

**LAPORAN PENELITIAN KOLABORASI  
TAHUN ANGGARAN 2014**



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

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

**M. Khaerudin, Ph.D.**

**Rustam Asnawi, Ph.D.**

**Dibiayai Dana DIPA BLU Universitas Negeri Yogyakarta Tahun  
Anggaran 2013 Sesuai dengan Surat Perjanjian Pelaksanaan  
Penelitian Dosen Fakultas Teknik Universitas Negeri Yogyakarta**

Nomor : .....

---

**FAKULTAS TEKNIK  
UNIVERSITAS NEGERI YOGYAKARTA  
2014**



DEPARTEMEN PENDIDIKAN DAN KEBUDAYAAN  
UNIVERSITAS NEGERI YOGYAKARTA  
FAKULTAS TEKNIK  
Alamat: Karangmalang Yogyakarta 55281  
Telp. 586168 pos. 292, 276, Telp & Fax (0274) 586734

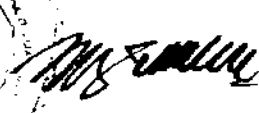


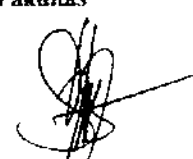
## HALAMAN PENGESAHAN LAPORAN PENELITIAN

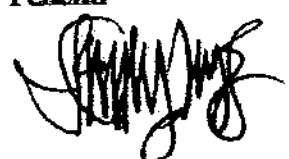
1. **Judul** : Pengembangan Sistem Pengujian Terkomputerisasi untuk Kompetensi Bidang Kejuruan Siswa SMK di Daerah Istimewa Yogyakarta
2. **Ketua Pelaksana Penelitian** :
  - a. Nama Lengkap : Dr. Haryanto, M.Pd., M.T.
  - b. Tempat, Tanggal Lahir : Kebumen, 10 Maret 1962
  - c. Jabatan Fungsional : Lektor
  - d. Program Studi : Pendidikan Teknik Elektro
  - e. Jurusan : Pendidikan Teknik Elektro
  - f. Alamat Rumah : Perum Tirto Permai 17 Nogotirto Gamping Sleman
  - g. Telpon/Faks/HP : 08164224572
  - h. e-mail : haryanto.ftuny@gmail.com
  - i. Bidang Keahlian : Evaluasi Pendidikan Teknologi Kejuruan
3. **Jenis Penelitian** : Mandiri
4. **Jumlah Tim Peneliti** : Ketua : satu orang  
Anggota : dua orang
5. **Lokasi Penelitian** : Jurusan Pendidikan Teknik Elektro FT UNY
6. **Biaya Yang Diperlukan**
  - a. Sumber dari Fakultas : Rp 10.000.000,00
  - b. Sumber lain : Rp ---

Jumlah : Rp 10.000.000,00

Yogyakarta, 31 Okt 2014

Dean Fakultas Teknik  
  
(Dr. Moch. Bruni Triyono)  
NIP.19560216 1986031 003

BP Fakultas  
  
(DR. Siti Hamidah)  
NIP. 19530820 197903 2 001

Peneliti  
  
(Dr. Haryanto, M.Pd., M.T.)  
NIP. 196203101986011001

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

**Haryanto. M. Khaerudin, Rustam Asnawi**

### **ABSTRAK**

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.

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 listrik dasar sekolah menengah dan mengembangkan perangkat lunak. Adapun langkah-langkah yang dibutuhkan adalah: (1) Analisis kebutuhan sistem, (2) Disain sistem, (3) implementasi, dan (4) Pengujian. 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 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 digunakan adalah dasar listrik.

Hasil penelitian yang didapat berdasar analisis data dan kajian terhadap hasil pengembangan produk adalah sebagai berikut: (1) Model pengujian dengan terkomputerisasi 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 IRT 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) Model pengujian terkomputerisasi dengan algoritma IRT dapat mendeskripsikan kemampuan siswa berdasar butir-butir tes yang diterima dan nilai-nilai estimasi hasil analisis.

## **PRAKATA**

Segala puji bagi Allah s.w.t. bahwa hanya dengan karunia dan rahmat-Nya kami dengan sabar setelah melalui berbagai kesulitan dan kemudahan dapat menyelesaikan penelitian yang berjudul:

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

Penyelesaian dan penyusunan laporan penelitian ini tidak terlepas dari bantuan berbagai pihak. Oleh karena itu, terima kasih kami ucapkan kepada:

1. Direktur DP2M Ditjen Dikti Kemendikbud beserta staf Jakarta.
2. Rektor UNY Yogyakarta.
3. Dekan Fakultas Teknik UNY Yogyakarta.
4. Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat UNY Yogyakarta.
5. Ketua Jurusan Pendidikan Teknik Elektro FT UNY Yogyakarta.
6. Karyawan LPPM dan Diknik Elektro FT UNY Yogyakarta.
7. Para mahasiswa S1 Jurusan Diknik Elektro FT UNY Yogyakarta yang telah terlibat dalam penelitian ini.
8. Berbagai pihak yang tak dapat disebutkan satu per satu yang telah membantu kelancaran penelitian kami.

Kami menyadari tak ada karya manusia yang sempurna. Kepada para pembaca, saran yang konstruktif sangat diharapkan. Semoga Allah s.w.t memberi pahala dan berbagai kenikmatan yang banyak serta rizqi yang barokah dunia dan akhirat, kepada pihak-pihak yang telah membantu dalam penyelesaian dan penyempurnaan penulisan laporan penelitian ini, *aamiin ya robbal 'aalamiin*.

Yogyakarta, 31 Oktober 2014

Peneliti,

## DAFTAR ISI

JUDUL	i
HALAMAN PENGESAHAN	ii
ABSTRAK	iii
PRAKATA	iv
DAFTAR ISI	v
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Identifikasi Masalah	2
C. Rumusan Masalah	2
D. Tujuan dan Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	4
A. Pengujian Berbantuan Komputer	4
B. Assesmen Pembelajaran	10
C. Pembelajaran Berbasis Kasus	11
BAB III METODE PENELITIAN	11
A. Model Pengembangan	11
B. Uji Coba Produk	11
BAB IV ANALISIS DATA DAN PEMBAHASAN	14
A. Skenario Tes	14
B. Penetapan Kriteria	15
C. Estimasi Kemampuan	16
D. Fungsi Informasi Butir	19
E. Fungsi Informasi Tes	19
F. User Interface untuk Testee	20
G. Laporan	20
H. Hasil Evaluasi Program	21
BAB V SIMPULAN DAN SARAN	22
A. Simpulan tentang Produk	22
B. Keterbatasan Penelitian	23
C. Saran	25
DAFTAR PUSTAKA	27
LAMPIRAN	29

# BAB I

## PENDAHULUAN

### A. Latar Belakang Masalah

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 terus dilakukan penelitian, khususnya dalam pengembangan sistem pengujian dan penilaian hasil belajar berbantuan komputer.

Sekolah berkualitas pada umumnya dapat 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.

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.

### **B. Identifikasi Masalah**

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.

### **C. Rumusan Masalah**

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

- a. Bagaimanakah sistem inferensi dalam program assessmen pembelajaran dan penilaian pembelajaran berbantuan komputer memilih butir-butir tes yang sesuai dengan kemampuan siswa?
- b. Bagaimanakah akurasi kinerja program assessmen pembelajaran dan penilaian berbantuan komputer dalam mendeskripsikan kemampuan siswa dan penilaian pelaksanaan program pembelajaran?

#### **D. Tujuan dan Manfaat Penelitian**

##### **1. Tujuan Penelitian**

Tujuan penelitian ini adalah untuk menghasilkan:

- a. 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.
- b. Program pengujian hasil belajar dan penilaian pendidikan berbantuan komputer yang teruji dan memiliki akurasi yang baik untuk mendeskripsikan kemampuan siswa.

##### **2. Manfaat Penelitian**

Adapun manfaat penting yang diperoleh adalah perbaikan pada perangkat pembelajaran, strategi, metode, dan proses pembelajaran mata kuliah Sistem Kendali Fuzzy diharapkan bermanfaat untuk mengembangkan perangkat pembelajaran model robot yang representatif untuk mendukung proses pembelajaran yang dapat mengarah pada terbentuknya kualitas pembelajaran dan kualitas keilmuan mahasiswa.



## BAB 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 adaptif 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. 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 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 yang adaptif (*adaptive 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 adaptive dan non-pengujian adaptive, (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 CAT 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, penekanan penelitian ini difokuskan pada proses pemilihan butir tes dan estimasi kemampuan peserta tes yang dilakukan dengan algoritma kecerdasan buatan. 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 kecerdasan buatan (AI) yang 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 estimasi 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 menjawab 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 kurang akurat 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  $\theta$  menjawab benar butir ke- $i$
- $\theta$  : 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 respon 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. Assessmen 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) memberi-

kan 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 *assessment* 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?

## **BAB III**

### **METODE PENELITIAN**

#### **A. Model Pengembangan**

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.

#### **B. Uji Coba Produk**

##### **1. 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.

## 2. 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 penggalan 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 di gunakan adalah dasar listrik.

### **3. 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.

### **4. 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.

### **5. 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.

## BAB IV

### ANALISIS DATA DAN PEMBAHASAN

#### 1. 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:

- a. Pilih 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. Tampilkan 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. Pilih soal berikutnya berdasarkan pola skor. (Kemungkinan skor yang diperoleh adalah 00, 01, 10, atau 11)
  - jika skor = 00 → ambil soal dengan  $b$  awal dikurangi 0.2,
  - jika skor 01 atau 10 → ambil soal lain dalam rentang seperti pada 2 butir pertama
  - Jika skor 11 → ambil soal dengan  $b$  awal ditambah 0.2
- d. Tampilkan 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. Hitung theta (3 skor) dengan rumus MLE, dan hitung SEM
- g. Pilih soal berikutnya berdasarkan jawaban butir ke-3
  - jika salah (skor=0) → ambil soal dengan  $b$  terakhir dikurangi 0.20

- Jika benar (skor=1) → ambil soal dengan b terakhir ditambah 0.20
- h. Tampilkan soal yang terpilih, ambil respon dari testee, selanjutnya *confirm* dengan kunci untuk memperoleh skor.
- i. Hitung nilai theta dari 4 skor terakhir. Demikian juga SEM.
- j. \_\_\_\_\_
- k. LOOP (ulang):
  - \* ) Ulangi Langkah f-i dan akhiri tes jika: (1) jumlah butir soal sudah mencapai 15 atau (2) nilai SEM lebih kecil atau sama dengan 0.33
  - \*\* ) Catat seluruh transaksi, nilai theta terakhir, dan nilai SEM terakhir
  - \*\*\* ) Khusus untuk content based, ditambahkan syarat pemerataan sk secara proporsional

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 menefiti scanning jawaban siswa yang dilakukan oleh M Roisul dan untuk Mahasiswa Fajar meneliti tes esay.

## 2. Penetapan kriteria

Penetapan kriteria dilakukan sebelum pelaksanaan ujian. Ada dua pilihan dalam penetapan kriteria, yaitu: (1) menggunakan default atau (2) merubah nilai default. Jika tidak ada perubahan nilai, maka pertanyaan berikut dijawab No. Gambar di bawah ini memberikan informasi bahwa cara-cara penggunaan default dan merubah nilai default. Disamping itu, juga mengenai rentang kemampuan yang ingin ditetapkan apakah -3 sampai +3 atau yang lain.

QUESTION  
DO YOU WANT TO CHANGE  
ANY CRITERIA?

YES → [ ]

NO → [ ]

Component	DEFAULT *	ALTERNATIVE (example)	REMARK
Initial THETA	0.00 ± 0.50	-1.00 ± 0.50	Misal ada info ttg perkiraan kemampuan tesler 150 di bawah rata-2 → -1.00 ± 0.50
Content Based	No	Yes	Yes= SK/KD dijadikan pertimbangan dalam NEXT ITEM → <i>content balancing</i>
Scale Transformation	UMEAN=1 USCALE=0	UMEAN=55 USCALE=15	Skala asli dengan interval -3 s/d +3 diubah menjadi 10 s/d 100
Time Constraint	No	Yes	Yes= maximum duration setiap butir soal di-AKTIF-kan
Max Item	15	20	Stopping rules-1
Min SEM	0.30	0.40	Stopping rules-2
D Value	1.7	1.0	Konstanta untuk fungsi logistik. Dengan D=1.7, kurva akan mendekati OGIVE
Next_Value	0.2	0.1	Penambahan/pengurangan nilai b dalam pemilihan NEXT ITEM

### 3. Estimasi Kemampuan (Theta)

Metode estimasi theta yang digunakan dalam studi ini adalah Maximum Likelihood Estimation (MLE). Contoh gambaran 1:

# THETA ESTIMATION (2-PL)

EDRES-versio

## INFORMATION:

Item	b	a	response
1	-1.00	1.00	1
2	0.00	1.20	0
3	1.00	0.80	1

## TWO-PARAMETER MODEL

$$P = 1 / (1 + \exp(-A(T - B)))$$

## CALCULATION

Result:  
THETA = 0.323827

ITERATION - 1		THETA = 1.000							
Item	b	a	U	P	Q	a(u-P)	a*a(PQ)		
1	-1.00	1.00	1	0.8808	0.1192	0.1192	0.1050		
2	0.00	1.20	0	0.7685	0.2315	(0.9222)	0.2562		
3	1.00	0.80	1	0.5000	0.5000	0.4000	0.1600		
sum=						(0.4030)	0.5212		
DELTA THETA (s) =			(0.7733)						
THETA (s+1) =			0.2267						

ITERATION - 2		THETA = 0.2267							
Item	b	a	U	P	Q	a(u-P)	a*a(PQ)		
1	-1.00	1.00	1	0.7732	0.2268	0.2268	0.1753		
2	0.00	1.20	0	0.5676	0.4324	(0.6811)	0.3534		
3	1.00	0.80	1	0.3501	0.6499	0.5199	0.1456		
sum=						0.0656	0.6744		
DELTA THETA (s) =			0.0973						
THETA (s+1) =			0.3239						

ITERATION - 3		THETA = 0.3239							
Item	b	a	U	P	Q	a(u-P)	a*a(PQ)		
1	-1.00	1.00	1	0.7898	0.2102	0.2102	0.1660		
2	0.00	1.20	0	0.5960	0.4040	(0.7152)	0.3467		
3	1.00	0.80	1	0.3680	0.6320	0.5056	0.1488		
sum=						0.0006	0.6616		
DELTA THETA (s) =			0.0009						
THETA (s+1) =			0.3248						

SEM = 1.229446

Contoh gambaran 2:

### THETA ESTIMATION (2-PL)

My-Version

	Item-1	Item-2	Item-3
b	-1	0	1
a	1	1.2	0.8
c	0	0	0
-3	0.119203	0.026597	0.039166
-2.99	0.120257	0.026909	0.039468
-2.98	0.121319	0.027225	0.039772
-2.97	0.122389	0.027545	0.040079
-2.96	0.123467	0.027868	0.040388
-2.95	0.124553	0.028195	0.040699
-2.94	0.125648	0.028526	0.041013
-2.93	0.126751	0.02886	0.041328
-2.92	0.127862	0.029199	0.041646
-2.91	0.128981	0.029541	0.041967
-2.9	0.130108	0.029887	0.04229

SCORE  
PATTERN:  
101

RESULT:  
I(max)= 0.117429448  
THETA= 0.32

0.004544496  
0.004618565  
0.004693762  
0.004770102  
0.004847602  
0.004926276  
0.005006140  
0.005087210  
0.005169501  
0.005253030  
0.005337813

+B7 \* (1-C7) \* (D7)

1/(1+EXP(-B\$5\*(A7-B\$4)))

TWO-PARAMETER MODEL

$P = 1 / (1 + \text{EXP}(-A(T - B)))$

0.29	0.784147	0.586133	0.361698
0.3	0.785835	0.58904	0.363547
0.31	0.787513	0.591942	0.365401
0.32	0.789182	0.594837	0.367258
0.33	0.790841	0.597726	0.369119
0.34	0.79249	0.600608	0.370983
0.35	0.79413	0.603483	0.372852
0.36	0.79576	0.606351	0.374725
0.37	0.79738	0.609212	0.376601
0.38	0.798991	0.612065	0.378481
0.39	0.800592	0.61491	0.380365

0.117383094  
0.117406345  
0.117421792  
0.117429448  
0.117429330  
0.117421453  
0.117405836  
0.117382501  
0.117351470  
0.117312768  
0.117266419

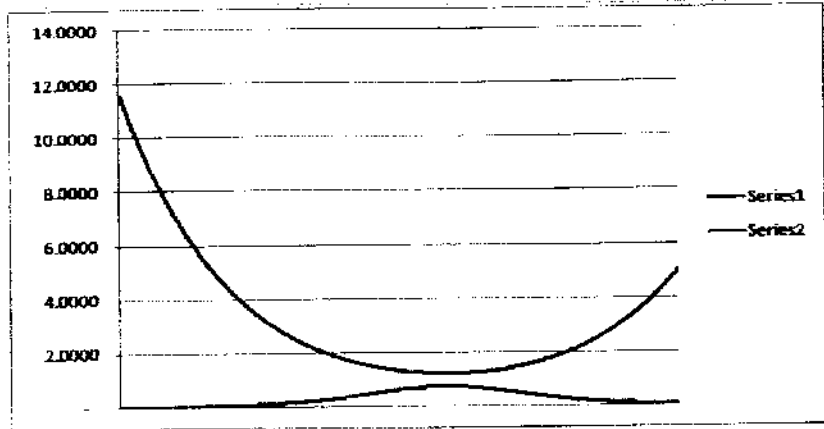
#### 4. Fungsi informasi butir (Item Information Function=IIF)

Item Information Function (IIF)

	ITEM-3	
	IIF	SEM
b =	0.5	0.5
a =	1	1
c =	0	0
-3	0.0075	11.5533
-2.99	0.0076	11.4561
-2.98	0.0077	11.3596
-2.97	0.0079	11.2640
-2.96	0.0080	11.1692
-2.95	0.0082	11.0752
-2.94	0.0083	10.9820
-2.93	0.0084	10.8895
-2.92	0.0086	10.7979
-2.91	0.0087	10.7071
-2.9	0.0089	10.6170
-2.89	0.0090	10.5277
-2.88	0.0092	10.4391
-2.87	0.0093	10.3513
-2.86	0.0095	10.2643
-2.85	0.0097	10.1780
-2.84	0.0098	10.0924
-2.83	0.0100	10.0076
-2.82	0.0102	9.9235

$$= \frac{(2.89 * [B55^2] * (1 - B56))}{((B56 + \text{EXP}(1.7 * B55 * (SA7 - B54))) * (1 + \text{EXP}(-1.7) * B55 * (SA7 - B54)))^2}$$

→ 1/(SQRT(IIF))

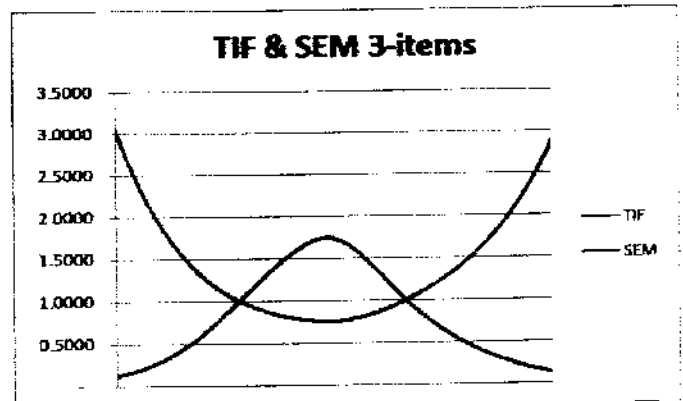


#### 5. Test Information Function & Standard Error Measurement (TIF & SEM)

- TIF diperoleh dengan menjumlahkan seluruh IIF dalam satu tes.
- $SEM = 1/(\sqrt{TIF})$
- Contoh:

Test Information Function (TIF) - 3 Items

	Item-1	Item-2	Item-3	TIF	SEM
b =	.1	0	1		
a =	1	1.2	0.8		
c =	0	0	0		
-3	0.0903	0.0091	0.0080	0.1074	3.051594
-2.99	0.0918	0.0093	0.0081	0.1091	3.027142
-2.98	0.0932	0.0095	0.0082	0.1109	3.002907
-2.97	0.0947	0.0097	0.0083	0.1127	2.978887
-2.96	0.0962	0.0099	0.0084	0.1145	2.95508
-2.95	0.0978	0.0101	0.0085	0.1164	2.931484
-2.94	0.0993	0.0103	0.0086	0.1182	2.908098
-2.93	0.1009	0.0105	0.0087	0.1202	2.88492
-2.92	0.1025	0.0107	0.0089	0.1221	2.861948
-2.91	0.1041	0.0109	0.0090	0.1241	2.839181
-2.9	0.1058	0.0112	0.0091	0.1261	2.816617
-2.89	0.1075	0.0114	0.0092	0.1281	2.794255
-2.88	0.1092	0.0116	0.0094	0.1301	2.772092
-2.87	0.1109	0.0119	0.0095	0.1322	2.750129





## 6. User interface untuk testee

Setelah proses login, peserta tes (testee) menjawab pertanyaan atau soal dengan tampilan sebagai berikut:

Demikian seterusnya, dan tes akan berakhir ketika kriteria stopping rules terpenuhi.

### CAT #1

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

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

Soal no 2 dari 15

## 7. Laporan

Setiap akhir pelaksanaan tes, untuk setiap peserta diperoleh laporan sebagai berikut:

# CBT & CAT Ujian Pendidikan Kesetaraan

Report

## CAT & CBT - Dashboard

Terima kasih telah menyelesaikan ujian dengan CAT System.

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

Informasi tersebut dapat digunakan untuk menelusuri dan mengevaluasi apakah algoritma valid secara psikometris dan matematis.

## 8. 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:

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

- c. 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.

## **BAB V**

### **SIMPULAN DAN SARAN**

#### **1. Simpulan tentang Produk**

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

- a. 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.
- b. 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.
- c. Ketepatan sistem inferensi dalam memilih butir-butir tes yang sesuai dengan respons jawaban siswa diwujudkan pada nilai.
- d. Kemampuan model pengujian dengan algoritma penalaran dalam mendeskripsikan kemampuan siswa berdasar butir-butir tes yang diterima dan nilai-nilai estimasi hasil analisis,

e. 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.

## 2. Keterbatasan Penelitian

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

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

Program pengujian 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 ( $\theta$ ) dan  $p(\theta)$  ke

p untuk menggambarkan kemampuan yang sebenarnya, keakuratannya masih perlu ditingkatkan dalam penelitian ini.

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

Jumlah basis pengetahuan yang digunakan untuk membangun sistem inferensi dalam penelitian ini adalah empat *rule*. Jumlah basis pengetahuan berpengaruh terhadap ketepatan hasil inferensi. Belum ada aturan untuk jumlah basis pengetahuan yang baik. Menurut peneliti, jumlah basis pengetahuan sebaiknya yang memenuhi kecukupan untuk membangun sistem inferensi. Dalam penelitian ini, basis pengetahuan yang dibangun dibatasi pada empat *rule*, karena telah memenuhi kemungkinan-kemungkinan kombinasi klasifikasi tingkat kesulitan butir yang dibentuk.

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

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

### **3. Saran Pemanfaatan dan Pengembangan Lebih Lanjut**

Program pengujian 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.

Klasifikasi tingkat kesulitan butir tes dalam penelitian ini dibagi dalam tiga kelompok, yaitu: tingkat kesulitan tinggi, tingkat kesulitan sedang, dan tingkat kesulitan mudah. Oleh sebab itu, jumlah butir-butir tes yang memenuhi syarat untuk dijadikan basis data *bank* soal jumlahnya harus memenuhi ketiga kelompok tersebut. Jika banyak soal (*N*) yang akan diujikan adalah 20 butir tes, maka dalam basis data *bank* soal minimal adalah butir sulit dan butir mudah masing-masing 20 butir tes dan 40 butir tes sedang, sehingga total 80 butir tes. Dengan rincian 20 butir tes tingkat kesulitan tinggi, 40 butir tes tingkat kesulitan sedang, dan 20 butir tes tingkat kesulitan rendah. Syarat butir tes yang bagus minimal memiliki parameter tingkat kesulitan butir dengan kisaran nilai -3 sampai 3 untuk tes modern dan daya beda

butir 0,0 hingga 0,4 serta tebakan 0,00 sampai 0,2 (jika ada 45 opsi jawaban per butir).

Program 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* (4<sup>th</sup> ed.). New York: Longman Inc.
- Brennan, R.L. (2006). *Educational measurement* (4<sup>th</sup> 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). *Pengetolakan 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* (6<sup>th</sup> 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 (3<sup>d</sup> 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: McGraw 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.



UNIVERSITAS NEGERI YOGYAKARTA

# FAKULTAS TEKNIK

Alamat : Kampus Karangmalang, Yogyakarta 55281

## LAPORAN PELAKSANAAN SEMINAR HASIL PENELITIAN

1. Nama Peneliti : Dr. Hariyanto, dkk
2. Jurusan : Pendidikan Teknik Elektro
3. Fakultas : Fakultas Teknik
4. Status Penelitian : a. Dosen Muda  
b. Hibah  
c. Mandiri  
d. Kelompok  
e. Kolaborasi
5. Judul Penelitian : Pengembangan sistem pengujian terkomputerisasi untuk kompetensi bidang kejuruan Sewa SMK di DIY
6. Pelaksanaan : Tanggal : 24 Oktober 2014  
Jam : 09.00 WIB – selesai.
7. Tempat : RE1
8. Dipimpin oleh : Ketua : K. Ima Ismara MPd, M Kes  
Sekretaris : Drs. Nur Kholis, MPd.
9. Peserta : a. Konsultan : 1 orang  
b. Nara Sumber : 1 orang  
c. BPP : 1 orang  
d. Peserta lain : 33 orang  
Jumlah : 34 orang
10. Hasil Seminar : Setelah mempertimbangkan penyajian, penjelasan, argumentasi serta sistematika dan tata tulis, seminar berkesimpulan :  
Laporan Instrumen Penelitian tersebut di atas:  
a. Diterima tanpa revisi/pembenahan  
b. Diterima dengan revisi/pembenahan  
c. Dibenahi, untuk diseminarkan ulang
11. Catatan: Teori IKT di buat rinci

Sekretaris

Drs. Nur Kholis, M.Pd  
NIP. 19681026 199403 1 003

Mengetahui  
Badan Pertimbangan Penelitian

Dr. Hariyanto, M.Pd, M.T.  
NIP. 19620310 198601 1 001

Ketua Sidang,

K. Ima Ismara, M.Pd, M.Kes  
NIP. 19610911 199001 1 001



UNIVERSITAS NEGERI YOGYAKARTA

# FAKULTAS TEKNIK

Alamat : Kampus Karangmalang, Yogyakarta 55281

## DAFTAR HADIR SEMINAR HASIL PENELITIAN

Hari, tanggal : Jumat, 24 Oktober 2014

Nama Peneliti :

Fak./Jurusan : Teknik/ Pendidikan Teknik Elektro

Judul Penelitian :

Kelas/Kelompok :

USULAN

INSTRUMEN

LAPORAN HASIL

NO	NAMA	GELAR	TANDA TANGAN
1	Djeman Mardapi	Prof. H., Ph.D	1
2	Sunyoto	M.Pd	2
3	Soeharto	M.SOE., Ed.D.	3
4	Setya Utama	M.Pd	4
5	K. Ima Ismara	M.Pd., M.Kes(Ind)	5
6	Edy Supriyadi	Dr., M.Pd.	6
7	Sunomo	Drs., MT.	7
8	Samsul Hadi	Dr., M.Pd., MT	8
9	Giri Wiyono	Dr., MT	9
10	Mutaqin	M.Pd, MT.	10
11	Sardjiman Djojopernoto	M.Pd	11
12	Mulian Jamin Alwi	Drs.	12
13	Ahmad Sujadi	M.Pd	13
14	Basrowi	M.Pd	14
15	Sunaryo Soenarto	Dr., M.Pd.	15
16	Istanto Wahyu Djatmiko	Dr., M.Pd.	16
17	Drs. Pangat Hendro Sutomo	Drs.	17
18	Haryanto	Dr., M.Pd., MT	18
19	Nyoman Astra	Drs.	19
20	Djoko Laras Budyo Taruno	Dr., M.Pd.	20
21	Herlambang Sigit Pramono	ST., M.Cs.	21
22	Rustam Asnawi	MT., Ph.D	22
23	Totok Heru Tri Maryadi	M.Pd.	23
24	Sukir	Drs., MT	24



UNIVERSITAS NEGERI YOGYAKARTA  
**FAKULTAS TEKNIK**

Alamat : Kampus Karangmalang, Yogyakarta 55281

NO	NAMA	GELAR	TANDA TANGAN
25	Zamtinah	M.Pd	25
26	Nur Kholis	Drs., M.Pd	26
27	Muhamad Ali	M.T.	27
28	Nurhening Yuniarti	M.T.	28
29	Hartoyo	M.Pd, MT.	29
30	Deny Budi Hertanto	M.Kom	30
31	Moh. Khairudin	M.T., Ph.D	31
32	Didik Hariyanto	M.T.	32
33	Sigit Yatmono	M.T.	33
34	Toto Sukisno	S.Pd, M.Pd.	34
35	Yuwono Indro Hatmojo	S.Pd., M.Eng.	35
36	Ilmawan Mustaqim	S.PdT., MT.	36
37	Ariadie Chandra Nugraha	ST., MT.	37
38	Faranita Surwi	ST., MT.	38
39	Andik Asmara	S.Pd.	39
40	Alex Sandria Jaya W	S.Pd	40
41	Eko Prianto	S.Pd, M.Eng	41
42	Sebastian A. Nugroho	S.T., M.T.	42.

Yogyakarta, 24 Oktober 2014  
Pimpinan Sidang

K. Ima Ismara, M.Pd, M.Kes(Ind)  
NIP. 19610911 199001 1 001