomputerisasi) karena dalam

pelaksanaan pengujian tidak lagi menggunakan “kertas dan pensil”. *Adaptive*, bahwa tingkat kesulitan butir-butir tes yang muncul didasarkan pada hasil analisis dengan algoritma dalam sebuah program, agar sesuai dengan kemampuan peserta. Kesesuaian butir soal didasarkan pada respons jawaban yang diberikan peserta tes.

Berbagai analisis statistik mengenai kemampuan respons butir yang digunakan antara lain Bayesian dan Maximum Likelihood. Analisis statistik yang demikian disebut dengan analisis non-monotonik. Kelemahan analisis penalaran non-monotonik adalah munculnya persepsi terbaru dapat merubah persepsi yang telah ada, sehingga keputusan tindakan yang diberikan selalu berubah-ubah. Akibat yang dirasakan adalah beban komputasi program semakin berat, sehingga semakin banyak butir tes yang dikerjakan akan semakin lambat kemunculan butir tes berikutnya. Hal itu disebabkan, perlakuan analisis mempersyaratkan dimulai dari butir tes yang pertama hingga butir tes yang terakhir. Oleh sebab itu, hasil akhir analisis memerlukan tabel nilai baku (nilai kritis) untuk menentukan apakah harus menerima keputusan dugaan (hipotesis) atau tidak. Hal tersebut akan memperlambat kerja komputer.

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 upaya membuat mesin komputer dapat “belajar” sehingga dapat bekerja dengan “kecerdasannya” seperti yang dimiliki manusia. Unsur pokok AI adalah adanya “*Knowledge Base (KB)*” atau basis pengetahuan, “*Data Base (DB)*” atau basis data dan “*Inference Engine (IE)*” atau mesin inferensi (Luger, 2005). Terkait dengan permasalahan pengujian hasil pembelajaran, komputer diprogram dengan pemberian *knowledge* untuk mengetahui kemampuan peserta tes melalui rpsons jawaban yang diberikan, untuk memilih butir tes yang tepat. Analisis yang dilakukan komputer didasarkan pada penalaran monotonik melalui sebuah program AI.

Pendekatan AI dalam penerapannya memiliki bermacam-macam metode, salah satunya adalah logika *fuzzy* (Terano, Asai, & Sugeno, 1987;

Zimmermann, 1991). Logika *fuzzy* banyak digunakan untuk keperluan pengambilan keputusan terhadap suatu permasalahan yang bersifat kualitatif maupun kuantitatif (Yan, Ryan, & Power, 1994). Penerapan dalam pengujian hasil belajar adalah untuk memilih tingkat kesulitan butir-butir tes yang tepat dengan kemampuan peserta tes, agar diperoleh hasil keputusan berupa penilaian yang benar.

Tujuan penelitian yang diharapkan adalah untuk menghasilkan:

1. Model sistem inferensi program CAT dengan algoritma logika *fuzzy* yang dapat memilih butir-butir tes yang tepat dengan kemampuan siswa.
2. Perangkat lunak berupa program CAT dengan algoritma logika *fuzzy* yang mampu digunakan untuk mendeskripsikan kemampuan siswa.

Metode Penelitian

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 lunak model CAT dengan algoritma logika *fuzzy*. Model yang digunakan untuk mengembangkan produk CAT dengan algoritma logika *fuzzy*, mengacu model Pressman (1997) dan Rolston (1988). Tahap kedua adalah menerapkan produk dalam situasi yang sebenarnya 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).

Penelitian dilakukan di: (a) Laboratorium Komputer Pendidikan Teknik Elektro FT UNY, untuk proses pengembangan/*development* produk program CAT. (b) SMA Negeri 6 Yogyakarta Daerah Istimewa Yogyakarta, sebagai tempat untuk implementasi/*research* produk program CAT. (c) Waktu penelitian, dilakukan pada bulan Juli 2007 sampai dengan Juni 2008 untuk pembuatan program CAT. Bulan Juli sampai dengan September 2008 untuk implementasi atau pengujian program CAT.

Subjek yang digunakan dalam penelitian untuk uji coba meliputi: (a) 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. (b) Siswa kelas XII IPA, untuk uji fungsionalitas dan unjuk kerja produk program CAT.

Data yang diperlukan dalam penelitian ini meliputi: (a) Data kualitatif berupa fungsi-fungsi yang diperlukan program, calon pengguna program, dan standar minimal perangkat keras yang digunakan program, serta hasil uji internal terhadap fungsionalitas program. (b) Data kuantitatif berupa tanggapan guru terhadap produk program CAT dan hasil unjuk kerja program dalam proses pengujian hasil belajar siswa.

Penelitian ini meliputi dua kegiatan, yaitu: (a) Pengembangan produk model CAT dengan logika *Fuzzy* dan (b) Implementasi hasil produk dalam situasi nyata di kelas. Teknik pengumpulan data meliputi: (a) Untuk pengembangan produk, menggunakan: (1) Observasi, mengenai: ketepatan instruksi (*syntax error*), ketepatan proses (*run time error*), dan ketepatan hasil (*logic error*), serta verifikasi dan validasi produk (*alpha-beta testing*). (2) Kuesioner dan wawancara, mengenai: kelengkapan dan ketepatan fungsi perangkat lunak. (3) Dokumentasi, mengenai: materi, bentuk dan model tes. (b) Untuk implementasi produk, menggunakan: (1) Observasi, mengenai: kebenaran, ketepatan, fungsionalitas, dan kemampuan produk dalam menguji kemampuan siswa. (2) Dokumentasi, mengenai: data nilai prestasi belajar siswa.

Teknik analisis data yang digunakan adalah deskriptif kuantitatif dan evaluatif. Teknik analisis deskriptif kuantitatif, dilakukan untuk menguji kelayakan produk *software* dalam memilih butir-butir tes yang tepat sesuai dengan tingkat kemampuan siswa yang diukur menggunakan tes terkomputerisasi dengan algoritma logika *fuzzy*. Teknik analisis deskriptif evaluatif dilakukan untuk menentukan kelayakan, kemampuan dan efektivitas kerja produk dalam fungsinya mendeskripsikan kemampuan siswa.

Model CAT dengan Algoritma Logika Fuzzy

Hambleton, Swaminathan, & Rogers (1991) dalam bukunya “*Fundamentals of Item Response Theory*” menyebutkan bahwa yang dimaksud dengan pengujian adaptif terkomputerisasi merupakan “*would be to give every examinee a test that is “tailored” or adapted, to the examinee’s ability level*”. Dikatakan juga bahwa tes dengan panjang butir yang telah *fixed*, tidak efisien jika diberikan kepada semua peserta tes. Hal itu disebabkan untuk peserta dengan kemampuan yang beragam mestinya cukup diberikan hanya dengan beberapa soal yang sesuai dengan levelnya. Ia mengatakan bahwa tes dapat dipendekkan tanpa kehilangan pengujian yang presisi, jika butir tes disiapkan (*administered*) sesuai dengan kemampuan masing-masing peserta tes. Untuk itu, tentunya memerlukan informasi yang lengkap tentang kemampuan peserta tes, sehingga masing-masing peserta tes akan memperoleh satu set soal yang sesuai dengan kemampuannya (*administered a unique set of items*).

Masters & Keeves (1999) menyebutkan bahwa di dalam adaptif tes, butir tes dipilih dari *bank* soal berdasar pada aturan (*rule*) pemilihan butir tes yang ditetapkan sebelumnya. Komputer dengan kemampuannya yang *programmable* mempunyai kemudahan yang tinggi untuk digunakan dalam proses pengujian yang adaptif (*adaptive testing*). Linn (1989) menyebutkan ada empat generasi pengujian terkomputerisasi yaitu: generasi (1) pengujian terkomputerisasi (*computerized testing/CT*), (2) pengujian adaptif terkomputerisasi (*computerized adaptive testing/CAT*), (3) pengukuran kontinu (*continuous measurement/CM*), dan (4) pengukuran cerdas (*intelligent measurement/IM*).

Baeck, Fogel, & Michalewicz, (1997), secara garis besar mengemukakan bahwa keempat generasi tersebut menunjukkan adanya evolusi komputasi dibidang alat ukur dan pengukuran yang dimulai dari klasikal konvensional (generasi 1 dan 2), hingga modern cerdas (generasi 3 dan 4). Menurut Goldberg (1989) dan Mitchell (1997), tujuan yang diinginkan peningkatan (evolusi) komputasi adalah untuk membuat mesin komputer bukan hanya sebagai alat hitung, tetapi juga optimalisasi daya kerja mesin yang memiliki kemampuan dan kecerdasan seorang pakar (*teaching, scoring,*

testing, dan *interpreting*) yang humanis (manusiawi = mampu mempelajari karakter kemampuan *testee* agar dapat menilai dengan benar).

Pengertian adaptif dalam penelitian ini adalah kemampuan untuk menyesuaikan diri dengan kondisi, keperluan, dan keadaan tertentu. Kemampuan yang dimaksud diprogramkan kepada perangkat keras komputer melalui algoritma logika *fuzzy*. Kemampuan tersebut adalah kemampuan komputer untuk memilih butir-butir tes yang sesuai dengan kemampuan peserta tes (*the examinee's ability*), yang dianalisis dengan logika *fuzzy* berdasar respons jawaban peserta pada butir tes sebelumnya.

Penelitian ini, menggunakan komputer untuk pengujian dengan mendasarkan pada generasi ke 4, melalui pendekatan teknik kecerdasan logika *Fuzzy*, yang berorientasi pada: (a) *item bank*, (b) kemudahan administrasi butir tes, (c) *item selection*, (d) estimasi kemampuan, (e) pengaturan tingkat kesulitan butir tes, (f) kemudahan layanan pengujian, (g) kemudahan perakitan butir tes ("*tailor*"), (h) penilaian hasil tes, (i) kecepatan pengolahan hasil, dan (j) kecepatan penyajian informasi hasil pengujian.

Model pengujian secara modern dengan IRT dibedakan berdasarkan jumlah parameter butir tes, yaitu model satu parameter (*Rasch model*), dua parameter, dan tiga parameter (Hambleton, Swaminathan, & Rogers, 1991). Van der Linden & Hambleton (1997), menyebutkan parameter-parameter tersebut adalah kesulitan item, daya beda item, dan tebakan. Model IRT untuk butir tes dikotomi dua parameter (kesulitan item, daya beda item) adalah sebagai berikut (van der Linden & Hambleton (1997); Hambleton, Swaminathan, & Rogers (1991); Lord (1980)):

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \text{ atau } P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \text{ i = butir tes}$$

ke 1, 2, 3, ... , n

$P_i(\theta)$: probabilitas peserta dengan kemampuan θ untuk menjawab butir ke-i dengan benar.

Θ : tingkat kemampuan peserta uji.

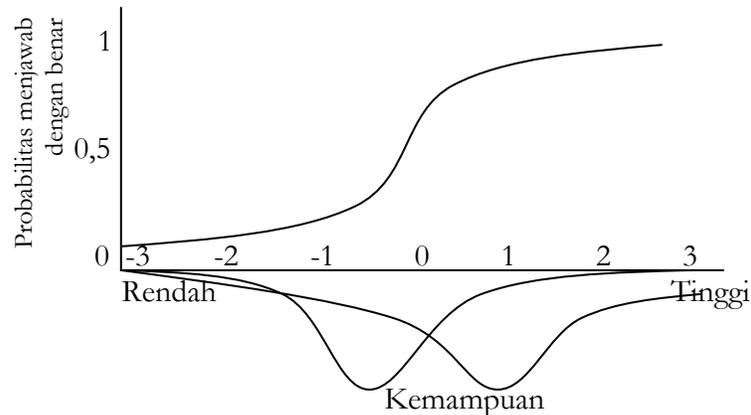
n : banyaknya butir tes.

e : nilai transcendental yang besarnya 2,718.

b_i : tingkat kesulitan butir.

a_i : daya beda butir.

Gambar 1 menunjukkan karakteristik kurva dan distribusi kemampuan dari dua kelompok, yaitu kelompok berkemampuan rendah dan kemampuan tinggi. Dari gambar terlihat bahwa kelompok dengan kemampuan tinggi memiliki probabilitas yang lebih besar untuk menjawab butir soal dengan benar dibandingkan kelompok dengan kemampuan rendah. Dengan demikian, karakteristik butir tidak bergantung pada kelompok.



Gambar 1. Kurva Karakteristik Butir dan Distribusi Kemampuan (Diadopsi dari Hambleton, Swaminathan, dan Rogers (1991))

Asumsi model Rasch didasarkan pada persyaratan objektivitas spesifik (Hambleton, Swaminathan, & Rogers, 1991). Asumsi ke-: (1) Adanya *unique ordering (unidimensionality)* menurut kemampuan orang dan kesulitan butir tes. (2) Sebagai implikasi dari asumsi ke-1, yaitu *local independence*, nilai kemungkinan orang menjawab benar suatu butir tes tidak bergantung pada jawaban butir tes lainnya. (3) Juga implikasi asumsi ke-1, yaitu *equality of discrimination*, rasio gangguan yang direpresentasikan oleh kemiringan (*slope*) maksimal kurva karakteristik butir diasumsikan sama untuk semua butir tes. (4) *Uni dimensionality* sekumpulan butir tes dalam suatu perangkat uji hanya mengukur satu dimensi. (5) Perilaku menebak yang acak (*random guessing behavior*). Model Rasch menyatakan bahwa untuk

setiap butir tes, nilai kemungkinan jawaban benar secara asymptotik cenderung nol sejalan dengan tingkat kemampuan orang yang menurun. (6) Bila kemampuan orang meningkat, nilai kemungkinan jawaban benar terhadap suatu tes mendekati satu (kebalikan asumsi ke-empat).

Sistem logika *fuzzy* merupakan salah satu cabang ilmu komputer yang mempelajari mengenai kecerdasan buatan (*artificial intelligence/AI*). Prinsip dasar sistem cerdas (Luger, (2005); Nilsson, (1980)) adalah membuat komputer melalui teknik pemrograman tertentu agar mampu berpikir, mengambil keputusan yang tepat dan bertindak, dengan cara-cara seperti yang dilakukan oleh manusia.

Pemilihan logika *fuzzy* digunakan sebagai pendekatan dalam analisis butir tes, karena logika *fuzzy* cocok dan sesuai untuk solusi permasalahan yang memetakan nilai-nilai kualitatif mengenai kemampuan kognitif seseorang (rendah, sedang, dan tinggi) ke dalam nilai-nilai kuantitatif. Dalam penelitian ini kualifikasi butir tes diklasifikasikan berdasarkan tingkat kesulitan butir tes yaitu: mudah, sedang, dan sulit. Klasifikasi tingkat kesulitan butir yang baik telah ditetapkan 0,3 hingga 0,8. Di samping itu, juga didasarkan pada tingkat daya beda butir tes, yaitu: rendah, sedang, dan tinggi dengan nilai 0,3 hingga 0,8. Tingkat kesulitan butir, daya beda butir, dan respons jawaban peserta tes melalui algoritma logika *fuzzy* digunakan sebagai dasar inferensi (pengambilan keputusan) untuk menentukan butir tes yang tepat diberikan kepada peserta tes.

Kemampuan logika *fuzzy* dalam menggambarkan kemampuan siswa menggunakan penalaran monoton. Melalui penalaran monoton, akan diperoleh tingkat kepastian mengenai kemampuan siswa berdasarkan respons yang diberikan selama proses pengujian. Faktor kepastian ini merupakan satu kesatuan logika *fuzzy* dalam rangka menggambarkan kemampuan siswa. Jika ada 2 daerah *fuzzy* direlasikan dengan implikasi sederhana, yaitu:

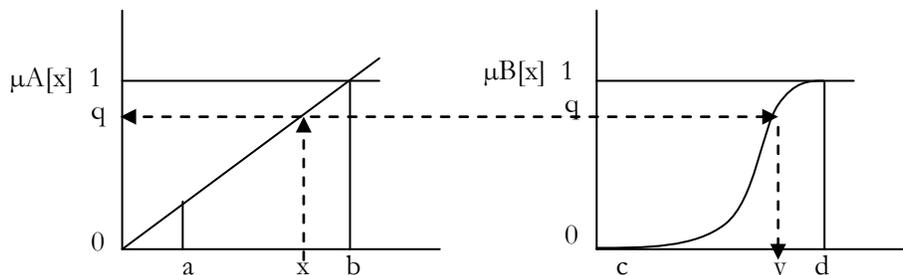
Jika x adalah A , maka y adalah B .

Fungsi transfernya dinyatakan dengan notasi $y = f((x,A),B)$.

Secara matematis digambarkan sebagai berikut:

$$\mu_A[x] = (x-a)/(b-a) = q \text{ sehingga } \mu_B[y] = 1-2[(d-y)/(d-c)]^2 = q$$

Dalam hal ini diketahui nilai x untuk mencari nilai y melalui nilai q sebagai penalaran monoton (Gambar 2 menunjukkan grafik proses tersebut).



Gambar 2. Penalaran Monoton

Sistem inferensi atau disebut juga logika *fuzzy control* (FLC), merupakan sistem mekanisme logika *fuzzy* dalam proses pengambilan keputusan. Penelitian ini menggunakan sistem inferensi (FLC) model Tsukamoto (Yan, Ryan, & Power. (1994)). Algoritma logika *fuzzy* untuk mendapatkan *output*, menurut metode ini ada empat tahapan, yaitu:

- a. Pembentukan himpunan *fuzzy* (fuzzifikasi). Dalam hal ini variabel *input* dan variabel *output* dibagi menjadi satu atau lebih himpunan *fuzzy*. Pembentukan himpunan *fuzzy* didasarkan pada fungsi keanggotaan *fuzzy*.
- b. Aplikasi fungsi implikasi (aturan), yaitu; penerapan fungsi basis aturan yang didasarkan pada basis pengetahuan. Menurut metode Tsukamoto, fungsi implikasi yang digunakan adalah Min (nilai terkecil).
- c. *Inferensi*, penegasan keputusan berdasar komposisi aturan. Komposisi aturan (*rule base*) merupakan kumpulan aturan yang digunakan sebagai dasar untuk melakukan inferensi. Apabila sistem terdiri dari beberapa aturan, maka inferensi diperoleh dari kumpulan dan korelasi antar aturan.
- d. *Defuzzifikasi* adalah penegasan hasil inferensi berdasar pada nilai rata-rata terbobot.

Input dari proses defuzzifikasi adalah suatu himpunan *fuzzy* yang diperoleh dari mekanisme inferensi terhadap komposisi aturan-aturan *fuzzy*. Sedangkan *output* yang dihasilkan defuzzifikasi merupakan suatu bilangan pada domain himpunan *fuzzy* tersebut. Sehingga jika diberikan suatu himpunan *fuzzy* dalam *range* tertentu, maka harus dapat diambil suatu nilai *crispy* tertentu sebagai *output*.

Nilai keluaran hasil analisis logika *fuzzy* berupa tingkat kesulitan butir yang selanjutnya diberikan kepada siswa. Proses selanjutnya adalah mengambil butir tes dari *bank* soal yang memiliki tingkat kesulitan sesuai dengan hasil analisis inferensi logika *fuzzy*. Selama ujian berlangsung, waktu dihitung terus untuk menentukan sudah berapa lama siswa menggunakan waktu dalam menjawab butir tes (W). Di samping itu, juga dihitung hasil perolehan untuk menentukan penguasaan peserta terhadap materi yang diujikan (KM). KM dari masing-masing siswa diestimasi dengan menggunakan maksimum Likelihood. Kemampuan yang diperoleh berupa skor -3 hingga +3. Di sekolah, hasil kemampuan tersebut dikonversi menjadi nilai 0 hingga 10 (jika nilai maksimum 10) atau 0 hingga 100 (jika nilai maksimum 100, yang digunakan dalam disertasi ini).

Penilaian terhadap kemampuan siswa, terlebih dahulu dilakukan dengan menghitung nilai $p_i(\theta)$ dan $q_i(\theta)$ dari setiap butir tes. Dalam penelitian ini, menggunakan dua parameter, yaitu: daya beda (a_i) dan indeks kesulitan (b_i), sehingga perhitungan nilai $p_i(\theta)$ (peluang peserta tes dengan kemampuan $[\theta]$ menjawab butir tes ke- i benar) dan $p_i(\theta)$ (peluang peserta tes dengan kemampuan $[\theta]$ menjawab butir tes ke- i salah) adalah:

$$p_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}}$$

$$p_i(\theta) + q_i(\theta) = 1$$

Nilai kemampuan (θ) diambil dengan rentang -3,00 hingga 3,0 dengan step 0,5 (Hambleton, R.K., Swaminathan, H., & Rogers, H.J. 1991: 28). Selanjutnya dengan diketahui nilai $p_i(\theta)$, $p_i(\theta)$, dan θ dapat dihitung nilai Likelihood $L(U|\theta)$ dengan formula:

$$L(U | \theta) = \prod_{i=1}^n p_i^u q_i^{1-u}$$

Keterangan:

n : banyak butir tes

u : jawaban siswa pada butir tes

Untuk mengetahui kemampuan siswa peserta tes, dihitung terlebih dahulu menggunakan *Maximum Likelihood Estimation* $L(\theta | U)$ dengan formula:

$$L(\theta | U) = \frac{L(U | \theta)}{\sum L(U | \theta)}$$

Berdasar hasil $L(\theta | U)$ untuk nilai θ dari -3,0 hingga 3,0 maka estimasi kemampuan siswa peserta tes adalah nilai θ dari hasil $L(\theta | U)$ yang paling tinggi (maksimum). Jadi estimasi kemampuan siswa peserta tes ditentukan dengan formula:

$$\text{Estimasi } \theta = \text{Maximum } [L(\theta | U)]$$

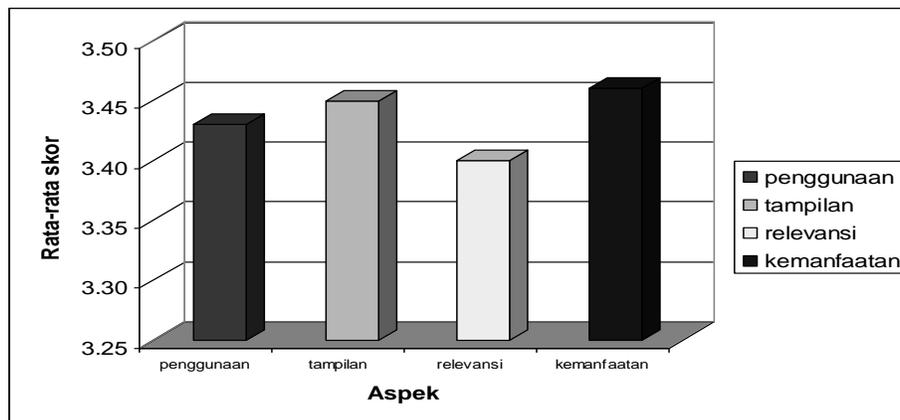
Estimasi tersebut mengandung makna bahwa peluang peserta tes dengan kemampuan $[\theta]$ menjawab butir tes $\max L(\theta | U) \times 100\%$ benar. Di sisi lain, berarti peluang peserta tes dengan kemampuan $[\theta]$ menjawab butir tes $[1 - \max L(\theta | U)] \times 100\%$ salah.

Hasil Penelitian dan Pembahasan

Penilaian tanggapan untuk mengetahui gambaran secara fisik mengenai program CAT, dilakukan melalui angket terhadap guru bidang studi matematika. Pengambilan data dilakukan setelah guru mengoperasikan program CAT dan menggunakannya untuk berbagai kegiatan sesuai dengan kewenangannya.

Skor tanggapan guru terhadap program CAT pada masing-masing aspek berada pada kisaran >3 sampai dengan 4. Tanggapan guru terhadap program CAT secara rinci, untuk aspek performansi operasional/penggunaan program memperoleh skor 3,43 (sangat baik). Tanggapan guru terhadap aspek performansi tampilan program memperoleh skor 3,45 (sangat baik). Tanggapan guru terhadap aspek relevansi hasil program dengan materi tes memperoleh skor 3,40 (sangat baik). Tanggapan terhadap aspek kemanfaatan program dalam mendukung tugas-tugas guru memperoleh skor 3,46 (sangat baik). Hal itu menunjukkan bahwa

tanggapan guru terhadap program CAT dengan algoritma logika *fuzzy* secara keseluruhan dinilai sangat baik, atau layak digunakan.

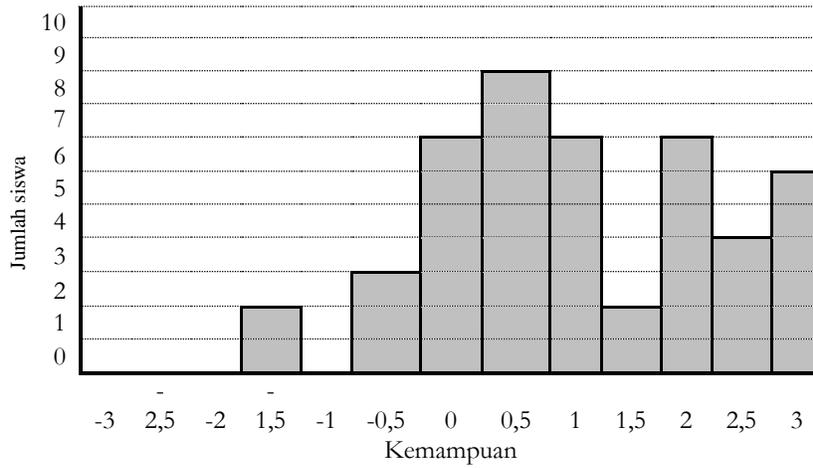


Gambar 3. Tanggapan Guru terhadap Program CAT

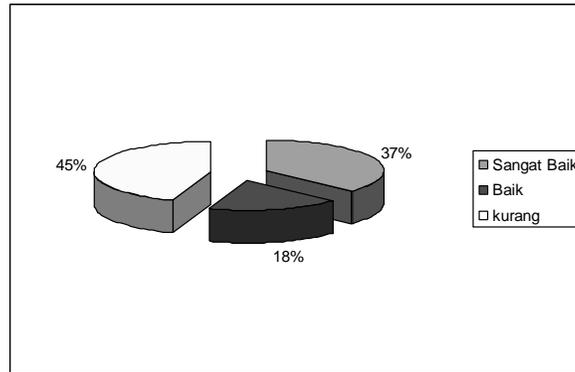
Keterangan:

Tanggapan 0 s.d 1 : Kurang; Tanggapan >1 s.d 2 : Cukup;
Tanggapan >2 s.d 3 : Baik; Tanggapan >3 s.d 4 : Sangat baik

Pengujian *beta* digunakan untuk mengetahui kemampuan kerja program CAT dalam memprediksi kemampuan siswa melalui butir-butir tes yang dikerjakan. Dalam hal ini, program telah diisi *bank* soal yang dilengkapi dengan parameter, yaitu: indeks kesulitan dan daya beda. Berdasar hasil pengujian dapat dideskripsikan sejumlah kelompok kemampuan seperti terlihat pada Gambar 4. Jika kemampuan -3 hingga <1 dinyatakan belum tuntas dengan kategori kurang baik, maka banyaknya siswa yang belum tuntas terdiri dari 17 orang. Jika kemampuan ≥ 1 hingga <2 dinyatakan tuntas dengan kategori baik, maka banyaknya siswa yang tuntas terdiri dari 7 orang. Jika kemampuan ≥ 2 hingga 3 dinyatakan tuntas dengan kategori sangat baik, maka banyaknya siswa yang tuntas terdiri dari 14 orang.



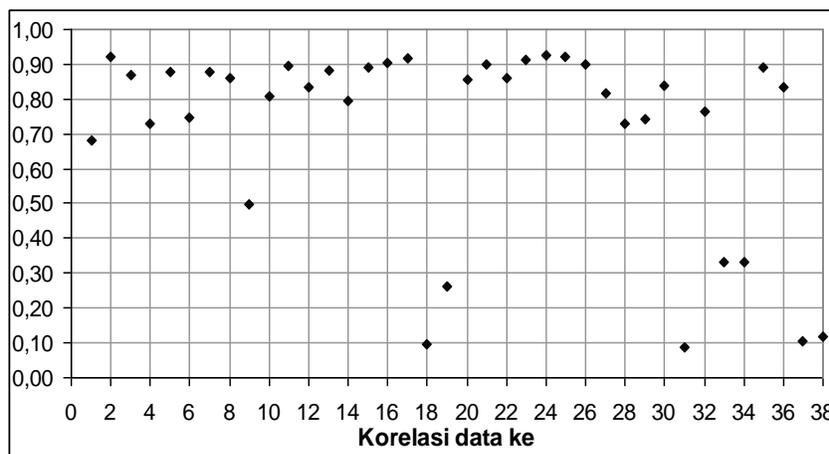
Gambar 4. Kemampuan Siswa Hasil Belajar.



Gambar 5. Persentase Hasil Tes Tiga Kelompok Kemampuan

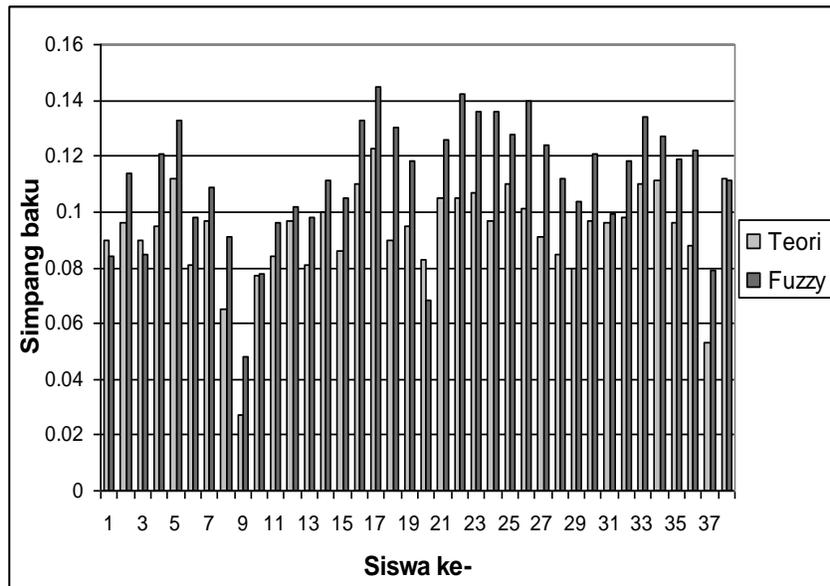
Deskripsi ketepatan kerja program CAT mengenai pemilihan butir-butir tes yang diberikan kepada siswa digambarkan dengan analisis nilai rata-rata korelasi dan simpang baku secara teori dengan hasil kerja inferensi *summary* program CAT. Hasil analisis korelasi mengenai tingkat kesulitan butir-butir tes diperoleh nilai korelasi tertinggi (r) sebesar 0,93 dan terendah (r) sebesar 0,09 dan nilai rata-rata korelasi (r) sebesar 0,72. Hal itu me-

nunjukkan bahwa program CAT dengan algoritma logika *fuzzy* berhasil dengan baik memilih butir-butir tes yang sesuai dengan kemampuan siswa. Artinya, bahwa butir-butir tes yang dipilih oleh program CAT *fuzzy* dari *bank* soal antara analisis secara teori dengan hasil kerja program CAT *fuzzy* memiliki relasi yang tinggi (rata-rata $r = 0,72$).



Gambar 6. Korelasi Tingkat Kesulitan Butir antara Analisis Teori dengan Hasil *Fuzzy*

Hasil analisis simpang baku (sb) secara teori mengenai tingkat kesulitan butir-butir tes diperoleh nilai tertinggi $sb=0,12$ dan terendah $sb=0,03$ serta nilai rata-rata $sb=0,09$. Hasil analisis sb secara *fuzzy* mengenai tingkat kesulitan butir-butir tes diperoleh nilai tertinggi $sb=0,15$ dan terendah $sb=0,05$ serta nilai rata-rata $sb=0,11$. Hal itu juga menunjukkan bahwa program CAT *fuzzy* berhasil dengan baik memilih butir-butir tes yang sesuai dengan kemampuan siswa. Artinya, bahwa butir-butir tes yang dipilih oleh program CAT *fuzzy* dari *bank* soal antara analisis secara teori dengan hasil kerja program CAT *fuzzy* memiliki perbedaan yang sangat kecil (rata-rata teori $sb=0,09$ dan rata-rata *fuzzy* $sb=0,11$).



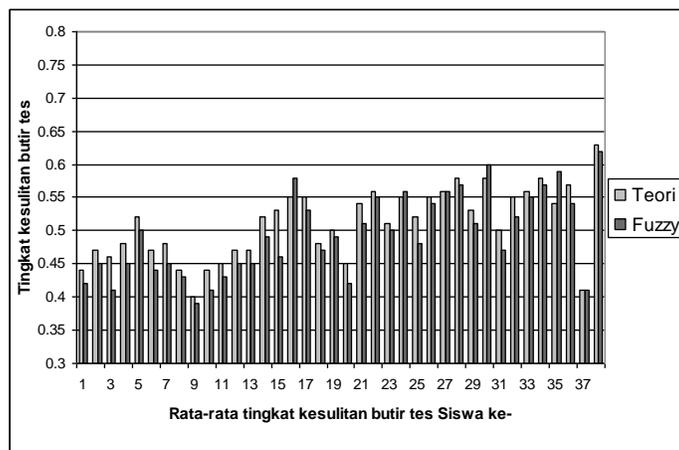
Gambar 7. Perbandingan Simpang Baku Hasil Analisis antara Teori dengan Hasil Program CAT (*Fuzzy*)

Hasil analisis MLE juga diperoleh grafik fungsi *Likelihood* mengenai deskripsi kemampuan siswa. Grafik fungsi kemampuan tersebut diperoleh dari analisis terhadap tingkat kesulitan butir yang diambil secara acak dari *bank* soal. Nilai kemampuan digambarkan melalui kurva dengan batas $-3 < \theta < +3$ hal itu menunjukkan bahwa kemampuan siswa berdistribusi normal. Hal itu disebabkan: (1) grafik kemampuan selalu ada di atas sumbu absis. (2) bentuknya simetrik terhadap $\theta = \text{rata-rata } (\mu)$. (3) grafiknya mendekati sumbu absis $\theta = \mu \pm 3(sb^2)$.

Berdasar hasil analisis data secara keseluruhan, menunjukkan bahwa sistem inferensi yang dibangun dengan menggunakan algoritma logika *fuzzy* pada program CAT telah berhasil dengan baik memilih butir-butir tes yang tepat dan sesuai dengan kemampuan siswa. Dalam hal itu, siswa dengan kemampuan tinggi menerima butir-butir tes dengan tingkat kesulitan tinggi. Siswa dengan kemampuan sedang menerima butir-butir tes dengan tingkat

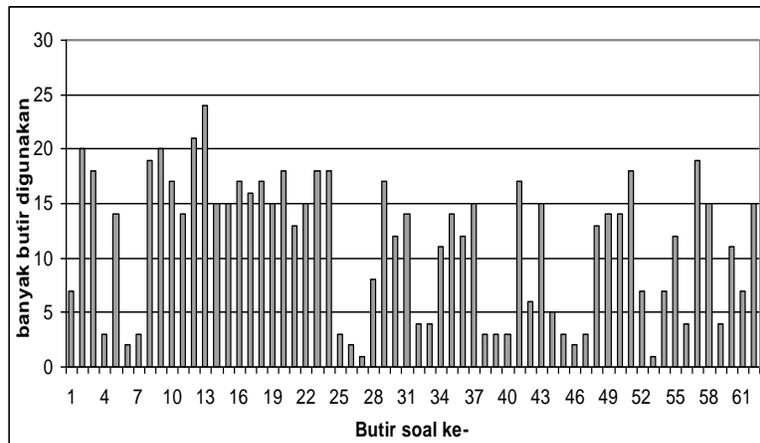
kesulitan sedang. Siswa dengan kemampuan rendah menerima butir-butir tes dengan tingkat kesulitan rendah. Di sisi lain, banyak butir yang diterima siswa antara yang satu dengan yang lain juga beragam (tidak sama). Siswa dengan kemampuan tinggi atau siswa dengan kemampuan rendah menerima banyak butir lebih sedikit dibandingkan dengan siswa berkemampuan sedang. Hal itu menunjukkan bahwa program CAT yang dibangun dengan algoritma logika *fuzzy* juga telah berhasil dengan baik mengatur banyak butir yang diberikan sesuai dengan kemampuan siswa (adaptif).

Program CAT yang dikembangkan juga mampu menampilkan kemasan butir tes dengan letak pilihan jawaban teracak. Hal ini dimaksudkan agar antar siswa tidak dapat saling bekerja sama dalam menjawab tes. Keunikan dan sekaligus nilai lebih dari produk program CAT ini adalah penggunaan teknik kecerdasan buatan (*artificial intelligence*) logika *fuzzy* untuk pengembangan sistem inferensi. Karena berdasar hasil kajian terhadap sejumlah jurnal penelitian luar maupun dalam negeri, hal itu belum pernah diteliti. Produk program CAT ini juga mampu: mengemas butir-butir tes secara otomatis, melakukan administrasi butir-butir tes, menyimpan hasil tes, dan melakukan penskoran.

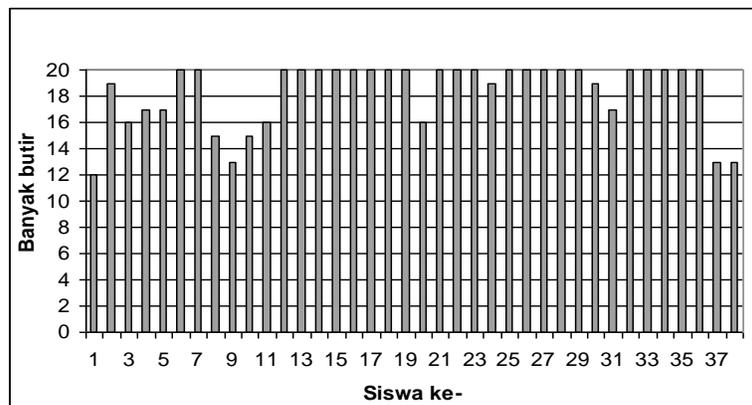


Gambar 8. Perbandingan Rata-rata Tingkat Kesulitan Butir Tes antara Teori dengan Hasil Program CAT (*Fuzzy*)

Basis data *bank* soal yang disediakan dalam program CAT berjumlah 62 butir tes. Masing-masing butir tes memiliki dua parameter, yaitu: tingkat kesulitan dan daya beda butir. Berdasarkan hasil tes, konfigurasi butir-butir tes yang digunakan oleh program CAT untuk siswa terdapat pada Gambar 9. Sedangkan banyaknya butir tes yang dikemas oleh program CAT yang diberikan kepada masing-masing siswa terdapat pada Gambar 10.



Gambar 9. Banyaknya Butir Tes yang Digunakan Program CAT



Gambar 10. Banyak Butir Tes yang Dikemas Program CAT untuk Siswa

Berdasar hasil kerja program di atas dapat dikatakan bahwa kinerja program CAT dengan algoritma logika *fuzzy* mampu melaksanakan tugas dengan baik. Tugas tersebut adalah (1) Mengemas butir-butir tes dengan tingkat kesulitan yang sesuai dengan kemampuan siswa. (2) Variasi jumlah butir-butir tes yang diterima siswa sesuai dengan kemampuannya. (3) Mendeskripsikan kemampuan siswa berdasar butir-butir tes yang diterima.

Kemampuan program CAT dengan algoritma logika *fuzzy* seperti tersebut diuraikan di atas, telah sesuai dengan yang diharapkan oleh teori respon butir. Bahwa setiap siswa menerima tingkat kesulitan butir-butir tes yang sesuai dengan karakteristik informasi butir. Di sisi lain, setiap siswa juga akan menerima jumlah butir tes yang berbeda-beda sesuai dengan kemampuannya. Hal itu sesuai dengan tuntutan sifat adaptabilitas tes dari program CAT. Produk program CAT dengan algoritma logika *fuzzy* yang dikembangkan telah mampu bekerja sesuai dengan tuntutan tersebut.

Sifat adaptabilitas program menuntut adanya sistem inferensi untuk menentukan keputusan bahwa setiap siswa harus menerima jumlah butir tes yang tepat, dan setiap siswa harus menerima butir tes yang tepat pula dengan karakteristik kemampuannya. Logika *fuzzy* yang digunakan sebagai sistem inferensi produk program CAT telah berhasil dengan baik dalam melakukan tugas-tugasnya untuk pengukuran kemampuan kognisi siswa. Hal itu berarti bahwa program CAT dengan algoritma logika *fuzzy* dapat digunakan sebagai penunjang untuk keperluan evaluasi hasil belajar siswa.

Simpulan

Simpulan yang dapat ditarik berdasar analisis data dan kajian terhadap hasil pengembangan produk adalah sebagai berikut:

- a. Model CAT dengan algoritma logika *fuzzy* 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.
- b. Sistem inferensi model CAT dengan algoritma logika *fuzzy* 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.

- c. Ketepatan sistem inferensi dalam memilih butir-butir tes yang sesuai dengan respons jawaban siswa diwujudkan pada nilai rata-rata kesalahan/simpang baku 0,093 secara teori dan 0,112 secara fuzzy. Di samping itu, berdasar analisis grafis dan analisis korelasi butir-butir tes antara analisis teori dengan hasil kerja program CAT fuzzy memiliki relasi yang tinggi, dengan nilai rata-rata $r = 0,72$.
- d. Kemampuan model CAT dengan algoritma logika fuzzy dalam mendeskripsikan kemampuan siswa berdasar butir-butir tes yang diterima dan nilai-nilai estimasi hasil analisis, diperoleh: (a) 34% siswa memperoleh predikat kemampuan tuntas sangat baik, (b) 21% siswa memperoleh predikat kemampuan tuntas baik, dan (c) 45% siswa memperoleh predikat kemampuan belum tuntas/kurang baik.
- e. Hasil temuan karya disertasi ini, adalah: (1) Model CAT dengan algoritma logika fuzzy digunakan untuk membentuk sistem inferensi yang bagian-bagiannya terdiri dari: fuzzifikasi, basis pengetahuan, implementasi basis pengetahuan, dan defuzzifikasi. (2) Sistem inferensi model CAT dengan algoritma logika fuzzy: (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.

Saran

Program CAT menggunakan algoritma logika fuzzy 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 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, sedang, dan mudah. Oleh sebab itu, jumlah butir-butir tes yang memenuhi syarat untuk dijadikan *bank* soal jumlahnya harus memenuhi ketiga kelompok tersebut. Jika banyak soal (N) yang akan diujikan adalah 20 butir tes, maka dalam *bank* soal minimal adalah 20 butir tes x 3 yakni 60 butir tes. Dengan rincian 20 butir tes tingkat kesulitan tinggi, 20 butir tes tingkat kesulitan sedang, dan 20 butir tes tingkat kesulitan rendah. Syarat butir tes yang bagus minimal memiliki parameter tingkat kesulitan butir dan daya beda butir dengan kisaran nilai 0,3 hingga 0,8 (secara klasik) dan -3 sampai 3 untuk tes modern.

Program CAT dalam pelaksanaannya harus dilakukan di dalam laboratorium komputer. Dengan demikian, disarankan agar sekolah harus memiliki minimal sebuah laboratorium komputer yang memenuhi standar minimal. Di samping itu, siswa yang akan diuji kemampuannya seyogyanya dilatih terlebih dahulu cara mengoperasikan dan bekerja dengan komputer.

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

- Baeck, T., Fogel, D.B., & Michalewicz, Z. (1997). *Handbook on evolutionary computation*. New York: IOP Press.
- Baumgartner, T.A., & Jackson, A.S. (1995). *Measurement for evaluation in physical education and exercise science (5th ed.)*. New York: WCB Brown & Benchmark Publishers.
- Bostock, S.J. (1997). *Designing web-based instruction for active learning*. Englewood Cliffs: Educational Technology Publications.

- Borg, W.R., & Gall, M.D. (1983). *Educational research: An introduction* (4th ed.). New York: Longman Inc.
- Brennan, R.L. (2006). *Educational measurement* (4th ed.). New York: American Council on Education.
- Chee, T.S., & Wong, A.F.L. (2003). *Teaching and learning with technology*". Singapore: Prentice Hall.
- Echols, J.M., & Shadily, H. (1986). *Kamus Inggris Indonesia*. Jakarta: PT. Gramedia.
- Eiben, A.E., & Michalewicz, Z. (1999). *Evolutionary computation*. New York: IOP Press.
- Goldberg, D.E. (1989). *Genetic algorithms in search: Optimization & machine learning*. New York: Addison-Wesley.
- Hambleton, R.K., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory*. New Delhi: Sage Pub. Inc.
- Hornby, AS. (1984). *Oxford advanced learner's dictionary of current English*. London: Oxford University Press.
- Lin, R.L. (1989). *Educational measurement* (3rd ed.). New York: American Council on Education, Macmillan Publishing Company.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Englewood Cliffs: Lawrence Erlbaum Associates, Publishers.
- Luger, G.F. (2005). *Artificial intelligence, structure and strategies for complex problem solving* (5th ed). New York: Addison Wesley.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan non-tes*. Yogyakarta: Mitra Cendekia Press.
- Masters, G.G., & Keeves, J.P. (1999). *Advances in measurement in educational research and assessment*. New York: Pergamon Press.
- Mitchell, T.M. (1997). *Machine learning*. New York: McGraw-Hill Companies, Inc.

- Nilsson, N.J. (1980). *Principles of artificial intelligence*. New York: Tioga Publishing Company.
- Pressman, R.S. (1997). *Software engineering: A practitioner's approach*. New York: Mc Graw hill Book, Inc.
- Program Pascasarjana. (2008). *Pedoman tesis dan disertasi Program Pascasarjana Universitas Negeri Yogyakarta*. Yogyakarta: Program Pascasarjana.
- Rolston, D.W. (1988). *Principles of artificial intelligence and expert systems development*. Singapore: McGraw Hill Book, Co.
- Terano, T., Asai, K., & Sugeno, M. (1992). *Fuzzy systems theory and its applications*. New York: Academic Press, Inc.
- Tim Prima Pena (2007). *Kamus besar bahasa Indonesia*. Jakarta: Gita Media Press.
- Towndrow, P.A., & Vallence, M. (2004). *Using IT in the language classroom: A guide for teachers and students in Asia* (3rd ed.). Singapore: Longman Pearson Education South Asia Pte. Ltd.
- Van der Linden, W.J., & Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer Verlag.
- Wang, L.X. (1994). *Adaptive fuzzy systems and control: Design and stability analysis*. Englewood Cliffs: Prentice Hall Inc.
- Widrow, B., & Stearns, S.D. (1985). *Adaptive signal processing*. Englewood Cliffs: Prentice Hall Inc.
- Yan J., Ryan M., & Power J. (1994). *Using fuzzy logic*. Englewood Cliffs: Prentice Hall Inc.
- Zimmermann, H.J. (1991). *Fuzzy set theory: And its applications* (2nd ed.). London: Kluwer Academic Publishing.