

PERBANDINGAN TES BERBASIS KOMPUTER DAN TES KERTAS PENSIL DALAM PENDIDIKAN

Oleh:
Sri Hartini*)

ABSTRACT

The test is one of important functions of educational system. It is used to give many information for students, teachers, parents and decision makers. Based on that information, they make many decisions better. In one hand, it can help them to make decision about everything are related learning and education as to decide a good class for the student, possibility to continue on learning and the other hand to give many feedback about effectively learning and educational process. Now, test paper pencil are based on Behaviorism Theory. Students are responded to stimulation from environment but really they need test more than just can make to relate stimulus and respon (S-R Bond). They need to improve how can make cognitive structure, how can use their knowledge to applied and how can do to solve many problems that they're found. We need Cognitive Theory beside Behaviorism Theory.

Test based on computer (TBoC) are improved by the Cognitive Theory. It give a new perspective to the teacher and the student. It has values added more than test based on paper-pencil (TBoPP). There are many indicators of values added of TBoC:

- 1. To make a good feeling of the testee.*
- 2. Mean score of Adaptif Test is higher more than Paper-Pencil Test.*
- 3. Score accuracy with computer is higher more than human manual procedures.*
- 4. Computer can make complicated tasks to simple tasks.*

Test Based on Computer has limited too beside values added, therefore we have to remember many condition for succeeding. The conditions we need are: Computer knowledge, Computer Platform and Speed of finishing the test.

We can't effectively on applied Test based on Computer if without three conditions above.

Keywords: *Comparison, Test Based on Computer, Test based on Paper Pencil, Behaviorism Theory, Cognitive Theory.*

*) Dosen pada FKIP-Universitas Slamet Riyadi

PENDAHULUAN

Tes merupakan salah satu fungsi sistem pendidikan yang penting. Tes menghasilkan informasi yang memungkinkan siswa, pendidik, orang tua maupun penentu kebijakan membuat keputusan yang lebih baik, seperti kelas manakah yang akan dimasuki siswa,

apakah mungkin siswa melanjutkan ke pelajaran berikutnya. Di samping itu tes juga memberikan umpan balik tentang efektivitas layanan pendidikan. Tes membantu guru dan siswa menentukan bagaimanakah sebaiknya siswa belajar.

Tes yang ada sekarang ini adalah kumpulan hasil dari teori belajar dan

model pengukuran yang telah dikembangkan untuk memenuhi tuntutan sosial dan pendidikan dari masa lampau. Tes ini berakar pada tradisi teori perilaku (Greeno et al, 1996). Meskipun konsep inti dari teori dan model tersebut masih bermanfaat untuk tujuan tertentu, ia perlu diperkaya atau ditambah untuk keperluan tes yang lebih baru.

Tes tradisional mempunyai beberapa keterbatasan. Keterbatasan pertama adalah tes tidak memfokuskan pada banyak aspek kognisi dan tidak didesain untuk mengakses perbedaan tingkat pemahaman siswa. Misalkan, aspek yang tidak diakses oleh tes tradisional meliputi organisasi pengetahuan, representasi masalah, penggunaan strategi, keterampilan pengaturan diri (Glaser, 1997). Keterbatasan kedua adalah berkenaan dengan kegunaan tes untuk memperbaiki proses pembelajaran. Sebagian besar tes tradisional memberi informasi terbatas yang dapat digunakan oleh guru untuk mengidentifikasi kondisi pembelajaran yang dapat memperbaiki prestasi siswa (Dalziel, 2001). Tes semacam ini tidak mengungkap apakah siswa menggunakan strategi yang salah untuk memecahkan masalah maupun tidak memahami konsep penting dalam bidang studi.

Teori kognitif memfokuskan pada bagaimana orang mengembangkan struktur pengetahuan yang meliputi konsep yang berhubungan dengan disiplin bidang studi (ranah pengetahuan) dan prosedur untuk penalaran dan pemecahan masalah (NRC, 2001). Dalam

teori kognitif, mengetahui bermakna lebih dari akumulasi informasi faktual, tetapi berarti dapat mengintegrasikan pengetahuan, keterampilan dan prosedur sedemikian rupa agar berguna untuk menyelesaikan masalah.

Tujuan penting tes tidak hanya menentukan apa yang diketahui seseorang, tetapi juga mengakses bagaimana dan kapan ia menggunakan apa yang diketahuinya. Tes sebaiknya didasari dengan teori kognitif modern yang menganggap bahwa belajar sebagai akuisisi struktur pengetahuan konseptual dan prosedural yang terintegrasi. Siswa yang telah mempelajari konsep dan keterampilan bidang studi tertentu telah memperoleh banyak skema struktur pengetahuan. Demikian juga siswa yang telah belajar memecahkan masalah dan keterampilan dalam suatu bidang telah memperoleh sekumpulan prosedur kognitif (glaser et al, 1991).

Pada berbagai tingkatan belajar, terdapat perbedaan integrasi pengetahuan, perbedaan tingkat keterampilan prosedural, perbedaan kecepatan akses memori dan perbedaan representasi tugas (task) yang dilaksanakan. Dengan demikian pengukuran kompetensi harus didasarkan pada tes dari struktur pengetahuan dan prosedur pemrosesan informasi tersebut.

Kenyataannya, banyak teknologi pengujian telah didesain untuk dilaksanakan setelah butir-butir tes dibuat. Analisis kesukaran butir, indeks diskriminasi, prosedur penskalaan dan analisis faktor dilaksanakan setelah butir

tes dibuat (Mc Dorbald, 1999). Diperlukan lebih banyak teori sebelum dan selama pembuatan butir-butir tes. Harus diketahui tentang sifat-sifat kognitif dari kompetensi yang diakses, struktur dan proses mental ketika seseorang memperoleh keterampilan. Pengukuran kompetensi sebaiknya didesain untuk mengakses tidak hanya pengetahuan faktual tetapi juga strategi kognitif, model mental dan organisasi pengetahuan.

Ada dua pendekatan pada desain tes. Sugrue et al (1998) membedakan antara pendekatan terpusat pada konstruk (*construct-oriented*) dan terbusat pada isi (*content-oriented*). Pada pendekatan konstruk, aspek kognitif dari pengetahuan siswa yang akan diakses didefinisikan secara jelas. Tugas tes dan metode penskoran diselaraskan dengan definisi konstruk tersebut. Sedangkan pada pendekatan isi, perhatian utama adalah siswa dapat mengerjakan tugas tes yang mengandung isi pengetahuan yang diakses. Spesifikasi tes terdiri dari beberapa butir untuk tiap cakupan isi.

Sekarang ini komputer tidak hanya mempengaruhi rancangan dan struktur isi kurikulum sekolah dari keseluruhan proses pembelajaran, tetapi juga mempengaruhi pembuatan tipe-tipe tes yang digunakan untuk penilaian hasil belajar. Tes berbasis komputer dipandang sebagai alternatif yang praktis dibandingkan dengan tes kertas pensil (Hamilton et al, 2001). Dengan munculnya komputer berkecepatan tinggi dan kapasitas penyimpanan yang

besar, serta harga yang relatif murah memungkinkan tes berbasis komputer akan menjadi lebih berkembang di masa mendatang. Keuntungan menggunakan tes berbasis komputer adalah: (1) memberi hasil penskoran yang otomatis dan pemberian umpan balik yang cepat ke siswa, (2) memungkinkan hasil analisis statistik yang cepat, (3) memungkinkan banyak jenis tes dapat diambil dari bank soal yang sebelumnya sudah disimpan di hard disk, dan (4) memungkinkan siswa mengakses dari tempat yang jauh (Jefferies et al, 2000).

Tes terus berubah dari bentuk tes konvensional pilihan ganda ke tes kinerja yang lebih kompleks. Telah banyak dilakukan studi perbandingan tes pilihan ganda kertas dan pensil dan pilihan ganda berbasis komputer. Studi memperlihatkan tes berbasis komputer dan tes pilihan ganda kertas dan pensil berfungsi serupa (Choi & Tinkler, 2002; Mead & Drasgow, 1993; Russell, 1999; Schaeffer et al, 1998). Dua tipe tes menghasilkan rerata, simpangan baku, reliabilitas dan kesalahan baku pengukuran yang serupa. Tetapi masih terdapat kemungkinan ancaman terhadap komparabilitas.

Tes berbasis komputer terus berkembang sesuai dengan kemajuan perangkat keras komputer. Metode untuk mengkategorikan skor harus dimodifikasi untuk mengakomodasi perubahan tersebut (Huff, 2001). Tes pilihan ganda yang dilaksanakan melalui komputer tidak lagi dianggap sebagai tes yang inovatif (Davies, 2001). Banyak yang

beranggapan bahwa tes pilihan ganda tidak dapat mencerminkan keluasan dan kedalaman, keterampilan berfikir tingkat tinggi dan kemampuan lain lainnya (Jodein, 2001).

Teknologi komputer telah memberikan kemungkinan untuk mengembangkan tipe baru, khususnya tes yang dapat mengakses hasil belajar tingkat tinggi (Hamilton et al, 2001). Beberapa tipe tes berbasis komputer dapat dikategorikan sebagai uraian, persamaan numerik, respon numerik ganda, pemecahan masalah, pemodelan grafis, select-and-place, drop and connect, create-tree, highlighting text, pengurutan informasi, menentukan hubungan, membuat hipotesis dan tes kinerja berbasis simulasi komputer (Zenisky & Sireci, 2001; Mislevy et al, 1998). Tipe-tipe tes ini berpotensi diterapkan pada berbagai macam program pengujian.

PEMBAHASAN

1. Perangkat Lunak Pengembangan

Perangkat lunak untuk mengembangkan CBT (*computer-based testing*) dan CAL (*computer-aided learning*) multimedia umumnya memberikan metafora yang memungkinkan pengembang mengorganisasi unsur-unsur media dan menentukan interaksi. Vaughan (1993) mendeskripsikan tiga pengantaraan metafora yang meliputi metafora halaman (yang digunakan oleh Hypercard, Supercard, Toolbook, visual basic, oracle media object dan

apple media tool); dan metafora presentasi berbasis waktu (yang digunakan oleh action, cinemation dan director).

Dalam penelitian ini pengembangan program CBT dilakukan dengan menggunakan sistem authoring macromedia authorware dan bahasa pemrograman Delphi. Perangkat lunak ini dirancang bagi orang yang ingin menghasilkan program pembelajaran multimedia. Authorware merupakan salah satu bahasa authoring yang populer untuk mengembangkan CBT dan CAL multimedia. Kelebihan authorware terletak pada penggabungan teks, gambar, efek transisi, tombol dan ikon, kemudahan melompat ke halaman lain, dan pemberian umpanbalik (Dalgarno, 1998).

Pada pemrograman bahasa tradisional, seseorang harus mempelajari perintah dan hukum-hukum bagaimana menggunakan perintah tersebut. Pemrograman authorware berbeda. Pada authorware tidak digunakan perintah baris-per-baris tetapi digunakan diagram alir tempat simbol melakukan fungsi tertentu. Diagram alir tidak hanya piranti untuk merancang program, tetapi kenyataannya ia adalah program itu sendiri.

2. Klasifikasi Tes

Bentuk tes yang paling populer adalah bentuk dimana peserta tes diminta memilih satu respon dari beberapa pilihan yang sudah

ditetapkan. Karena peserta tes memilih satu jawaban, bukan membuat, tes demikian disebut tes respon-dipilih (*selected-response*). Tes demikian meliputi pilihan-ganda, penjodohan dan benar-salah. Alternatifnya, tes dapat mengharuskan peserta tes untuk mengembangkan jawabannya sendiri. Tes bentuk ini disebut tes respon-dibuat (*constructed-response*). Respon umumnya terdiri dari suatu bentuk yang kualitasnya dapat dinilai secara akurat (Stecher et al, 1997).

Ciri utama yang membedakan semua tes respon-dibuat adalah diperlukan penilai manusia untuk menilai respon. Seseorang harus

mengulas tiap jawaban, membandingkannya dengan standar dan memberikan penilaiannya. Penskoran manusia lebih lambat dan lebih mahal daripada penskoran yang dilakukan oleh mesin.

Terdapat bermacam cara mengklasifikasi tipe tes. National Center for Research in Vocational Education menyebutkan ada empat kategori tes: tes tulis, tugas kinerja, proyek dan portofolio (Rahn et al, 1995). Seperti diperlihatkan pada tabel 1, kategori tes tulis meliputi respon-dipilih dan respon-dibuat, sedangkan tiga kategori lainnya termasuk tes respon-dibuat.

Tabel 1. Kategori Tes

Kategori	Tipe Respon	
	Dipilih	Dibuat
Tes tulis		
Pilihan ganda, benar-salah, penjodohan	v	
Open-ended		v
Uraian, berbasis masalah		v
Tugas kinerja		v
Proyek		v
Portofolio		v

Sistem klasifikasi ini didasarkan utamanya pada format bagaimana pertanyaan dihadirkan dan bagaimana respon dihasilkan. Tetapi, tes respon-dipilih dan respon-dibuat berbeda dalam beberapa hal, yang meliputi kekompleksan pembuatan, administrasi penskoran, waktu yang diperlukan untuk mengerjakannya, biaya pembuatannya dan tuntutan kognitif.

3. Penskoran Butir Tes Berbantuan Komputer

Ada dua komponen dalam kerangka kerja penskoran: (1) respon yang diberikan oleh peserta tes dan (2) ungkapan jawaban.

Respon Peserta

Respon dibangkitkan oleh "aksi" maupun "tidak ada aksi" yang dilakukan oleh peserta dan dapat bervariasi pada spektrum kekompleksan yang luas. Pada salah satu ujung spektrum adalah respon karakter sederhana (misal "A"), jawaban singkat maupun komponen-kejadian (misal, *tombol nilai = True*). Sebagian besar tes berbantuan komputer modern memberikan kontrol yang dapat menangkap respon aksi tertentu dengan menggunakan indikator keadaan atau nilai hasil yang dikembalikan lagi sebagai respon. Misal, untuk menyimpan apakah peserta menggunakan *mouse* untuk mengklik obyek kontrol tertentu seperti *check box* yang ditampilkan pada layar komputer, kita dapat mencatat sebagai variabel respon dari keadaan obyek kontrol tersebut

(contoh *kontrol keadaan = True* maupun *kontrol keadaan = False*, sesuai dengan keadaan "on" maupun "off"). Kita juga dapat menyimpan hasil dari tombol keyboard yang ditekan maupun masukan teks. Pada ujung lain dari spektrum, terdapat lebih banyak struktur data yang ditangkap sebagai input (misal, teks jawaban uraian, rumus matematika, program komputer) yang memerlukan transformasi untuk menyederhanakan data.

Ungkapan Jawaban

Kunci jawaban adalah pola respon yang dapat dihubungkan ke respon peserta tes dengan aturan tertentu. Aturan dapat berupa "true" maupun "false" dan kita dapat menetapkan nilai ke salah satu atau kedua alternatif tersebut. Misalkan, penskoran dikotomi dari butir pilihan ganda umumnya menggunakan logika membandingkan nilai respon (karakter maupun keadaan) dengan jawaban yang sudah tersimpan untuk tiap butir. Jika respon sesuai dengan jawaban, ditetapkan sebuah nilai (biasanya satu). Jika respon tidak sesuai dengan jawaban, peserta tes tidak mendapatkan nilai (nilai nol). Ada tiga bagian ungkapan jawaban: (1) respon yang dievaluasi, (2) kunci jawaban yang digunakan untuk membandingkan dengan respon, dan (3) tata bahasa logika (Boolean) yang melaksanakan perbandingan dan mengevaluasi respon.

4. Temuan Terkait dengan Penerapan Test Berbasis Komputer dan Kertas Pensil

Schaeffer *et al* (1993) melakukan penelitian untuk menentukan hubungan antara tes *Graduate Record Examination* (GRE) berbasis komputer dan kertas-pensil dengan butir-butir tes yang sama. GRE terdiri dari tiga bagian yaitu verbal yang terdiri dari 38 butir pilihan ganda; analitik yang terdiri dari 30 butir pilihan ganda; dan kuantitatif yang terdiri dari 25 butir pilihan ganda. Penelitian menggunakan sampel 1000 orang yang mengambil GRE versi komputer dan 200 orang yang mengambil GRE versi kertas-pensil. Data penelitian memperlihatkan bahwa peserta tes dapat mengerjakan GRE versi komputer tanpa kesulitan yang berarti, dan reaksi terhadap tes berbasis komputer amat menyenangkan. Tidak terdeteksi adanya pengaruh mode-butir yang berarti. Tidak terdeteksi adanya pengaruh mode-tes untuk bagian verbal dan analitik, tetapi ditemukan pengaruh mode-tes untuk bagian kuantitatif. Konversi penyetaraan GRE versi komputer hampir menyerupai versi kertas-pensil untuk ketiga bagian.

Choi dan Tinkler (2002) mengevaluasi komparabilitas skor butir-butir pilihan ganda tes berbasis komputer dan kertas-pensil yang diadministrasikan ke siswa kelas tiga dan sepuluh dalam pelajaran matematika dan membaca. Koefisien korelasi antara tes berbasis komputer

dan kertas-pensil mempunyai rentang dari 0,93 untuk tes membaca kelas 10 sampai 0,96 untuk tes matematika kelas 3. Tetapi, terdapat perbedaan rerata pada estimasi kesukaran butir untuk mode yang berbeda. Perbedaan rerata pada kesukaran butir lebih besar untuk kelas 3 daripada kelas 10. Selain itu, perbedaan rerata lebih besar untuk tes membaca daripada tes matematika.

Jodoin (2001) membandingkan butir-butir tes inovatif dengan tes pilihan ganda. Tipe pilihan ganda disebut juga tipe tradisional. Tipe inovatif terdiri dari dua jenis: DC (*Drop and Connect*) dan CT (*Create Tree*). Respon peserta dikalibrasi dengan metode IRT. Digunakan model tiga parameter untuk butir yang diskor dikotomi. Digunakan model graded response untuk butir yang diskor politomi. Program PARSCALE digunakan untuk menganalisis data. Butir tipe inovatif memberikan informasi yang lebih banyak daripada tipe pilihan ganda. Tetapi tipe inovatif memerlukan waktu lebih lama daripada tipe tradisional.

Schaeffer *et al* (1998) membandingkan skor GRE tipe tes adaptif komputer dan tipe kertas-pensil. Komparabilitas adalah penting sebab program GRE menawarkan kedua mode tersebut. Peserta tes yang diambil secara sukarela diminta mengerjakan salah satu mode tes. Digunakan analisis varian untuk membandingkan skor dari dua tipe tes. Penskoran tes adaptif menggunakan

metode pensekoran 80% yaitu yang dianalisis adalah peserta yang menjawab sedikitnya 80% dari jumlah butir tes. Hasil penelitian memperlihatkan bahwa rerata skor tes adaptif lebih tinggi dari pada rerata skor tes tulis untuk ketiga variabel.

Boyle (2002) membandingkan skor yang diperoleh dari butir-butir pilihan ganda dan dari butir-butir respon ganda dari tes berbasis komputer. Penggunaan pertanyaan pilihan ganda berbasis komputer telah menghasilkan tes yang reliabel, meskipun skor yang rendah. Penggunaan pensekoran negatif menyebabkan hal tersebut. Karena pertambahan tersebut lalu digunakan pertanyaan respon ganda. Ada dua tipe tes berbantuan komputer: tipe pilihan ganda (MCQ) dan tipe respon ganda (MRQ). Jumlah masing-masing tipe adalah 30 butir. Soal diberikan ke mahasiswa yang mengambil mata kuliah mineralogi. Digunakan analisis statistik deskriptif. Hasil penelitian memperlihatkan bahwa pertanyaan MCQ mudah diterapkan, dan menghasilkan asesmen yang reliabel. Tetapi jika digunakan pensekoran negatif, skor yang dihasilkan rendah. Sedangkan pensekoran negatif pada MRQ menghasilkan skor yang lebih baik.

Bennett *et al* (2000) membandingkan perbedaan antara pensekoran yang dilakukan oleh manusia dengan yang dilakukan oleh komputer. Untuk menguji akurasi

pensekoran Bennett menjalankan sistem pensekoran terhadap respon dari 1864 peserta tes. Tes terdiri dari 33 butir. Tipe tes adalah *mathematical expression*. Sistem pensekoran otomatis dibandingkan dengan sistem pensekoran yang dilakukan oleh manusia. Akurasi pensekoran adalah 99,62% atau kesalahan 38 respon per 10000 respon. Hasil ini memperlihatkan salah satu keunggulan tes berbantuan komputer yaitu pensekoran dilakukan secara otomatis.

Rosenquit *et al* (2000) membandingkan asesmen hands-on dan asesmen berbasis komputer. Sebanyak 40 siswa SLTP yang mengambil pelajaran fisika telah berperan dalam penelitian ini. Setengah kelas melakukan asesmen hands-on pertama dan dua minggu berikutnya melakukan asesmen versi komputer. Separuh siswa lainnya melakukan sebaliknya. Hasil penelitian menunjukkan bahwa kedua kelompok tidak berbeda secara signifikan ($F=0,0031$) saat mengerjakan asesmen hands-on dan versi komputernya. Kedua metode asesmen, hands-on dan komputer dapat dipertukarkan.

Van de Vijver dan Harsveld (1994) melakukan studi terhadap 326 peserta didik di *Royal Military Academy* di Belanda. Satu kelompok mengambil versi komputer dari *General Aptitude Test Battery* (GATB) dan kelompok lain mengambil bentuk kertas-pensil dari tes ini. Mereka menyimpulkan

bahwa tes yang sederhana lebih dipengaruhi oleh efek medium daripada tugas yang lebih kompleks.

Tes akan menjadi adil atau tidak adil jika peserta tes yang berkemampuan sama dalam pengetahuan yang sedang diukur menjawab secara berbeda pada pertanyaan tertentu disebabkan oleh pengalaman maupun pengetahuan yang tidak terkait dengan pengetahuan yang diukur tersebut. Standards (AERA, APA & NCME, 1999) menyebut pengalaman maupun keterampilan yang tidak terkait dengan konstruk yang diukur sebagai varian konstruk tak-relevan (*construct-irrelevant*). Seperti semua tes, tes berbasis komputer berusaha mengukur konstruk target semaksimal mungkin. Pada saat yang sama tes berbasis komputer dapat memunculkan faktor gangguan, yang menyebabkan kontaminasi pada skor yang disebabkan konstruk tak-relevan.

Varian konstruk tak-relevan pada tes berbasis komputer dapat berupa pengetahuan komputer, platform komputer dan kecepatan.

Pengetahuan komputer. Jika kemampuan berinteraksi dengan komputer diperlukan untuk mengerjakan tes dengan baik, tetapi tes tidak dirancang untuk mengukur fasilitas komputer, maka pengetahuan komputer akan mempengaruhi kinerja tes. Penelitian Choi dan Tinkler (2002) menunjukkan bahwa untuk peserta tes yang lebih dewasa pengalaman

komputer tidak mempengaruhi kinerja tes. Untuk tipe pilihan ganda sampai sekarang dinyatakan bahwa perbedaan pengetahuan komputer hanya sedikit mempengaruhi skor tes (Taylor *et al*, 1998). Pada tingkatan perguruan tinggi teknik, semua mahasiswa sudah terbiasa bekerja dengan komputer, sehingga diperkirakan konstruk tak-relevan ini tidak berpengaruh secara signifikan pada kinerja tes.

Platform komputer. Perbedaan bermacam-macam platform komputer dapat memunculkan konstruk tak-relevan. Peserta tes yang mengenal dengan baik sistem operasi Macintosh dapat mengalami kesulitan dengan sistem operasi Windows, dan sebaliknya. Demikian juga, seseorang yang terbiasa bekerja dengan monitor besar akan mengalami kesulitan ketika menghadapi komputer yang mempunyai monitor lebih kecil. Bridgeman *et al* (2001) menemukan bahwa memperkecil layar dapat mengurangi kinerja tes membaca tetapi dengan penurunan yang tidak signifikan. Juga Powers dan Potenza (1999) menemukan bukti bahwa tulisan pada komputer laptop tidak kompatibel dengan tulisan pada komputer desktop yang mempunyai keyboard dan tampilan layar yang lebih baik. Penelitian yang akan dilaksanakan oleh penulis akan menggunakan komputer dengan sistem operasi yang sama dan layar monitor yang berukuran sama.

Kecepatan. Bentuk lain konstruk tak-

relevan adalah kecepatan yang mengacu pada kemampuan bekerja cepat ketika mengerjakan tes. Russell (1999) menemukan bahwa kecepatan menekan keyboard merupakan prediktor dari skor siswa pada tes open-ended sains dan bahasa. Tetapi sebaliknya, skor pada tes open-ended matematika berkorelasi kecil dengan kecepatan menekan keyboard. Ini kemungkinan disebabkan jawaban terhadap butir tes matematika tidak terlalu sering menggunakan tombol keyboard.

SIMPULAN

Atas dasar uraian di atas, maka dapat diambil suatu kesimpulan bahwa test Berbasis Komputer memiliki keunggulan dibanding test Kertas Pensil. Keunggulan-keunggulan test Berbasis Komputer, ditunjukkan oleh indikator-indikator sebagai berikut :

1. Menimbulkan reaksi yang menyenangkan pada diri teste
2. Rerata skor test adaptif lebih tinggi daripada rerata skor test tulis
3. Akurasi penskoran yang dilaksanakan dengan komputer lebih tinggi dibanding dengan yang dilakukan secara manual oleh manusia.
4. Medium seperti komputer memiliki efek menyederhanakan tugas-tugas yang kompleks.

Disamping keunggulan-keunggulan tersebut Test Berbasis Komputer memiliki sejumlah keterbatasan. Perlu diingat bahwa ada

faktor-faktor yang harus diperhatikan agar test Berbasis Komputer dapat terlaksana sesuai dengan tujuan yang diharapkan. Faktor-faktor tersebut adalah

1. Pengetahuan komputer,
2. Platform komputer dan
3. Kecepatan mengerjakan test.

Jika ketiga faktor tersebut tidak terpenuhi niscaya penerapan test Berbasis Komputer tidak akan efektif.

DAFTAR PUSTAKA

- Dalziel, J. (2001). *Enhancing web-base learning with computer assisted assessment:: pedagogical and technical considerations*. Proceeding of the Fourth Annual Computer Assisted Assessment Conference. Loughborough, July 2001.
- Glaser R. (1991). Expertise and assessment. Dalam MC Wittrock & FL Baker (Eds) *Testing on cognition* (pp 17-30). Englewood Cliffs, NJ: Prentice Hall.
- Glaser R & Silver E (1994). Assessment, testing and instruction. Dalam L. Darling hammand (Ed). *Review of research in education* (Vol 20 pp. 393-419). Washington DC: American educational research association.

- Glaser R (1997). *Assessment and education: access and achievement* (CSE Tech. Rep. no 435). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing.
- Greeno JG Collins A.M & Resnick LB (1996). Cognition and learning. Dalam DC Berliner & RC Calfee (Eds). *Handbook of educational psychology* (pp 15-45) New York: Macmillan.
- Hamilton LS Klein SP & Lorie W (2001). *Using web-based testing for large-scale assessment*. Santa Monica: RAND.
- Huff KL (2001). *Overcoming unique challenges to a complex performance assessment: a novel approach to standard setting*. Makalah yang disampaikan di pertemuan tahunan national council on measurement in education. Atlanta, GA.
- Jefferies P, Constable I, Kiely B, Richardson D & Abraham A (2000). *Computer aided assessment using Web CT*. proceeding of the fourth annual computer assisted assessment conference, loughborough, June 2000.
- Jodoin M.G (2001). *An empirical examination of IRT information for innovative item format in a computer based certification testing program*. Makalah yang disampaikan di pertemuan tahunan National Council on Measurement in Education, Atlanta, GA.
- McDonald RP (1999). *Test theory: A unified treatment*. Mahwah N.J: Lawrence Erlaum Associate Publishers.
- Mislevy RJ, Steinberg LS, Breyer FJ, Almood RG, Johnson L (1998). *A cognitive task analysis with implications for designing: A simulation-based performance assessment* (CSE Tech. Rep. No. 487). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing.
- Samejima F (1997). The graded response model. In WJ Van der Linden & Hambleton RK (Eds). *Handbook of modern item response theory*. New York: Springer.
- Sagrué B (1994). *Specifications for the design of problem-solving assessments in science* (CSE Tech. Rep. No. 387). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing.

Zemisky AL & Sireci SG (2001). Feasibility review of selected performance assessment item type for the computerized uniform CPA exam. *Laboratory of Psychometric and Evaluative Research Report No. 405*. School of Education, University of Massachusetts, Amherst, MA.