

PENGEMBANGAN INSTRUMEN PENILAIAN BERBASIS QUIZSTAR UNTUK MENGUKUR KEMAMPUAN KOGNITIF PESERTA DIDIK SMA PADA MATERI POKOK TEORI KINETIK GAS

DEVELOPING ASSESSMENT INSTRUMENT BASED QUIZSTAR IN THEORY OF KINETIC GAS TO MEASURE COGNITIVE ABILITIES SENIOR HIGH SCHOOL STUDENTS

Oleh: Hendra Oktavianta⁽¹⁾; Edi Istiyono⁽²⁾

hendra.rave@gmail.com; edi_istiyono_uny@yahoo.com, edi_istiyono@uny.ac.id

⁽¹⁾Mahasiswa Pendidikan Fisika; ⁽²⁾Dosen Jurusan Pendidikan Fisika

Abstrak

Penelitian ini bertujuan untuk mengetahui 1) kelayakan instrumen penilaian berbasis *Quizstar* yang dikembangkan untuk mengukur kemampuan kognitif peserta didik SMA, 2) penguasaan materi teori kinetik gas peserta didik. Penelitian ini merupakan penelitian pengembangan dengan model 4-D (*Four D Models*) menurut pendapat Thiagarajan yang terdiri atas tahap *define, design, develop, dan disseminate*. Hasil penelitian ini: (1) Instrumen penilaian berbasis *Quizstar* layak untuk mengukur kemampuan kognitif peserta didik dengan telah memenuhi validitas isi dengan *expert judgment* dan telah mendapatkan bukti empiris *fit* dengan model *Rasch*. Seluruh item dalam kriteria baik karena tingkat kesukaran berada pada rentang -1,65 sampai dengan 1,78. Instrumen penilaian berbasis *Quizstar* telah memiliki koefisien reliabilitas sebesar 0,73 yang termasuk kategori reliabel. Berdasarkan fungsi informasi, instrumen penilaian berbasis *Quizstar* sangat tepat digunakan untuk mengukur kemampuan kognitif peserta didik yang berkemampuan dari -2,3 sampai 1,1. Media penilaian berbasis *Quizstar* mendapatkan nilai CVI sebesar 1,00 yang termasuk kategori sangat baik dengan reliabilitas sebesar 0,868 yang termasuk kategori sangat reliabel. Berdasarkan analisis angket respon peserta didik didapatkan hasil pada kategori cukup, baik, dan sangat baik berturut – turut 11%, 52%, dan 36%. (2) Kemampuan peserta didik mendapatkan kategori kurang, cukup, baik, dan sangat baik berturut – turut 21%, 44%, 25%, dan 9%.

Kata Kunci: pengembangan instrumen, *Quizstar*, teori kinetik gas

Abstract

This research aims to determine 1) the feasibility assessment instrument based Quizstar for measuring cognitive abilities of high school students, 2) mastery of the kinetic theory of gases learners. This research was the development of the model 4-D (Four D Models) in the opinion of Thiagarajan consisting of stages define, design, develop, and disseminate. The results of this study are: (1) assessment instrument based Quizstar feasible for measuring cognitive ability of learners to have met the content validity by expert judgment and has gained empirical evidence fit the Rasch models. The items difficulty indexes were between -1.65 and 1.78, which means the items were good. Assessment instruments based Quizstar have had reliability coefficient of 0.73 that included reliable category. Based on the information function, assessment instruments very appropriately used to measure cognitive abilities of learners capable of -2.3 to 1.1. Assessment instruments based Quizstar get CVI value of 1.00 which included a very good category with the reliability of 0.868 which includes a very reliable category. The result analysis of learners questionnaire responses showed in enough, good, and very good categories respectively – were 11%, 52%, and 36%.

(2) The ability of learners get less, enough, good, and very good category respectively – were 21%, 44%, 25% 9%.

Keyword: developing assessment instrument, *Quizstar*, theory of kinetic gas

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.

Pengukuran kognitif diperlukan instrumen tes yang memiliki validitas dan reliabilitas. Instrumen tes yang memiliki validitas dan reliabilitas yang tinggi sangat baik untuk mengukur kemampuan kognitif peserta didik.

Dalam pelaksanaan pendidikan banyak keputusan yang harus dibuat oleh seorang guru antara lain yang menyangkut proses pembelajaran, hasil belajar dan seleksi bimbingan (Sudaryono, 2012:36). Untuk memantau proses, kemajuan, dan memperbaiki hasil belajar peserta didik, diperlukan penilaian. Penilaian dapat dilakukan secara lisan maupun tertulis. Secara garis besar ada dua bentuk soal tes tertulis, yaitu: memilih jawaban dan menyuplai jawaban. Soal tes tertulis yang jawabannya dengan memilih jawaban antara lain: pilihan ganda, dua pilihan (benar-salah, ya-tidak), menjodohkan, dan sebab-akibat.

Pada kenyataannya tes berbentuk pilihan ganda lebih banyak digunakan dari pada bentuk tes yang lain. Hal ini karena tes pilihan ganda memiliki kelebihan - kelebihan, antara lain: (1) materi yang diujikan dapat mencakup sebagian besar bahan pembelajaran, (2) hasil jawaban peserta didik dapat dikoreksi dengan mudah dan cepat, (3) jawaban setiap pertanyaan sudah pasti benar atau salah, sehingga penilaian objektif. Walaupun terdapat juga kelemahan pilihan ganda, yaitu: (1) kemungkinan peserta didik untuk melakukan tebakan jawaban cukup besar dan (2) proses berpikir peserta didik tidak dapat dilihat dengan nyata (Nana Sujana, 1990:49). Oleh karena itu, tes pilihan ganda banyak digunakan untuk tes-tes dalam skala besar, misalnya ujian nasional dan ujian masuk perguruan tinggi.

Tes berbentuk pilihan ganda yang baik memiliki jawaban yang salah (*distracters*) yang sesuai dipikirkan peserta didik. Selain itu, peserta didik juga melihat panjang jawaban (Adams dan Wieman, 2010:11). Oleh karena itu, panjang kalimat tiap *distracters* diusahakan sama.

Salah satu tes yang dipapakai untuk mengukur kemampuan kognitif adalah tes prestasi. Tes prestasi memiliki validitas dan reliabilitas yang tinggi sehingga sangat baik untuk mengukur kemampuan kognitif peserta didik (Sparrow dan Davis, 2000: 118), . Oleh karena itu, tes prestasi banyak digunakan dalam penelitian yang berkaitan dengan proses belajar mengajar dalam sistem pendidikan.

Berdasarkan hasil wawancara yang dilakukan pada beberapa guru di SMA di Daerah Istimewa Yogyakarta, instrumen penilaian yang biasa digunakan berupa tes pilihan ganda atau tes uraian yang dikerjakan di kertas (*paper based test*). Hal ini disebabkan karena adanya kecenderungan mempertahankan kebiasaan lama, menulis di kertas.

Paper Based Test merupakan sistem ujian dengan menggunakan kertas sebagai media tes. Samsul Hadi (2013:11) menyatakan bahwa *paper based test* memiliki kelemahan yaitu memerlukan waktu pengadministrasian yang lebih lama, memerlukan kertas yang cukup banyak, memerlukan ruang khusus untuk menyimpan data tes, memerlukan tenaga serta peralatan yang memadai untuk memindai dan memperoleh skor hasil tes. Salah satu alternatif yang dapat dilakukan untuk mengatasi permasalahan tersebut, adalah menerapkan ujian berbasis computer atau *computerized based test*.

Computerized Based Test merupakan sistem ujian dengan memanfaatkan teknologi komputer sebagai media test. Kuswari Hernawati (2007:3) 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.

Salah satu cara untuk membuat penilaian berbantuan komputer adalah memanfaatkan fasilitas di internet berupa *Contain Management System* yang telah diprogram dalam bentuk website diantaranya adalah *Quizstar* (<http://quizstar.4teachers.org/>), *Quibblo* (<http://www.quibblo.com>), *Proprofs* (<http://www.proprofs.com>), dan *ClassMarker* (<http://www.classmarker.com>). Pada penelitian ini *Contain Management System* yang digunakan adalah *Quizstar*. *Quizstar* dipilih karena memiliki beberapa keunggulan dibandingkan *Contain Management System* yang lain yaitu memiliki fitur yang cukup lengkap yang memudahkan dalam melakukan penilaian. Selain itu *Quizstar* tersedia secara gratis sehingga pengguna dalam hal ini guru tidak perlu menganggarkan biaya khusus. Seperti website pada umumnya *Quizstar* dapat diakses dengan komputer yang tersambung dengan jaringan internet.

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 permasalahan di atas, perlu dikembangkan instrumen penilaian untuk mengukur kemampuan kognitif peserta didik berbantuan komputer. Oleh karena itu, dilakukan penelitian yang berjudul “Pengembangan Instrumen Penilaian Berbasis *Quizstar* untuk Mengukur Kemampuan Kognitif Peserta Didik SMA pada Materi Pokok Teori Kinetik Gas”

METODE PENELITIAN

Jenis Penelitian

Penelitian ini merupakan penelitian pengembangan dengan model 4-D (*Four D Models*) menurut Thiagarajan (1975:5) yang terdiri dari tahap pendefinisian (*define*), tahap perancangan (*design*), tahap pengembangan (*develop*) dan tahap penyebaran (*disseminate*).

Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di SMA N 4 Yogyakarta, SMA N 5 Yogyakarta, SMA N 6 Yogyakarta, dan SMA N 11 Yogyakarta. Penelitian dilakukan pada bulan Mei sampai Juni 2016.

Subyek Penelitian

Subyek penelitian ini adalah peserta didik kelas XI SMA yang telah menerima materi teori kinetik gas. Pada subjek uji coba terbatas dilakukan di SMA N 6 Yogyakarta dengan jumlah 120 orang peserta didik. Subyek uji coba skala luas dilakukan di SMA N 4 Yogyakarta, SMA N 5 Yogyakarta, SMA N 6 Yogyakarta, dan SMA N 11 Yogyakarta dengan jumlah 318 orang peserta didik.

Intrumen Penelitian

1. Kisi-kisi Soal

Kisi-kisi soal disusun berdasarkan pada Kurikulum 2013, pada mata pelajaran fisika SMA Kelas XI tahun pelajaran 2015/2016. Pokok bahasan 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).

2. Soal Pilihan Ganda

Soal yang dikembangkan adalah soal objektif pilihan ganda yang terdiri dari dua paket. Soal dikemas dalam aplikasi *Quizstar*.

3. 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.

4. 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 *Quizstar*.

Teknik Analisis Data

1. Validitas Isi Soal

Validitas isi butir soal dapat ditunjukkan

dengan statistik yang diusulkan oleh Aiken (1985) dalam Syaifuddin Azwar (2015:112-114).

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 instrument yang diukur. Butir soal dikatakan memiliki validitas isi yang baik dan mendukung validitas isi tes secara keseluruhan apabila nilai V lebih dari 0,5.

2. Validitas Soal

a. Kecocokan Butir Soal dengan Model Rasch

Penetapan *fit* butir soal secara keseluruhan dan *fit* peserta didik secara keseluruhan dikembangkan Adam dan Khoo (1996:28) 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.

b. 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, 2011:11). 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

c. 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: 36), tingkat kesukaran bernilai baik jika berada pada rentang -2 sampai +2.

d. 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.

3. Kelayakan Media Penilaian Berbasis *Quizstar*

a. Menghitung nilai *Content Validity Ratio* (CVR)

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

(Lawshe, 1975: 567)

N_e = jumlah validator yang menyetujui

N = jumlah total validator

b. Menghitung nilai *Content Validity Index* (CVI)

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

c. Kategori Hasil Perhitungan CVR dan CVI

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)

4. Kelayakan Media Penilaian Berbasis *Quizstar*

Reliabilitas instrumen media ditentukan dengan mencari *Interclass Correlation Coefficient* (ICC). Nilai ICC dianalisis dengan SPSS versi 23. Fleiss (dalam B. C. Craven dan A. R. Morris, 2010:210) 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

5. Angket Respon Peserta Didik

Data berupa nilai yang terkumpul dari angket angket respon peserta didik dikonversikan menjadi data kuantitatif. Berdasarkan kriteria skala lima, diperoleh kriteria penilaian angket respon peserta didik yang disajikan pada Tabel 3 sebagai berikut.

Tabel 3. Kriteria Kelayakan Angket Respon Peserta Didik

No	Rentang Skor Kuantitatif	Kategori
1	$X > 3,4$	Sangat Tinggi
2	$2,8 < X \leq 3,4$	Tinggi
3	$2,22 < X \leq 2,8$	Cukup
4	$1,6 < X \leq 2,22$	Rendah
5	$X \leq 1,6$	Sangat Rendah

6. Kemampuan Peserta Didik

Data berupa nilai peserta didik dikonversikan menjadi data kuantitatif yang disajikan dalam bentuk persentase tiap kategori dengan skala penilaian dengan langkah- langkah seperti pada analisis angket respon peserta didik. Sebelum nilai peserta didik dikonversikan menjadi data kuantitatif, terlebih dahulu dilakukan penyetaraan nilai antara kelompok 1 (paket A) dan kelompok 2 (paket B). Menurut Livingston (2004:15) penyetaraan nilai antara kelompok 1 (paket A) dan kelompok 2 (paket B). dilakukan dengan menggunakan persamaan berikut:

$$Y = \left(\frac{SD(Y)}{SD(X)} \right) X + \left[mean(Y) - \left(\frac{SD(Y)}{SD(X)} \right) mean(X) \right] \quad (2)$$

Berdasarkan kriteria skala lima, diperoleh kriteria penilaian kemampuan kognitif peserta didik yang disajikan pada Tabel 6.

Tabel 6. Kriteria Kemampuan Kognitif Peserta Didik

No	Rentang Skor Kuantitatif	Kategori
1	$X > 80$	Sangat Tinggi
2	$60 < X \leq 80$	Tinggi
3	$40 < X \leq 60$	Cukup
4	$20 < X \leq 40$	Rendah
5	$X \leq 20$	Sangat Rendah

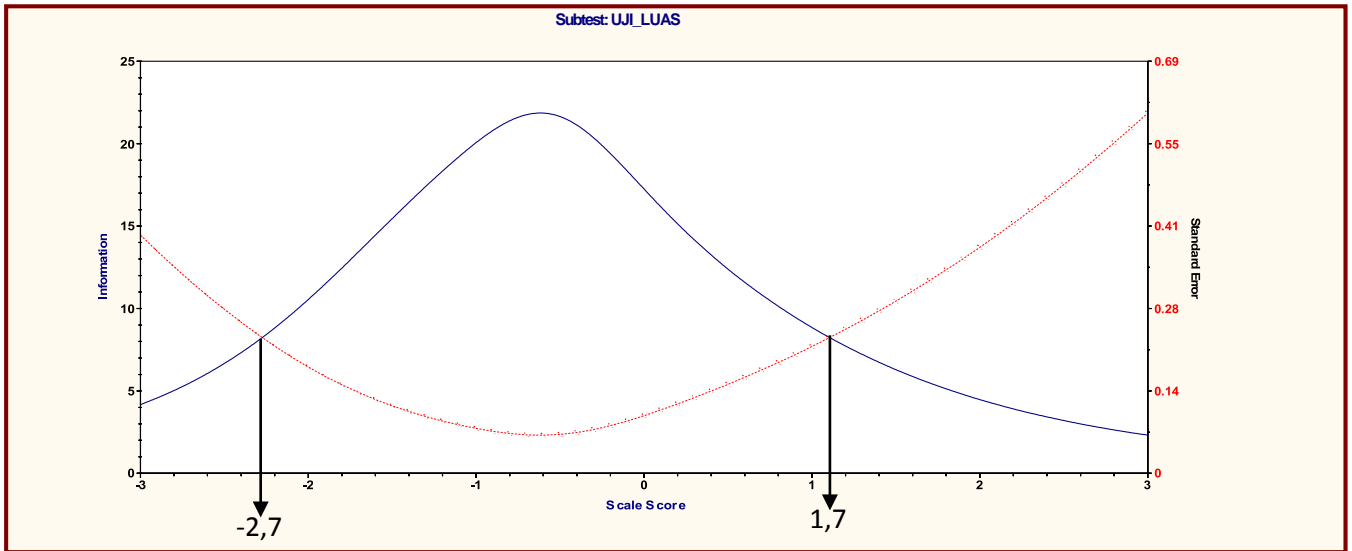
HASIL PENELITIAN DAN PEMBAHASAN

Tabel 7. Hasil Estimasi untuk Butir Soal dan Testi menurut Model Rasch Hasil Uji Coba Terbatas

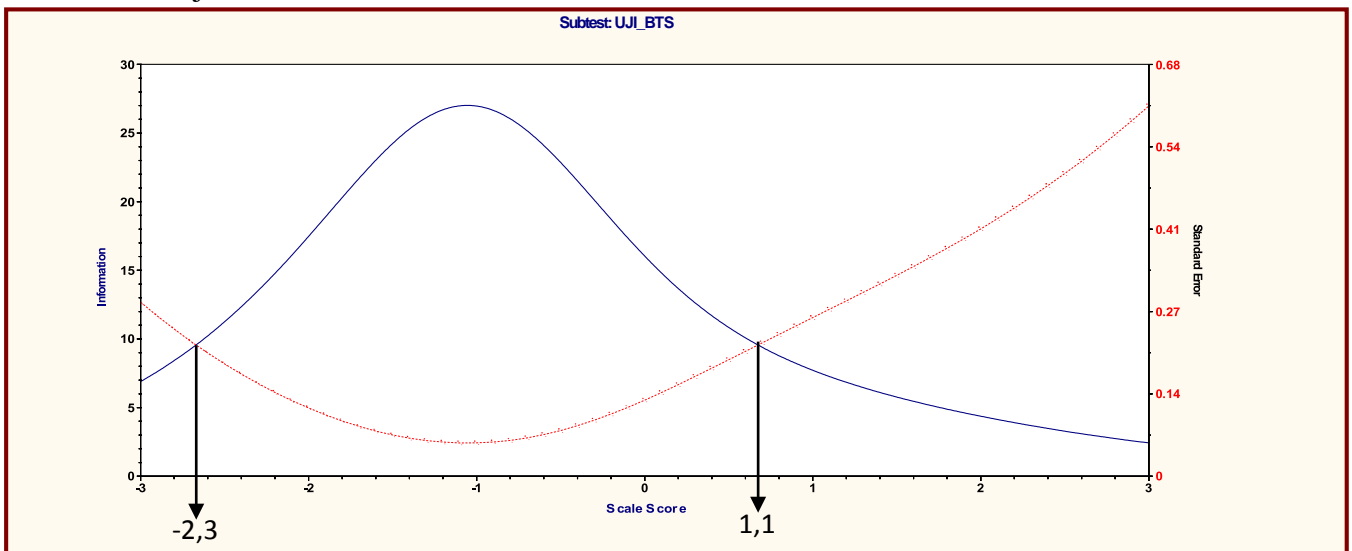
Uraian	Estimasi untuk item	Estimasi untuk testi
Nilai rata-rata dan simpangan baku	$0,00 \pm 0,86$	$-0,76 \pm 0,52$
Nilai rata-rata dan simpangan baku yang sudah disesuaikan	$0,00 \pm 0,83$	$-0,76 \pm 0,45$
Reliabilitas	0,93	0,71
Nilai rata-rata dan simpangan baku INFIT MNSQ	$1,00 \pm 0,04$	$1,00 \pm 0,10$
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 7. Hasil Estimasi untuk Butir Soal dan Testi menurut Model Rasch Hasil Uji Coba Luas

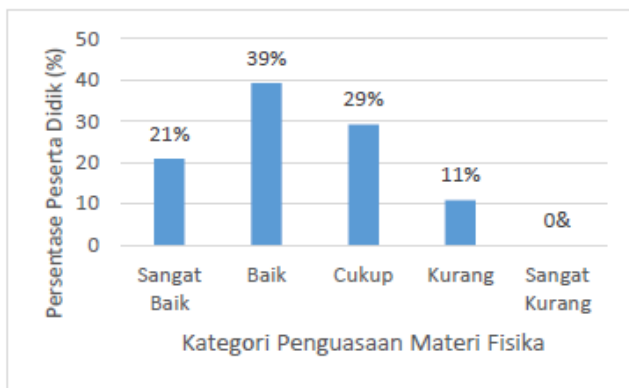
Uraian	Estimasi untuk item	Estimasi untuk testi
Nilai rata-rata dan simpangan baku	$0,00 \pm 0,72$	$-0,95 \pm 0,52$
Nilai rata-rata dan simpangan baku yang sudah disesuaikan	$0,00 \pm 0,71$	$-0,95 \pm 0,44$
Reliabilitas	0,96	0,73
Nilai rata-rata dan simpangan baku INFIT MNSQ	$1,00 \pm 0,04$	$1,00 \pm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



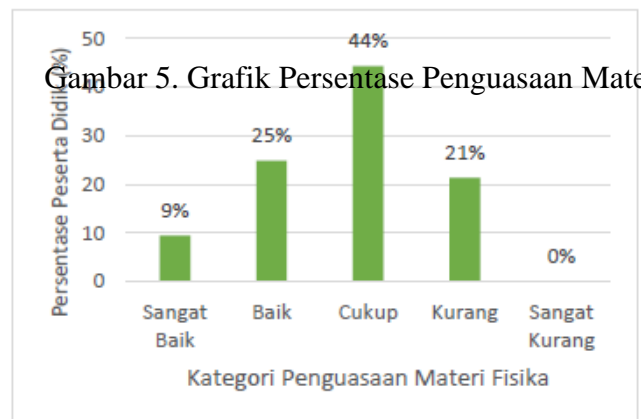
Gambar 1. Fungsi Informasi dan SEM pada Uji Coba Terbatas



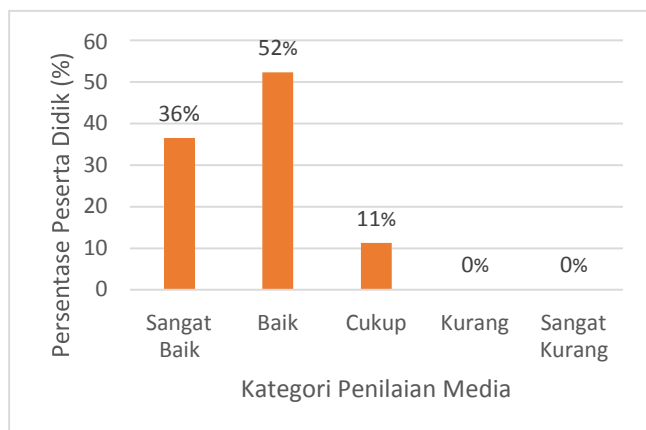
Gambar 2. Fungsi Informasi dan SEM pada Uji Coba Luas



Gambar 4. Grafik Persentase Penguasaan Materi Peserta Didik pada Uji Coba Terbatas



Gambar 5. Grafik Persentase Penguasaan Materi Peserta Didik pada Uji Coba Luas



Gambar 6. Grafik Persentase Angket Peserta Didik

Instrumen isi soal divalidasi isi untuk mengetahui kecocokan isi melalui *expert judgment* oleh ahli. Berdasarkan analisis diperoleh nilai koefisien validitas V Aiken terendah sebesar 0,78 dan tertinggi 1,00. Berdasarkan hasil tersebut keseluruhan butir yang dihasilkan sebagai produk awal dinyatakan valid dan dapat digunakan untuk uji coba terbatas.

Penetapan *fit* butir soal secara keseluruhan dikembangkan Adam dan Khoo (1996:28) berdasarkan nilai rata-rata INFIT MNSQ beserta simpangan. Jika rata-rata INFIT MNSQ sekitar 1 dan simpangan bakunya 0, atau rata-rata INFIT t mendekati 0 dan simpangan bakunya 1, maka keseluruhan tes *fit* dengan model. Berdasarkan Tabel 7 analisis data, nilai rata-rata INFIT MNSQ 1,00 dan simpangan baku 0,04 (sekitar 0), maka keseluruhan tes *fit* dengan model *Rasch*.

Penetapan *fit* setiap butir soal terhadap model mengikuti kaidah Adam dan Khoo (1996:28), suatu butir soal *fit* terhadap model jika nilai INFIT MNSQ antara 0,77 sampai 1,30. Nilai INFIT MNSQ pada uji coba terbatas berkisara antara 0,98 sampai dengan 1,05 sehingga semua butir soal sebanyak 72 *fit* dengan model *Rasch*.

Setelah melalui tahap penetapan *fit* setiap butir soal terhadap model maka tahap selanjutnya adalah penentuan tingkat kesukaran. Menurut Hambleton & Swaminathan (1985: 36), tingkat kesukaran bernilai baik jika berada pada rentang -2 sampai +2. Berdasarkan analisis tingkat kesukaran butir soal terletak antara antara -1,99 sampai dengan 1,83 sehingga semua butir soal bernilai baik.

Berdasarkan hasil analisis dengan Bilog diperoleh fungsi informasi dan *standard error measurement* (SEM). Berdasarkan fungsi informasi dan SEM yang dinyatakan Gambar 1, maka tes ini cocok untuk peserta didik dengan kemampuan $-2,7 \leq \theta \leq 1,7$.

Penetapan *fit* butir soal secara keseluruhan dikembangkan Adam dan Khoo (1996:28) berdasarkan nilai rata-rata INFIT MNSQ beserta simpangan. Jika rata-rata INFIT MNSQ sekitar 1 dan simpangan bakunya 0, atau rata-rata INFIT t mendekati 0 dan simpangan bakunya 1, maka keseluruhan tes *fit* dengan model. Berdasarkan Tabel 8 analisis data, nilai rata-rata INFIT MNSQ 1,00 dan simpangan baku 0,04 (sekitar 0), maka keseluruhan tes *fit* dengan model *Rasch*.

Penetapan *fit* setiap butir soal terhadap model mengikuti kaidah Adam dan Khoo (1996:28), suatu butir soal *fit* terhadap model jika nilai INFIT MNSQ antara 0,77 sampai 1,30. Nilai INFIT MNSQ pada uji coba terbatas berkisara antara 0,93 sampai dengan 1,09 sehingga semua butir soal sebanyak 72 *fit* dengan model *Rasch*.

Selain untuk menguji kecocokan, output program *Quest* juga menampilkan estimasi reliabilitas. Berdasarkan hasil analisis dengan program *Quest* juga didapatkan reliabilitas instrumen (tes) tersebut diestimasikan sebesar 0,73 yang dinyatakan pada Tabel 8. Nilai reliabilitas ini tergolong kategori reliabel.

Setelah melalui tahap penetapan *fit* setiap butir soal terhadap model maka tahap selanjutnya adalah penentuan tingkat kesukaran. Menurut Hambleton & Swaminathan (1985: 36), tingkat kesukaran bernilai baik jika berada pada rentang -2 sampai +2. Berdasarkan analisis tingkat kesukaran butir soal terletak antara antara -1,65 sampai dengan 1,78 sehingga semua butir soal bernilai baik.

Berdasarkan hasil analisis dengan Bilog diperoleh fungsi informasi dan *standard error measurement* (SEM). Berdasarkan fungsi informasi dan SEM yang dinyatakan Gambar 2, maka tes ini cocok untuk peserta didik dengan kemampuan $-2,3 \leq \theta \leq 1,1$.

Kelayakan media penilaian berbasis *Quizstar* dianalisis dengan *Content Validity Ratio* (CVR) dan *Content Validity Index* (CVI). Berdasarkan analisis didapatkan nilai CVI sebesar 1,00 yang termasuk kategori sangat baik. Instrumen media dianalisis reliabilitas dengan menggunakan koefisien korelasi antar kelas (*Interclass Correlation Coefficient*, ICC) menggunakan program SPSS 23. Setelah dilakukan analisis media penilaian berbasis *Quizstar* memperoleh nilai reliabilitas 0,868 yang termasuk kategori istimewa atau sangat reliabel.

Selain penilaian dari validator, kelayakan media penilaian berbasis *Quizstar* dapat dilihat respon peserta didik. Berdasarkan hasil analisis didapatkan persentase penilaian media *Quizstar*. Berdasarkan analisis didapatkan persentase penilaian media *Quizstar* oleh peserta didik yang dapat dilihat pada Gambar 6. Penilaian media penilaian berbasis *Quizstar* oleh peserta didik didapatkan hasil bahwa 11% pada kategori cukup, 52% pada kategori baik, dan 36% pada kategori sangat baik. Berdasarkan hasil tersebut dapat dikatakan bahwa rata – rata peserta didik menilai media penilaian berbasis *Quizstar* dalam kategori baik.

Analisis kemampuan peserta didik dilihat dari nilai yang didapatkan peserta didik. Sebelum nilai peserta didik dikonversikan menjadi data kuantitatif, terlebih dahulu dilakukan penyetaraan nilai antara kelompok 1 (paket A) dan kelompok 2 (paket B).

Berdasarkan hasil pada uji coba terbatas kemampuan peserta didik peserta didik 11% pada kategori kurang, 29% pada kategori cukup, 39% pada kategori baik dan 21% pada kategori sangat baik yang dapat dilihat pada Gambar 3. Berdasarkan hasil tersebut dapat dikatakan bahwa rata – rata kemampuan peserta didik pada uji terbatas dalam kategori baik.

Berdasarkan hasil pada uji luas kemampuan peserta didik peserta didik 21% pada kategori kurang, 44% pada kategori cukup, 25% pada kategori baik dan 9% pada kategori sangat baik yang dapat dilihat pada Gambar 4. Berdasarkan hasil tersebut dapat dikatakan rata –

rata kemampuan peserta didik pada uji luas dalam kategori cukup.

SIMPULAN DAN SARAN

Simpulan

Berdasarkan hasil penelitian dan analisis terhadap temuan-temuan selama penelitian, maka diperoleh kesimpulan sebagai berikut :

1. Instrumen penilaian berbasis *Quizstar* dikembangkan layak untuk mengukur kemampuan kognitif peserta didik SMA pada materi pokok teori kinetik gas karena memiliki karakteristik sebagai berikut:
 - a. Instrumen penilaian telah memenuhi validitas isi dengan *expert judgment* dan telah mendapatkan bukti empiris *fit* dengan model *Rasch*.
 - b. Seluruh butir soal dalam kriteria baik karena tingkat kesukaran berada pada rentang antara -1,65 sampai dengan 1,78.
 - c. Instrumen penilaian berbasis *Quizstar* telah memiliki koefisien reliabilitas sebesar 0,73 yang termasuk kategori reliabel.
 - d. Berdasarkan fungsi informasi, instrumen penilaian berbasis *Quizstar* sangat tepat digunakan untuk mengukur kemampuan kognitif peserta didik yang berkemampuan dari -2,3 sampai 1,1.
 - e. Berdasarkan analisis angket respon peserta didik didapatkan hasil pada kategori cukup, baik, dan sangat baik berturut – turut 11%, 52%, dan 36%.
2. Kemampuan peserta didik mendapatkan kategori kurang, cukup, baik, dan sangat baik berturut – turut 21%, 44%, 25%, dan 9%.

Saran

Berdasarkan hasil penelitian, maka saran dari peneliti untuk penelitian selanjutnya antara lain:

1. Perlu dikembangkan penelitian sejenis pada pokok bahasan lain dalam mata pelajaran fisika.
2. Perlu dikembangkan penelitian sejenis dalam bentuk soal lain seperti uraian, benar – salah atau campuran.

3. Perlu dilakukan pembuatan penilaian berbasis komputer dengan menggunakan aplikasi yang berbeda untuk menghasilkan sistem penilaian yang lebih baik.

DAFTAR PUSTAKA

- Adam, R.J. & Khoo S.T. (1996). *Acer Quest: The Interactive Test Analysis System*. Victoria: The Australia Council for Educational Research, 2010, Hlm 1-24
- Adams, Wendy K. & Wieman, Carl E. (2010). Development and Validation of Instruments to Measure Learning of Expert-Like Thinking. *International Journal of Science Education*, 1-24
- Bambang Subali. (2011). *Panduan Analisis Data Pengukuran Pendidikan Untuk Memperoleh Bukti Empirik Kesahihan Menggunakan Program Quest*. Yogyakarta: Lembaga Penelitian dan Pengabdian pada Masyarakat UNY.
- Craven, BC & AR Morris. (2010). Modified Ashworth Scale Reliability for Measurement of Lower Extremity Spasticity among Patients with SCI. *Spinal Cord* (48), 207-213
- Hambleton, R.K., & Swaminathan, H. (1985). *Item Respon Theory Principles and Applications*. USA: Kluwer.Nijhoff Publishing.
- Kuswari Hernawati. (2007). Evaluasi dan Penilaian Interaktif Berbasis Web. *Skripsi*. Yogyakarta: FMIPA Universitas Negeri Yogyakarta.
- Lawshe, C.H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology* (28), 563-575.
- Livingston, S. A. (2004). *Equating Test Scores (Without IRT)*. USA: Educational Testing Service.
- Nana Sudjana. (1990). *Penilaian Hasil Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.
- Özden, M. Y., Erturk, I., & Sanli, R. (2004). Students' perception of online assessment: A case study. *Journal of Distance Education*, 19, 77-93.
- Saifuddin Azwar. (2015). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Samsul Hadi. (2013). *Pengembangan Computerized Adaptive Test Berbasis Web*. Yogyakarta: Aswaja Pressindo.
- Sara S. Sparrow & Stephanie M. Davis. (2010). Recent Advances in Assessment of Intelligence and Cognition. *J. Child Psychol Psychiat*, 41(1), 117-131
- Sudaryono. (2012). *Dasar-Dasar Evaluasi Pembelajaran*. Yogyakarta: Graha Ilmu.
- Thiagarajan, S. , D.S. , Semmel & M.I, Semmel. (1974). *Instructional Development for Training Center of Expectional Children*. Minepolish: Indiana University.