

ARTIKEL PENELITIAN KOLABORASI

TAHUN ANGGARAN 2014

**Pengembangan Sistem Pengujian Terkomputerisasi untuk
Kompetensi Bidang Kejuruan Siswa SMK di DIY**



Oleh:

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

M. Khaerudin, Ph.D.

Rustam Asnawi, Ph.D.

JURUSAN PENDIDIKAN TEKNIK ELEKTRO

FAKULTAS TEKNIK

UNIVERSITAS NEGERI YOGYAKARTA

OKTOBER 2014

I. Pendahuluan

Keberadaan sejumlah komputer di sekolah menunjukkan adanya komitmen pemerintah dalam meningkatkan kualitas pendidikan berbasis TI. Namun demikian, sejauh ini efektivitas dan efisiensi peranan TI dalam peningkatan kualitas proses pembelajaran/pendidikan untuk meningkatkan kualitas hasil belajar masih perlu dilakukan penelitian, khususnya sistem pengujian dan penilaian hasil belajar berbantuan komputer.

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 umum yang melanjutkan ke jenjang pendidikan berikutnya, dan persentase lulusan sekolah menengah kejuruan/vokasi yang masuk ke dunia usaha dan atau dunia industri, serta rendahnya persentase tingkat *drop out* atau siswa putus sekolah. Untuk itu, permasalahan yang masih perlu dikaji adalah penggunaan TI untuk meningkatkan kualitas pencapaian hasil belajar dengan melalui pendalaman materi atau tes.

Target keberhasilan peningkatan kualitas pendidikan, khususnya dalam pembelajaran adalah melalui *assessment* berbantuan 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 Base Test* (CBT), komputer diprogram untuk mengolah urutan tampilan butir-butir soal ujian dan menerima pilihan jawaban dari peserta (Chee & Wong. 2003 96; Towndrow & Vallence. 2004: 244). Namun sejauh ini, permasalahan mengenai penggunaan dan pengujian atau pengukuran hasil belajar berbantuan TI yang tepat merupakan permasalahan yang menarik untuk diteliti secara terus menerus.

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 antara lain 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.

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 dan 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 banyak diketahui ketepatan program berbantuan komputer dengan algoritma penalaran untuk mendeskripsikan kemampuan siswa dan penilaian proses program pembelajaran/pendidikan.

Berdasar permasalahan di atas dapat dirumuskan sebagai berikut: (1) Bagaimanakah sistem inferensi dalam program assessmen pembelajaran dan penilaian pembelajaran berbantuan komputer memilih butir-butir tes yang sesuai dengan kemampuan siswa? (2) Bagaimanakah akurasi kinerja program assessmen pembelajaran dan penilaian berbantuan komputer dalam mendeskripsikan kemampuan siswa dan penilaian pelaksanaan program pembelajaran?

Adapun tujuan penelitian ini adalah untuk menghasilkan: (1) Sistem pengujian hasil belajar dan penilaian pendidikan melalui program berbantuan komputer dalam memilih butir-butir tes yang sesuai dengan kemampuan siswa dan melakukan penilaian terhadap keberhasilan belajar. (2) Program pengujian hasil belajar dan penilaian pendidikan berbantuan komputer yang teruji dan memiliki akurasi yang baik untuk mendeskripsikan kemampuan siswa.

II. Kajian Teori

A. Pengujian Berbantuan Komputer

Hambleton, Swaminathan, & Rogers (1991: 145-146) dalam bukunya "*Fundamentals of Item Response Theory*" menyebutkan bahwa yang dimaksud dengan pengujian terkomputerisasi merupakan "*would be to give every examinee a test that is "tailored" or adapted, to the examinee's ability level*". Selanjutnya juga dikatakan bahwa tes dengan panjang butir yang telah *fixed*, tidak efisien jika diberikan kepada semua peserta tes. 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 butir tes yang sesuai dengan kemampuannya (*administered a unique set of items*). Masters & Keeves (1999: 129) menyebutkan bahwa di dalam adaptif tes, butir tes dipilih dari basis data *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 (*testing*).

Berdasar kajian terhadap sejumlah jurnal penelitian mengenai pengujian adaptive, dapat dikemukakan bahwa penerapan komputerisasi untuk penilaian dan pengujian hasil belajar adalah sebagai berikut: (a) kemudahan layanan pengujian, kecepatan pengolahan

hasil dan kecepatan penyajian informasi hasil pengujian, (b) Analisis respons butir soal, (c) Analisis butir tes model monotonous maupun polytomous, dikotomi dan non-dikotomi, (d) Analisis butir tes model unidimensi maupun multidimensi, (e) Kemudahan administrasi terhadap kumpulan butir-butir soal, atau sering dikenal dengan istilah *bank* soal, (f) Pemilihan butir soal yang diberikan kepada peserta tes, (g) Analisis butir tes untuk konstruksi tes, (h) Pengujian statistik dengan berbagai algoritma dan rumus-rumus yang ada, (i) Dasar pengaturan jenis tingkat kesulitan tes, (j) Analisis studi banding pengujian adaptif dan non-pengujian adaptif, (k) Generalisasi nilai hasil pengujian, dan (l) Dasar estimasi kemampuan peserta tes. Adapun pendekatan analisis atau algoritma yang digunakan adalah non-monotonik atau statistikal.

1. Pemilihan Butir Tes

Salah satu hal pokok yang perlu mendapat perhatian dalam penyusunan dan pengembangan instrumen tes, di samping prosedur penulisan butir tes, adalah prosedur analisis dan seleksi butir tes. Hal ini menjadi teramat penting karena kualitas instrumen tes juga ditentukan oleh kualitas butir-butir tes di dalamnya. Dalam instrumen tes (Wood dan Skurnik, 1969: 55), hanya butir-butir yang ditulis dengan mengikuti *blue-print* dan bimbingan kaidah penulisan butir yang benar sajalah yang akan berfungsi sebagaimana seharusnya dan dapat mendukung validitas skala secara keseluruhan. Selanjutnya butir-butir yang telah ditulis dengan cara yang benar masih perlu diuji untuk menentukan fungsi butir secara benar seperti yang diharapkan.

Leung, Chang, & Hau (2002: 11), dari hasil penelitiannya dalam jurnal mengenai pemilihan butir tes mengemukakan bahwa kendali pemilihan butir, meminimalkan *Test-Overlap*, dan penggunaan kelompok butir yang efisien adalah sebagian dari isu yang penting di dalam mendisain pengujian secara adaptif terkomputerisasi (cerdas). Metode memilih butir yang digunakan adalah Sympson-Hetter (SH), secara parsial memberi solusi dengan baik melalui prosedur probabilistik yang mencoba untuk pengendalian dengan hasil *high-discrimination* materi yang efektif di dalam meningkatkan pemanfaatan keseluruhan kelompok butir, tanpa mengorbankan efisiensi dan efektivitas dalam penilaian kemampuan.

Davis & Dodd (2003: 10), berdasar hasil penelitiannya dalam jurnal CAT, menyebutkan bahwa pemilihan butir ditujukan untuk ketepatan materi tes yang akan diberikan kepada peserta tes. Untuk itu diperlukan parameter yang jelas tentang karakteristik materi yang akan diberikan. Hasil penelitian Hau, & Chang (2001: 13), mengenai pemilihan butir lebih menekankan pada kemudahan administrasi dan biaya yang murah dalam penyelenggaraannya.

Prosedur seleksi atau pemilihan butir menyangkut beberapa tahap kerja. Prosedur yang paling sederhana meliputi dua tahap (Azwar, 2003: 55). Tahap pertama, analisis dan seleksi butir berdasarkan evaluasi kualitatif. Evaluasi ini melihat: kesesuaian butir yang

ditulis dengan *blue-print* dan indikator perilaku yang hendak diungkap. Kesesuaian butir yang telah ditulis dengan kaidah penulisan yang benar. *Social desirability* yang tinggi dari butir-butir yang ditulis perlu diperhatikan.

Tahap kedua, adalah prosedur seleksi butir berdasarkan data empiris (data hasil uji coba butir pada kelompok subjek yang karakteristiknya setara dengan subjek yang hendak dikenai pengujian) dengan melakukan analisis kuantitatif terhadap parameter-parameter butir. Pada tahap ini paling tidak dilakukan seleksi butir berdasarkan daya pembeda, dan tingkat kesulitan butir. Seleksi butir yang lengkap meliputi: analisis validitas dan reliabilitas tes, distribusi jawaban, aplikasi analisis faktor, bias butir, dan fungsi informasi.

Berdasar kajian teori dan hasil penelitian tersebut di atas, proses pemilihan butir tes dan estimasi kemampuan peserta tes dilakukan dengan komputer. Di samping itu, juga kemudahan layanan pengujian, kecepatan pengolahan hasil dan kecepatan penyajian informasi hasil pengujian terprogram serta kualitas program pembelajaran. Pendekatan algoritma ini menerapkan salah satu teknik komputasi dipandang sangat cocok untuk menyelesaikan permasalahan yang bersifat kualitatif maupun kuantitatif yang melibatkan permasalahan numeris (angka/bilangan) maupun alpha-numeris (karakter/string).

2. Mekanisme Analisis Butir Tes

Silverius (1991: 4) mengatakan bahwa hasil tes diharapkan dapat memberikan gambaran atau informasi yang akurat tentang tingkat kemampuan seseorang terhadap penguasaan sesuatu materi. Winkel (2004: 531) menyebutkan: hasil tes dapat memberi informasi yang sangat berguna bagi pengambilan keputusan serta dapat digunakan sebagai salah satu dasar untuk mengadakan penilaian dalam rangka menentukan kebijakan. Oleh sebab itu, kualitas kebijakan dipengaruhi oleh kualitas informasi hasil tes. Kualitas informasi hasil tes ditentukan oleh kualitas tes. Selanjutnya kualitas tes didapat dari butir tes yang dirakit memenuhi rambu-rambu dan aturan yang telah ditentukan.

Untuk menguji kualitas setiap butir tes, perlu dilakukan analisis butir tes. Tujuan utama analisis butir tes untuk mendapatkan informasi tentang karakteristik setiap butir tes, baik secara kualitatif maupun secara kuantitatif (Surapranata, 2004: 1). Hasil analisis butir tes dapat digunakan untuk menguji apakah suatu soal diperkirakan akan berfungsi (analisis kualitatif) atau telah berfungsi (analisis empirik) dengan baik.

Analisis kualitatif, pada dasarnya adalah penelaahan butir tes ditinjau dari segi kaidah penulisan soal, yaitu: 1) isi dan materi, 2) konstruksi, dan 3) bahasa. Telaah materi, digunakan untuk mengetahui apakah isi atau materi yang ditanyakan sesuai dengan tujuan pertanyaan yang tersirat dalam indikator. Telaah konstruksi tes, untuk melihat apakah kaidah-kaidah penulisan tes telah sesuai. Telaah Bahasa, untuk melihat apakah bahasa

yang digunakan sudah jelas dan komunikatif, sehingga mudah dimengerti serta tidak menimbulkan pengertian dan penafsiran yang berbeda (Surapranata, 2004: 1-10).

Analisis kuantitatif (Surapranata, 2004: 10)., dasarnya adalah menganalisis respons tes yang diberikan oleh *testee* (peserta uji). Ada dua cara analisis kuantitatif, yaitu dengan teori tes klasik dan dengan teori tes modern, yaitu *item response theory* /IRT (Hambleton, Swaminathan, & Rogers, 1991: 2). Penggunaan IRT akan menghasilkan esitmasi kemampuan yang lebih akurat dibandingkan dengan teori tes klasik.

Karakteristik soal pada umumnya meliputi tingkat kesulitan dan daya pembeda. Daya beda menurut teori tes klasik adalah proporsi peserta tes yang menjawab butir benar, sedang daya beda adalah besarnya koefisien korelasi *point biserial*. Tingkat kesulitan dan daya beda butir tes berdasarkan teori respons butir yaitu bergerak dari -3 sampai +3 sesuai dengan penggunaan distribusi normal..

a) Tingkat Kesulitan Tes (p)

Secara klasik tingkat kesulitan tes (p) diperoleh dengan beberapa cara, antara lain: skala kesulitan linier, skala bivariat, indeks Davis, dan proporsi men-jawab benar serta dengan item respon teori (IRT). Prinsipnya terdapat dua cara, yaitu: secara teori tes klasik dan secara IRT. Secara tes klasik (Surapranata, 2004: 12; Friedenberg, 1995: 262; Silverius, 1991: 166) klasifikasi tingkat kesulitan yang baik bervariasi dari 0,3 sampai 0,7. Menurut IRT tingkat kesulitan butir tes yang baik bervariasi dari -3 sampai +3.

b) Daya Beda Tes (d)

Daya beda suatu butir tes berfungsi untuk menentukan dapat tidaknya suatu butir membedakan kemampuan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada pada kelompok itu. Tujuan pengujian daya beda adalah untuk melihat kemampuan butir tes dalam membedakan kemampuan antara peserta yang berkemampuan tinggi dengan peserta yang berkemampuan rendah (Surapranata, 2004: 23; Friedenberg, 1995: 271; Silverius, 1991: 172). Seperti halnya dengan daya beda butir, juga terdapat cara tes klasik dan IRT dalam menentukan butir tes yang baik.

3. Teori Respons Butir

Model analisis butir dengan teori tes klasik memiliki asumsi yang lemah, yakni: asumsi yang dapat dengan mudah dipenuhi oleh kebanyakan data tes. Adapun kelemahan tes klasik (Hambleton, Swaminathan, and Rogers, 1991: 2), adalah: (a) statistik yang digunakan untuk analisis butir tes bergantung pada sampel yang digunakan. (b) skor yang diperoleh peserta tes tergantung pada tingkat kesulitan butir tes (c) konsep tes paralel menurut tes klasik dalam prakteknya sulit dipenuhi, (d) teori tes klasik tidak memperhatikan

pola respons peserta tes, sehingga hasilnya akurag akurag dan (e) indeks kesalahan baku pengukuran diasumsikan sama untuk semua peserta tes

Kelemahan teori tes klasik tersebut dapat diatasi dengan menggunakan teori respons butir (*item response theory/IRT*). Pada teori ini estimasi kemampuan peserta tes berdasarkan pola respons atau jawaban peserta tes, sehingga lebih banyak informasi yang digunakan, dan hasilnya akan lebih akurat. Teori respons butir menggunakan distribusi normal, sehingga hasil penskorannya harus dikonversi ke skala baku (Hambleton, Swaminathan, & Rogers, 1991:9). Konversi skor dari skala pada distribusi normal ke skala baku akan memudahkan bagi setiap orang untuk menafsirkan hasilnya.

Berdasarkan jumlah parameter yang diestimasi, teori respons butir diklasifikasi menjadi tiga. Parameter yang digunakan ada yang satu, yaitu tingkat kesulitan butir, sering disingkat dengan model 1-p dan disebut dengan Rasch model. Model dua parameter, yaitu parameter tingkat kesulitan dan daya abeda, dikenal dengan model 2-p. Model yang ketiga adalah model 3-p, yaitu yang memiliki tiga parameter. Parameter tersebut adalah tingkat kesulitan, daya beda dan tampak dugaan. (Hambleton, Swaminathan, & Rogers, 1991:12,; Van der Linden & Hambleton 1997: 4), Jadi ada tiga model pada teori respons butir dalam mengestimasi parameter kemampuan peserta tes.

Persamaan matematik teori respons butir untuk butir tes dikotomi dua parameter, yaitu parameter tingkat kesulitan butir, daya beda butir, adalah sebagai berikut (van der Linden & Hambleton (1997: 5); (Hambleton, Swaminathan, & Rogers, 1991: 15); (Lord, 1980: 14)):

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad \text{atau} \quad P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad i = \text{butir tes ke } 1, 2, 3, \dots, n-1, n$$

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

Asumsi model Rasch didasarkan pada persyaratan objektivitas spesifik (Hambleton, Swaminathan, & Rogers, 1991: 9). Asumsi ke-: (a) adanya *unique ordering (unidimensionality)* menurut kemampuan orang dan kesulitan butir tes. (b) *local independence*, peluang peserta tes menjawab benar suatu butir tes tidak bergantung pada peluang menjawab benar butir lain (c) besarnya daya beda antar butir homogen (d) tes hanya mengukur satu dimensi (e) tidak ada dugaan dalam menjawab tes, karena tingkat kesulitan tes sama dengan tingkat kemampuan peserta tes.

Model analisis butir tes dengan teori respons butir yang digunakan dalam penelitian ini adalah model 1-p, yaitu dengan parameter tingkat kesulitan. Semua butir tes yang akan

digunakan dikalibrasi terlebih dahulu untuk menentukan besarnya parameter butir. Butir soal yang sudah dikalibrasi tersebut disimpan dalam basis data *bank* soal yang terkomputerisasi. Pemilihan butir tes dianalisis dengan menggunakan algoritma cerdas yang mendasarkan pada respons jawaban peserta ujian pada butir tes sebelumnya.

B. Assesmen Pembelajaran

Assesmen pembelajaran sebagai komponen instruksional bertujuan untuk mengetahui keberhasilan pembelajaran menjadi berharga, bermutu atau bernilai (Winkel, 2004: 531). Fokus Penilaian adalah individu, yaitu prestasi belajar yang dicapai dalam kelompok atau kelas (Mardapi, 2008:8). Baumgartner & Jackson (1995: 15) menyebut evaluasi adalah penilaian yang digunakan untuk mengukur ketercapaian tujuan belajar, yakni: (1) banyak materi yang diserap siswa, (2) perkembangan kemampuan siswa menghadapi persoalan nyata, dan (3) kualitas pelaksanaan program pengajaran. Penilaian untuk mengetahui pencapaian tujuan atau aktivitas yang dilakukan melalui kegiatan belajar, disebut penilaian hasil. Penilaian untuk mengetahui suatu kegiatan program pembelajaran dapat berhasil mencapai tujuan yang telah ditetapkan disebut penilaian proses (Winkel, 2004: 540).

Evaluasi produk menurut Mardapi (2008:11) dibedakan ke dalam dua hal, yaitu: penilaian formatif yang diberikan selama proses pembelajaran, dan penilaian sumatif yang diberikan pada akhir setelah periode pembelajaran. Penilaian formatif sering disebut dengan *assessment for learning* dan penilaian sumatif disebut juga dengan istilah *assessment of learning*. Hasil penilaian formatif digunakan untuk perbaikan strategi pembelajaran, dan hasil penilaian sumatif sebagai dasar kebijakan untuk membuat keputusan tentang peringkat kemampuan hasil belajar peserta didik.

Tujuan assesmen pembelajaran menurut Winkel (2004: 538) dan Baumgartner & Jackson (1995:179), meliputi: (1) untuk menimbulkan motivasi peserta didik, (2) memberikan umpan balik kepada peserta didik, (3) memberikan umpan balik kepada pendidik, (4) memberikan informasi kepada orang tua/lembaga pengirim atau *stakeholders*, (5) memberikan informasi seleksi atau pernyataan kelulusan, dan (6) memberikan pertanggungjawaban terhadap pelaksanaan kegiatan.

C. Pertanyaan Penelitian

Pertanyaan yang diajukan dalam penelitian ini adalah:

1. Bagaimanakah membangun model sistem inferensi dalam program assesmen dan penilaian pembelajaran berbantuan komputer dengan algoritma *cerdas* dalam:
 - a. Memilih butir-butir tes yang sesuai dengan kemampuan siswa?
 - b. Mendeteksi kemungkinan jawaban salinan dari peserta tes?

2. Bagaimanakah (a) akurasi/ketepatan program, (b) kinerja program pengujian dan penilaian berbantuan komputer dalam mendeskripsikan kemampuan siswa dan penilaian pelaksanaan program pembelajaran/pendidikan?

III. Metode Penelitian

Penelitian ini menggunakan pendekatan *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-butir tes mata pelajaran matematik sekolah menengah dan mengembangkan perangkat lunak.

A. Desain Uji Coba

Uji coba produk diterapkan untuk mengetahui kemampuan produk dalam melakukan proses pengujian keberhasilan siswa terhadap topik pelajaran yang telah diajarkan guru. Proses pengujian dilakukan menggunakan bantuan unit komputer dan dilaksanakan di lab komputer. Sebelum program digunakan, guru melakukan seting (konfigurasi) soal yang antara lain: memasukkan banyaknya soal (N) yang akan diujikan, memasukkan lama waktu batas pelaksanaan ujian (T), dan mendaftar sejumlah siswa peserta yang akan diuji agar dapat *login* dan mengerjakan tes di komputer.

Pada saat peserta uji menghadapi program cerdas dengan algoritma cerdas dan siap untuk mengerjakan soal-soal, pertama komputer akan menampilkan butir tes ke-1 dengan tingkat kesulitan sedang yang diambil secara acak dari basis data *bank* soal. Selanjutnya komputer mencatat lama waktu mengerjakan dan respon jawaban peserta untuk menentukan tingkat kesulitan butir tes ke-2 dan seterusnya. Jawaban benar akan dibobot 1 dan jawaban salah dibobot 0. Benar dan salah respon jawaban peserta juga digunakan untuk penentuan skor yang diperoleh. Penghitungan waktu digunakan untuk menentukan lama waktu yang telah digunakan peserta uji dalam mengerjakan tes. Besarnya penguasaan materi yang diperoleh dan lama waktu (T) yang dipakai peserta tes digunakan untuk menentukan klasifikasi kemampuan (KM) peserta tes.

B. Subjek Coba

Penelitian *research and development* ini dalam pelaksanaannya dilakukan di: (a) Laboratorium Komputer JPTE UNY, untuk proses pengembangan /*development* produk program cerdas. (b) SMK di Daerah Istimewa Yogyakarta, sebagai tempat untuk implementasi/*research* produk produk program cerdas dalam situasi yang sebenarnya. (c) Waktu penelitian, dilakukan pada bulan April 2014 sampai dengan Mei 2014 untuk pembuatan program. Bulan Mei sampai dengan September 2014 dilakukan implementasi atau pengujian program tahap pertama.

Subjek yang digunakan dalam penelitian untuk uji coba meliputi: (a) Guru bidang mata pelajaran dasar listrik untuk proses penggalian informasi dan identifikasi kebutuhan sistem yang diperlukan dalam program dan menilai produk program. 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. Mata pelajaran yang diugnakan adalah dasar listrik.

C. Jenis Data

Data yang diperlukan dalam penelitian ini 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 cerdas dan hasil unjuk kerja program dalam proses pengujian hasil belajar siswa.

D. Instrumen Pengumpul Data

Penelitian ini meliputi dua kegiatan, yaitu: (a) pengembangan produk 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.

E. Teknik Analisis Data

Teknik analisis data yang digunakan dalam rangka menjawab rumusan masalah pada BAB I yang kemudian dirinci dalam pertanyaan penelitian yang diajukan pada BAB II 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.

IV. HASIL PENELITIAN

A. Skenario tes

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

1. Pilih 2 butir soal dengan nilai b dalam interval $-0.5 < b < +0.5$ (default),
2. Tampilkan kedua butir soal tersebut satu per satu. Ambil respon dari testee, selanjutnya analisis respon untuk memperoleh skor. Skor=1 jika benar, skor=0 jika salah.
3. Pilih soal berikutnya berdasarkan pola skor.
4. Tampilkan soal yang terpilih,
5. Hitung theta (3 skor) dengan rumus MLE , dan hitung SEM
6. Pilih soal berikutnya berdasarkan jawaban.
7. LOOP (ulang): Lakukan kembali butir b sampai f .
8. akhiri tes jika: (1) jumlah butir soal sudah mencapai maks atau (2) nilai SEM lebih kecil atau sama dengan 0.33

Implementasi ini dilakukan oleh mahasiswa Kim Fajrin dengan menggunakan satu parameter atas RASCH model. Hasil penelitiannya menyebutkan bahwa butir” yang sudah terkalibrasi dengan IRT mampu berfungsi untuk mengukur kemampuan siswa lebih akurat. Untuk dua Mahasiswa yang lainnya meneliti scanning jawaban siswa yang dilakukan oleh M Roisul dan untuk Mahasiswa Fajar meneliti tes esay.

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

B. Hasil evaluasi terhadap program

Secara umum algoritma (*initial/the first item, estimation & next item, dan stopping rules*) sudah berjalan sesuai dengan rancangan. Namun demikian, masih ada sedikit masalah dengan perhitungan atau kalkulasi sebagai berikut:

1. Hasil estimasi theta dengan MLE adalah benar dengan membandingkan hasil perhitungan Excel, bahkan lebih akurat karena menggunakan 4 digit desimal
2. 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.

3. Di samping itu, tampilan pada laporan perlu diurutkan berdasarkan nomor urut butir soal yang tampil, serta perlu ditambahkan field sumber untuk menelusuri butir soal yang dipilih.

V. SIMPULAN DAN SARAN

A. Simpulan tentang Produk

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

1. Model pengujian dengan algoritma penalaran 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.
2. Sistem inferensi model pengujian 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.
3. Ketepatan sistem inferensi dalam memilih butir-butir tes yang sesuai dengan respons jawaban siswa diwujudkan pada nilai.
4. Kemampuan model pengujian dengan algoritma penalaran dalam mendeskripsikan kemampuan siswa berdasar butir-butir tes yang diterima dan nilai-nilai estimasi hasil analisis,
5. Hasil temuan karya disertasi ini, adalah: (1) model pengujian dengan algoritma penalaran digunakan untuk membentuk sistem inferensi. (2) Sistem inferensi model pengujian: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.

B. Saran

Program pengujian 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 pengujian 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 pengujian ini, seyogyanya dilatih terlebih dahulu cara mengoperasikan dan bekerja dengan komputer. Hal itu dimaksudkan agar dalam mengerjakan ujian terkomputerisasi dengan program pengujian, 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 pengujian 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 pengujian 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.
- Baeck, T., Fogel, D.B., & Michalewicz, Z. (1997). *Handbook on evolutionary computation*. New York: IOP Press.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- van der Linden, W.J., & Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer Verlag.