

Akurasi pendekatan *Classical Test Theory* dan pendekatan *Item Response Theory* dalam menganalisis soal UAS Fisika semester genap kelas X SMA Negeri di Purworejo Tahun Pelajaran 2013/2014

Suhariyono, Sriyono, Nur Ngazizah

Program Studi Pendidikan Fisika
Universitas Muhammadiyah Purworejo
Jalan K.H.A. Dahlan 3, Purworejo, Jawa Tengah
email: suhariyono87@gmail.com

Intisari – Telah dilakukan penelitian untuk mengetahui: (1) reliabilitas, tingkat kesukaran dan indeks daya beda berdasarkan *Classical Test Theory* (2) *goodnes of fit*, *thrushole* dan indeks daya beda berdasarkan *Item Response Theory* (3) *Standart Error of Measurement* pada *Classical Test Theory* dan *Item Response Theory*. Teknik pengumpulan data dengan metode dokumentasi yaitu mengumpulkan lembar jawab UAS. Analisis menggunakan program *ITEMAN* dan *Program BIGSTEP*. Hasil analisis (1) Koefisien reliabilitas untuk soal KTSP sebesar 0,806, soal kurikulum 2013 sebesar 0,576. Tingkat kesukaran untuk soal KTSP adalah sebesar 0,374, soal kurikulum 2013 adalah sebesar 0,364. Indeks daya beda untuk soal KTSP sebesar 0,370, soal kurikulum 2013 adalah sebesar 0,270 (2) *Goodnes of fit* soal KTSP adalah 14,29% yang tidak fit dengan model dan 85,71% yang fit dengan model, sedangkan soal kurikulum 2013 adalah 14,29% yang tidak fit dengan model dan 85,71% yang fit dengan model. Tingkat kesukaran (*Thrushole*) diperoleh hasil untuk soal KTSP adalah 9,1% tergolong mudah dan 90,9% tergolong sedang. Pada soal kurikulum 2013 terdapat 3,23% tergolong mudah, dan 96,77% soal tergolong sedang. Indeks daya beda soal untuk KTSP diperoleh 0,299 dan pada soal kurikulum 2013 diperoleh 0,160. (3) *Standard Error of Measurement (SEM)* dari program *ITEMAN* diperoleh untuk soal KTSP sebesar 2,48, sedang untuk soal kurikulum 2013 sebesar 2,49. Sedang dari program *BIGSTEP* diperoleh soal KTSP sebesar 0,12 sedangkan untuk soal kurikulum 2013 sebesar 0,16. Dari hasil di atas dapat disimpulkan bahwa tingkat akurasi analisis menggunakan pendekatan *Item Response Theory* lebih tinggi daripada dengan pendekatan *Item Response Theory*.

Kata kunci: *Classical Test Theory*, *Item Response Theory*, *Standar Error of Measurement*.

I. PENDAHULUAN

Kualitas sebuah tes dapat dilihat dengan menggunakan analisis kualitatif (teoritik) dan kuantitatif (empiris), tes dikatakan baik jika telah memenuhi persyaratan penyusunan dari sisi materi, konstruksi dan bahasa. Sedangkan secara kuantitatif dilakukan dengan analisis *item*. Banyak metode dalam bidang psikometrika yang digunakan untuk menganalisis *item* dari pendekatan klasik hingga modern.

Teori tes klasik telah lama menjadi acuan oleh guru dalam menganalisis alat ukur atau instrumen di dunia pendidikan khususnya di kabupaten Purworejo. Dalam teori ini bergantung pada sampel, semakin banyak sampel maka semakin stabil. Begitu juga pada koefisien reliabilitas dan validitas tes akan menjadi lebih tinggi apabila kelompok subjek merupakan kelompok yang kemampuannya heterogen (bervariasi besar) dan tes yang sama akan mempunyai koefisien yang lebih rendah apabila dikenakan pada kelompok yang kemampuannya relatif homogen.

Pendekatan modern masih sangat jarang digunakan oleh para guru untuk menganalisis alat ukur atau instrumen pendidikan. Padahal di dalam pendekatan modern bukan tergantung pada sampel namun bergantung pada kemampuan soal dan kemampuan peserta didik. Bagaimana soal dapat mengungkap kemampuan peserta didik dan peserta didik dengan kemampuan tertentu dapat menyelesaikan soal tertentu dengan benar. Digunakan model matematis untuk menghubungkan karakteristik *item* dengan kemampuan responden. Hubungan tersebut digambarkan melalui kurva karakteristik *item*. Berbeda dengan teori tes klasik yang memfokuskan informasi pada level tes, dalam teori respon *item* memfokuskan pada informasi pada level

item sehingga dapat menutupi kekurangan yang terdapat pada teori tes klasik.

Penerapan model *Item Respons Theory* didasarkan atas beberapa asumsi postulat, yaitu: (1) kinerja seorang pada suatu *item* dapat diprediksikan oleh seperangkat faktor yang disebut *traits*, *latent traits*, atau kemampuan; dan (2) hubungan antara kinerja peserta dengan suatu *item* dan seperangkat kemampuan (abilitas) laten yang mendasarinya dapat digambarkan oleh suatu fungsi yang menarik secara monotonik yang disebut *item characteristic function* atau *item characteristic curve*.

Analisis menggunakan teori tes klasik maupun *Item Response Theory* tentu ada beberapa hal yang di ukur seperti reliabilitas, tingkat kesukaran dan indeks daya beda. Reabilitas adalah tingkat ketepatan dan keajegan skor tes. Tingkat kesukaran adalah peluang menjawab benar suatu soal pada tingkat kemampuan tertentu, sedangkan indeks daya beda adalah kemampuan suatu butir soal dapat membedakan antara peserta didik yang menguasai materi dengan peserta didik yang belum menguasai materi. Meski secara teoritik – fundamental berbeda dengan teori tes klasik, *Item Response Theory* memiliki hubungan yang erat. Oleh sebab itu, bagi para pembaca yang telah mengenal teori tes klasik, hubungan tersebut dapat dijadikan dasar dalam memahami *Item Response Theory*. Setelah mempelajarinya secara lebih mendalam, barulah dapat diketahui manfaat keunggulan *Item Response Theory* dibandingkan teori tes klasik (*Classical Test Theory*). Sejauh ini para ahli pengukuran psikologi dan pendidikan belum memberikan perhatian yang serius dalam menyadari

dan menyambut gelombang perkembangan teori pengukuran.

Kabupaten Purworejo terdapat Sekolah Menengah Atas yang berjumlah 42 sekolah baik yang negeri maupun swasta. Namun untuk analisis soal Ulangan Akhir Semester belum banyak dilakukan, biasanya hanya menggunakan pendekatan klasik saja. Namun di dalam penelitian ini nanti akan disajikan menggunakan pendekatan klasik maupun modern, baik cara menganalisis, program yang digunakan, maupun hasil dari kedua proses tersebut, maka akan dilakukan penelitian judul “ *Akurasi pendekatan Classical Test Theory dan pendekatan Item Response Theory dalam menganalisis soal UAS Fisika semester genap kelas X SMA Negeri di Purworejo tahun pelajaran 2013/2014*”

II. LANDASAN TEORI

A. Penilaian

Penilaian merupakan alih bahasa dari istilah *assessment* yang berarti suatu kegiatan untuk memberikan berbagai informasi secara berkesinambungan dan menyeluruh tentang proses dan hasil yang telah dicapai peserta didik (Arifin, 2013:4). Sedangkan menurut Gronlund penilaian adalah suatu proses yang sistematis dari pengumpulan, analisis, dan interpretasi informasi/data untuk menentukan sejauh mana peserta didik telah mencapai tujuan pembelajaran. Sementara Nitko (1996) dalam Arifin menjelaskan “ *assessment is abroad term defined as a process for obtaining information that is used for making decisions about students....*”

B. Teori Tes Klasik

Teori Tes Klasik atau disebut teori skor murni klasik (Allen & Yen, 1979:57) didasarkan pada suatu model aditif, yakni skor amatan merupakan penjumlahan dari skor sebenarnya dan skor kesalahan pengukuran. Jika dituliskan dengan pernyataan matematika, maka kalimat tersebut menjadi

$$X = T + E \quad (1)$$

Keterangan :

X : skor amatan

T : skor sebenarnya

E : skor kesalahan pengukuran (*error score*)

C. Teori Respon Butir

Pendekatan Teori Respons Butir merupakan pendekatan alternatif yang dapat digunakan dalam menganalisis suatu tes. Ada dua prinsip yang digunakan pada pendekatan ini, yakni prinsip relativitas dan prinsip probabilitas. Pada prinsip relativitas, unit dasar dari pengukuran bukanlah siswa atau butir tetapi lebih kepada kemampuan siswa relatif terhadap butir. Jika β_n merupakan indeks dari kemampuan siswa ke n pada trait yang diukur, dan δ_i merupakan indeks dari tingkat kesulitan dari butir ke- i relatif yang terkait dengan kemampuan yang diukur, maka bukan β_n atau δ_i yang merupakan unit pengukuran, tetapi lebih kepada perbedaan antara kemampuan dan dari siswa relatif terhadap tingkat kesulitan butir atau ($\beta_n - \delta_i$) perlu dipertimbangkan. Sebagai alternatifnya perbandingan antara kemampuan terhadap tingkat kesulitan dapat digunakan. Jika kemampuan dari siswa melampaui tingkat kesulitan butir, maka respons siswa diharapkan benar, dan jika kemampuan

siswa kurang dari tingkat kesulitan butir, maka respons siswa diharapkan salah (Keeves & Alagumalai, 1999:24).

D. Program ITEMAN

Program ini dapat digunakan untuk: (1) menganalisis data file (format ASCII) jawaban butir soal yang dihasilkan melalui manual *entry* data atau dari mesin *scanner* ; (2) menskor dan menganalisis data soal pilihan ganda dan skala Likert untuk 30.000 siswa dan 250 butir soal; (3) menganalisis sebuah tes yang terdiri dari 10 skala (subtes) dan memberikan informasi tentang validitas setiap butir (daya pembeda, tingkat kesukaran, proporsi jawaban pada setiap option), reliabilitas (KR-20/Alpha), *standar error of measurement, mean, variance*, standar deviasi, *skew*, kurtosis untuk jumlah skor pada jawaban benar, skor minimum dan maksimum, skor median, dan frekuensi distribusi skor.

E. Program BIGSTEPS

BIGSTEPS adalah perangkat lunak (*software*) yang diciptakan secara khusus untuk analisis dan kalibrasi butir soal (*item test*). Software ini dikembangkan dan diterbitkan pertama kali tahun 1992 oleh Benjamin D. Wright & John Linacre dari University of Chicago Amerika Serikat. Hasil (*output*) dari analisis menggunakan software ini antara lain informasi yang berkaitan dengan skor yang diperoleh, estimasi tingkat kemampuan peserta test (*measure of ability*) dan estimasi tingkat kesukaran butir soal (*measure of difficulty*) dalam bentuk skala baku yang disebut LOGIT (*log odd unit*), indeks kesalahan pengukuran (*standard error of measurement*), kecocokan antara data dan model (*infit* dan *oufit*), korelasi daya pembeda (*point biserial*), serta berbagai informasi lain yang berkaitan dengan butir soal dan responden.

F. Pustaka

Lilis Tri Ariyana (2011) meneliti tentang analisis butir soal ulangan akhir semester gasal IPA kelas IX SMP di Kab. Grobogan. Hasil penelitian mengemukakan bahwa soal memiliki tingkat kesukaran sedang, daya beda yang baik, efektifitas pengecoh berfungsi serta *reliable* dengan kategori tinggi.

Nopitalia (2010) meneliti tentang analisis soal tes buatan guru biologi Madrasah Tsanawiyah Se-Jakarta Selatan berdasarkan aspek kognitif Taksonomi Bloom. Hasil analisisnya bahwa soal-soal didominasi pada ranah kognitif C1 (pengetahuan) dengan persentase 60,26%, C2 (pemahaman) 38,46% dan C4 (analisis) hanya 1,28%.

Abdul kodir (2013) meneliti tentang akurasi soal UAS mata pelajaran Fisika kelas X semester genap SMA Negeri di Purworejo Tahun Pelajaran 2012/2013 dengan hasil cukup akurat dalam mengukur kompetensi siswa dengan standard error pengukuran 0,09. Banyaknya soal UAS yang tidak fit dengan model rasch adalah sebanyak 4 (11,42%). Yaitu butir nomor 1, 12, 21, 32. Besarnya proporsi kemampuan siswa rendah ($\delta < -2$), sedang ($-2 \leq \delta \leq 2$), dan tinggi ($\delta > 2$).

III. METODE PENELITIAN

Penelitian ini pada Ujian Akhir Semester genap mata pelajaran Fisika SMA Negeri di Kabupaten Purworejo tahun pelajaran 2013/2014. Waktunya dari bulan Desember 2013 hingga Agustus 2014.

Teknik pengumpulan data yang dilakukan dalam penelitian ini menggunakan metode dokumentasi.

Dokumentasi yang dilakukan adalah dengan mengumpulkan lembar jawab UAS semester genap mata pelajaran IPA Fisika kelas X SMA Negeri di Kabupaten Purworejo tahun pelajaran 2013/2014,

untuk metode klasik peneliti menggunakan program *ITEMAN*. Dimana hasil korelasi poin biserialnya dapat dituliskan dengan

$$r_{pbi} = \frac{\bar{Y}_p - \bar{Y}_t}{St} \sqrt{\frac{P}{(1-P)}} \text{ atau } r_{bis} = \frac{\bar{Y}_p - \bar{Y}_t}{St} \sqrt{\frac{P(1-P)}{U}} \quad (2)$$

keterangan

Y_p : mean skor pada kriterion siswa yang menjawab benar soal.

Y_t dan St : mean dan standard deviasi kriterion seluruh siswa.

p : proporsi siswa yang menjawab benar soal.

U : ordinat kurva normal.

Untuk menganalisis secara modern peneliti, Dalam penelitian ini, akan digunakan model satu parameter (model *Rasch*) dengan bantuan program *BIGSTEPS* versi 2.30. Pemilihan model Rasch didasarkan pada pernyataan Ivailo Partchev (2004: 10) yang menyebutkan bahwa untuk butir soal dikotomi dapat menggunakan model satu parameter (1P). Menurut Hambleton, Swaminathan, & Rogers (1991: 12-13) persamaan matematik fungsi karakteristik butir model Rasch dapat dinyatakan seperti berikut ini.

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}} \quad i = 1, 2, \dots, n \quad (3)$$

keterangan:

$P_i(\theta)$: peluang menjawab benar peserta yang berkemampuan θ pada butir ke-i

b_i : tingkat kesukaran butir ke-i

e : bilangan transendental yang bernilai 2,718

n : banyaknya butir tes

D : konstanta bernilai 1,7 sebagai simpangan baku distribusi logistik

IV. HASIL DAN PEMBAHASAN

1. Data hasil analisis soal UAS Semester Genap mata pelajaran Fisika kelas X SMA Negeri di Purworejo tahun pelajaran 2013/2014.

a. Berdasarkan penelitian dengan program *ITEMAN* maka didapatkan data sebagai berikut.

1) Data KTSP

Tabel 1.
Scale Statistics soal UAS KTSP

Scale	0	Minimum	2.000
N of Items	33	Maximum	30.000
N of Examinees	425	Median	12.000
Mean	12,332	Alpha	0,806
Variance	31,831	SEM	2.487
Std. Dev.	5.642	Mean P	0.374
Skew	0.679	Mean Item-Tot.	0.370
Kurtosis	0,002	Mean Biserial	0.487

2) Data kurikulum 2013

Tabel 2.

Scale Statistics soal UAS kurikulum 2013

Scale:	0	Minimum	2,000
N of Items	33	Maximum	27,000
N of Examinees	218	Median	11,000
Mean	11,271	Alpha	0,576
Variance	14,730	SEM	2,499
Std. Dev.	3,838	Mean P	0,364
Skew	0,535	Mean Item-Tot.	0,270
Kurtosis	0,611	Mean Biserial	0,360

b. Berdasarkan penelitian dengan program *BIGSTEP* maka diperoleh data sebagai berikut

1. Data KTSP

SUMMARY OF 425 MEASURED (NON-EXTREME) PERSONS							
	SCORE	COUNT	MEASURE	ERROR	MNSQ	INFIT	MNSQ OUTFIT
MEAN	12.3	33.0	-.64	.42	1.00	-.1	1.02
S.D.	3.6	.0	.92	.06	.13	1.2	.25
RMSE	.42	ADJ. S.D.	.81	PERSON SEP	1.93	PERSON SEP REL.	.79

Gambar 1.

Interval kepercayaan soal UAS KTSP

2. Data kurikulum 2013

SUMMARY OF 218 MEASURED (NON-EXTREME) PERSONS							
	SCORE	COUNT	MEASURE	ERROR	MNSQ	INFIT	MNSQ OUTFIT
MEAN	11.3	31.0	-.66	.41	1.00	.0	1.00
S.D.	3.8	.0	.64	.04	.12	.9	.20
RMSE	.41	ADJ. S.D.	.49	PERSON SEP	1.18	PERSON SEP REL.	.58

Gambar 2.

Interval kepercayaan soal UAS kurikulum 2013

2. Data yang tidak fit dari hasil analisis soal UAS Semester Genap mata pelajaran Fisika kelas X SMA Negeri di Purworejo tahun pelajaran 2013/2014.

a. Data hasil penelitian menggunakan program *ITEMAN* maka didapatkan hasil sebagai berikut.

1) Peserta KTSP

Tabel 3.

Kecocokan dengan model klasik KTSP

Statistik	Nomor Butir	Jumlah
Fit	1,2,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,24,25,26,27,28,29,30,31,32,33,34,35	33
Tidak Fit	3,23	2
Jumlah		35

2) Peserta kurikulum 2013

Tabel 4.

Kecocokan dengan model klasik kurikulum 2013

Statistik	Nomor Butir	Jumlah
Fit	1,2,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,23,24,26,27,29,30,31,32,33,34,35	31
Tidak Fit	3,22,25,28	4
Jumlah		35

asil penelitian menggunakan program *BIGSTEP* maka didapatkan data sebagai berikut.

1) Peserta KTSP

Tabel 5.

Kecocokan dengan model *Rasch* KTSP

Statistik	Nomor Butir	Jumlah
Fit	1,2,4,5,6,7,9,10,11,12,13,14,16,17,18,19,20,21,22,24,25,26,28,29,31,32,33,34,35,	30
Tidak Fit	3,8,15,23,27	5
Jumlah		35

2) Peserta kurikulum 2013

Tabel 6.

Kecocokan dengan model *Rasch* kurikulum 2013

Statistik	Nomor Butir	Jumlah
Fit	1,2,4,5,6,7,8,9,10,11,12,13,14,15,16,18,19,20,21,24,26,27,29,30,31,32,34,35,	30
Tidak Fit	3,22,25,28,33	5
Jumlah		35

3) *Pembahasan Hasil Penelitian*

Telah dilakukan penelitian dengan judul “ Akurasi pendekatan *Classical Test Theory* dan pendekatan *Item Response Theory* dalam menganalisis soal UAS Fisika semester genap kelas X SMA Negeri di Purworejo tahun pelajaran 2013/2014”. Penelitian ini dimulai pada bulan Desember 2013 hingga Agustus 2014 mulai dari pengajuan proposal, pengambilan data, analisis hingga pembuatan laporan dalam bentuk skripsi.

Sampel dalam penelitian ini adalah 5 sekolah SMA Negeri di Purworejo, 3 sekolah dengan KTSP dan 2 sekolah dengan kurikulum 2013. Jumlah respondennya 425 siswa dari KTSP dan 218 siswa dari kurikulum 2013.

Pengolahan data dalam penelitian ini dengan menggunakan program *ITEMAN* untuk *Classical Test Theory* dan program *BIGSTEP* untuk *Item Response Theory*. Hal yang diungkap dalam penelitian ini adalah Koefisien reliabilitas, tingkat kesukaran, dan indeks daya beda pada pendekatan klasik dan *Goodness of fit*, *thrushole* dan indeks daya beda pada pendekatan modern, serta *Standard Error of Measurement* baik dalam pendekatan klasik maupun pendekatan modern.

Hasil dari analisis program *ITEMAN* menunjukkan bahwa koefisien reliabilitas soal UAS KTSP sebesar 0,806 dan koefisien reliabilitas soal UAS kurikulum 2013 sebesar 0,576. Dari angka tersebut dapat disimpulkan bahwa reliabilitas soal UAS KTSP termasuk sangat tinggi dan reliabilitas soal UAS kurikulum 2013 termasuk sedang (Guilford, 1956: 145) menyatakan bahwa $0,80 \leq \alpha \leq 1,00$ (sangat tinggi), $0,60 \leq \alpha \leq 0,80$ (tinggi), $0,40 \leq \alpha \leq 0,60$ (sedang)

Tingkat kesukaran dari hasil analisis menggunakan *ITEMAN* menunjukkan bahwa pada soal UAS KTSP mempunyai nilai 0,374 dan pada soal UAS kurikulum 2013 bernilai 0,364. Dari angka tersebut dapat disimpulkan bahwa soal UAS program KTSP maupun kurikulum 2013 mempunyai tingkat kesukaran sedang karena rentang tingkat kesukaran $0,70 \leq TK \leq 1,00$ mudah, $0,30 \leq TK \leq 0,70$ sedang dan $0,00 \leq TK \leq 0,30$ sukar.

Indeks daya beda dari hasil analisis menggunakan *ITEMAN* menunjukkan bahwa pada soal UAS KTSP mempunyai nilai 0,370 dan pada soal UAS kurikulum 2013 bernilai 0,270. Dari angka tersebut dapat di jelaskan bahwa indeks daya beda pada soal UAS KTSP tergolong baik dan soal UAS kurikulum 2013 masih belum baik, ini dikarenakan indeks daya beda dikatakan baik apabila mempunyai nilai $\geq 0,3$.

Goodness of fit dari hasil analisis menggunakan *BIGSTEP* diperoleh data pada KTSP terdapat 5 soal yang tidak fit dengan model serta 30 soal yang fit dengan model dan pada kurikulum 2013 sama yaitu 5 soal tidak fit dengan model dan 30 soal yang fit dengan model. Soal fit dengan model yaitu soal yang mempunyai nilai $outfit -2,00 \leq fit \leq 2,00$.

Thrushole (tingkat kesukaran) dari hasil analisis menggunakan *BIGSTEP* pada soal UAS KTSP di dapatkan soal yang tergolong mudah ada 3 soal karena mempunyai nilai $outfit \leq -2,00$ dan ada 30 soal yang tergolong sedang karena mempunyai nilai $outfit -2,00 \leq fit \leq 2,00$. Pada soal UAS kurikulum 2013 didapatkan 1 soal yang tergolong mudah karena mempunyai nilai $outfit \leq -2,00$ dan 30 soal tergolong mudah karena mempunyai nilai $outfit -2,00 \leq fit \leq 2,00$.

Indeks daya beda dari hasil analisis menggunakan *BIGSTEP* menunjukkan dalam soal UAS KTSP mempunyai nilai 0,299 dan pada soal UAS kurikulum 2013 mempunyai nilai 0,160.

Standart Error of measurement berdasarkan analisis menggunakan *ITEMAN* diperoleh hasil 2,48 untuk soal UAS KTSP dan 2,49 untuk soal UAS kurikulum 2013. Berdasarkan analisis menggunakan *BIGSTEP* diperoleh hasil 0,12 untuk soal KTSP dan 0,16 untuk soal kurikulum 2013.

Tingkat akurasi pendekatan modern (*Item Response Theory*) lebih akurat dibandingkan dengan pendekatan klasik (*Classical Test Theory*). Kesimpulan ini diperoleh dari hasil analisis yang sudah dilakukan. Dimana *Standard Error of Measurement* dari pendekatan modern lebih kecil dari pendekatan klasik. Dilihat dari hasil analisis juga dapat ditarik kesimpulan bahwa soal UAS KTSP lebih reliabel jika dibandingkan dengan soal UAS kurikulum 2013.

V. KESIMPULAN

Hasil analisis (1) Koefisien reliabilitas untuk soal KTSP sebesar 0,806, soal kurikulum 2013 sebesar 0,576. Tingkat kesukaran untuk soal KTSP adalah sebesar 0,374, soal kurikulum 2013 adalah sebesar 0,364. Indeks daya beda untuk soal KTSP sebesar 0,370, soal kurikulum 2013 adalah sebesar 0,270 (2) *Goodnes of fit* soal KTSP adalah 14,29% yang tidak fit dengan model dan 85,71% yang fit dengan model, sedangkan soal kurikulum 2013 adalah 14,29% yang tidak fit dengan model dan 85,71% yang fit dengan model. Tingkat kesukaran (*Thrushole*) diperoleh hasil untuk soal KTSP adalah 9,1% tergolong mudah dan 90,9% tergolong sedang. Pada soal kurikulum 2013 terdapat 3,23% tergolong mudah, dan 96,77% soal tergolong sedang. Indeks daya beda soal untuk KTSP diperoleh 0,299 dan pada soal kurikulum 2013 diperoleh 0,160. (3) *Standard Error of Measurement (SEM)* dari program *ITEMAN* diperoleh untuk soal KTSP sebesar 2,48, sedang untuk soal kurikulum 2013 sebesar 2,49. Sedang dari

program *BIGSTEP* diperoleh soal KTSP sebesar 0,12 sedangkan untuk soal kurikulum 2013 sebesar 0,16. Dari hasil di atas dapat disimpulkan bahwa tingkat akurasi analisis menggunakan pendekatan *Item Response Theory* lebih tinggi daripada dengan pendekatan *Item Response Theory*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada:

1. Eko Setyadi Kurniawan, M.Pd.Si., sebagai reviewer jurnal ini.
2. Kepala SMA Negeri 2, 3, 4, 5,10 Purworejo yang telah memberikan ijin untuk melaksanakan penelitian.
3. Guru mata pelajaran Fisika dan peserta didik kelas SMA Negeri 2, 3, 4, 5,10 Purworejo yang telah banyak membantu dalam penelitian ini.

PUSTAKA

Artikel jurnal:

- [1] Musmuliadi. 2009. Hubungan model penskoran terhadap estimasi skor sesungguhnya berdasarkan teori response butir. *Jurnal Penelitian dan Evaluasi Pendidikan*, 13, 246-267

Buku:

- [1] Allen, M.J & Yen, W.M.1979. *Introduction to measurementTheory*. Monterey: Brook/Cole Publishing Company.
- [2] Hambleton, R.K & Swaminathan H.1985. *Item Response Theory Principles and Applications*. Boston: Kluwer_Nijhoff Publishing
- [3] Arifin,Z.2013.*Evaluasi Pembelajaran*. Yogyakarta:Rosda Karya