



Item Analysis Using Rasch Model in Semester Final Exam Evaluation Study Subject in Physics Class X TKJ SMK Negeri 2 Manokwari

Jefri Palimbong¹, Mujasam², & Alberto Y. T. Allo^{2*}

¹Mahasiswa Program Studi Pendidikan Fisika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Papua

²Dosen Program Studi Pendidikan Fisika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Papua

*Corresponding author: jhntn909@gmail.com

Abstract: The purpose of this research is to item analyze in semester final exam evaluation using Rasch model in terms of validity, reliability, and difficulty level study subject in physics class X TKJ SMK Negeri 2 Manokwari. This research was a qualitative research that is evaluation of learning outcomes using quantitative descriptive method with data collection techniques used documentation. The data of respondents 44 students with the number of items 30 questions, obtained the result are the validity of the conformity level of item 26 questions was fit and 4 questions not fit. The students reliability 0,37 is weak, the item reliability 0,83 is good and the reliability between students with item 0,42 is bad. The difficult level of item general in the medium category, the means is very good because it is not difficult and not easy. The conclusion item analysis using rasch model in semester final exam evaluation were valid, reliable, and the difficulty level of item is very good.

Keywords: item analysis, evaluate learn, rasch model.

Analisis Butir Soal Menggunakan Model Rasch pada Ujian Evaluasi Belajar Akhir Semester Mata Pelajaran Fisika Kelas X TKJ SMK Negeri 2 Manokwari

Abstrak: Tujuan penelitian ini adalah untuk menganalisis butir soal ujian evaluasi semester menggunakan model Rasch ditinjau dari validitas, reliabilitas, dan tingkat kesukaran butir soal pada pelajaran fisika kelas X TKJ SMK Negeri 2 Manokwari. Penelitian ini merupakan penelitian kualitatif yakni evaluasi hasil belajar yang menggunakan metode deskriptif kuantitatif dengan teknik pengumpulan data menggunakan dokumentasi. Data responden sebanyak 44 peserta didik dengan jumlah butir soal sebanyak 30 soal diperoleh hasil yaitu validitas tingkat kesesuaian butir soal sebanyak 26 soal yang fit dan 4 soal yang tidak fit. Reliabilitas peserta didik 0,37 bermakna lemah, reliabilitas butir soal 0,83 bermakna bagus dan reliabilitas antara peserta didik dengan butir soal 0,42 bermakna buruk. Tingkat kesukaran soal secara umum dalam kategori yang sedang, berarti soal tersebut sudah sangat baik karena tidak terlalu sukar dan tidak terlalu mudah. Kesimpulannya bahwa analisis butir soal menggunakan model rasch pada ujian evaluasi belajar akhir semester dinyatakan valid, reliabel, dan tingkat kesukaran butir soal sangat baik.

Kata kunci: analisis butir soal, evaluasi belajar, model rasch.

PENDAHULUAN

Peraturan Pemerintah No. 19 tahun 2005 tentang Standar Nasional Pendidikan, evaluasi hasil belajar dilakukan dengan penilaian hasil belajar yang bertujuan untuk memantau proses pembelajaran, kemajuan dan perbaikan hasil belajar serta untuk menilai dan mengukur tingkat ketercapaian kompetensi peserta didik (BNSP, 2009). Evaluasi berasal

dari kata *evaluation* yang berarti mengukur dan juga menilai. Mengukur berarti membandingkan sesuatu dengan cara memberikan suatu ukuran angka terhadap hasil belajar yang diperoleh peserta didik, sedangkan menilai yaitu mengambil satu keputusan ukuran baik dan juga buruk terhadap sesuatu yang dinilai. Kegiatan menganalisis butir soal dilakukan untuk meningkatkan mutu soal serta mengidentifikasi kekurangan-kekurangan butir tes yang telah dibuat dalam rangka perbaikan pembelajaran (Arikunto, 2012). Kualitas instrument soal sangat menentukan hasil akhir pembelajaran berupa pengusaan peserta didik terhadap tujuan pembelajaran yang dirumuskan (Yusuf & Widyaningsih, 2018).

Proses analisis butir soal merupakan kegiatan yang dilakukan untuk menelaah dan mengkaji setiap butir soal melalui pengumpulan informasi dari jawaban peserta didik agar diperoleh soal yang bermutu sebelum soal-soal tersebut digunakan (Ratumanan & Laurens, 2011). Analisis butir soal dapat dilakukan secara kualitatif yakni yang berkaitan dengan isi dan bentuknya, dan secara kuantitatif yakni yang berkaitan dengan ciri-ciri statistiknya. Analisis kuantitatif ini mencakup pada analisis validitas, reliabilitas, tingkat kesukaran, daya pembeda soal serta tingkat pengecoh yang dimiliki oleh soal tersebut (Rahayu, dkk, 2014).

Butir soal yang valid berarti instrumen tersebut dapat digunakan untuk mengukur apa yang seharusnya ingin diukur. Sedangkan instrumen yang reliabel merupakan instrumen yang bila diujikan beberapa kali pada objek yang sama dalam waktu yang berbeda akan menghasilkan data yang sama (Sugiyono, 2014). Tingkat kesukaran soal berkaitan dengan kemampuan peserta didik untuk dapat menjawab atau dapat menyelesaikan soal tersebut dengan baik dan benar. Dalam penelitian ini untuk menganalisis butir soal hasil Evaluasi Belajar Akhir Semester (EBAS) peserta didik menggunakan model Rasch untuk menguji validitas, reliabilitas, serta *person* dan *item* secara sekaligus. Pemodelan Rasch dapat digunakan untuk pengukuran validitas dan reliabilitas secara langsung berdasarkan prinsip probabilitas (Tennant, dkk, 2004).

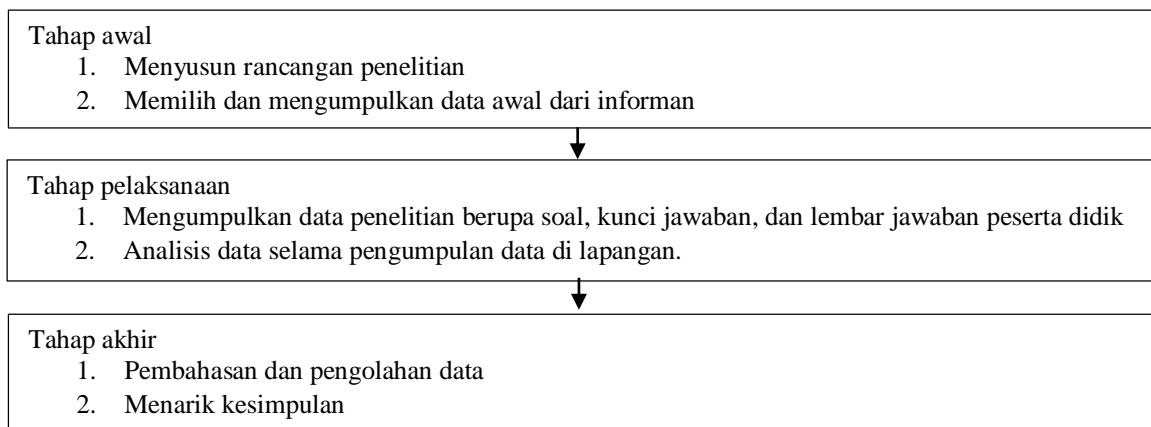
Keunggulan model Rasch antara lain: mampu melakukan prediksi terhadap data yang hilang, yang didasarkan pada pola respons yang sistematis; mampu menghasilkan nilai pengukuran standar *error* untuk instrumen yang digunakan yang dapat meningkatkan ketepatan perhitungan; dan kalibrasi yang dilakukan sekaligus dalam tiga hal yaitu skala pengukuran, responden, dan butir soal (Sumintono & Widhiarso, 2015). Keunggulan dari model Rasch tersebut sangat cocok digunakan dalam penelitian ini karena untuk mengevaluasi kemampuan butir soal EBAS mata pelajaran fisika terhadap peserta didik. Penggunaan model Rasch lebih efektif diterapkan dibandingkan dengan analisis klasik (Fisher, 1993). Hasil observasi menunjukkan soal fisika yang digunakan di SMK Negeri 2 Manokwari menggunakan soal-soal yang disusun sendiri oleh guru bidang studi fisika yang tidak divalidasi oleh guru fisika lainnya dan soal-soal tersebut belum pernah diujicobakan terlebih dahulu kepada peserta didik sehingga soal-soal tersebut belum dapat diketahui tingkat kualitasnya.

Berdasarkan penjelasan di atas masalah yang akan diteliti dalam penelitian ini yaitu bagaimana analisis butir soal EBAS dengan menggunakan model Rasch ditinjau dari validitas, reliabilitas, dan tingkat kesukaran butir soal pada pelajaran fisika kelas X TKJ SMK Negeri 2 Manokwari.

METODE PENELITIAN

Penelitian ini merupakan penelitian kualitatif yakni pada jenis penelitian evaluasi hasil belajar. Penelitian evaluasi dimaksudkan untuk menilai suatu program agar tersedia informasi yang diperlukan untuk pengambilan keputusan. Metode penelitian

menggunakan metode deskriptif kuantitatif, yang mempunyai tujuan untuk menggambarkan peristiwa yang terjadi dengan menggunakan statistik. Fokus penelitian pada soal pilihan ganda ujian EBAS berjumlah 30 soal, pada pelajaran fisika kelas X TKJ SMK Negeri 2 Manokwari dengan jumlah responden 44 peserta didik. Analisis data menggunakan teknik analisis deskriptif kuantitatif. Data yang diteliti dianalisis dengan menggunakan model Rasch dengan menggunakan aplikasi Winstep® versi 3.73. Tahapan penelitian secara ringkas sebagaimana pada Gambar 1.



Gambar 1. Tahapan dalam penelitian

Pengumpulan data dalam penelitian ini menggunakan metode dokumentasi. Metode dokumentasi adalah metode yang digunakan untuk mendapatkan data mengenai hal-hal atau variabel yang berupa catatan, transkip, buku, surat kabar, majalah, prasasti notulen rapat, agenda dan sebagainya (Arikunto, 2001). Dengan metode dokumentasi penelitian ini akan memperoleh data-data tentang hasil belajar yang berupa (a) Lembar soal Ujian EBAS mata pelajaran Fisika, (b) Lembar jawaban peserta didik, dan (c) Lembar kunci jawaban soal. Dari hasil lembar dokumentasi yang didapatkan selanjutnya lembar soal tersebut dianalisis secara kuantitatif yakni pada validitas, reliabilitas dan tingkat kesukaran butir soal.

Analisis validitas yang dihasilkan dengan program Winstep® yaitu analisis berupa validitas konstruk dan validitas konten, analisis validitas konten meliputi tingkat kesesuaian soal yang berfungsi untuk melihat kualitas tingkat kesesuaian butir dengan model. Informasi yang diberikan berupa informasi kesesuaian butir soal dengan kriteria yaitu dengan melihat nilai *outfit mean square*, *outfit z-standard*, dan *point measure correlation*. Butir soal dikatakan valid atau diterima jika telah memenuhi minimal 2 kriteria dan diperbaiki jika memenuhi salah satu dari ketiga kriteria tersebut, serta dibuang bila tidak ada yang memenuhi kriteria tersebut. Nilai kesesuaian butir sangat dipengaruhi oleh banyaknya data, semakin besar sampel yang digunakan maka semakin baik juga tingkat kesesuaianya. Nilai kriteria yang dipakai dalam melihat tingkat kesesuaian butir (Sumintono & Widhiarso, 2015):

- Nilai *Outfit Mean Square* (MNSQ) yang diterima: $0,5 < \text{MNSQ} < 1,5$
- Nilai *Outfit Z-Standard* (ZSTD) yang diterima: $-2,0 < \text{ZSTD} < +2,0$
- Nilai *Point Measure Correlation* (Pt Measure Corr): $0,4 < \text{Pt Measure Corr} < 0,85$

Sumintono & Widhiarso, (2015) menyatakan bahwa pemodelan Rasch dapat melakukan analisis sampai ke tingkat butir soal, tingkat individu, dan tingkat instrumen

secara lebih rinci dalam hal ini akan membimbing guru dalam pembuatan soal, ataupun dalam mengambil kebijakan yang tepat, logis, dan ilmiah berdasarkan panduan kriteria kualitas instrumen sebagai berikut:

- a. *Person* dan *item measure* menunjukkan nilai rata-rata peserta didik dan item, nilai yang paling kecil menunjukkan tingkat kemampuan peserta didik dan soal yang rendah dengan nilai *logit* 0,0.
- b. Nilai *alpha crombach* nilai ini digunakan untuk mengukur reliabilitas yakni interaksi antara soal dan item, kriteria yang digunakan yakni ($< 0,5$: Buruk, $0,5 - 0,6$: Jelek, $0,6 - 0,7$: Cukup, $0,7-0,8$: Bagus, $> 0,8$: Istimewa).
- c. Nilai *person reliability* dan *item reliability* merupakan nilai reliabilitas peserta didik dan nilai reliabilitas item soal. Kategori yang digunakan yaitu: $< 0,67$: Lemah, $0,67 - 0,80$: Cukup, $0,8 - 0,90$: Bagus, $0,91 - 0,94$: Bagus sekali, $> 0,94$: Istimewa.
- d. *Separation* merupakan pengelompokan peserta didik dan *item*. Semakin besar nilai *separation* maka kualitas instrumen akan semakin baik. Persamaan (1) menunjukkan persamaan yang digunakan dalam melihat pengelompokan peserta didik dan item.

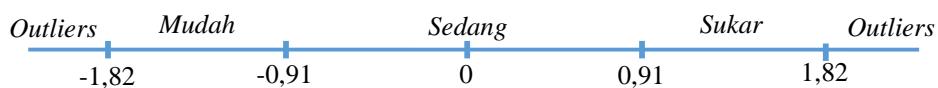
$$H = \frac{[(4 \times \text{Separation})+1]}{3} \quad (1)$$

Keterangan:

H : Nilai strata (Pengelompokan)

Separation : Nilai *separation* untuk responden yang dihasilkan

Tingkat kesukaran butir soal (*item measure*) memberikan tampilan nilai *logit* dari setiap butir soal secara terperinci. Hasil *outuput* dari tabel ini memberikan informasi butir soal yang diurutkan dari yang memiliki nilai *logit measure* yang tertinggi ke nilai *logit* yang terendah yang menunjukkan bahwa soal diurutkan dari butir soal yang tersukar sampai pada yang butir soal yang termudah. Pengelompokan tingkat kesukaran soal dapat ditentukan berdasarkan nilai rata-rata *logit* yang dijumlahkan dengan nilai standar deviasi (SD). Nilai ini berguna untuk mengidentifikasi kelompok *item* (*separation*). Pada penelitian ini didapatkan nilai separation hasil dari penjumlahan nilai rata-rata *logit* $0,00 + 0,91$ SD = $0,91$, sehingga akan diperoleh kelompok soal sebagaimana pada Gambar 2. (Sumintono & Widhiarso, 2015).



Gambar 2. Tingkat kesukaran butir soal

HASIL DAN PEMBAHASAN

1. Validitas

Hasil analisis validitas soal yang dihasilkan berupa validitas konstruk dan validitas konten. Hasil analisis yang dilakukan dengan model Rasch didapatkan hasil pada *output tables item dimensionality*, hasil analisis dapat dilihat pada Gambar 3.

	-- Empirical --		Modeled
Total raw variance in observations	=	37.4 100.0%	100.0%
Raw variance explained by measures	=	7.4 19.8%	18.8%
Raw variance explained by persons	=	1.2 3.1%	3.0%
Raw Variance explained by items	=	6.2 16.7%	15.8%
Raw unexplained variance (total)	=	30.0 80.2% 100.0%	81.2%
Unexplned variance in 1st contrast	=	4.3 11.5% 14.3%	
Unexplned variance in 2nd contrast	=	2.7 7.3%	9.2%
Unexplned variance in 3rd contrast	=	2.4 6.3%	7.9%
Unexplned variance in 4th contrast	=	2.3 6.2%	7.8%
Unexplned variance in 5th contrast	=	1.8 4.8%	6.0%

Gambar 3. Hasil *output* analisis validitas konstruk

Gambar 3 menunjukkan hasil validitas konstruk pada nilai *Raw variance explained by measures* secara empiris diperoleh 19.8% sedangkan model Rasch memprediksi 18.8% dalam hal ini validasi konstruk secara empiris hampir sama dengan nilai yang di prediksi oleh model Rasch. Penilaian validasi konstruk tersebut dimaknai kurang baik karena dikatakan penilaian validasi konstruk baik apabila *Raw variance explained by measures* > 40%. Sedangkan nilai *Unexplained variance* yang didapatkan semuanya < 15% dimaknai kurang baik. Validasi konstruk yang diperoleh dengan penilaian kurang baik dikarenakan dalam penelitian ini validasi konstruk tidak dilaksanakan oleh guru fisika kelas X, guru tersebut membuat soal sendiri dan tidak meminta kepada rekan guru fisika yang lain untuk melakukan validasi konstruk. Validasi konstruk dapat dilakukan dengan melibatkan beberapa validator sehingga diperoleh hasil yang lebih baik (Widyaningsih & Yusuf, 2018). Namun demikian walaupun tidak dilakukan validasi konstruk, di dalam model Rasch validasi konstruk dapat diprediksi hasilnya. Pemodelan Rasch sangat efektif diterapkan karena dapat diperoleh hasil analisis validitas yang dapat dipercaya (Sari, dkk, 2016). Model Rasch juga lebih mudah digunakan karena telah disertai dengan analisis secara langsung menggunakan aplikasi komputer (Mair & Hatzinger, 2007). Hasil analisis validitas konten yang dapat dilihat tingkat kesesuaian butir soal ditunjukkan pada *output tables item fit* sebagaimana pada Tabel 1.

Tabel 1. Hasil analisis validitas tingkat kesesuaian butir soal

Nomor	Soal ke	Outfit MNSQ	Outfit ZSTD	Keterangan
1	S4	1,44	1,7	Butir Soal Fit
2	S26	1,24	0,9	
3	S3	1,19	1,1	
4	S12	1,18	0,6	
5	S24	1,16	0,8	
6	S15	1,16	1,1	
7	S27	1,15	1,2	
8	S13	1,11	0,7	
9	S30	1,10	0,5	
10	S6	1,10	0,8	
11	S14	1,01	0,1	
12	S18	0,95	-0,2	
13	S11	1,02	0,2	
14	S21	0,99	0,0	
15	S8	0,95	-0,4	

Nomor	Soal ke	Outfit MNSQ	Outfit ZSTD	Keterangan
16	S20	0,96	-0,4	Butir Soal Tidak Fit
17	S19	0,95	-0,5	
18	S2	0,91	-0,5	
19	S28	0,93	-0,3	
20	S7	0,89	-0,6	
21	S5	0,90	-0,4	
22	S17	0,86	-0,9	
23	S25	0,78	-0,9	
24	S23	0,85	-0,9	
25	S29	0,86	-1,2	
26	S10	0,59	-0,6	
27	S16	1,90	1,6	
28	S9	1,54	1,7	
29	S22	0,80	-2,3	
30	S1	0,76	-2,9	

Tabel 1 menunjukkan hasil analisis validitas tingkat kesesuaian butir soal yaitu dari 30 butir soal EBAS diperoleh sebanyak 26 soal yang dinyatakan fit dan 4 soal yang tidak fit karena tidak memenuhi kriteria *Outfit MNSQ* dan *Outfit ZSTD*. Apabila soal telah dinyatakan fit artinya memenuhi kriteria dan dapat menjamin bahwa tingkat pemahaman peserta didik memang diuji melalui butir-butir soal yang sesuai dan berkualitas.

2. Reliabilitas

Hasil analisis reliabilitas butir soal ditunjukkan pada *summary statistic* dapat dilihat pada Tabel 2.

Tabel 2. Nilai reliabilitas peserta didik dan *item* soal

Variabel	Rata-rata logit (SD)	Separation	Reliabilitas	α Crombach
Peserta didik	-0,74	0,76	0,37	0,42
Butir soal	0,00	2,23	0,83	

Tabel 2 menunjukkan hasil nilai reliabilitas peserta didik yaitu 0,37 yang berarti bahwa reliabilitas peserta didik lemah karena nilai tersebut $< 0,67$. Sedangkan nilai reliabilitas soal didapatkan nilai 0,83 yang berarti bahwa reliabilitas dari soal tersebut bagus yaitu berada pada rentang 0,80 - 0,90. Sedangkan nilai *alpha cronbach* yaitu 0,42 yang artinya nilai reliabilitas antara peserta didik dan butir soal berkategori buruk. Penilaian tersebut didasarkan pada nilai *alpha cronbach* yang dapat mengukur interaksi antara *person* dan butir-butir soal secara keseluruhan. Ditentukan pada nilai *separation* dan nilai *alpha cronbach*. Pengelompokan peserta didik dan butir soal dapat ditentukan sebagaimana pada persamaan (1).

$$H = \frac{[(4 \times \text{Separation}) + 1]}{3}$$

$$H_{\text{Peserta didik}} = \frac{[(4 \times 0,76) + 1]}{3} = \frac{4,04}{3} = 1,34$$

$$H_{\text{Butir soal}} = \frac{[(4 \times 2,23) + 1]}{3} = \frac{9,92}{3} = 3,30$$

Didapatkan nilai $H_{\text{Peserta didik}}$ sebanyak 1,34 yang dibulatkan 1, yang bermakna *separation* peserta didik kurang baik karena kemampuan kognitif peserta didik hanya didominasi satu kemampuan saja. *Separation* peserta didik dikatakan baik bila kemampuan kognitif peserta didik dibagi dalam kemampuan tinggi, sedang, dan rendah. Untuk menjelaskan lebih jauh mengapa *separation* peserta kurang baik dibutuhkan hasil analisis item dan person pada Tabel 5. $H_{\text{Butir Soal}}$ sebanyak 3,30 yang dibulatkan 3, yang bermakna *separation* soal bernilai baik karena membagi soal dalam 3 kelompok yaitu tinggi, sedang, dan rendah.

3. Tingkat Kesukaran Butir Soal

Berdasarkan hasil analisis data didapatkan hasil tingkat kesulitan soal pada hasil *output item measure*. Tingkat kesukaran butir soal dapat dikelompokkan dengan mengkombinasikan nilai rata-rata *logit* dan nilai Standar Deviasi (SD). Nilai ini berguna untuk mengidentifikasi kelompok *item (separation)*. Dari hasil ini didapatkan nilai rata-rata *logit measure* yaitu 0,00 dan nilai SD sebesar 0,91.

Tabel 4. Hasil *item measure* analisis kesukaran butir soal

<i>Item Statistics</i>	<i>Measure</i>	Keterangan
Mean	0,00	Berkategori sedang
S.D.	0,91	

Tabel 5. Hasil analisis item dan person

Kategori	Inisial Peserta Didik	Jumlah	Persentase
Berkemampuan Tinggi	A20	1	2%
Berkemampuan di atas Rata-rata	A21, A03	2	5%
Berkemampuan rata-rata	A11, A23	2	5%
Berkemampuan di bawah rata-rata	A09, A10, A17, A24, A05, A16, A07, A13, A15, A25, A04, B05, B14, B15, B16, A06, A08, A14, A19, B01, B09, B17, A01, A18, B02, B07, B08, A02, A26, A27, B06, B13, A22, B03, B11, A12, B10	37	84%
Berkemampuan sangat rendah	B04, B12	2	5%
Total		44	100%

Hasil analisis *item* dan *person* menunjukkan kemampuan peserta didik didominasi pada kategori berkemampuan di bawah rata-rata sebesar 84% artinya tingkat kemampuan peserta didik untuk menjawab soal dengan benar masih kurang karena peserta didik kurang memahami konsep dengan benar. Hasil analisis *item* dan *person* bersesuaian dengan hasil $H_{\text{Peserta didik}}$ yang hanya didominasi satu tingkatan kemampuan peserta didik. Penggunaan model Rasch dapat mengukur tingkat perbedaan kemampuan responden dalam menjawab item soal yang diberikan (Andersen, 1973).

Tabel 6. Deskripsi hasil analisis tingkat kesukaran soal

Kategori	Nomor Butir Soal	Jumlah	Percentase
<i>Outlier</i>	10, 30	2	7%
Sukar	16, 12	2	7%
Sedang	9, 11, 26, 4, 14, 25, 5, 24, 28, 18, 7, 2, 13, 23, 15, 6, 8, 29, 21, 22, 19, 1	22	73%
Mudah	20, 27, 17, 3,	4	13%
Total		30	100%

Hasil analisis tingkat kesukaran soal didominasi oleh soal yang berkategori sedang berdasarkan model Rasch butir soal tersebut termasuk mempunyai tingkat kesukaran yang baik karena soal tersebut lebih didominasi oleh soal yang berkategori sedang yang artinya soal tidak terlalu sukar dan tidak terlalu mudah. Sedangkan kategori *outlier* maksudnya butir soal tersebut tidak dapat digunakan atau dibuang saja. Pemodelan Rasch dapat memberikan informasi yang detail tentang tingkat kesulitan instrumen yang diberikan (Curtis, 2007). Lebih lanjut Ismail, dkk, (2010) bahwa melalui pemodelan Rasch dapat diperoleh instrumen soal yang baik untuk mengukur tingkat pemahaman responden dalam penelitian sehingga sangat direkomendasikan penggunaanya terutama dalam bidang pendidikan.

SIMPULAN DAN SARAN

Analisis butir soal EBAS menggunakan model Rasch pada materi fisika kelas X TKJ SMK Negeri 2 Manokwari, diperoleh validitas tingkat kesesuaian butir soal sebanyak 26 soal yang fit dan 4 soal yang tidak fit. Reliabilitas peserta didik 0,37 bermakna lemah, reliabilitas butir soal 0,83 bermakna bagus dan reliabilitas antara peserta didik dengan butir soal 0,42 bermakna buruk. Tingkat kesukaran soal secara umum dalam kategori yang sedang, berarti soal tersebut sudah sangat baik karena tidak terlalu sukar dan tidak terlalu mudah. Kesimpulannya bahwa analisis butir soal EBAS menggunakan model Rasch dinyatakan valid, reliabel, dan tingkat kesukaran butir soal sangat baik.

Saran dalam penelitian ini yaitu soal-soal yang dibuat oleh guru fisika sebaiknya terlebih dahulu divalidasi dan diujicobakan ke peserta didik serta peserta didik yang telah mengikuti ujian EBAS fisika diberi remidial.

DAFTAR PUSTAKA

Andersen, E. B. (1973). A Goodness of Fit Test for the Rasch Model. *Psychometrika*, 38(1), 123-140.

Arikunto, S. (2001). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Arikunto, S. (2012). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

BNSP. (2009). *Peraturan Pemerintah Republik Indonesia Nomor 19 Tahun 2005 Tentang Standar Nasional Pendidikan*. Jakarta: BNSP.

Curtis, D. D. 2007. X-ray Your Data with Rasch. *International Education Journal*, 8(2), 249-259.

Fisher, A. G. (1993). The Assessment of IADL Motor Skills: An Application of Many-faceted Rasch Analysis. *American Journal of Occupational Therapy*, 47(4), 319-329.

Ismail, I., Idrus, R. M., & Johari, S. S. M. (2010). Acceptance on Mobile Learning Via SMS: a Rasch Model Analysis. *International Journal of Interactive Mobile Technologies (iJIM)*, 4(2), 10-16.

Mair, P., & Hatzinger, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, 20(9), 1-20.

Rahayu, Dwi T., Purnomo H. B., & Sukidin. (2014). Analisis Tingkat Kesukaran dan Daya Beda pada Soal Ujian Tengah Semester Ganjil Bentuk Pilihan Ganda Mata Pelajaran Ekonomi Kelas X di SMA Negeri 5 Jember Tahun Ajaran 2012-2013. *Jurnal Edukasi UNEJ*, 1(1).

Ratumanan, T. G., & Laurens, T. (2011). *Penilaian Hasil Belajar pada Tingkat Satuan Pendidikan*. Ambon: Unesa University Press.

Sari, D. R., Sekarwana, N., Hinduan, Z. R., & Sumintono, B. (2016). Analisis Tingkat Kepuasan Masyarakat terhadap Dimensi Kualitas Pelayanan Tenaga Pelaksana Eliminasi Menggunakan Pemodelan Rasch. *JSK*, 2(1), 47-55.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Bandung: Trim Komunikata.

Sugiyono. (2014). *Metode Penelitian Kombinasi*. Bandung: AlfaBeta.

Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value in Health*, 7, S22-S26.

Widyaningsih, S. W., & Yusuf, I. (2018). Analisis Soal Modul Laboratorium Fisika Sekolah I Menggunakan Racsh Model, *Gravity: Jurnal Ilmiah Penelitian dan Pembelajaran Fisika*, 4(1).

Yusuf, I., & Widyaningsih, S. W. (2018). Profil Kemampuan Mahasiswa dalam Menyelesaikan Soal HOTS di Jurusan Pendidikan Fisika Universitas Papua, *Jurnal Komunikasi Pendidikan*, 2(1), 42-49.