
OPTIMASI PENGUKURAN HASIL BELAJAR MELALUI *COMPUTERIZED INTELLIGENT MEASUREMENT MODEL FUZZY*

Dr. Haryanto, M.Pd., M.T.

Pendidikan Teknik Elektro FT UNY Yogyakarta

haryanto.ftuny@gmail.com

Abstrak. Makalah ini dimaksudkan untuk menyampaikan hasil penelitian dalam upaya meningkatkan akurasi pengukuran hasil belajar melalui pendekatan *Computerized Intelligent Measurement* (CIM) dengan model logika fuzzy. Hal itu bertujuan untuk memperoleh penilaian hasil belajar yang optimal dan berkeadilan. Optimal karena pengukuran dilakukan secara efektif dan efisien, sedangkan berkeadilan karena butir-butir tes dipilih melalui komputasi cerdas model logika fuzzy, yang sesuai dengan karakteristik peserta tes. Model sistem pengukuran tersebut sangat diperlukan untuk membantu proses evaluasi pembelajaran yang lebih bermakna, agar pengambilan keputusan hasil evaluasi belajar seoptimal mungkin dan dapat dipertanggungjawabkan.

Penelitian yang dilakukan dengan pendekatan *Research and Development* (R & D) ini, menerapkan sistem *computerized intelligent measurement* (CIM) model logika fuzzy. Materi tes yang diperlukan distandarisasi melalui serangkaian uji coba dan analisis butir dengan *item response theory* (IRT) untuk memperoleh parameter dan karakteristik butir yang selanjutnya dijadikan basis data *item bank*. Sistem inferensi dalam CIM dilakukan dengan algoritma logika fuzzy. Logika fuzzy berfungsi untuk keakurasian, keadilan dan ketepatan pemilihan butir-butir tes berdasar respon jawaban benar-salah peserta tes. Berdasar data respon jawaban, estimasi kemampuan peserta tes di analisis dengan maksimum Likelihood. Hasil yang diperoleh adalah bahwa pengukuran hasil belajar dengan pendekatan CIM model logika fuzzy mampu: (1) Memilih butir tes dengan akurat dan tepat sesuai dengan karakteristik peserta tes. (2) mengemas banyak butir tes sesuai dengan tingkat kemampuan peserta tes sehingga pengukuran menjadi lebih efektif dan efisien. (3) melakukan estimasi kemampuan peserta tes dengan baik. Simpulan temuan penelitian bahwa CIM model logika fuzzy mampu meningkatkan akurasi hasil pengukuran sehingga dapat menghasilkan penilaian yang optimal.

Kata kunci: *Computerized Intelligent Measurement, Logika Fuzzy, Item Response Theory, Item Bank.*

A. 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: 44; Bates & Poole, 2003: 55; Chee & Wong, 2003: 118). Permasalahan yang perlu dikaji adalah seberapa besar penggunaan TI dapat meningkatkan kualitas hasil pendidikan?

Target keberhasilan peningkatan kualitas pendidikan, khususnya dalam pembelajaran melalui pengukuran hasil belajar berbasis TI, juga telah mulai diupayakan (Chee & Wong, 2003: 96). Penggunaan TI untuk keperluan tes menurut Chee & Wong (2003: 96) dan Towndrow & Vallence (2004: 244), 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: 96; Towndrow & Vallenge, 2004: 244). Namun sejauh ini, permasalahan 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 maupun acak, 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: 145). Peningkatan kualitas tes dimaksudkan untuk menyesuaikan tingkat kesulitan butir tes dengan kemampuan peserta tes, dengan tujuan untuk mengetahui kemampuan siswa pada mata pelajaran tertentu. Pemilihan butir tes dengan menggunakan komputasi melalui model fuzzy disebut dengan istilah "*Computerized Adaptive Test/CAT*". Disebut *computerized* (terkomputerisasi) karena dalam pelaksanaan pengujian tidak lagi menggunakan "kertas dan pensil" tapi

pakai komputer. *Adaptive* berarti tampilan butir-butir soal tes didasarkan pada hasil analisis dengan algoritma dalam sebuah program, agar butir-butir tes yang diberikan kepada peserta tes 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: 13). 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 melalui logika *fuzzy* untuk keperluan *testing* yaitu *Computerized Intelligence Measurement* (CIM) memiliki prospek yang baik sehingga perlu dikaji dan diteliti lebih lanjut.

Berdasar uraian di atas, terdapat beberapa permasalahan yang perlu diteliti sebagai berikut:

1. Keefektifan dan efisiensi penggunaan TI untuk mendukung pencapaian kualitas pendidikan belum banyak dikaji khususnya untuk mengungkap pengukuran hasil-hasil pembelajaran.
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 pada CAT.
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 logika *fuzzy* untuk mendeskripsikan kemampuan siswa dan penilaian proses program pembelajaran/pendidikan.
8. Belum diketahui hasil kerja program berbantuan komputer dengan logika *fuzzy* 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 logika *fuzzy*. Untuk itu, maka penelitian ini dibatasi pada:

1. Ketepatan program komputer CIM dengan logika *fuzzy* dalam memilih butir-butir tes yang tepat dengan kemampuan siswa.
2. Kinerja program komputer CIM dengan logika *fuzzy* dalam mendeskripsikan kemampuan siswa dan penilaian pelaksanaan program pembelajaran/pendidikan.

Berdasar identifikasi dan batasan masalah di atas maka masalah penelitian dapat dirumuskan sebagai berikut:

1. Bagaimanakah hasil kerja program CIM dengan logika *fuzzy* dalam memilih butir-butir tes yang sesuai dengan kemampuan siswa dalam pengukuran hasil belajar siswa?
2. Bagaimanakah kinerja program CIM dalam mendeskripsikan kemampuan siswa dan penilaian pelaksanaan program pembelajaran/pendidikan?

Tujuan penelitian ini adalah untuk menghasilkan:

1. Sistem pengukuran hasil belajar melalui program berbantuan komputer dengan algoritma *fuzzy* untuk memilih butir-butir tes yang sesuai dengan kemampuan siswa dan melakukan penilaian terhadap hasil belajar.
2. Kinerja program pengukuran hasil belajar berbantuan CIM dengan logika *fuzzy* 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.

B. Kajian Pustaka

1. Pengukuran Hasil Belajar Berbasis CIM

Hambleton, Swaminathan, & Rogers (1991) 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 karakter 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 tes adaptif, 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*).

Pengertian adaptif dalam pengukuran adalah kemampuan untuk menyesuaikan diri dengan kondisi, keperluan, dan keadaan tertentu. Kemampuan yang dimaksud diprogramkan kepada perangkat keras komputer melalui logika *fuzzy*. Kemampuan tersebut adalah kemampuan komputer untuk memilih butir-butir tes yang sesuai dengan karakter 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 pendekatan teknik kecerdasan logika *Fuzzy*, yang berorientasi pada: (a) *item bank*, (b) *item selection*, (c) estimasi kemampuan, (d) pengaturan tingkat kesulitan butir tes, (e) kemudahan layanan pengujian, (f) kemudahan perakitan butir tes (“*tailor*”), (g) penilaian hasil tes.

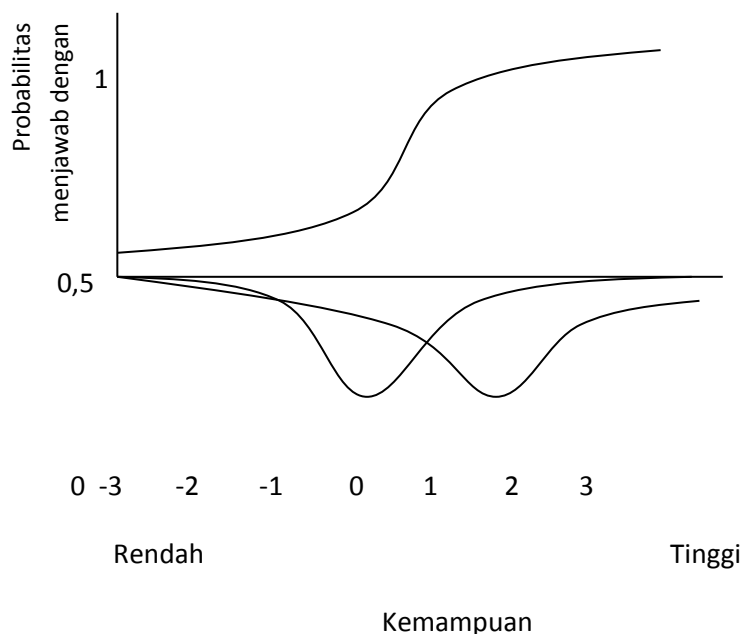
Model pengujian secara modern dengan IRT dibedakan berdasarkan jumlah parameter butir tes, yaitu model satu parameter (*Rasch model*), dua parameter, dan tiga parameter (Hambleton, Swaminathan, & Rogers, 1991). Van der Linden & Hambleton (1997), menyebutkan parameter-parameter tersebut adalah kesulitan 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)):

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

atau

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad i = \text{butir tes ke } 1, 2, 3, \dots, n \quad (1)$$

- $P_i(\theta)$: 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:
Kurva Karakteristik Butir dan Distribusi Kemampuan
(Diadopsi dari Hambleton, Swaminathan, dan Rogers (1991))

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.

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 mensyaratkan 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).

2. Computerized Intelligence Measurement (CIM) dengan Logika Fuzzy

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 dalam pengukuran dengan CIM, 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 -3 hingga +3. Di

samping itu, juga didasarkan pada tingkat daya beda butir tes, yaitu: rendah, sedang, dan tinggi. 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 transferinya dinyatakan dengan notasi $y = f((x,A),B)$. (2)

Secara matematis digambarkan sebagai berikut:

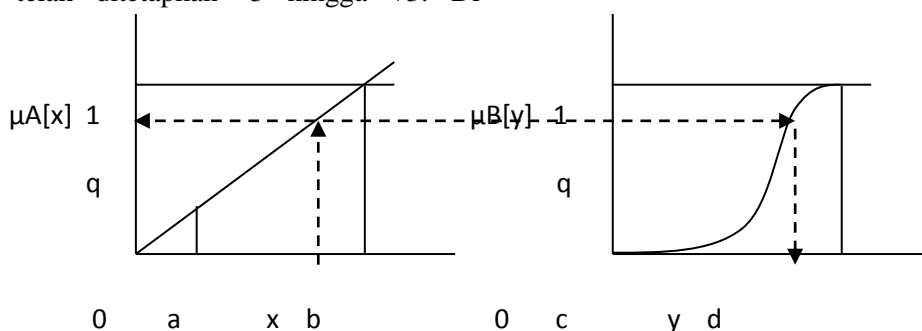
$$\mu_A[x] = (x-a)/(b-a) = q$$

sehingga

$$\mu_B[y] = 1-2[(d-y)/(d-c)]^2 = q$$

(3)

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)).

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).

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 karakter kemampuan $[\theta]$ menjawab butir tes ke-i benar) dan $q_i(\theta)$ (peluang peserta tes dengan karakter kemampuan $[\theta]$ menjawab butir tes ke-i salah) adalah:

$$p_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}} \quad (4)$$

$$p_i(\theta) + q_i(\theta) = 1 \quad (5)$$

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 $p_i(\theta)$, $q_i(\theta)$, dan θ dapat dihitung nilai Likelihood $L(U|\theta)$ dengan formula:

$$L(U|\theta) = \prod_{i=1}^n p_i^u q_i^{1-u} \quad (6)$$

Keterangan:

n: banyak butir tes

u: jawaban siswa pada butir tes

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

$$L(\theta|U) = \frac{L(U|\theta)}{\sum L(U|\theta)} \quad (7)$$

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 karakter kemampuan siswa peserta tes ditentukan dengan formula:

$$\text{Estimasi } \theta = \text{Maximum } [L(\theta | U)] \quad (8)$$

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

C. Metode Penelitian

1. Jenis 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 butir tes mata pelajaran matematik SMA dan mengembangkan perangkat lunak model CIM dengan mengacu model Pressman (1997, 37) dan Rolston (1988, 140).

2. Subjek Penelitian

Penelitian ini dilakukan di: (a) Laboratorium Komputer PEP Pascasarjana UNY, untuk proses pengembangan /*development* produk program CAT. (b) Siswa sekolah menengah sebagai tempat untuk implementasi/*research* produk program CAT dalam situasi yang sebenarnya. (c) 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: (a) Guru bidang 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. (b) siswa kelas XI, untuk uji fungsionalitas dan unjuk kinerja produk program CAT. Mata pelajaran yang diugnakan adalah matematika.

3. Jenis Data

Data yang diperlukan meliputi: (a) Data kualitatif berupa fungsi-fungsi yang diperlukan program, calon pengguna program, dan standard 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.

4. Instrumen Penelitian

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.

5. Teknik Analisis Data

Teknik analisis data yang digunakan dalam rangka menjawab rumusan masalah pada pendahuluan 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 CIM dengan logika *fuzzy* untuk proses pengukuran 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 siswa.

D. Diskusi dan Pembahasan

1. Skenario CIM

Proses yang terjadi di dalam program CIM dimulai dari peserta tes melakukan login, mengerjakan tes, hingga memperoleh hasil tes. Adapun skenarionya sebagai berikut:

- a. 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 .
- b. 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.
- c. Memilih soal berikutnya berdasarkan pola skor. (Kemungkinan skor yang diperoleh adalah 00, 01, 10, atau 11)
- d. Menampilkan soal yang terpilih, ambil respon dari testee, selanjutnya *confirm* dengan kunci untuk memperoleh skor.
- e. Kemungkinan skor yang diperoleh adalah 000, 001, 010, 100, ..., dan seterusnya, atau 111
- f. Menghitung THETA (3 skor) dengan rumus MLE, dan hitung SEM
- g. Memilih soal berikutnya berdasarkan jawaban butir ke-3
- h. Menampilkan soal yang terpilih, ambil respon dari testee, selanjutnya *confirm* dengan kunci untuk memperoleh skor.
- i. Menghitung nilai THETA dari 4 skor terakhir. Demikian juga SEM.
- j. LOOP (mengulang):
 - 1) Mengulangi Langkah f-i dan AKHIRI TES jika: (1) hasil estimasi sudah tercapai maksimum atau (2) nilai SEM lebih kecil atau sama dengan 0.01
 - 2) Mencatat transaksi, nilai THETA terakhir yang paling maksimum, dan nilai SEM terakhir
 - 3) Khusus untuk CONTENT BASED CAT, ditambahkan syarat PEMERATAAN SK secara proporsional

2. Estimasi Kemampuan (Theta)

Metode estimasi theta yang digunakan dalam studi ini adalah *Maximum Likelihood Estimation* (MLE).

PROCEEDING
Seminar Nasional Psikometri

Sistem yang dikembangkan mempunyai tiga bagian. Bagian pertama untuk admin, bagian kedua adalah untuk pengelola, dan bagian ketiga untuk peserta tes. Karakteristik sistem ini secara singkat adalah sebagai berikut: Bagian admin dapat mengelola isi semua tabel. Bagian pengelola dapat memanipulasi semua tabel, kecuali tabel *User*. Pengguna biasa hanya bisa mengelola butir soal. Bagian peserta tes dapat mengubah secara tidak sengaja isi tabel yang terkait dengan hasil tes ketika ia

mengirimkan jawaban atas butir soal yang disajikan sistem. Semua bagian dapat diakses pengguna setelah *login*. Untuk menjamin keamanan sistem *username* dan *password* dienkripsi dan semua halaman selalu *redirect* ke halaman *login* dan tidak dapat di-*bypass*.

Hasil CBT dari peserta tes ditunjukkan dalam Tabel 1. Pada Tabel 1 terlihat bahwa kemampuan peserta tes pada CBT hanya ditentukan berdasarkan persentase jumlah jawaban benar dari soal yang diberikan sistem secara random.

Tabel 1. Hasil CBT

No.	Kode Butir	Daya Beda	Tingkat Kesulitan	Tebakan	Skor Jawaban	θ (Skala 100)
1	11	1.67	0.7	0.15	0	28.571
2	50	0.75	0.03	0.88	1	
3	43	0.69	0.44	0.52	0	
4	62	1.28	0.38	0.05	0	
5	22	0.92	0.2	0.97	0	
6	59	1.58	1.04	0.44	1	
7	18	1.55	1.69	0.82	0	

Hasil CIM dari peserta tes berkemampuan rendah ditunjukkan dalam Tabel 2. Hasil pengolahan data Tabel 2 secara grafik memperoleh riwayat hasil

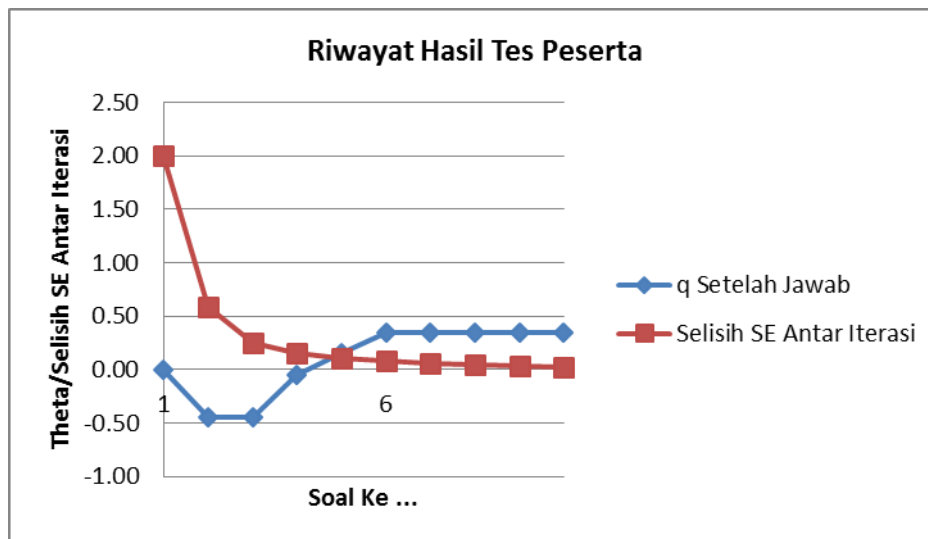
peserta tes seperti ditunjukkan pada Gambar 3. Gambar 3 menunjukkan bahwa untuk lima kali peserta menjawab salah estimasi kemampuan peserta tes sudah konvergen.

Tabel 2. Hasil Simulasi CIM pada Siswa Berkemampuan Rendah

N o.	Kode Butir	Daya Beda	Tingkat Kesulitan	Tebakan	Skor Jawaban	θ Awal	θ Setelah Jawab	$P(\theta)$	$Q(\theta) = 1 - P(\theta)$	IIF	SE (θ)	Selis ih SE Antar Iterasi
1	93	0.13	0	0.3	0	0.00	0.00	0.50	0.50	0.25	2.00	2.00
2	87	0.61	-0.45	0.81	1	0.00	-0.45	0.50	0.50	0.25	1.41	0.59
3	6	0.6	-0.25	0.13	0	-0.	-0.4	0.5	0.42	0.2	1.1	0.25

PROCEEDING
Seminar Nasional Psikometri

						45	5	8		4	6	
4	63	0.91	-0.05	0.55	1	-0.45	-0.05	0.50	0.50	0.25	1.00	0.16
5	69	1.37	0.15	0.71	1	-0.05	0.15	0.50	0.50	0.25	0.90	0.11
6	83	0.95	0.35	0.1	1	0.15	0.35	0.50	0.50	0.25	0.82	0.08
7	88	1.36	0.55	0.39	0	0.35	0.35	0.58	0.42	0.24	0.76	0.06
8	76	1.67	0.75	0.11	0	0.35	0.35	0.66	0.34	0.22	0.71	0.04
9	86	1.79	0.95	0.26	0	0.35	0.35	0.73	0.27	0.19	0.68	0.03
10	28	0.49	1.15	0.55	0	0.35	0.35	0.80	0.20	0.16	0.67	0.01



Gambar 3. Riwayat Hasil Tes Peserta CIM Berkemampuan Rendah

Deskripsi ketepatan kerja program CIM mengenai pemilihan butir-butir tes yang diberikan kepada siswa digambarkan dengan analisis nilai rata-rata, korelasi dan simpang baku secara teori dengan hasil kerja logika *fuzzy*. 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 menunjukkan bahwa program CIM dengan algoritma logika *fuzzy* berhasil dengan baik

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

Gambar 4: Korelasi Tingkat Kesulitan Butir antara Analisis Teori dengan Hasil *Fuzzy*

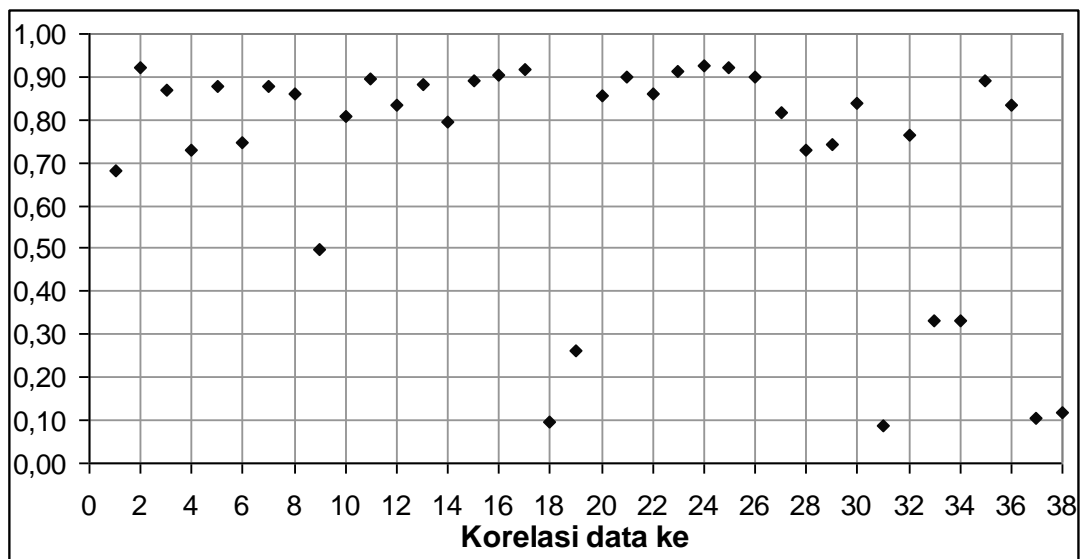
E. Kesimpulan

1. Sistem *bank* soal yang dapat menampung butir soal yang bisa digunakan untuk berbagai keperluan tes dapat dibuat dengan entitas jenjang pendidikan, kelas, mata pelajaran, KI, KD, indikator, butir, peserta tes, sekolah, waktu, detail tes, dan pengguna.
2. CBT dapat dikembangkan dengan menyajikan soal secara *random*, menguji jawaban peserta, menghitung jawaban benar & salah, mengecek alokasi waktu yang tersedia. Bila waktu habis atau semua soal telah disajikan, selanjutnya dihitung kemampuan akhir peserta tes.
3. CIM dapat dikembangkan dengan cara peserta tes diberi soal dengan tingkat kesulitan sedang dengan asumsi kemampuan awalnya (θ awal) juga sedang. Kemudian dihitung: 1) kemampuan (θ) setelah menjawab

berdasarkan daya beda (a), tingkat kesulitan (b), dan tebakan semu (c) butir soal, 2) probabilitas menjawab benar berdasarkan kemampuan tersebut ($P(\theta)$), 3) probabilitas menjawab salah ($Q(\theta)$), 4) fungsi informasi butir ($I_i(\theta)$), 5) kesalahan baku ($SE(\theta)$), dan 6) selisih kesalahan baku antar penyajian soal. Proses diulang sampai selisih kesalahan baku antar penyajian soal sekecil mungkin, soal atau waktu habis.

2. Keterbatasan Penelitian

- a. 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.
- b. Jenis dan model tes yang biasa



- c.
- d. 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.

- e. Program CIM 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.
- f. 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 jumlahnya. Jumlah basis data bank soal yang baik tentunya yang memenuhi batas kecukupan, yakni jika soal yang diujikan 50 butir maka jumlah minimal basis data bank soal adalah 300 butir dan masing-masing tingkat kesulitan memiliki rentang -3 sampai +3.
 - g. Jumlah basis pengetahuan yang digunakan dalam logika fuzzy 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.
 - h. Pengujian berbasis komputer dengan model CIM 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.
 - i. Keterbatasan lain adalah dalam penggunaan model CIM 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 CIM menyediakan fasilitas pass untuk melompat/skip ke butir tes yang selanjutnya. Butir tes yang di skip tidak akan dihitung dalam penentuan skor.

3. Saran Pemanfaatan dan Pengembangan Lebih Lanjut

- a. Program CIM menggunakan algoritma penalaran merupakan salah satu alternative 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.
- b. Program CIM 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 CIM 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 CIM ini, seyogyanya dilatih terlebih dahulu cara mengoperasikan dan bekerja dengan komputer. Hal itu dimaksudkan agar dalam mengerjakan ujian terkomputerisasi dengan program CIM, siswa telah terbiasa (familier) menggunakan komputer.
- c. 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 CIM 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

- Anderson, L.W., & Krathwohl, D.R. (2001). *A taxonomy for learning, teaching, and assessing*. New York: Addison Wesley Longman. Inc.
- Azwar, S. (1996). *Tes prestasi, fungsi dan pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2003). *Penyusunan skala psikologi*. Yogyakarta: Pustaka Pelajar.
- 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.
- 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.
- Clarke, S. (2005). *Formative assessment in action weaving the elements together*. London: Hodder Murray.
- Davis, L.L., & Dodd, B.G. (2003). Item exposure constraints for testlets in the verbal reasoning section of the MCAT. *American Psychological Measurement*. vol: 27 no: 5 Sep 2003.
- Depdikbud, Ditjen Dikdasmen, & Dit Dikmenum. (1997). *Pengelolaan pengujian bagi guru mata pelajaran*. Jakarta: Depdikbud.
- Echols, J.M., & Shadily, H. (1986). *Kamus Inggris Indonesia*. Jakarta: PT. Gramedia.
- Friedenberg, L. (1995). *Psychological testing: Design, analysis, and use*. Boston: Allyn & Bacon .
- Gronlund, N.E., & Linn, R.L. (1990). *Measurement and evaluation in teaching (6th ed.)*. New York: Mc Millan Publishing Company.
- Hambleton, R.K., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory*. New Delhi: Sage Pub. Inc.
- Hau, K.T., & Chang, H.H. (2001) Item selection in CAT: Should more discriminating items be used first? *Journal of Evaluation & Measurement*. Vol: 38 no: 3.
- Hornby, AS. (1984). *Oxford advanced learner's dictionary of current English*. London: Oxford University Press.
- Leung, C.K., Chang, H.H., & Hau, K.T. (2002). Item selection in CAT: Improving the a-stratified design with the sympon-hetter algorithm. *Journal of APM*. vol 26 no.4 Des 2002.
- Linn, 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*

PROCEEDING
Seminar Nasional Psikometri

- 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.
- Pressman, R.S. (1997). *Software engineering: A practitioner's approach*. New York: Mc Graw hill Book, Inc.
- Rolston, D.W. (1988). *Principles of artificial intelligence and expert systems development*. Singapore: McGraw Hill Book, Co.
- Stark, J.S., & Thomas, A. (1994). *Assessment and program evaluation*. New York: Simon & Schuster Custom Publishing.
- Steward, D.V. (1987). *Software engineering, with systems analysis and design*. New York: Wadsworth, Inc.
- Silverius, S. (1991). *Evaluasi hasil belajar dan umpan balik*. Jakarta: Grasindo Gramedia Widiasarana Indonesia.
- Surapranata, S. (2004). *Analisis, validitas, reliabilitas dan interpretasi hasil test*. Bandung: PT Remaja Rosdakarya.
- Suryabrata, S. (2000). *Pengembangan alat ukur psikologis*. Yogyakarta: Andi Offset.
- 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.
- Tim Pustaka Yustisia. (2007). *Panduan lengkap KTSP (kurikulum tingkat satuan pendidikan)*. Yogyakarta: Pustaka Yustisia.
- Towndrow, P.A., & Vallenge, 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.
- Weinberg G.M., & Geller. (1985). *Computer information systems: An introduction to data processing*. New York: Little Brown & Company. Ltd.
- Widrow, B., & Stearns, S.D. (1985). *Adaptive signal processing*. Englewood Cliffs: Prentice Hall Inc.
- Woolfolk, A. (2007). *Educational psychology (10th ed.)*. New York: Pearson Education, Inc.