

## **PENGEMBANGAN SISTEM PENILAIAN HASIL BELAJAR MATA PELAJARAN MENGANALISIS RANGKAIAN LISTRIK BERBASIS *COMPUTERIZED ADAPTIVE TESTING***

Kamaruddin  
SMK Negeri 2 Sumbawa Besar  
kamar.kamaruddin@gmail.com

Haryanto  
Universitas Negeri Yogyakarta  
haryanto.ftuny@gmail.com

### **Abstrak**

*Penelitian dan pengembangan ini bertujuan untuk: (1) menghasilkan instrumen tes terstandar untuk mata pelajaran Menganalisis Rangkaian Listrik yang tersimpan dalam bank soal elektronik; (2) menghasilkan perangkat lunak sistem penilaian berbasis CAT (SPBCAT) yang mampu melaksanakan tes secara adaptif; dan (3) SPBCAT yang mampu melakukan proses pengujian dengan unjuk kerja yang layak. Penelitian dan pengembangan ini dilaksanakan dalam empat tahap, yaitu: (1) definisi; (2) desain; (3) pengembangan; dan (4) pengujian. Kesimpulan dari hasil penelitian ini adalah: (1) instrumen tes terstandar yang berhasil dikembangkan berbentuk 105 butir pilihan ganda, tersusun dalam empat KD, dengan reliabilitas empiris, masing-masing 0,9086, 0,9067, 0,9087, dan 0,9086, dan tersimpan dalam bank soal elektronik; (2) SPBCAT mampu melaksanakan tes secara adaptif, dan menampilkan hasil tes dengan tepat; (3) perangkat lunak SPBCAT telah mampu menampilkan unjuk kerjanya secara layak dalam menguji 28 peserta tes secara serempak, dan mendapat penilaian sangat baik oleh pengguna.*

**Kata kunci:** CAT, Menganalisis, Rangkaian, Listrik

## **DEVELOPMENT OF LEARNING OUTCOMES ASSESSMENT SYSTEMS FOR ELECTRIC CIRCUITS ANALYZING SUBJECTS BASED ON *COMPUTERIZED ADAPTIVE TESTING***

### **Abstract**

*This research and development aim to: (1) produce standardized test instruments for Electric Circuits Analyzing subjects which is stored in an electronic item bank; (2) produce a software of learning outcomes assessment system based on CAT (SPBCAT) which able to performs test adaptively; and (3) SPBCAT which is able to perform adaptive testing process with decent performance. This research and development were conducted in four stages, i.e.: (1) the definition; (2) the design; (3) the development; and (4) the testing. The conclusions of this research are: (1) the standardized test instrument which has been successfully developed in the forms of 105 items of multiple choice, arranged in four KD, with empirical reliability is 0.9086, 0.9067, 0.9087, and 0.9086, and stored in the electronic item bank; (2) SPBCAT has been able to perform test adaptively, displays test result precisely; (3) SPBCAT has been able to show its performance properly by present test items simultaneously to 28 students and got a very good appraisal from its user*

**Keywords:** CAT, Analyzing, Circuit, Electric.

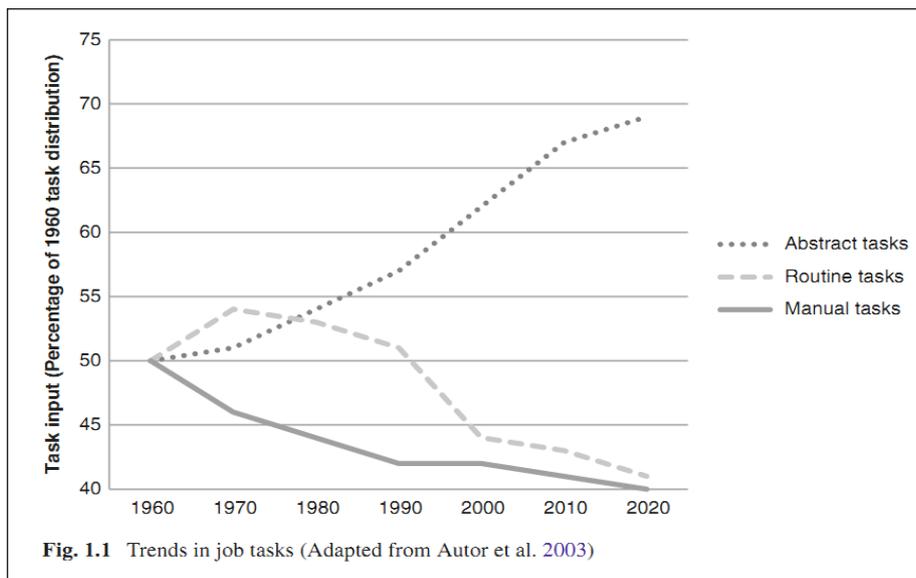
**PENDAHULUAN**

Undang-undang nomor 20 tahun 2003, dalam pasal 35 dan 36 menyebutkan bahwa pengembangan kurikulum pada semua jenjang pendidikan dilakukan dengan mengacu pada Standar Nasional Pendidikan dengan prinsip diversifikasi sesuai dengan satuan pendidikan, potensi daerah, dan peserta didik. Kompetensi lulusan sebagai salah satu dari delapan acuan pengembangan kurikulum yang terdapat dalam Standar Nasional Pendidikan merupakan kualifikasi kemampuan lulusan yang mencakup pengetahuan, keterampilan, dan sikap.

Undang-Undang nomor 20 tahun 2003 pasal 18 dan penjelasan pasal 15 menyatakan bahwa Sekolah Menengah Kejuruan (SMK) sebagai salah-satu sub-sistem dari Sistem Pendidikan Nasional, merupakan bentuk pendidikan pada tingkat menengah yang memiliki tujuan utama, menyiapkan tamatannya untuk bekerja pada bidang pekerjaan tertentu. Meski-

pun demikian, bukan berarti bahwa SMK hanya menekankan tamatannya untuk terampil bekerja, tetapi mereka juga diharuskan untuk menjadi manusia yang berilmu pengetahuan dan beradab.

Sudira (2012, p.21) menyatakan bahwa pengembangan kompetensi siswa SMK hendaknya tidak menekankan pada penguasaan praktik saja atau pada penguasaan teori saja. Hendaknya praktik dan teori dikuasakan bersama-sama secara proporsional. Pergeseran secara substansial dalam struktur ketenagakerjaan juga perlu menjadi perhatian, khususnya bagi pendidik dan penyelenggara pendidikan kejuruan. Sebuah studi yang dilakukan oleh Autor, Levy, & Murnane (2003) menunjukkan bahwa dari tahun 1960 sampai hari ini, telah terjadi peningkatan dalam pekerjaan abstrak dan juga penurunan baik dalam pekerjaan rutin maupun manual. Tren pergeseran struktur ketenagakerjaan tersebut ditunjukkan oleh gambar 1.



Gambar 1. Tren Pergeseran Substansial di dalam Struktur Ketenagakerjaan (Griffin, Care, & McGaw, 2011, p.3)

Berdasarkan hasil wawancara pada kegiatan pra-survei yang dilakukan di beberapa SMK yang memiliki kompetensi keahlian Teknik Instalasi Tenaga Listrik (TITL) dapat dirangkum beberapa permasalahan terkait dengan penguasaan kompetensi siswa baik teori kejuruan maupun praktik kejuruan. Beberapa masalah tersebut antara lain: (1) siswa kurang

fokus untuk menguasai materi kejuruan; (2) jika guru memberikan permasalahan yang sedikit berbeda dari yang dicontohkan maka sebagian besar siswa akan mengalami kebingungan dalam menyelesaikan permasalahan tersebut; dan (3) siswa sulit untuk mengenali simbol-simbol yang digunakan dalam kelistrikan.

Penulis melanjutkan pengamatan pada silabus untuk mata pelajaran Menganalisis Rangkaian Listrik dan membandingkan dengan tes yang dibuat oleh guru. Berdasarkan pengamatan terhadap dua dokumen tersebut, ditemukan adanya materi lain dalam lembar tes. Sementara materi-materi tersebut tidak terdapat dalam silabus mata pelajaran Menganalisis Rangkaian Listrik ini.

Instrumen tes yang baik dan terstandar untuk kegiatan penilaian, merupakan hal yang penting dan bersifat urgen dalam penyelenggaraan pendidikan khususnya pendidikan kejuruan, karena hasil penilaian tersebut memberikan informasi sebagai acuan pengambilan keputusan. Berry dalam Berry & Adamson (2011, p.5) menyatakan bahwa penilaian memiliki dua fungsi yaitu: (1) untuk membuat pertimbangan dari kinerja individu atau keefektifan sistem; dan (2) untuk meningkatkan pembelajaran.

Penilaian yang baik dan akurat membutuhkan instrumen tes hasil belajar yang memenuhi standar. Penyusunan instrumen tes terstandar dapat dilakukan melalui beberapa tahap. Menurut Mardapi (2008, p.88) terdapat sembilan tahap dalam mengembangkan instrumen tes prestasi yaitu: (1) menyusun spesifikasi tes; (2) menulis butir-butir tes; (3) menelaah butir-butir tes; (4) melakukan uji coba instrumen tes; (5) menganalisis butir-butir tes; (6) memperbaiki butir-butir tes; (7) merakit butir-butir tes; (8) melaksanakan pengujian; dan (9) menafsirkan hasil pengujian. Instrumen tes yang telah memenuhi standar, tentunya akan menjadi langkah awal bagi sekolah dalam mengembangkan sistem penilaian berbantuan komputer jika sekolah tersebut memiliki fasilitas lab komputer dan minimal didukung oleh sistem jaringan lokal (*Local Area Network*).

Jhonson & Jhonson (2002, p.9) menyebutkan bahwa setidaknya terdapat empat pihak yang berkepentingan dengan hasil penilaian ini, yaitu: (1) siswa dan orang tuanya; (2) guru; (3) pihak pengelola sekolah; dan (4) pembuat kebijakan. Informasi yang diperoleh dari proses penilaian dapat digunakan oleh siswa sebagai acuan untuk melakukan perbaikan atau untuk melanjutkan ke materi pada kompetensi yang lain. Informasi yang diperoleh dari kegiatan penilaian ini, akan memiliki nilai yang tinggi

jika disampaikan kepada pengguna informasi dengan segera setelah kegiatan tes berakhir.

Ketersediaan informasi yang cepat dan akurat dari hasil penilaian sangat tergantung pada sarana Teknologi Informasi dan Komunikasi (TIK). Komputer sebagai bagian dari sarana TIK berperan penting dalam mengolah dan menyajikan informasi tersebut. TIK menawarkan begitu banyak kemungkinan yang luar biasa untuk pengajaran dan pembelajaran, termasuk penilaian berbasis teknologi. Csapó, et al. (2011, p.144) menyatakan bahwa penilaian berbasis teknologi dalam bidang pendidikan memiliki beberapa kelebihan antar lain: (1) mampu melakukan proses penilaian yang lebih efisien; (2) menawarkan metode baru yang dapat direalisasikan; dan (3) memungkinkan analisis canggih dari data yang tersedia, sehingga dapat memberikan umpan balik yang cepat bagi peserta didik dan stakeholder. Lebih lanjut Csapó, et al. (2011, pp.144-145) menyatakan bahwa masih terdapat banyak kendala dalam mengimplementasikan penilaian berbasis teknologi untuk skala besar.

Ketersediaan sarana TIK merupakan salah satu syarat untuk dapat melaksanakan penilaian berbantuan komputer. Berdasarkan hasil pengamatan yang dilakukan pada kegiatan pra-survei diperoleh informasi bahwa sebagian besar SMK di DIY sudah memiliki laboratorium komputer. Keberadaan infrastruktur TIK ini memberikan peluang yang sangat luas bagi SMK untuk dapat menerapkan sistem penilaian berbantuan komputer.

Berdasarkan penjelasan di atas, dapat dipahami bahwa sebagai salah satu kompetensi dari Dasar Kompetensi Kejuruan, mata pelajaran Menganalisis Rangkaian Listrik merupakan materi yang sangat penting, mengingat sebagai mata pelajaran dalam kelompok Dasar Kompetensi Kejuruan, penguasaan siswa terhadap materi ini sangat mempengaruhi kemampuan siswa menguasai kompetensi berikutnya. Untuk mengetahui pencapaian siswa dalam menguasai kompetensi dasar ini dapat dilakukan dengan beberapa cara antara lain melalui proses penilaian. Penilaian sebagai cara yang sangat efektif untuk mengukur pencapaian kompetensi siswa ini memerlukan suatu instrumen yang valid dan reliabel sehingga meng-

hasilkan informasi yang akurat. Informasi yang akurat akan memiliki nilai yang tinggi, jika disampaikan kepada tujuannya dengan tepat dan segera. Sehubungan dengan hal tersebut, dibutuhkan dukungan TIK yang mudah diakses, dan mampu mengolah serta menyajikan informasi kepada semua siswa, guru, dan stakeholder pendidikan.

Pemanfaatan TIK dalam bidang pendidikan khususnya untuk kegiatan penilaian hasil belajar memiliki banyak kelebihan. Namun demikian, TIK untuk kegiatan penilaian hasil belajar ini belum dimanfaatkan secara optimal. Hal ini dapat dipahami, karena di sebagian besar SMK tidak memiliki perangkat lunak yang mampu menyimpan butir-butir tes dalam sistem data-base bank soal, menyajikan butir-butir tes kepada siswa, dan memberikan informasi kepada siswa tentang hasil tes yang telah mereka ikuti dengan segera setelah mereka menyelesaikan tes tersebut. Dengan memperhatikan paparan di atas, maka Penulis berpandangan bahwa pengembangan instrumen tes terstandar untuk mata pelajaran Menganalisis Rangkaian Listrik untuk memudahkan siswa dalam mengukur dan menilai pemahaman dan penguasaan mereka terhadap kompetensi ini dan bagi guru dalam mengukur pencapaian tujuan pengajarannya merupakan hal yang bersifat penting dan mendesak.

Penelitian dan Pengembangan ini bertujuan untuk: (1) menghasilkan instrumen tes terstandar berbantuan komputer untuk mata pelajaran Menganalisis Rangkaian; dan (2) menghasilkan perangkat lunak sistem penilaian hasil belajar mata pelajaran Menganalisis Rangkaian Listrik berbasis CAT (SPBCAT) dengan unjuk kerja yang layak.

### **Pendidikan Kejuruan**

Berdasarkan penjelasan Undang-Undang nomor 20 tahun 2003 pasal 15, pendidikan kejuruan merupakan pendidikan menengah yang mempersiapkan peserta didik terutama untuk bekerja dalam bidang tertentu. Falk & Surata (2011, p.54) mendefinisikan pendidikan kejuruan sebagai suatu kegiatan belajar formal yang menekankan kualitas untuk mendapatkan sebuah keahlian. Pendidikan kejuruan memerlukan sebuah komponen pengalaman yang

berisikan keterampilan kerja. Pada akhir masa pendidikannya, peserta didik yang memenuhi standar kelulusan, akan mendapat pengakuan “kompeten” yang diperkuat dengan sertifikat kompetensi.

Berdasarkan Panduan BSNP tentang Penyusunan Kurikulum diperoleh penjelasan yang lebih rinci tentang tujuan pendidikan menengah kejuruan yaitu untuk meningkatkan kecerdasan, pengetahuan, kepribadian, akhlak mulia, serta keterampilan untuk hidup mandiri dan mengikuti pendidikan lebih lanjut sesuai dengan kejuruan. Empat hal pokok yang dapat disarikan penjelasan tersebut bahwa pendidikan kejuruan bertujuan: (1) mempersiapkan lulusan untuk dapat segera mendapat pekerjaan; (2) memberikan keterampilan kepada peserta didik; (3) memberikan bekal pengetahuan akademik; (4) memungkinkan orang untuk menangani tuntutan pekerjaan mereka dengan kompeten.

Djojonegoro (1998, p.36) menyebutkan bahwa pendidikan kejuruan bertujuan mempersiapkan siswa sebagai calon tenaga kerja dan mengembangkan ekstensi siswa, untuk kepentingan siswa, masyarakat, bangsa dan negara. Selanjutnya menurut penjelasan Undang-Undang nomor 20 tahun 2003 pasal 15 menyatakan bahwa pendidikan kejuruan merupakan pendidikan menengah yang mempersiapkan siswa terutama untuk bekerja dalam bidang tertentu.

### **Kompetensi Siswa SMK (Bidang TITL)**

Menurut penjelasan Undang-Undang nomor 20 tahun 2003 pasal 35 ayat 1 menjelaskan bahwa kompetensi lulusan merupakan kualifikasi kemampuan lulusan yang mencakup sikap, pengetahuan, dan keterampilan sesuai dengan standar nasional yang telah disepakati. Dalam konteks penelitian ini, kompetensi siswa SMK bidang listrik yang dimaksud adalah kemampuan unjuk kerja keahlian, yang dibentuk melalui pengetahuan, keterampilan dan sikap tentang pekerjaan bidang Teknik Instalasi Tenaga Listrik (TITL).

Berdasarkan Permendiknas nomor 28 tahun 2009 tentang Standar Kompetensi Kejuruan SMK/MAK pada Kompetensi Keahlian TITL, terdapat 19 (sembilan belas)

standar kompetensi, yang dikelompokkan menjadi Dasar Kompetensi Kejuruan dan Kompetensi Kejuruan. Dasar Kompetensi Kejuruan (DKK) terdiri atas lima Standar Kompetensi yaitu: (1) menganalisis rangkaian listrik; (2) menggunakan hasil pengukuran; (3) menafsirkan gambar teknik listrik; (4) melakukan pekerjaan mekanik dasar; (5) menerapkan keselamatan dan kesehatan kerja (K3).

Terdapat beberapa alasan pemilihan standar kompetensi ini sebagai fokus pengembangan antara lain: (1) menganalisis rangkaian listrik merupakan salah satu dasar kompetensi kejuruan yang merupakan pondasi pengetahuan siswa untuk melangkah ke kompetensi berikutnya; (2) materi dalam tiap standar kompetensi dalam kurikulum TITL tidak dapat dipisahkan dari materi yang terdapat dalam standar kompetensi menganalisis rangkaian listrik ini; (3) Program Studi Keahlian Teknik Ketenagalistrik terbagi menjadi lima kompetensi keahlian, dan standar kompetensi menganalisis rangkaian listrik merupakan kompetensi yang harus dikuasai siswa dalam kelima kompetensi keahlian tersebut; dan (4) penulis adalah tenaga pengajar SMK pada Kompetensi Keahlian Teknik Instalasi Tenaga Listrik dan telah mengajar untuk standar kompetensi menganalisis rangkaian listrik ini sejak awal tahun 2005, sehingga diharapkan penulis dapat mengembangkan penilaian hasil belajar pada materi ini dengan lebih baik.

### Teori Respon Butir

Hambleton, Swaminathan, & Rogers (1991, p.5) menyebutkan beberapa kelebihan dari IRT yaitu: (1) karakteristik butir yang tidak tergantung pada kelompok (*group-independent*); (2) skor menggambarkan kemampuan peserta tes yang tidak tergantung pada tes (*test-independent*); (3) model yang diekspresikan pada level butir bukan pada level tes; (4) model yang tidak harus memerlukan tes paralel untuk menilai reliabilitasnya; dan (5) model yang memberikan ukuran presisi untuk masing-masing skor kemampuan.

IRT memiliki tiga model yang paling populer dan cocok untuk data jawaban butir dikotomi (*dichotomous*) untuk model 1PL,

2PL, dan 3PL secara berturut ditunjukkan pada persamaan 1, 2, dan 3 sebagai berikut:

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

$$P_i(\theta) = \frac{e^{D a_i(\theta)}}{1 + e^{D a_i(\theta - b_i)}} \quad i = 1, 2, \dots, n \quad (2)$$

$$P_i(\theta) = c_i + (1 - c_i) \frac{1}{1 + e^{-D a_i(\theta - b_i)}} = \quad (3)$$

Di mana:

- $P_i(\theta)$  : probabilitas menjawab benar pada soal ke- $i$  oleh peserta dengan kemampuan  $\theta$
- $D$  : 1,7
- $e$  : 2,718
- $a_i$  : parameter daya beda soal ke- $i$  (*slope*)
- $b_i$  : parameter tingkat kesulitan soal ke- $i$  (*threshold*)
- $c_i$  : parameter tebakan soal ke- $i$

Hulin, Drasgow, & Parsons, (1983, pp.34-35) menyatakan bahwa butir-butir dengan parameter- $a$  bernilai kecil mampu menyediakan beberapa informasi tentang kemampuan dalam jangkauan  $\theta$  yang lebar. Sebaliknya butir-butir dengan parameter- $a$  bernilai besar menyediakan informasi yang kuat tentang nilai  $\theta$  pada daerah di dekat nilai parameter- $b$ , tetapi hanya menyediakan sedikit informasi tentang  $\theta$  untuk daerah yang jauh dari nilai parameter- $b$ .

Parameter kedua yang digunakan dalam IRT adalah parameter- $b$ . Parameter- $b$  adalah titik pada skala kemampuan di mana seorang peserta tes memiliki peluang untuk menjawab benar butir tersebut sebesar 0,5 untuk model 1PL dan 2PL, dan sebesar  $(1 + c)/2$  untuk model 3PL (Hambleton & Jones, 1993, p.41, DeMars, 2010, p.5, dan Baker, 2001, p.30).

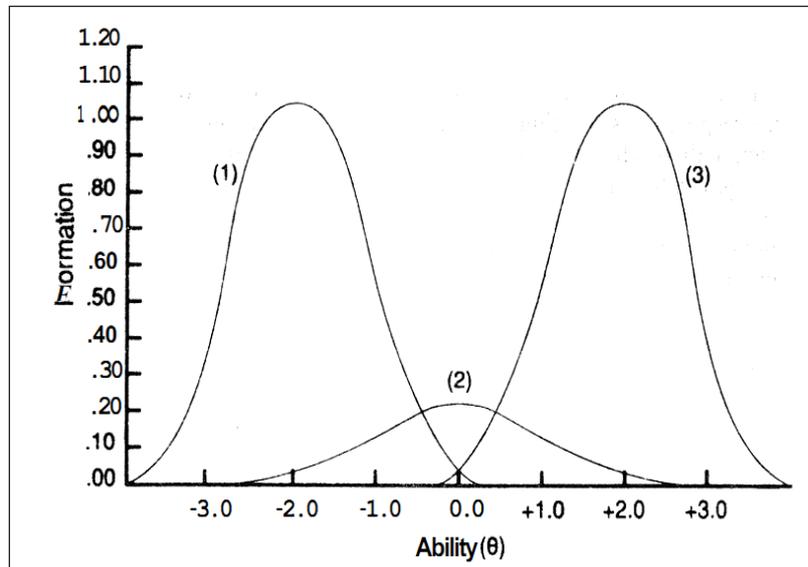
Parameter ketiga dalam IRT yang digunakan pada model 3PL parameter- $c$  yang disebut sebagai parameter tingkat kesempatan semu (*pseudo-level-chance*). Hulin, Drasgow, & Parsons (1983, p.31) menyatakan bahwa parameter- $c$  diinterpretasikan sebagai probabilitas jawaban benar dari peserta tes berkemampuan sangat rendah.

Besaran penting berikutnya dalam IRT adalah fungsi informasi butir (FIB) yang din-

yatakan dengan  $I_i(\theta)$  merupakan fungsi yang menyediakan informasi oleh butir ke- $i$  pada  $\theta$ . Wendler & Walker (2006, p.456) menyatakan bahwa setiap butir memiliki informasi, yang menyatakan seberapa mampu butir tersebut dapat membedakan antara peserta tes yang

berkemampuan sama pada tingkat kemampuan yang berbeda.

Kurva karakteristik butir yang dihasilkan oleh fungsi informasi butir dari tiga buah butir yang berbeda digambarkan sebagai berikut:



Gambar 7. Kurva Karakteristik Butir dari Tiga Butir yang Berbeda (Sumber: Bachman, 1990, p.208)

Fungsi informasi butir untuk model 3PL dinyatakan dengan persamaan berikut (Hambleton, Swaminathan, & Rogers, 1991, p.91, dan Jones, Smith, & Talley, 2006, p.500):

$$I_i(\theta) = \frac{2.89a_i^2(1 - c_i)}{(c_i + e^{1.7a_i(\theta - b_i)})(1 + e^{-1.7a_i(\theta - b_i)})} \quad (4)$$

Berdasarkan persamaan di atas Hambleton, Swaminathan, & Rogers (1991, p.91), Bachman (1990, p.207) dan DeMars (2010, p.80) menjelaskan bahwa: (1) nilai informasi butir akan mencapai nilai tertinggi jika nilai tingkat kesulitan butir semakin mendekati  $\theta$ ; (2) nilai informasi butir akan lebih memuncak jika nilai parameter- $a$  tinggi dan menjadi lebih datar jika nilai parameter- $a$  rendah; (3) nilai informasi butir akan naik jika parameter- $c$  mendekati nol.

### Fungsi Informasi Tes (FIT)

Hambleton, Robin, & Dehui Xing (2000, p.572), Jones, Smith, & Talley (2006, p.456), dan Froelich (2009, p.297) menyatakan bahwa fungsi Informasi tes merupakan penjumlahan

dari fungsi informasi butir dari masing-masing yang menyusun tes tersebut. Fungsi informasi dari sebuah tes dinyatakan oleh Hambleton, Swaminathan, & Rogers (1991, p.94), DeMars (2010, p.82), dan Froelich (2009, p.297) dengan persamaan berikut:

$$I(\theta) = \sum_{i=1}^n I_i(\theta) \quad (5)$$

### Standard Error of Measurement (SEM)

Simms & Watson (2007, p.252) menyatakan bahwa dalam IRT, SEM sama dengan invers dari akar kuadrat informasi tes pada setiap titik disepanjang skala kemampuan. Suatu pengukuran yang baik ditandai dengan kecilnya kesalahan pengukuran, yaitu nilai SEM yang kecil. Dengan kata lain, nilai SEM dapat diperkecil dengan menambahkan jumlah butir dalam tes. Persamaan SEM dinyatakan oleh Jones, Smith, & Talley (2006, p.500), dan Embretson & Reise (2000, p.185) sebagai berikut:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (6)$$

### **Pengujian Berbasis Komputer (CBT)**

Kemajuan teknologi dalam empat puluh tahun terakhir, yang ditandai dengan munculnya perangkat komputer telah merevolusi hampir setiap aspek dalam pengujian psikologis dan penilaian, (Drasgow & Mattern, 2006, p.59). Hal ini dapat dipahami karena perkembangan teknologi komputer pada generasi baru memungkinkan penyimpanan butir tes dalam sebuah *data base* bank soal dalam jumlah yang besar. Komputer mampu melakukan perhitungan yang rumit dengan cepat sehingga mampu menampilkan butir tes sesuai dengan jawaban peserta tes, bahkan mampu memberikan saran cerdas kepada pengguna berdasarkan basis pengetahuan yang ada pada program yang dijalankan dalam komputer tersebut.

Model Pengujian Berbasis Komputer (CBT) telah dikembangkan dan diimplementasikan dengan berbagai bentuk. Folk & Smith (2002, p.41) menyebutkan tiga model CBT, yaitu (1) *Computerized Adaptive Testing* (CAT); (2) *Linear on-the-Fly Testing* (LOFT); dan (3) *Computerized Mastery Testing* (CMT); Parshall et.al. (2002, pp.10-12) membagi metode CBT menjadi 4 (empat) macam program pengujian yaitu: (1) *Computerized Fixed Tests* (CFT); (2) *Automated Test Assembly* (ATA); (3) *Computerized Adaptive Testing* (CAT); (4) *Computerized Classification Tests*.

### **Computerized Adaptive Testing**

Van der Linden (2005, p.211) menyatakan bahwa pengujian adaptif sangat efisien. Simulasi komputer yang luas telah menunjukkan bahwa untuk menghasilkan perkiraan kemampuan yang sama akuratnya, tes adaptif hanya membutuhkan 40-50% dari butir yang diperlukan untuk tes berukuran tetap (*fixed test*). CAT merupakan pengembangan lebih lanjut dari pengujian adaptif. VerticalNews.com (26 Oktober 2011) menyebutkan bahwa CAT merupakan area paling inovatif di dalam pendidikan, yang menyediakan pengalaman penilaian yang susunan butir-butir tesnya disesuaikan dengan tingkat kemampuan peserta tes berdasarkan hasil analisis jawaban peserta tes selama proses pengujian berlangsung. CAT memberikan banyak solusi dari kesulitan yang ada pada pengujian adaptif tanpa bantuan komputer. Larson &

Madsen (2011, p.32) menyatakan bahwa CAT mampu memberikan solusi yang sangat menjanjikan untuk menyediakan tes dengan ketelitian yang memadai dan ukuran yang umum dalam mengukur perbedaan kemampuan individu.

Pengujian menggunakan metode CAT memiliki banyak kelebihan dibandingkan pengujian menggunakan kertas-dan-pensil. Kantrowitz, Dawson, & Fetzer (2011, p.227) menyatakan bahwa CAT mengkombinasikan ilmu pengetahuan dan teknologi untuk membantu menyampaikan sebuah tujuan pengujian dan sekaligus mengamankan pengujian tersebut. Wainer, Bradlow, & Xiaohui, (2007, p.10) menyatakan bahwa CAT dapat menghasilkan tingkat kepresisian yang sama dengan tes berukuran tetap (*fixed test*) yang memiliki sekitar dua kali jumlah butir. Pendapat lain tentang kelebihan CAT disampaikan oleh Rudner (1998a), antara lain: (1) CAT dapat memberikan tes berdasarkan permintaan (“*on demand*”) dan penskoran diberikan dengan segera; (2) tidak diperlukan lembar jawaban maupun administrator tes terlatih; dan (3) pengujian dilakukan secara individual, sehingga, seorang peserta tes tidak harus menunggu yang lain selesai, sebelum melanjutkan ke bagian berikutnya.

Pemilihan butir CAT dilakukan dengan mengikuti tingkat kemampuan peserta tes masing-masing. Wainer, Bradlow, & Xiaohui (2007, p.10) menyatakan bahwa, pengujian CAT dimulai dengan menyajikan butir dengan tingkat kesulitan sedang kepada peserta tes. Pemilihan butir pertama ini merupakan hal yang penting dalam memulai CAT. Embretson & Reise (2000, p.266) menyatakan bahwa jika dapat diasumsikan bahwa populasi peserta tes terdistribusi normal, maka pilihan yang wajar untuk memulai CAT adalah dengan butir kesulitan sedang, misalnya butir dengan parameter kesulitan antara -0,5 dan +0,5. Pemilihan butir kedua dan beberapa butir berikutnya dalam algoritma CAT dilakukan dengan memperhatikan jawaban peserta tes yang pada butir sebelumnya. Jika peserta tes menjawab benar butir tersebut, maka akan diberikan butir yang lebih sulit, sebaliknya akan diberikan butir yang lebih mudah (Hulin, Drasgow, & Parsons, 1983,

p.218; Wainer, Bradlow, & Xiaohui, 2007, p.10, dan K. Wong et.al., 2010, p.24). Lord (1980, p.153) menyatakan bahwa prosedur ini diulang sampai jawaban peserta tes berpola, yaitu setidaknya ada satu jawaban benar dan satu jawaban salah, sehingga algoritma pemilihan butir adaptif selanjutnya dilakukan berdasarkan estimasi kemampuan sementara peserta tes. Estimasi sementara ini digunakan untuk menghitung kontribusi informasi relatif dari calon butir, dan ditentukan dari jawaban butir sebelumnya.

Hambleton, Swaminathan, & Rogers (1991, p.148) menyebutkan dua pendekatan yang umum digunakan untuk mengestimasi kemampuan, yaitu *maximum likelihood* dan Bayesian. Embretson & Reise (2000, p.267) menyatakan bahwa pemilihan butir dengan memiliki informasi yang paling psikometrik pada tingkat estimasi kemampuan peserta tes, biasanya cocok dengan pendekatan *maximum likelihood*, sedangkan strategi pemilihan butir yang membuat standard error peserta tes terkecil biasanya cocok dengan pendekatan Bayesian. Pendekatan *maximum likelihood* untuk pemilihan butir dan penskoran CAT didasarkan pada fungsi *likelihood* dengan persamaan sebagai berikut (Hulin, Drasgow, & Parsons, 1983, p.48, Hambleton, Swaminathan, & Rogers, 1991, p.34, dan Embretson & Reise, 2000, p.161):

$$L(u_1, u_1, \dots, u_n | \theta) = \prod_{j=1}^n P_j^{u_j} Q_j^{1-u_j} \quad (7)$$

Di mana:

- $u_j$  : skor jawaban butir dengan nilai 1 atau 0;
- $j$  : 1, 2, ..., n
- $P_j$  : probabilitas untuk menjawab dengan benar butir ke-j, dari seorang peserta tes yang dipilih secara acak dengan kemampuan  $\theta$
- $Q_j$  : 1 -  $P_j$

Terdapat beberapa aturan berhenti yang dapat digunakan untuk menghentikan perulangan CAT. Embretson & Reise (2000, pp.267-268) menyebutkan dua aturan penghentian yang umum digunakan dalam CAT, yaitu: (1)

mengulang pemberian butir sampai tingkat kesalahan baku pengukuran di bawah beberapa nilai yang dapat diterima, prosedur ini dikenal sebagai tes adaptif dengan panjang tidak tetap (*variable-length adaptive testing*); dan (2) memberikan butir tes dalam jumlah yang tetap, yang dikenal sebagai tes adaptif dengan panjang tetap (*fixed-length adaptive testing*). Linacre (2000, p.11) menyatakan bahwa CAT dapat berhenti jika: (1) butir dalam bank soal telah habis; (2) panjang tes maksimum telah tercapai; (3) ukuran kemampuan telah diestimasi dengan kepresisian yang cukup; (4) ukuran kemampuan yang cukup jauh dari kriteria lulus-gagal; dan (5) Peserta tes menunjukkan perilaku test yang tidak wajar.

Berdasarkan dua pendapat di atas dan dari penjelasan sebelumnya, secara rinci prosedur CAT dalam penelitian pengembangan ini dinyatakan dengan langkah sebagai berikut: (1) pilih sebuah butir dari bank soal dengan tingkat kesulitan sedang antara -0,5 dan +0,5; (2) tampilkan soal terpilih; (3) periksa jawaban; (4) berikutnya, pilih butir yang lebih sulit jika jawaban benar atau lebih mudah jika jawaban salah; (5) ulangi langkah nomor 2, 3, dan 4 sampai jawaban berpola; (6) jika jawaban berpola, lakukan perhitungan estimasi kemampuan dengan menggunakan prosedur *maximum likelihood*; (7) hitung nilai informasi butir, dari tiap butir tes yang belum diberikan; (8) berikan butir tes dengan nilai informasi butir yang paling tinggi; (9) estimasi kemampuan baru peserta tes () berdasarkan keseluruhan pola jawaban; (10) hitung standar kesalahan ; (11) ulangi langkah 6 sampai dengan 10 hingga standar Kesalahan , stabil; (12) berikan skor hasil tes berdasarkan peserta tes.

### Bank Soal

Idealnya sebuah bank soal harus memiliki jumlah butir yang memadai pada setiap tingkat kemampuan untuk mencapai tingkat kepresisian perkiraan  $\theta$  yang diinginkan (Weiss, 1982, p.478). Ketika kriteria bank soal ini terpenuhi, maka semua peserta tes, mendapatkan sebuah tes yang menyesuaikan secara wajar dengan kemampuan mereka dan semua peserta tes dapat diukur secara akurat (Embretson & Reise, 2000, p.264).

Wright & Stone (1999, p.107) menyatakan bahwa bank soal merupakan sekumpulan butir-butir yang disusun dengan hati-hati dan dikalibrasi bersama-sama. Demikian pula Shaik (2006, p.34) menyatakan bahwa bank soal adalah koleksi dari butir-butir yang telah dikalibrasi yang tersimpan dalam sebuah *data base* yang dioperasikan dan dikelola oleh sistem manajemen *data base* (DBMS), seperti Oracle dan Microsoft SQL. Upaya untuk mendapatkan parameter-parameter-butir disebut sebagai proses kalibrasi butir. Shaik (2006, p.34) menyatakan bahwa dibutuhkan sampel yang besar untuk proses kalibrasi butir. Sebuah model pengukuran digunakan untuk mengkalibrasi butir-butir dalam bank soal. Sampel kalibrasi digunakan untuk mengestimasi parameter-butir dan menilai model fit terhadap data. Segall (2005, p.436) menyatakan bahwa untuk model 3PL, program pengujian skala besar cenderung untuk menggunakan sampel yang mengandung 500 atau lebih jawaban per item untuk memperkirakan parameter item. Jones, Smith, & Talley (2006, p.501) menyatakan bahwa model respon butir multi-parameter mungkin tidak sesuai untuk kasus-kasus di mana  $N \leq 200$  observasi. Namun, model respon butir satu parameter yang memaksa  $a_i = 1$  dan  $c_i = 0$  telah terbukti berguna dalam situasi jumlah sampel yang kecil.

## METODE PENELITIAN

### Jenis Penelitian

Penelitian ini merupakan penelitian pengembangan dengan menggunakan empat tahap yaitu: (1) definisi; (2) desain; (3) Pengembangan; dan (4) pengujian.

### Waktu dan Tempat Penelitian

Penelitian dilaksanakan pada bulan Agustus 2013 sampai dengan bulan Februari 2014. Pengujian instrumen tes hasil belajar mata pelajaran Menganalisis Rangkaian Listrik dilaksanakan pada 10 SMK di Daerah Istimewa Yogyakarta yang memiliki Kompetensi Kahlian Teknik Instalasi Tenaga Listrik. Pelaksanaan uji coba terbatas dan uji coba diperluas terhadap perangkat lunak SPBCAT dilaksanakan di SMK Negeri 2 Pengasih Kulonprogo.

### Target/Subjek Penelitian

Subjek uji coba terbatas adalah lima guru dan 20 siswa kelas XI TITL pada Kompetensi Keahlian TITL SMK Negeri 2 Pengasih. Subjek uji coba diperluas adalah 28 siswa kelas XI TITL pada sekolah yang sama.

### Prosedur

Kegiatan penelitian diawali dengan tahap definisi. Dalam tahap ini dilakukan analisis kebutuhan, analisis awal-akhir, analisis pengguna, analisis materi, analisis teknologi, dan analisis bahasa pemrograman. Hasil dari tahap definisi selanjutnya digunakan sebagai acuan dalam tahap desain produk.

Terdapat empat jenis jenis produk yang didesain, dalam penelitian ini yaitu: (1) desain instrumen tes hasil belajar mata pelajaran Menganalisis Rangkaian Listrik; (2) desain instrumen penilaian kinerja perangkat lunak SPBCAT; (3) desain sistem basis data; dan (4) desain perangkat lunak SPBCAT.

Pengembangan instrumen tes hasil belajar mata pelajaran Menganalisis Rangkaian Listrik dilakukan dengan langkah: (1) menjabarkan indikator pencapaian hasil belajar yang didasarkan pada kompetensi dasar yang ingin dicapai; (2) mengembangkan indikator menjadi kisi-kisi tes berdasarkan standar kompetensi, dan kompetensi dasar yang ada dalam mata pelajaran Menganalisis Rangkaian Listrik; (3) mengembangkan kisi-kisi tes menjadi butir-butir tes pada setiap kompetensi dasar; (4) menyusun butir-butir menjadi instrumen tes; dan (4) membuat kunci jawaban.

Instrumen penilaian kinerja perangkat lunak SPBCAT dikembangkan menjadi tiga macam menurut fungsi dan penggunaannya, yaitu: (1) instrumen pengujian keberfungsian perangkat lunak SPBCAT yang digunakan oleh validator ahli pemrograman web untuk menguji validitas dan kinerja perangkat lunak yang dikembangkan; (2) angket penilaian guru terhadap perangkat lunak SPBCAT yang digunakan untuk menggali penilaian guru terhadap kinerja perangkat lunak SPBCAT sebagai pengguna pertama (*first user*); (3) angket penilaian siswa terhadap perangkat lunak SPBCAT yang digunakan untuk menggali penilaian siswa terhadap kinerja perangkat lunak

SPBCAT sebagai pengguna akhir (end user). Langkah-langkah pengembangan instrumen penilaian kinerja perangkat lunak SPBCAT adalah sebagai berikut: (1) mengembangkan indikator-indikator kinerja perangkat lunak SPBCAT pada masing-masing aspek menjadi kisi-kisi penilaian; (2) mengembangkan kisi-kisi penilaian menjadi butir-butir pernyataan; dan (3) menyusun butir-butir pernyataan yang telah dikembangkan menjadi instrumen penilaian kinerja perangkat lunak SPBCAT.

Sistem basis data merupakan bagian yang tidak terpisahkan dari perangkat lunak SPBCAT. Pengembangan sistem basis data dilakukan dengan langkah-langkah sebagai berikut: (1) membuka halaman browser, dan masuk ke dalam halaman sistem basis data; (2) membuat sistem basis data dengan nama yang sesuai dengan tujuan penelitian; (3) membuat tabel dengan kode nama yang sesuai dengan kategori/kelompok data yang akan disimpan; (4) menambahkan kolom tabel (table-field) dan memberikan nama kolom-kolom tersebut dengan kode nama yang sesuai dengan kelompok data yang akan diisikan; (5) memilih tipe data yang akan disimpan oleh masing-masing kolom tabel; (6) menuliskan ukuran data masing-masing kolom; dan (7) menambahkan atribut masing-masing kolom jika diperlukan.

Perangkat lunak SPBCAT yang telah melalui proses desain, selanjutnya dikembangkan sehingga dapat menjalankan tugas utamanya yaitu menyajikan butir-butir tes secara adaptif. Selain kemampuan utama di atas, perangkat lunak SPBCAT dikembangkan untuk mampu melakukan beberapa fungsi pendukung antara lain melakukan verifikasi pengguna pada saat login, menampilkan hasil ujian siswa, mengaktifkan/menonaktifkan pengguna, dan beberapa tugas pendukung lainnya. Pengembangan perangkat lunak SPBCAT dalam penelitian ini dilakukan sebagai berikut: (1) mengembangkan perangkat lunak SPBCAT serta komponen pendukung lainnya berdasarkan rancangan; (2) selama proses pengembangan, perangkat lunak SPBCAT senantiasa diujicobakan untuk memastikan dapat bekerja dengan benar, dan memastikan bahwa input yang diberikan telah dapat diproses dengan benar sehingga menghasilkan output yang sesuai dengan rancangan;

dan (3) melakukan pengujian kinerja perangkat lunak sehingga diperoleh produk perangkat lunak yang bebas dari kesalahan.

Produk-produk yang telah dikembangkan selanjutnya diujicobakan dilapangan. Uji coba produk dapat dikelompokkan menjadi dua, yaitu uji coba terhadap instrumen penilaian dan uji coba terhadap perangkat lunak SPBCAT. Uji coba instrumen penilaian terdiri atas dua tahap, yaitu (1) validasi instrumen tes hasil belajar dan instrumen penilaian kinerja perangkat lunak SPBCAT oleh ahli pengukuran; dan (2) uji coba lapangan terhadap instrumen tes hasil belajar. Uji coba terhadap perangkat lunak SPBCAT dilakukan oleh pihak lain selain pembuat program, dan disebut sebagai pengujian eksternal. Pengujian eksternal dibagi menjadi tiga tahap yaitu: (1) validasi ahli pemrograman web; (2) uji coba terbatas oleh kelompok guru sebagai pengguna pertama (first user) dan oleh kelompok siswa sebagai pengguna akhir (end user) dengan jumlah terbatas; dan (3) uji coba diperluas oleh kelompok siswa sebagai end user dengan jumlah pengguna yang sebenarnya.

### **Teknik Analisi Data**

Teknik pengumpulan data dari penelitian ini dilakukan dengan melakukan tes hasil belajar, menyebarkan angket, wawancara, dan observasi. Penelitian ini menggunakan tujuh buah instrumen pengumpulan data yaitu: (1) lembar validasi instrumen tes hasil belajar mata pelajaran Menganalisis Rangkaian Listrik; (2) lembar validasi instrumen pengujian keberfungsian perangkat lunak SPBCAT; (3) lembar validasi angket penilaian guru/siswa terhadap perangkat lunak SPBCAT; (4) Instrumen tes hasil belajar mata pelajaran Menganalisis Rangkaian Listrik; (5) instrumen pengujian keberfungsian perangkat lunak SPBCAT; (6) angket penilaian guru terhadap perangkat lunak SPBCAT; dan (7) angket penilaian siswa terhadap perangkat lunak SPBCAT. Rincian dari ke tujuh instrumen tersebut adalah sebagai berikut:

Analisis data dari penelitian ini bertujuan untuk mengetahui bahwa produk yang dikembangkan telah memenuhi persyaratan kevalidan dan kelayakan untuk digunakan. Terdapat

empat jenis analisis data yang dilakukan dalam penelitian ini, yaitu: (1) analisis data hasil validasi ahli pengukuran pendidikan; (2) analisis butir tes; (3) analisis data hasil penilaian ahli pemrograman web; dan (4) analisis data hasil penilaian pengguna.

Analisis data hasil validasi ahli pengukuran pendidikan mengacu pada instrumen validasi ahli pengukuran. Instrumen tersebut menggunakan dua kriteria, yaitu layak dan tidak layak terhadap butir yang dinilai. Hasil pengujian validasi selanjutnya dianalisis den-

gan langkah-langkah: (1) butir-butir yang dinyatakan tidak layak, disingkirkan dari instrumen; dan (2) butir-butir yang layak namun terdapat kesalahan dalam redaksi dan penulisan, baik pada pokok soal maupun pada pilihan jawaban, direvisi berdasarkan masukan dari validator

Analisis butir tes untuk mata pelajaran Menganalisis Rangkaian Listrik dilakukan menggunakan software BILOG-MG. Kriteria butir yang layak ditentukan berdasarkan tabel berikut

Tabel 1. Kriteria Penerimaan Butir Berdasarkan Nilai

Paramater/prob	Kriteria	Keterangan
SLOPE	0,3 s/d 2,00	Daya beda butir baik
THRESHOLD	-2,00 s/d +2,00	Tingkat kesulitan butir baik
ASYMPTOTE	0 s/d	Faktor tebakan baik
PROB	> 0,05	Butir fit model

Analisis data hasil validasi ahli pemrograman web mengacu pada instrumen pengujian keberfungsian perangkat lunak SPBCAT. Instrumen tersebut menggunakan dua kriteria, yaitu layak dan tidak layak berdasarkan unjuk kerja perangkat lunak SPBCAT saat pengujian. Perbaikan kode program akan dilakukan untuk setiap aspek yang tidak mampu ditunjukkan oleh perangkat lunak SPBCAT.

## HASIL PENELITIAN DAN PEMBAHASAN

Perangkat lunak SPBCAT yang dikembangkan dalam penelitian dan pengembangan ini telah melalui proses validasi ahli dan uji coba lapangan. Perangkat lunak SPBCAT telah mengalami perbaikan/revisi sesuai dengan saran dan masukan baik dari validator maupun dari responden di lapangan.

Instrumen tes terstandar yang telah dikembangkan untuk mata pelajaran Menganalisis Rangkaian Listrik berbentuk pilihan ganda yang memiliki lima pilihan jawaban. Instrumen tes dikembangkan dengan mengacu pada silabus yang digunakan pada Kompetensi Keahlian Teknik Instalasi Tenaga Listrik. Kisi-kisi tes dikembangkan dengan mengacu pada in-

dikator-indikator pencapaian kompetensi yang terdapat dalam silabus tersebut. Butir-butir tes yang dikembangkan dari kisi-kisi tes dikelompokkan berdasarkan KD yang ada dalam SK Menganalisis Rangkaian Listrik dalam mata pelajaran Menganalisis Rangkaian Listrik. Butir-butir tes hasil pengembangan telah divalidasi oleh ahli pengukuran sehingga butir tes telah sesuai dengan indikator-indikator dalam silabus, memenuhi ciri atribut butir yang hendak diukur, dan dapat mengukur setiap aspek berpikir seperti diuraikan dalam standar kompetensi, kompetensi dasar, maupun indikator yang terdapat dalam kurikulum.

Instrumen tes yang telah valid diujicobakan di lapangan dengan melibatkan 264 siswa kelas XI TITL pada Kompetensi Keahlian Teknik Instalasi Tenaga Listrik. Variasi jawaban siswa terhadap instrumen tes, diolah menggunakan perangkat lunak BILOG-MG MG Versi 3.0 dari *Scientific Software International, Inc.* Berdasarkan hasil analisis butir, diperoleh informasi reliabilitas empiris instrumen tes pada masing-masing KD, yaitu: (1) reliabilitas empiris instrumen tes KD 1 adalah 0,9086; (2) reliabilitas empiris instrumen tes KD 2 adalah 0,9067; (3) reliabilitas empiris instrumen tes KD 3 adalah 0,9087; dan (4) re-

liabilitas empiris instrumen tes KD 4 adalah 0,9086.

Berdasarkan hasil analisis menggunakan BILOG-MG, diperoleh karakteristik dari masing-masing butir. Karakteristik butir tes terdiri atas daya beda butir, tingkat kesulitan butir, faktor tebakan butir, dan nilai fit model. Berdasarkan hasil analisis, dapat diketahui bahwa butir-butir tes pada kompetensi dasar 1, 2, dan 4 paling cocok menggunakan model logistik satu parameter, dan butir-butir tes pada kompetensi dasar 3 paling cocok menggunakan model logistik 2 parameter. Dengan demikian butir-butir tes pada kompetensi dasar 1, 2, dan 4 hanya dijelaskan oleh sebuah parameter yaitu tingkat kesulitan butir, sedangkan butir-butir tes pada kompetensi dasar 3 dijelaskan oleh dua parameter, yaitu daya beda dan tingkat kesulitan butir.

Berdasarkan data dalam file *output* PH2 untuk masing-masing KD, diperoleh informasi tentang butir-butir yang fit model dengan nilai PROB masing-masing butir lebih besar dari 0,05. Jumlah butir yang memenuhi kriteria fit model untuk KD 1 sebanyak 26 butir dari 30 butir yang diujikan, KD 2 sebanyak 27 butir dari 30 butir yang diujikan, KD 3 sebanyak 27 butir dari 30 butir yang diujikan, dan KD 4 sebanyak 25 butir dari 30 butir yang diujikan. Distribusi butir yang dikelompok dalam kategori tingkat kesulitan rendah, sedang, dan tinggi pada masing-masing kompetensi dasar adalah: (1) KD 1, terdiri atas 27% butir dengan tingkat kesulitan rendah, 50% butir dengan tingkat kesulitan sedang, dan 23% butir dengan tingkat kesulitan tinggi; (2) KD 2, terdiri atas 26% butir dengan tingkat kesulitan rendah, 48% butir dengan tingkat kesulitan sedang, dan 26% butir dengan tingkat kesulitan tinggi; (3) KD 3, terdiri atas 26% butir dengan tingkat kesulitan rendah, 52% butir dengan tingkat kesulitan sedang, dan 22% butir dengan tingkat kesulitan tinggi; dan (4) KD 4, terdiri atas 24% butir dengan tingkat kesulitan rendah, 52% butir dengan tingkat kesulitan sedang, dan 24% butir dengan tingkat kesulitan tinggi.

Butir-butir yang memenuhi kriteria kelayakan berdasarkan hasil analisis butir menggunakan BILOG-MG, selanjutnya dituliskan kembali ke dalam bank soal. Bank soal meru-

upakan bagian dari sistem basis data yang terintegrasi dengan perangkat lunak SPBCAT. Bank soal yang telah dikembangkan mampu menyimpan pokok butir tes beserta pilihan jawaban, kunci jawaban, dan parameter dari butir-butir tersebut. Butir-butir tes dalam Bank soal juga dilengkapi oleh identitas butir berupa kode butir, tanda butir, dan nomor *ID* kompetensi dasar.

Bank soal mampu menyimpan tiga jenis parameter butir, yaitu daya beda, tingkat kesulitan, dan faktor tebakan. Kode butir dalam bank soal merupakan penanda butir yang bersifat unik untuk membedakan sebuah butir dengan butir yang lainnya. Tanda butir berfungsi sebagai penanda bahwa butir tersebut berstatus aktif atau tidak. Hanya butir berstatus aktif yang akan dimunculkan di lembar tes siswa. Nomor *ID* kompetensi dasar (*idKd*) menyimpan informasi tentang *ID* dari kompetensi dasar untuk butir tersebut, sehingga informasi tentang nomor dan judul standar kompetensi, serta nomor dan judul kompetensi dasar akan dapat diperoleh oleh pengguna bertipe guru yang sedang mengakses data dari bank soal.

Butir-butir tes dari bank soal dapat diakses oleh guru menggunakan perangkat lunak SPBCAT melalui menu “Menjelajahi Butir” atau menu “Membuat Tes”. Bagian dari butir tes yang ditampilkan kepada guru dari bank soal meliputi: pokok butir, pilihan jawaban, kode butir, daya beda, tingkat kesulitan, faktor tebakan, nomor standar kompetensi, nomor kompetensi dasar, dan kunci jawaban. Butir-butir tes dari bank soal juga dapat diakses oleh siswa pada saat mengikuti tes menggunakan perangkat lunak SPBCAT. Bagian dari butir tes yang ditampilkan kepada siswa peserta tes dari bank soal meliputi: pokok butir, pilihan jawaban, dan nomor standar kompetensi atau kompetensi dasar.

Hasil uji coba perangkat lunak SPBCAT pada *end user* baik pada uji coba terbatas maupun uji coba diperluas menunjukkan bahwa perangkat lunak mampu memilih butir-butir awal dari bank soal dengan tingkat kesulitan sedang dan menyajikan butir tes tersebut kepada peserta tes. Perangkat lunak SPBCAT mampu menampilkan pilihan jawaban dari butir tes secara acak untuk mengurangi peluang

kerjasama antar peserta tes. Perangkat lunak SPBCAT mampu menilai dengan segera jawaban yang diberikan peserta tes dan memutuskan apakah jawaban yang diberikan adalah benar atau salah berdasarkan kode jawaban butir. Perangkat lunak SPBCAT mampu memilih butir selanjutnya dari bank soal dengan tingkat kesulitan butir yang lebih tinggi jika butir sebelumnya dijawab benar, dan mampu memilih butir selanjutnya dari bank soal dengan tingkat kesulitan yang lebih rendah jika butir sebelumnya dijawab salah. Perangkat lunak SPBCAT akan mengulang proses pemilihan butir dengan tingkat kesulitan yang lebih tinggi atau lebih rendah berdasarkan jawaban peserta tes, hingga jawaban tersebut berpola, yaitu minimal satu jawaban benar dan satu jawaban salah.

Berdasarkan pola jawaban peserta tes, perangkat lunak SPBCAT mampu melakukan estimasi kemampuan sementara peserta tes menggunakan metode estimasi maksimum *likelihood*. Hasil estimasi kemampuan sementara () digunakan oleh perangkat lunak untuk memilih butir dari bank soal yang memiliki nilai fungsi informasi butir (FIB) tertinggi pada tersebut. perangkat lunak SPBCAT telah mampu mengulang proses penyajian butir dengan nilai FIB tertinggi hingga kondisi “persyaratan berhenti” (*stopping rule*) terpenuhi.

Kemasan butir yang berhasil disusun oleh perangkat lunak SPBCAT secara adaptif telah sesuai dengan kemampuan masing-masing peserta tes. Kesesuaian kemasan butir-butir tes dengan kemampuan peserta tes dapat dilihat dari nilai SEM yang relatif kecil. Berdasarkan rekaman data hasil tes siswa dalam uji coba terbatas dapat diketahui bahwa perangkat lunak SPBCAT pernah membuat kemasan butir tes secara adaptif dengan nilai SEM terkecil sebesar 0,311 dan terbesar sebesar 0,499. Nilai SEM yang kecil menunjukkan bahwa kemasan butir yang telah disusun secara adaptif menggunakan perangkat lunak SPBCAT ini telah memberikan hasil tes yang sangat tepat untuk peserta tes tersebut.

Perangkat lunak SPBCAT telah dikembangkan untuk mampu mengukur kemampuan peserta tes berdasarkan hasil tes yang telah dikerjakan. Perhitungan tingkat kemampuan peserta tes menggunakan metode MLE di-

lakukan dengan tingkat ketelitian 0,01. Hasil estimasi kemampuan peserta tes menggunakan metode MLE, merupakan skala logit dengan rentang -3 sampai dengan +3, dan belum dapat dipahami secara umum. Agar dapat dipahami secara umum, perangkat lunak SPBCAT telah dikembangkan untuk mampu mengkonversi nilai dalam skala logit tersebut menjadi nilai dalam skala 0 sampai dengan 100 dan dalam skala 1 sampai dengan 4. Selain dalam bentuk angka, perangkat lunak SPBCAT telah dikembangkan untuk mampu menampilkan predikat yang dicapai dalam bentuk huruf. Perangkat lunak SPBCAT juga telah dikembangkan untuk mampu menampilkan keterangan ketercapaian kompetensi peserta tes, yaitu kompeten dan belum kompeten.

Perangkat lunak SPBCAT telah diujicobakan pada *local area network* (LAN) di laboratorium KKPI-2 SMK 2 Pengasih. Dalam uji coba diperluas perangkat lunak SPBCAT telah terbukti dapat diakses oleh 28 komputer klien secara serempak dan telah terbukti mampu menyajikan butir-butir tes secara adaptif kepada 28 siswa dengan lancar dan tanpa masalah.

## SIMPULAN DAN SARAN

### Simpulan

Instrumen tes terstandar berbantuan komputer yang telah dikembangkan untuk mata pelajaran Menganalisis Rangkaian Listrik memiliki bentuk pilihan ganda. tiap-tiap KD memiliki reliabilitas empiris yang tinggi, yaitu lebih dari 0,9; (2) butir-butir tes pada KD 1, 2, dan 4 paling cocok dengan model logistik satu parameter, sedangkan butir-butir tes pada KD 3 paling cocok dengan model logistik dua parameter; (3) jumlah keseluruhan butir yang memenuhi kriteria kelayakan adalah 105 dan tersimpan dalam bank soal; (4) bank soal yang telah dikembangkan mampu menyimpan butir-butir tes terstandar yang dilengkapi dengan lima pilihan jawaban, kunci jawaban, daya beda butir, tingkat kesulitan butir, faktor tebakan butir, kode butir, tanda butir dan nomor *ID* kompetensi dasar.

Perangkat lunak SPBCAT telah mampu menunjukkan kinerjanya dalam kegiatan penilaian hasil belajar untuk mata pelajaran Men-

ganalisis Rangkaian Listrik, yaitu mampu: (1) menyajikan butir-butir tes berdasarkan algoritma CAT; (2) menghasilkan kemasan butir-butir tes yang telah sesuai dengan kemampuan masing-masing peserta tes dengan nilai SEM yang relatif kecil; (3) mengambil dan mengirim data dari sistem basis data dengan baik; dan (4) mengukur kemampuan akhir peserta tes menggunakan metode MLE dengan tingkat ketelitian 0,01.

Secara keseluruhan, perangkat lunak SPBCAT telah mampu menampilkan unjuk kerjanya secara layak meliputi: (1) menyajikan butir-butir tes secara adaptif kepada 28 siswa dengan lancar dan tanpa masalah; (2) bekerja dengan baik di bawah sistem operasi Windows; (3) mendeteksi tipe pengguna dan memberikan hak akses yang sesuai; (4) menerima dan mengolah masukan yang diberikan pengguna; (5) melakukan proses pengujian dengan menyajikan butir tes secara adaptif (6) mengolah hasil tes siswa dan menampilkan dengan segera setelah siswa menyelesaikan kegiatan tes tersebut; (7) menyimpan data yang dimasukkan pengguna atau data hasil olahan perangkat lunak ke dalam sistem basis data; dan (8) mengolah serta menyajikan data dari sistem basis data ke halaman pengguna.

### Saran

Perangkat lunak SPBCAT yang dilengkapi butir-butir tes terstandar dalam bank soal, telah teruji validitas dan reliabilitasnya, sehingga dapat dimanfaatkan pada kegiatan penilaian mata pelajaran Menganalisis Rangkaian Listrik.

Perangkat lunak SPBCAT telah didesain dan dikembangkan dengan memperhatikan kemampuan pengguna dalam mengoperasikan komputer pada tingkat dasar. Tombol dan menu navigasi yang digunakan dibuat dengan sederhana dan dilengkapi petunjuk (*tool tips*) yang akan muncul jika penunjuk *mouse* diarahkan di atas komponen tertentu pada halaman yang diakses pengguna. Desain yang sederhana dan didukung oleh *tool tip* tersebut memungkinkan setiap pengguna dapat mengoperasikan produk dengan mudah dan lancar.

### DAFTAR PUSTAKA

- Autor, D., Levy, F., & Murnane, R. (2003, Nopember). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118 (4), 1279–1333.
- Berry, R., & Adamson, B. (2011). Assessment reform past, present and future. Dalam Berry, R., & Adamson, B. (Eds.). *Assessment Reform in Education* (pp. 3-14). New York: Springer.
- DeMars, C. (2010). *Item response theory*. New York: Oxford University Press, Inc.
- Djojonegoro, Wardiman (1998). *Pengembangan sumber daya manusia melalui sekolah menengah kejuruan (SMK)*. Jakarta: PT Jayakarta Agung Offset.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists multivariate*. Mahwa: Lawrence Erlbaum Associates, Inc.
- Falk, I., & Surata, K. (2011). Where ‘the TVET system’ meets the performativity of vocational learning: Borderlands of innovation and future directions. Dalam Catts, R., Falk, I., & Wallace, R. (Eds.). *Vocational Learning Innovative Theory and Practice* (pp. 33-62). New York: Springer.
- Folk, V.G., & Smith, R.L. (2002). Models for delivery of CBTs. Dalam Mills, et.al. (Eds.). *Computer Based Testing Building the Foundation for Future Assessments*. (pp. 41-63). Mahwah: Lawrence Erlbaum Associates, Inc.
- Froelich, A.G. (2009). Methods from item response theory: Going beyond traditional validity and reliability in standardizing assessments. Dalam Shelley II, M.C., Yore, L.D., & Hand, B. (Eds.). *Quality Research in Literacy and Science Education International Perspectives and Gold Standards*. (pp. 287-302). New York: Springer.

- Hambleton, R.K., & Jones, R.W. (1993). *Comparison of classical test theory and item response theory and their applications to test development*. Diakses tanggal 14 Oktober 2013 dari <http://ncme.org/linkservid/66968080-1320-5CAE-6E4E546A2E4FA9E1/showMeta/0/>.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park: Sage Publications.
- Hulin, C.L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: application to psychological measurement*. Homewood: Dow Jones Trurn.
- Jones, P., Smith, R.W., & Talley, D. (2006). Developing test forms for small-scale achievement testing systems. Dalam Downing, S.M. & Haladyna, T.M. (Eds.). *Handbook of Test Development*. (pp. 487-526). Mahwah: Lawrence Erlbaum Associates, Inc.
- Jhonson, D. W., & Jhonson, R.T. (2002). *Meaningful assessment: A manageable and cooperative process*. Boston: Allyn and Bacon.
- Kantrowitz, T.M., Dawson, C.R., & Fetzer M.S. (2011). Computer adaptive testing (CAT): A faster, smarter, and more secure approach to pre-employment testing [Versi elektronik]. *Journal of Business Psychology*, 26:227–232. Diakses tanggal 20 Juni 2014 dari <http://search.proquest.com/docview/867330795/758370F8DF04076PQ/17?accountid=31324>
- Larson, J.W., & Madsen, H.S. (2011). Computerized adaptive language testing: Moving beyond computer-assisted testing [Versi elektronik]. *Calico Journal*, Vol 2, 3, 32-43. Diakses tanggal 20 Juni 2014 dari <http://search.proquest.com/docview/749990459/758370F8DF04076PQ/13?accountid=31324>.
- Parshall, C. G., et.al. (2002). *Practical considerations in computer-based testing*. New York: Springer.
- Segall, D.O. (2005). Computerized adaptive testing. Dalam Kempf-Leonard, K., et.al. (Eds.). *Encyclopedia of Social Measurement (Vol. 1)* (pp 429-438). Amsterdam: Elsevier Inc.
- Shaik, N. (2006). Computer-adaptive online exit surveys: Conceptual and methodological issues. Dalam Williams, D.D., Howell, S.L., & Hricko, M. (Eds.). *Online Assessment, Measurement, and Evaluation: Emerging Practices*. (pp. 28-44). Hershey: Information Science Publishing.
- Simms, L.J., & Watson, D. (2007). The construct validation approach to personality scale construction. Dalam Robins, R.W., Fraley, R.C., Krueger, R.F. (Eds.). *Handbook of Research Methods in Personality Psychology* (pp.240-258). New York: The Guilford Press.
- Sudira, Putu (2012). *Filosofi dan teori pendidikan vokasi dan kejuruan*. Yogyakarta: UNY Press.
- van der Linden, W.J. (2005). *Linear models for optimal test design*. New York: Springer Science Business Media, Inc.
- Wainer H., Bradlow, E.T., & Xiaohui Wang (2007). *Testlet response theory and its applications*. Cambridge: Cambridge University Press.
- Weiss, D. J. (1982). *Improving measurement quality and efficiency with adaptive testing*. *Applied Psychological Measurement*, 6, pp. 473-492. Diambil pada 28 Maret 2014 dari <http://iacat.org/sites/default/files/biblio/v06n4p473.pdf>.
- Wendler, C.L.W., & Walker, M.E., (2006). Practical issues in designing and maintaining multiple test forms for large-scale programs. Dalam Downing, S.M., & Haladyna, T.M. (Eds.). *Handbook of Test Development*. (pp. 445-468). Mahwah: Lawrence Erlbaum Associates, Inc.