



Pengembangan Instrumen Asesmen Pengetahuan Fisika Berbasis Komputer untuk Meningkatkan Kesiapan Peserta Didik dalam Menghadapi Ujian Nasional Berbasis Komputer

Suyoso *, Edi Istiono, Subroto

¹ Jurusan Pendidikan Fisika FMIPA UNY. Jl. Colombo No. 1, Karangmalang, Yogyakarta, Indonesia.

* Korespondensi Penulis. E-mail: suyososuyoso@gmail.com dan suyoso@uny.ac.id

Received: 10 January 2017; Revised: 10 March 2017; Accepted: 10 April 2017

Abstrak

Penelitian ini bertujuan untuk 1) menghasilkan instrumen penilaian berbasis komputer yang layak untuk mengukur kemampuan kognitif peserta didik SMA, 2) menguji efektivitas instrumen penilaian berbasis komputer yang dikembangkan untuk mengukur pengetahuan Fisika peserta didik SMA, 3) mengukur tingkat kesiapan peserta didik dalam menghadapi ujian nasional berbasis komputer. Penelitian ini menggunakan model *Research and Development (R&D)* yang terdiri atas tiga tahap utama yaitu: tahap pendahuluan, tahap pengembangan dan tahap evaluasi. Pada tahap pendahuluan telah diidentifikasi permasalahan penilain pembelajaran fisika berbasis komputer. Tahap pengembangan telah dilakukan pembuatan desain instrumen penilaian fisika berbasis komputer, validasi produk, revisi produk dan uji coba produk secara terbatas. Pada tahap evaluasi telah dilakukan uji coba secara luas, revisi produk, dan pembuatan produk akhir. Uji statistik deskriptif digunakan untuk mendeskripsikan hasil aplikasi penilaian pengetahuan berbasis komputer. Hasil penelitian ini: (1) Instrumen soal yang dikembangkan layak untuk mengukur kemampuan kognitif peserta didik SMA, (2) instrumen penilaian berbasis komputer lebih efektif dan efisien untuk mengukur kemampuan pengetahuan Fisika peserta didik SMA daripada *paper and pencil test*, (3) 58,67% peserta didik sudah siap menghadapi tes berbasis komputer, 34% peserta didik belum siap melakukan tes berbasis komputer, dan 7,33% tidak siap melakukan tes berbasis komputer.

Kata Kunci: Penilaian berbasis Komputer, Fisika

The Development of Computer-Based Assessment of Physics Science to Improve Readiness of Students in Dealing Computer-Based National Examanation

Abstract

This research aims to 1) produce computer-based assessment instruments eligible to measure the cognitive abilities of senior high school students, 2) test the effectiveness of computer-based assessment instrument developed to measure the knowledge of physics in senior high school students, 3) measuring the level of readiness of students in dealing computer-based national examanation. This research uses a model of Research and Development (R & D) which consists of three main stages: the preliminary stage, the stage of development and evaluation phase. At the preliminary stage has been identified problems of computer-based assessment of learning physics. At the development stage has been carried out the design manufacture of computer-based assessment instrument physics, product validation, product revisions and testing of products on a limited basis. In the evaluation stage has been tested extensively, product revision, and the final product. Descriptive statistical test used to describe the results of the application of computer-based knowledge assessment. The results research show that : 1) computer-based assessment instruments eligible to measure cognitive abilities of senior high school students, 2) computer-based assessment instruments are more effective and efficient to measure the students knowledge of physics in senior high school rather than paper and pencil test, 3) 58.67% of students are ready to face computer-based test, 34% of students not yet ready for the computer based test and 7.33% are not ready to perform the test based computer.

Keywords: Computer-Based Assessment, Physics

How to Cite: suyoso, s., Istiyono, E., & Subroto, S. (2017). Pengembangan instrumen asesmen pengetahuan fisika berbasis komputer untuk meningkatkan kesiapan peserta didik dalam menghadapi ujian nasional berbasis komputer. *Jurnal Pendidikan Matematika dan Sains*, 5(1). doi:<http://dx.doi.org/10.21831/jpms.v5i1.12461>

Permalink/DOI: DOI: <http://dx.doi.org/10.21831/jpms.v5i1.12461>

PENDAHULUAN

Pengukuran dalam bidang pendidikan mencakup beberapa ranah, diantaranya: pengukuran ranah kognitif (pengetahuan), ranah afektif (sikap) dan ranah psikomotor (keterampilan). Biasanya, pengukuran pada ranah kognitif dilakukan dengan tes. Ranah afektif diukur dengan kuesioner, wawancara dan pengamatan, sedangkan ranah psikomotor diukur melalui pengamatan.

Dalam pelaksanaan pendidikan banyak keputusan yang harus dibuat oleh seorang guru antara lain yang menyangkut proses pembelajaran, hasil belajar dan seleksi bimbingan (Sudaryono, 2012, p.36). Untuk penelitian dengan objek manusia yang di dalamnya mencakup unsur variabel yang terkait dengan manusia, tes merupakan alat ukur yang sering ditemui di bidang penelitian pendidikan, psikologi maupun sosiologi. Dengan tes, seorang peneliti dapat mengukur konstruk yang diinginkan. Melalui indikator yang dipilih, seorang peneliti kemudian dapat mengidentifikasi konstruk yang hendak diukur (Sukardi, 2008, p.138). Indikator-indikator tersebut digunakan sebagai pedoman penyusunan instrumen penilaian.

Dalam penelitian pendidikan yang berkaitan dengan metode, pendekatan, strategi pembelajaran, dan kegiatan yang berkaitan dengan proses belajar mengajar sering direfleksikan sebagai variabel bebas, sedangkan yang merupakan variabel terikat di antaranya adalah pencapaian hasil belajar dan efektivitas program. Tes prestasi memiliki validitas dan reliabilitas yang tinggi sehingga sangat baik untuk mengukur kemampuan kognitif peserta didik (Sparrow & Davis, 2000, p.118). Oleh karena itu, wajar apabila tes prestasi banyak digunakan dalam penelitian maupun dalam pembahasan yang berkaitan erat dengan proses belajar mengajar.

Dari hasil wawancara yang dilakukan dengan beberapa guru fisika Sekolah Menengah Atas (SMA) di Daerah Istimewa Yogyakarta dan sebagian di daerah Provinsi Jawa Tengah, sekiranya ini instrumen penilaian yang biasa digunakan berupa tes objektif pilihan majemuk atau tes uraian yang dikerjakan di kertas. Hal ini

disebabkan karena adanya kecenderungan mempertahankan kebiasaan lama, menulis di kertas. Guru belum tergerak untuk menerapkan penilaian berbantuan komputer atau *computerized based test*.

Computerized based test merupakan sistem ujian dengan memanfaatkan teknologi komputer sebagai media tes. Hernawati (2007, p.3) menyatakan bahwa *computerized based test* memiliki beberapa kelebihan, diantaranya: mengurangi waktu untuk pekerjaan penilaian tes dan membuat laporan tertulis, menghilangkan pekerjaan logistik seperti mendistribusikan, menyimpan dan tes menggunakan kertas. Selain itu hasil penilaian dapat langsung dilihat setelah penilaian dilakukan sehingga mengurangi beban guru dalam mengoreksi jawaban peserta didik. Selain itu, berdasarkan penelitian oleh Özden, Erturk, & Sanli (2004) peserta didik memberikan tanggapan yang positif terhadap tes berbasis komputer. Selain itu, tes berbasis komputer juga akurat untuk mengukur peserta tes dengan kemampuan sedang sampai tinggi (Santoso, 2010)

Berdasarkan hasil observasi pada beberapa SMA di Daerah Istimewa Yogyakarta, sebagian besar sekolah sudah memiliki fasilitas yang layak untuk melakukan tes berbasis komputer. Hal ini dipicu adanya ujian nasional dilakukan berbasis komputer. Walau pada akhirnya belum diimplementasikan secara menyeluruh, akan tetapi sebagian besar sekolah sudah memiliki sarana dan prasarana yang memadai untuk melakukan tes berbasis komputer.

Berdasarkan uraian tersebut, perlu dikembangkan instrumen penilaian berbasis komputer untuk meningkatkan kesiapan peserta didik dalam menghadapi ujian nasional berbasis komputer. Dengan demikian tujuan penelitian ini adalah untuk 1) menghasilkan instrumen penilaian berbasis komputer yang layak untuk mengukur kemampuan kognitif peserta didik SMA, 2) menguji efektivitas instrumen penilaian berbasis komputer yang dikembangkan untuk mengukur pengetahuan Fisika peserta didik SMA, dan 3) mengukur tingkat kesiapan peserta didik dalam menghadapi ujian nasional berbasis komputer.

Dengan adanya penelitian ini diharapkan peserta didik lebih mengenal berbagai macam penilaian berbasis komputer sehingga ketika menghadapi ujian berbasis komputer sudah tidak canggung lagi.

METODE

Jenis Penelitian

Jenis Penelitian ini menggunakan desain *Research and Development (R&D)*. Metode penelitian dan pengembangan adalah metode penelitian yang digunakan untuk menghasilkan produk tertentu dan menguji keefektifan produk tersebut (Sugiyono, 2011, p.333). Model *R&D* terdiri atas tiga tahap utama yaitu: tahap pendahuluan, tahap pengembangan dan tahap evaluasi.

Tempat, Waktu dan Subyek Penelitian

Penelitian ini dilaksanakan di SMAN 4 Yogyakarta, SMAN 5 Yogyakarta, SMAN 6 Yogyakarta, dan SMAN 11 Yogyakarta. Penelitian dilakukan pada bulan Mei sampai Juni 2016. Subyek penelitian ini adalah peserta didik kelas XI SMA yang telah menerima materi teori kinetik gas. Pada subjek uji coba terbatas dilakukan di SMAN 6 Yogyakarta dengan jumlah sekitar 100 orang peserta didik. Subyek uji coba skala luas dilakukan di SMAN 4 Yogyakarta, SMAN 5 Yogyakarta, SMAN 6 Yogyakarta, dan SMAN 11 Yogyakarta dengan jumlah subyek sekitar 300 orang peserta didik.

Instrumen Penelitian

Kisi-kisi Soal

Kisi-kisi soal disusun berdasarkan pada Kurikulum 2013, pada mata pelajaran fisika SMA Kelas XI tahun pelajaran 2015/2016. Pokok bahasan fluida dan teori kinetik gas semester genap yang mengarah pada kemampuan kognitif peserta didik. Soal berbentuk pilihan ganda yang terdiri dari dua paket dengan delapan *anchor item* (soal yang sama pada kedua paket). Soal pilihan ganda terkemas dalam aplikasi *Flash* untuk materi fluida dan *Quizstar* untuk materi teori kinetik gas.

Lembar Validasi Instrumen

Lembar validasi digunakan untuk melakukan penilaian oleh dosen ahli terhadap instrumen penilaian yang dikembangkan. Saran yang diberikan oleh validator menjadi dasar dilakukan

perbaikan pada instrumen penilaian yang dikembangkan.

Angket Respon Peserta Didik dan Guru

Angket ini merupakan instrumen yang berisi pendapat para pengguna, yakni guru dan peserta didik yang bersangkutan. Angket ini bertujuan untuk mengetahui respon pengguna terhadap aplikasi *Flash* dan *Quizstar*.

Teknik Analisis Data

Validitas Isi Soal

Validitas isi butir soal dapat ditunjukkan dengan statistik yang diusulkan oleh Aiken (1985) dalam Syaifuddin Azwar (2015, pp. 112-114). Formula *V* Aiken dirumuskan sebagai berikut:

$$V = \frac{\sum s}{[n(c - 1)]} \quad (1)$$

Keterangan:

V = koefisien validitas isi

n = banyak penilai

s = *r* - *lo*

r = angka yang diberikan seorang penilai

lo = angka penilaian validitas terendah

c = angka penilaian validitas tertinggi

Besar nilai *V* menunjukkan nilai koefisien validitas instrumen yang diukur. Butir soal dikatakan memiliki validitas isi yang baik dan mendukung validitas isi tes secara keseluruhan apabila nilai *V* lebih dari 0,5.

Validitas Soal

Kecocokan Butir Soal dengan Model PCM

Penetapan *fit* butir soal secara keseluruhan dan *fit* peserta didik secara keseluruhan dikembangkan Adam & Khoo (1996, p. 28) berdasarkan nilai rata-rata INFIT MNSQ (INFIT *Mean of Square*) beserta simpangan bakunya. Jika rata-rata INFIT MNSQ sekitar 1 dan simpangan bakunya 0, maka keseluruhan tes *fit* dengan model.

Penetapan *fit* tiap butir soal mengikuti kaidah bahwa *Item characteristic curve* (ICC) akan mendatar (*flat*) bila besarnya INFIT MNSQ untuk butir soal atau *e* lebih besar dari satuan logit > 1,30 atau < 0,77. Oleh karena itu, dalam program *Quest* suatu butir soal atau peserta didik *fit* terhadap model jika nilai INFIT MNSQ antara 0,77 sampai 1,30.

Reliabilitas

Estimasi reliabilitas menurut IRT (*Item Response Theory*) dihitung berdasarkan peserta

didik disebut dengan indeks sparasi person. Semakin tinggi indeks sparasi *person* semakin konsisten setiap butir soal pengukur digunakan untuk mengukur peserta didik yang bersangkutan sehingga disebut reliabilitas tes (Subali, 2011, p. 11). Adapun klasifikasi reliabilitas soal disajikan dalam Tabel 1.

Tabel 1. Tingkat Reliabilitas Berdasarkan Nilai Alpha

Alpha	Tingkat Reliabilitas
0,00 s.d. 0,19	Kurang Reliabel
0,20 s.d. 0,39	Agak Reliabel
0,40 s.d. 0,59	Cukup Reliabel
0,60 s.d. 0,79	Reliabel
0,80 s.d. 1,00	Sangat Reliabel

Tingkat Kesukaran Butir Soal

Tingkat kesukaran butir soal diperoleh dari interpretasi program *Quest* pada nilai *threshold* di lihat dari *item estimates*. Menurut Hambleton & Swaminathan (1985, p. 36), tingkat kesukaran bernilai baik jika berada pada rentang -2 sampai +2.

Fungsi Informasi dan SEM

Untuk mendapatkan fungsi informasi dan SEM digunakan program Bilog. Berdasarkan fungsi informasi dan SEM, maka tes cocok untuk peserta didik dengan rentang kemampuan (θ) tertentu.

Kelayakan Media Penilaian Berbasis Quizstar

Menghitung nilai *Content Validity Ratio* (CVR)

$$CVR = \frac{(N_e - \frac{N}{2})}{\frac{N}{2}} \quad (2)$$

(Lawshe, 1975, p. 567)

N_e = jumlah validator yang menyetujui
 N = jumlah total validator

Menghitung nilai *Content Validity Index* (CVI)

$$CVI = \frac{\text{jumlah CVR}}{\text{jumlah butir angket}} \quad (3)$$

Rentang hasil nilai CVR dan CVI adalah $-1 < 0 < 1$. Angka tersebut dikategorikan sebagai berikut.

- $-1 < x < 0$ = tidak baik
- 0 = baik
- $0 < x < 1$ = sangat baik (Lawshe, 1975)

Reliabilitas instrumen media ditentukan dengan mencari *Interclass Correlation Coefficient* (ICC). Nilai ICC dianalisis dengan SPSS versi 23. Fleiss (Craven & Morris, 2010, p. 210) mengategorikan tingkat reliabilitas menjadi beberapa kategori yang disajikan pada Tabel 2.

Tabel 2. Nilai ICC dan Klasifikasi

Nilai ICC	Klasifikasi
<0,40	Kurang Reliabel
0,40 – 0,75	Reliabel
>0,75	Sangat Reliabel

Angket Respon Peserta Didik

Data berupa nilai yang terkumpul dari angket respon peserta didik dikonversikan menjadi data kuantitatif. Menurut Azwar (1998: 163) lima level kemampuan memiliki rentang seperti dinyatakan pada Tabel 3

Tabel 3. Rentang Lima Level Kemampuan

Interval Kemampuan	Level
$M_i + 1,5SB_i < \theta$	Sangat Tinggi
$M_i + 0,5SB_i < \theta \leq M_i + 1,5SB_i$	Tinggi
$M_i - 0,5SB_i < \theta \leq M_i + 0,5SB_i$	Sedang
$M_i - 1,5SB_i < \theta \leq M_i - 0,5SB_i$	Rendah
$\theta < M_i - 1,5SB_i$	Sangat Rendah

HASIL DAN PEMBAHASAN

Hasil Penelitian

Hasil penelitian meliputi hasil estimasi butir soal dan testi baik pada uji coba terbatas maupun hasil uji coba skala luas. Hasil estimasi tersebut digunakan untuk mendeskripsikan kualiti-tas dan kelayakan perangkat tes yang dikembangkan. Hasil estimasi butir soal dan testi pada uji coba terbatas dan skala luas untuk perangkat tes berbasis *Flash* dan *Quizstar* dapat dilihat pada Tabel 4 sampai Tabel 7.

Tabel 4. Hasil Estimasi untuk Butir Soal dan Testi menurut Model *PCM* Hasil Uji Coba Terbatas berbasis *Flash*

Uraian	Estimasi untuk item	Estimasi untuk testi
Nilai rata-rata dan simpangan baku	0,00 ± 0,99	-0,89 ± 0,43
Nilai rata-rata dan simpangan baku yang sudah disesuaikan	0,00 ± 0,96	-0,89 ± 0,32
Reliabilitas	0,94	0,56
Nilai rata-rata	1,00 ± 0,04	1,00 ± 0,11

Uraian	Estimasi untuk item	Estimasi untuk testi
dan simpangan baku INFIT MNSQ		
Nilai rata-rata dan simpangan baku OUTFIT MNSQ	$1,00 \pm 0,30$	$0,99 \pm 0,18$
Nilai rata-rata dan simpangan baku INFIT t	$0,04 \pm 0,39$	$0,04 \pm 0,54$
Nilai rata-rata dan simpangan baku OUTFIT t	$0,99 \pm 0,08$	$0,00 \pm 0,74$

Tabel 5. Hasil Estimasi untuk Butir Soal dan Testi menurut Model *PCM* Hasil Uji Coba Terbatas berbasis *Quizstar*

Uraian	Estimasi untuk item	Estimasi untuk item
Nilai rata-rata dan simpangan baku	$0,00 \pm 0,86$	$-0,76 \pm 0,52$
Nilai rata-rata dan simpangan baku yang sudah disesuaikan	$0,00 \pm 0,83$	$-0,76 \pm 0,45$
Reliabilitas	0,93	0,71
Nilai rata-rata dan simpangan baku INFIT MNSQ	$1,00 \pm 0,04$	$1,00 \pm 0,10$
Nilai rata-rata dan simpangan baku OUTFIT MNSQ	$1,00 \pm 0,07$	$1,00 \pm 0,22$
Nilai rata-rata dan simpangan baku INFIT t	$0,12 \pm 0,60$	$0,01 \pm 0,99$
Nilai rata-rata dan simpangan baku OUTFIT t	$0,05 \pm 0,49$	$-0,03 \pm 0,83$

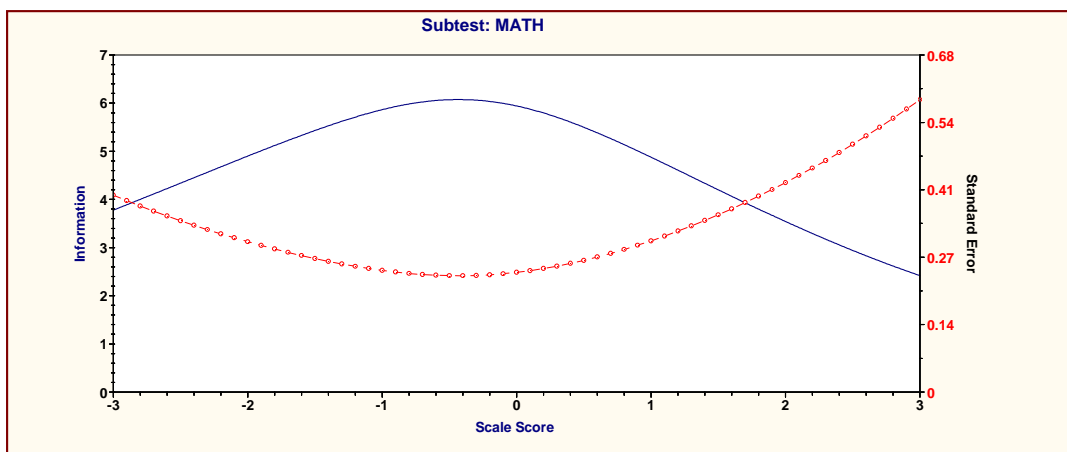
Tabel 6. Hasil Estimasi untuk Butir Soal dan Testi menurut Model *PCM* Hasil Uji Coba Luas berbasis *Flash*

Uraian	Estimasi untuk item	Estimasi untuk testi
Nilai rata-rata dan simpangan baku	$0,00 \pm 0,98$	$-0,89 \pm 0,41$

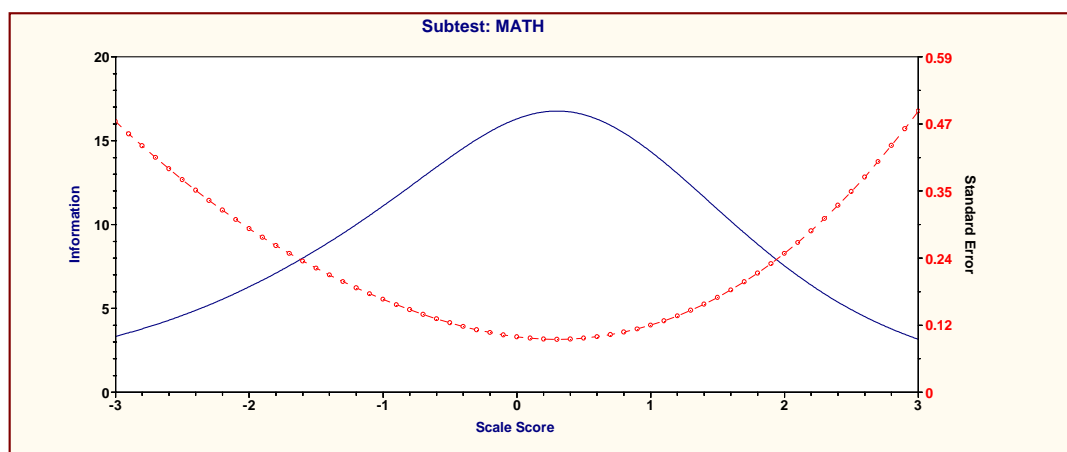
Uraian	Estimasi untuk item	Estimasi untuk testi
Nilai rata-rata dan simpangan baku yang sudah disesuaikan	$0,00 \pm 0,95$	$-0,89 \pm 0,30$
Reliabilitas	0,94	0,53
Nilai rata-rata dan simpangan baku INFIT MNSQ	$1,00 \pm 0,04$	$1,00 \pm 0,12$
Nilai rata-rata dan simpangan baku OUTFIT MNSQ	$1,00 \pm 0,07$	$1,00 \pm 0,21$
Nilai rata-rata dan simpangan baku INFIT t	$0,08 \pm 0,48$	$0,02 \pm 0,93$
Nilai rata-rata dan simpangan baku OUTFIT t	$0,04 \pm 0,40$	$0,00 \pm 0,85$

Tabel 7. Hasil Estimasi untuk Butir Soal dan Testi menurut Model *PCM* Hasil Uji Coba Luas berbasis *Quizstar*

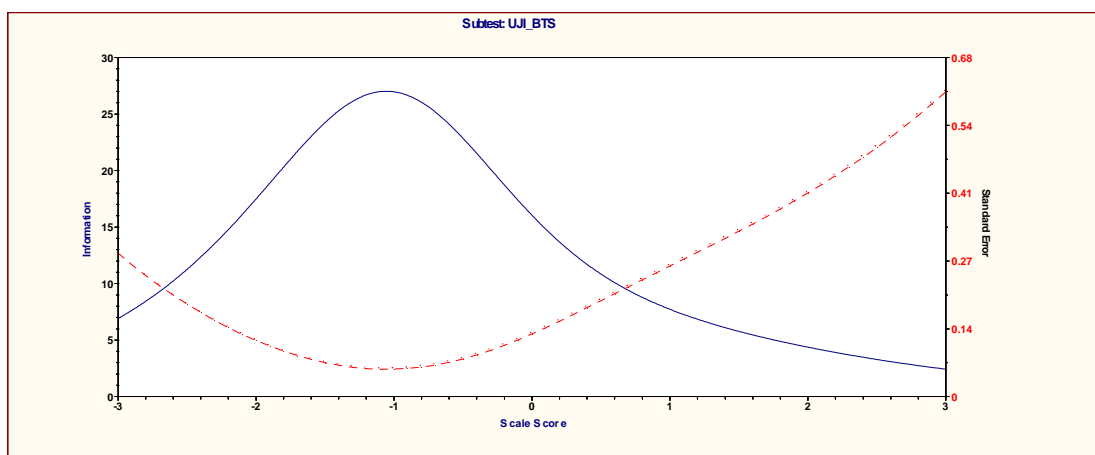
Uraian	Estimasi untuk item	Estimasi untuk testi
Nilai rata-rata dan simpangan baku	$0,00 \pm 0,72$	$-0,95 \pm 0,52$
Nilai rata-rata dan simpangan baku yang sudah disesuaikan	$0,00 \pm 0,71$	$-0,95 \pm 0,44$
Reliabilitas	0,96	0,73
Nilai rata-rata dan simpangan baku INFIT MNSQ	$1,00 \pm 0,04$	$1,00 \pm 0,07$
Nilai rata-rata dan simpangan baku OUTFIT MNSQ	$1,00 \pm 0,10$	$1,00 \pm 0,15$
Nilai rata-rata dan simpangan baku INFIT t	$0,02 \pm 0,86$	$0,02 \pm 0,54$
Nilai rata-rata dan simpangan baku OUTFIT t	$0,03 \pm 0,79$	$0,05 \pm 0,60$



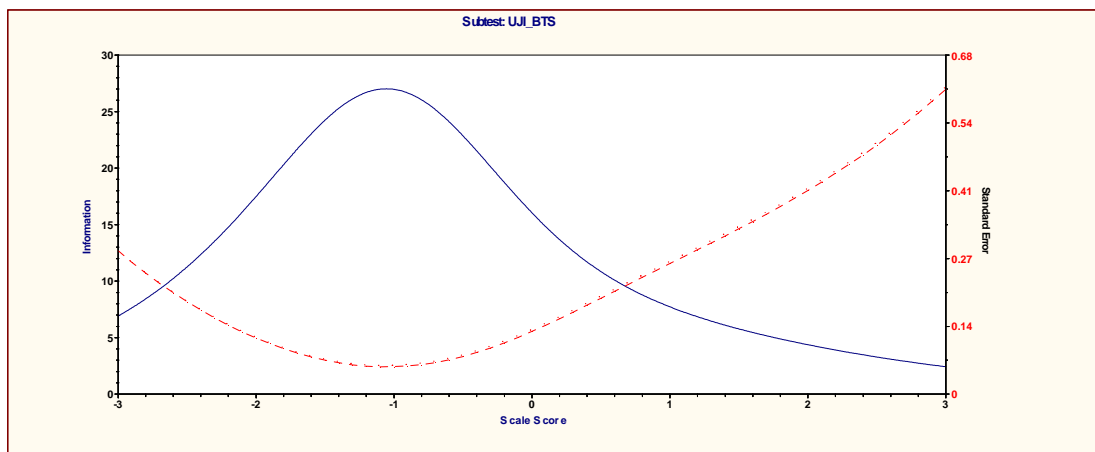
Gambar 1. Fungsi Informasi dan SEM pada Uji Coba Terbatas Berbasis *Flash*



Gambar 2. Fungsi Informasi dan SEM pada Uji Coba Luash Berbasis *Flash*



Gambar 3. Fungsi Informasi dan SEM pada Uji Coba Terbatas Berbasis *Quizstar*



Gambar 4. Fungsi Informasi dan SEM pada Uji Coba Luas Berbasis Quizstar

Pembahasan

Analisis Soal pada materi Fluida

Analisis uji terbatas menggunakan program *Quest* dan *Bilog*. Program *Quest* digunakan untuk mencari nilai *INFIT MNSQ*, nilai ini merupakan informasi diterima atau ditolaknya butir soal yang dikembangkan. Berdasarkan nilai *INFIT MNSQ* butir-butir yang dikembangkan terletak pada nilai diantara 0,98 sampai dengan 1,05. Butir dikatakan baik jika nilai *INFIT MNSQ* terletak antara 0,77 sampai dengan 1,30 (Adam & Khoo, 1996: 28), dengan demikian semua butir berdasarkan kriteria ini *fit* atau dapat diterima berdasarkan nilai *INFIT MNSQ* pada output program *Quest*.

Program *Bilog* digunakan untuk mencari grafik fungsi informasi dan *standard error measurement* (SEM). Berdasarkan fungsi informasi dan SEM yang dinyatakan dengan Gambar 1 instrumen yang dikembangkan cocok untuk peserta didik dengan tingkat kemampuan (θ) antara -2,7 sampai dengan 1,7.

Berdasarkan analisis hasil uji coba luas menggunakan program *Quest*, *difficulty* butir-butir terletak antara -2,80 sampai 1,32. Tingkat kesukaran (b) di luar batas $-2,0 \leq b \leq 2,0$ adalah butir kurang baik menurut Hambleton & Swaminathan (1991, p. 34). Oleh karena itu dilakukan eliminasi butir soal yang memiliki tingkat kesukaran kurang dari -2,0. Butir-butir yang memiliki nilai tingkat kesukaran kurang dari -2,0 adalah butir nomor 35, 38 dan 39 atau butir nomor 3, 25 dan 32 baik soal paket A ataupun paket B. Tiga butir soal tersebut merupakan *anchor item*. Berdasarkan nilai *INFIT MNSQ* butir-butir yang dikembangkan terletak pada nilai diantara 0,87 sampai dengan 1,12. Butir dikatakan baik jika nilai *INFIT MNSQ* terletak antara 0,77 sampai dengan 1,30

(Adam & Khoo, 1996, p. 28), dengan demikian semua butir berdasarkan kriteria ini *fit*.

Program *Bilog* digunakan untuk mencari grafik fungsi informasi dan *standard error measurement* (SEM). Berdasarkan fungsi informasi dan SEM yang dinyatakan dengan Gambar 2 Instrumen yang dikembangkan cocok untuk peserta didik dengan tingkat kemampuan (θ) antara -1,6 sampai dengan 2,0.

Analisis Soal pada materi Teori Kinetik Gas

Analisis uji terbatas menggunakan program *Quest* dan *Bilog*. Program *Quest* digunakan untuk mencari nilai *INFIT MNSQ*. Berdasarkan nilai *INFIT MNSQ* butir-butir yang dikembangkan terletak pada nilai diantara 0,90 sampai dengan 1,08. Butir dikatakan baik jika nilai *INFIT MNSQ* terletak antara 0,77 sampai dengan 1,30 (Adam & Khoo, 1996, p. 28), dengan demikian semua butir berdasarkan kriteria ini *fit*.

Program *Bilog* digunakan untuk mencari grafik fungsi informasi dan *standard error measurement* (SEM). Berdasarkan fungsi informasi dan SEM yang dinyatakan dengan Gambar 3. Instrumen yang dikembangkan cocok untuk peserta didik dengan tingkat kemampuan (θ) antara -2,8 sampai dengan 1,7.

Berdasarkan analisis hasil uji coba luas menggunakan program *Quest*, *difficulty* butir-butir terletak antara -1,99 sampai 1,83. Tingkat kesukaran (b) di luar batas $-2,0 \leq b \leq 2,0$ adalah butir kurang baik menurut Hambleton & Swaminathan (1991, p. 34). Berdasarkan nilai *INFIT MNSQ* butir-butir yang dikembangkan terletak pada nilai diantara 0,93 sampai dengan 1,09. Butir dikatakan baik jika nilai *INFIT MNSQ* terletak antara 0,77 sampai dengan 1,30

(Adam & Khoo, 1996, p. 28), dengan demikian semua butir berdasarkan kriteria ini *fit*.

Program *Bilog* digunakan untuk mencari grafik fungsi informasi dan *standard error measurement* (SEM). Berdasarkan fungsi informasi dan SEM yang dinyatakan dengan Gambar 4. Instrumen yang dikembangkan cocok untuk peserta didik dengan tingkat kemampuan (θ) antara -2,3 sampai dengan 1,1. Rentang kemampuan peserta didik tersebut termasuk dalam kategori sedang.

Penilaian Media

Aplikasi penilaian berbasis *flash* dan *Quizstar* berdasarkan penilaian dari guru mendapatkan nilai CVI sebesar 1,00 yang termasuk kategori sangat baik. Aplikasi berdasarkan analisis persentase kelayakan dari angket respon peserta didik didapatkan kategori sangat baik, baik, kurang baik dan tidak baik berturut-turut 33,94%, 63,82%, 6,79%, dan 0,41%.

Kesiapan Peserta Didik Menghadapi Tes Berbasis Komputer

Penetapan *fit* testi secara keseluruhan juga dapat dilihat dari rata-rata nilai *infitMNSQ* beserta simpangan bakunya (Adam & Kho, 1996). Pada analisis uji terbatas nilai rata-rata *infitMNSQ* sebesar 1,00 dengan standar deviasi sebesar 0,12 untuk *macromedia* dan rata-rata 1,00 dengan 0,07 untuk *Quizstar*. Sedangkan pada uji luas diperoleh nilai rata-rata *infitMNSQ* sebesar 1,00 dengan standar deviasi sebesar 0,13 untuk *macromedia* dan rata-rata 1,00 dengan standar deviasi 0,10 untuk *Quizstar*.

Hasil analisis nilai *infitMNSQ* untuk testi merepresentasikan bahwa peserta didik lebih siap mengerjakan tes berbasis *paper and pencil test* daripada tes berbasis komputer karena standar deviasi pada uji luas (berbasis komputer) lebih besar daripada uji terbatas (*paper and pencil test*). Berdasarkan hasil tersebut juga diketahui peningkatan kesiapan peserta didik dalam melakukan tes berbasis komputer dilihat dari penurunan standar deviasi tes berbasis *Quizstar* (dilakukan setelah tes berbasis *macromedia*) jika dibandingkan dengan tes berbasis *macromedia*. Selain itu berdasarkan angket peserta didik 34% peserta didik belum siap melakukan tes berbasis komputer dan 7,33% tidak siap melakukan tes berbasis komputer. Dengan adanya pengembangan instrumen berbasis komputer peserta didik mengenal berbagai macam bentuk tes berbasis komputer

dan meningkatkan kesiapan meng-hadapi tes berbasis komputer.

SIMPULAN

Berdasarkan hasil penelitian dan analisis terhadap temuan-temuan selama penelitian, maka diperoleh kesimpulan yaitu (1) instrumen soal yang dikembangkan layak untuk mengukur kemampuan kognitif karena telah memenuhi validitas isi dengan *expert judgment* dan telah mendapatkan bukti empiris *fit* dengan model *PCM*; (2) aplikasi penilaian berbasis komputer lebih efektif dan efisien dibandingkan dengan penilaian berbasis *paper and pencil test*; dan (3) 58,67% peserta didik sudah siap menghadapi tes berbantuan komputer, 34% peserta didik belum siap melakukan tes berbasis komputer dan 7,33% tidak siap melakukan tes berbasis komputer.

Kendala utama yang dihadapi dalam pelaksanaan penilaian berbasis komputer yaitu keterbatasan fasilitas penunjang seperti komputer, maka dalam hal ini pemerintah hendaknya dapat memberikan bantuan untuk memenuhi kebutuhan tersebut, agar proses penilaian berbasis komputer ini dapat dilakukan secara merata. Untuk penelitian di masa yang akan datang, hendaknya dilaksanakan penelitian serupa pada mata pelajaran lainnya dan pada setiap jenjang, sehingga kedepannya perangkat tes berbasis komputer akan semakin banyak tersedia dan dapat mempermudah guru untuk melaksanakan penilaian berbasis komputer pada mata pelajaran dan jenjang lainnya.

DAFTAR PUSTAKA

- Azwar, S. (2015). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Adams, R. J., & Khoo, S. T. (1996). *Quest: The interactive test analysis system version 2.1*. Victoria: The Australian Council for Educational Research
- Craven, B. C., & Morris, A. R. (2010). Modified ashworth scale reliability for measurement of lower extremity spasticity among patients with SCI. *Spinal Cord*, 48, 207-213.
- Sudaryono. (2012). *Dasar-dasar evaluasi pembelajaran*. Yogyakarta: Graha Ilmu
- Hambleton, R. K., & Swaminathan, H. (1985). *Item respon theory: Principles and*

- Applications*. Boston, MA: Kluwer Nijhoff Publishing
- Hernawati, K. (2007). *Evaluasi dan penilaian interaktif berbasis web*. Skripsi tidak diterbitkan. Yogyakarta: FMIPA Universitas Negeri Yogyakarta.
- Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- Özden, M. Y., Erturk, I., & Sanli, R. (2004). Students' perception of online assessment: A case study. *Journal of Distance Education*, 19, 77-93.
- Sparrow, S. S. & Davis, S. M. (2010). Recent advances in assessment of intelligence and cognition. *Journal of Children Psychology Psychiat*, 41 (1), 117-131.
- Santoso, A. (2010). *Pengembangan computerized adaptive testing* untuk mengukur hasil belajar mahasiswa Universitas Terbuka. *Jurnal Penelitian dan Evaluasi Pendidikan*, 14 (1), 62 – 83.
- Subali, B. (2011). *Panduan analisis data pengukuran pendidikan untuk memperoleh bukti empirik kesahihan menggunakan program Quest*. Yogyakarta: Lembaga Penelitian dan Pengabdian pada Masyarakat UNY.
- Sugiyono. (2011). *Metode penelitian administrasi dilengkapi dengan metode R&D*. Bandung: Alfabeta

PROFIL SINGKAT

Suyoso, lahir di Klaten, 10 Juni 1953, S1 Pendidikan Fisika IKIP Semarang (UNES) lulus tahun 1980, S2 Fisika UGM lulus tahun 1994, dosen Jurusan Pendidikan Fisika FMIPA UNY.