



Analisis Kualitas Butir Soal Fisika Pada Penilaian Akhir Semester

Siswanti¹, Sri Wahyu Widyaningsih², Sri Rosepda Br. Sebayang³

¹Mahasiswa Program Studi Pendidikan Fisika, FKIP, Universitas Papua

^{2,3}Dosen Program Studi Pendidikan Fisika, FKIP, Universitas Papua

antysiswanty@gmail.com

Info Artikel

Riwayat artikel

Dikirim: May 21, 2023

Direvisi : Oktober 30, 2023

Diterima: November 02, 2023

Kata Kunci:

Kualitas Butir Soal
Evaluasi Hasil Belajar
Model Rasch

ABSTRAK

Tujuan dari penelitian ini adalah untuk menganalisis kualitas butir soal fisika dengan menggunakan Model Rasch yang ditinjau berdasarkan tingkat validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas pengecoh soal pada Penilaian Akhir Semester Gasal kelas X IPA SMA YPK Oikoumene Manokwari tahun ajaran 2022/2023. Metode penelitian yang digunakan dalam penelitian ini adalah metode deskriptif kuantitatif yakni evaluasi hasil belajar. Teknik pengumpulan data menggunakan teknik dokumentasi, dengan sampel sebanyak 28 peserta didik dari kelas X IPA 1-X IPA 3. Penelitian ini difokuskan pada soal pilihan ganda yang terdiri dari 15 soal. Hasil analisis menunjukkan bahwa 1) berdasarkan tingkat validitasnya, 8 butir soal (60%) yang valid dan 7 butir soal (40%) yang tidak valid. 2) Berdasarkan reliabilitasnya, soal tersebut reliabel karena mempunyai nilai pada butir soal sebesar 0,01 berkriteria lemah, sedangkan untuk reliabilitas antara peserta didik dengan butir soal sebesar 0,94 yang berkriteria bagus sekali. 3) Berdasarkan tingkat kesukaran soalnya, masuk dalam kriteria kurang baik dengan jumlah 6 butir soal (40%) didominasi oleh soal yang mudah. 4) Berdasarkan daya pembeda soalnya, masuk dalam kriteria cukup dengan jumlah 12 soal (80%). 5) Berdasarkan efektivitas pengecoh soalnya, masuk dalam kriteria baik karena terdapat 8 butir soal (60%) dengan *option* pengecohnya berfungsi dengan baik. Kesimpulannya bahwa analisis kualitas butir soal menggunakan Model Rasch pada penilaian akhir semester gasal dapat dinyatakan valid, reliabel, tingkat kesukaran soal kurang baik, daya pembeda soal cukup dan efektivitas pengecoh soalnya baik.

ABSTRACT

The purpose of this study was to analyze the quality of the physics items using the Rasch Model which was reviewed based on the level of validity, reliability, level of difficulty, discriminating power and effectiveness of the problem distractors in the Final Assessment of Odd Semester class X IPA SMA YPK Oikoumene Manokwari in the 2022/2023 academic year. The research method used in this study is a quantitative descriptive method, namely the evaluation of learning outcomes. The data collection technique used documentation techniques, with a sample of 28 students from class X IPA 1-X IPA 3. This research focused on multiple choice questions consisting of 15 questions. The results of the analysis show that 1) based on the level of validity, 8 items (60%) are valid and 7 items (40%) are invalid. 2) Based on its reliability, the question is reliable because it has a value in the item of 0,01 which is a weak criterion, while for reliability between students and the item is 0,94 which has very good criteria. 3) Based on the level of difficulty of the questions, it is included in the unfavorable criteria with a total of 6 questions (40%) dominated by easy questions. 4) Based on the differentiating power of the questions, it is included in the sufficient criteria with a total of 12 questions (80%). 5) Based on the effectiveness of the distractor, it is included in the good criteria because there are 8 questions (60%) with the distractor option functioning properly. The conclusion is that the analysis of the quality of the

items using the Rasch Model at the end of the odd semester assessment can be declared valid, reliable, the level of difficulty of the items is not good, the discriminating power of the items is sufficient and the effectiveness of the distractors is good.

© 2023 Pendidikan Fisika, Universitas Papua, Provinsi Papua Barat, Indonesia.

PENDAHULUAN

Pendidikan adalah suatu cara atau sistem yang dilaksanakan untuk membentuk peserta didik guna meningkatkan Sumber Daya Manusia (SDM) yang berkualitas. Seorang pendidik tentunya memegang peranan penting dalam proses pembelajaran. Tugas utama pendidik dalam pembelajaran adalah merencanakan dan melaksanakan pembelajaran, serta menentukan proses penilaian dari hasil belajar melalui kegiatan evaluasi (Liesfi, 2016). Evaluasi berasal dari sebuah kata *evaluation* yang memiliki arti mengukur dan menilai (Prayetno, 2021). Sedangkan hasil belajar merupakan perubahan tingkah laku dari peserta didik yang terjadi setelah pembelajaran (Sudjana, 2009: 22). Oleh karena itu, dapat disimpulkan bahwa evaluasi hasil belajar merupakan keseluruhan kegiatan pengukuran (pengumpulan data dan informasi), pengolahan, penafsiran dan pertimbangan untuk membuat keputusan mengenai tingkat hasil belajar yang dinilai berdasarkan aspek kognitif, afektif dan psikomotorik yang dicapai oleh peserta didik setelah melakukan kegiatan belajar dalam upaya mencapai tujuan pembelajaran yang telah ditetapkan. Peran penting dilakukannya evaluasi dinilai dalam ketentuan Pemerintah No. 19 tahun 2005 mengenai Standar Nasional Pendidikan, penilaian evaluasi hasil belajar dilakukan dengan cara menilai hasil belajarnya, yang bermaksud untuk meninjau prosedur pembelajaran, perkembangan dan peningkatan hasil belajar, serta mengevaluasi nilai dan mengukur tingkat ketercapaian keterampilan peserta didik (Peraturan Pemerintah Republik Indonesia No. 19 Tahun 2005 Tentang Standar Nasional Pendidikan, 2005).

Dalam kegiatan evaluasi hasil belajar diperlukan alat atau teknik penilaian, sehingga pelaksanaannya jadi lebih terarah. Alat evaluasi dalam pendidikan yang digunakan untuk mengumpulkan data dapat berupa tes atau non tes. Salah satu contoh alat evaluasi hasil belajar pada proses pembelajaran adalah dengan melakukan Penilaian Akhir Semester (PAS). PAS merupakan bagian dari bentuk evaluasi serta mempunyai tujuan untuk mengukur dan juga menilai kompetensi keberhasilan dari peserta didik. Dalam penyusunan butir soal pendidik hendaknya sudah menganalisis soal tersebut sebelum diberikan kepada peserta didik agar dapat mengetahui kualitas dari suatu butir soal tersebut.

Analisis butir soal adalah suatu tindakan yang dilaksanakan untuk mengkaji serta menyelidiki tiap-tiap butir soal dengan mengumpulkan data terhadap jawaban dari peserta didik untuk memperoleh soal yang berkualitas sebelum soal-soalnya digunakan (Ratumanan, T. G., & Laurens, 2011). Kualitas butir soal benar-benar memastikan hasil akhir pembelajaran yang berupa kemampuan peserta didik tentang tujuan pembelajaran yang ditetapkan (Widyaningsih & Yusuf, 2018). Analisis butir soal dapat dilaksanakan secara kuantitatif yakni yang terkait dengan sifat statistiknya. Analisis kuantitatif meliputi analisis validitas, reliabilitas, tingkat kesukaran, daya pembeda dan tingkat pengecoh yang dimiliki oleh soal tersebut (Rahayu et al., 2014).

Sama seperti sekolah-sekolah pada umumnya, SMA YPK Oikoumene Manokwari menggunakan PAS pada semester gasal sebagai bentuk evaluasi hasil belajar yang telah dilakukan selama satu semester. Berdasarkan hasil wawancara dan observasi yang telah dilakukan terhadap guru mata pelajaran fisika di SMA YPK Oikoumene Manokwari, diketahui bahwa guru belum melakukan analisis butir soal, sehingga soal tersebut belum diketahui kualitasnya. Dalam penyusunan soal ujian, guru cenderung menggunakan soal-soal yang sudah pernah diberikan sebelumnya atau mengambil dari buku paket, Lembar Kerja Peserta Didik

(LKPD), dan beberapa soal dibuat sendiri oleh guru. Soal yang telah dibuat oleh guru tidak dianalisis atau diujicobakan terlebih dahulu.

Berdasarkan pada permasalahan yang telah diuraikan serta untuk membantu guru dalam mengkaji dan menyelidiki soal yang hendak diberikan agar soal yang digunakan tergolong pada soal yang baik dan berkualitas, maka peneliti tertarik untuk melakukan penelitian tentang analisis kualitas butir soal. Dengan menganalisis kualitas butir soal, diharapkan dapat memperoleh soal-soal berkualitas yang dapat mengukur hasil pembelajaran secara maksimal.

METODE PENELITIAN

Penelitian ini menggunakan metode deskriptif kuantitatif yang mana termasuk dalam jenis penelitian evaluasi hasil belajar. Dimana penelitian ini bertujuan untuk mendeskripsikan perihal yang telah terbentuk dengan menggunakan data statistik. Adapun tiga tahapan yang digunakan dalam metode penelitian evaluasi hasil belajar yaitu tahap awal, pelaksanaan dan tahap akhir. Tahap awal merupakan tahapan perencanaan penelitian dimana tahap ini meliputi penyusunan rancangan atau desain penelitian, pemilihan dan pengumpulan data awal dari informan. Tahap pelaksanaan adalah tahap berlangsungnya penelitian, tahap ini meliputi pemberian surat permohonan izin penelitian, pengumpulan data penelitian (lembar soal, kunci jawaban soal dan lembar jawaban peserta didik) dan menganalisis data sewaktu dilapangan. Tahap akhir meliputi pembahasan dan pengolahan data dan setelah itu menarik kesimpulan dari data yang dianalisis.

Teknik Pengumpulan data yang dirancang oleh peneliti yaitu memakai metode dokumentasi. Teknik dari metode yang dipakai hendaknya memperoleh data-data tentang hal-hal atau variabel-variabel yang berbentuk catatan, transkrip, buku, surat kabar, majalah, rapat, agenda dan lain sebagainya disebut metode dokumentasi (Arikunto, 2013). Metode dokumentasi pada penelitian ini yaitu dengan meminta lembar soal PAS mata pelajaran fisika, kunci jawaban soal, serta lembar jawaban peserta didik. Dari hasil lembar dokumentasi yang didapatkan selanjutnya lembar soal tersebut dianalisis secara kuantitatif yakni pada tingkat validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas pengecoh butir soal.

Penelitian ini difokuskan pada soal pilihan ganda PAS yang terdiri dari 15 soal, pada pelajaran fisika kelas X IPA SMA YPK Oikoumene Manokwari dengan jumlah sampel sebanyak 28 peserta didik terdiri dari kelas X IPA 1 yang berjumlah 10 peserta didik, kelas X IPA 2 sebanyak 8 peserta didik, dan kelas X IPA 3 sebanyak 10 peserta didik. Data yang diteliti dianalisis dengan model *Rasch* menggunakan aplikasi *Winstep*[®] versi 5.4.1.0 dengan berbantuan program *Microsoft Excel* 2016.

Hasil analisis validitas yang didapatkan dari program *Winstep*[®] yaitu analisis berupa validitas konstruk dan validitas isi (konten), dimana analisis validitas isi meliputi tingkat kesesuaian soal yang berfungsi untuk melihat kualitas tingkat kesesuaian butir soal dengan mode. Data 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 disebut 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 kesesuaiannya. Nilai kriteria yang digunakan dalam melihat tingkat kesesuaian butir (validitas isi) menurut (Sumintono, B., & 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): $0,4 < \text{Pt Measure Corr} < 0,85$.

Sedangkan untuk melihat nilai *item dimensionality* (validitas konstruk), terdapat kriteria yang telah ditetapkan oleh (Sumintono, B., & Widhiarso, 2015) yakni sebagai berikut:

- a. Nilai *Raw variance explained by measures*, kriteria yang dipakai yaitu: (<20%: buruk, ≥ 20%: cukup, ≥ 40%: baik, ≥ 60%: sangat baik).
- b. Nilai *Unexplained variance in contrasts* $1^{st} - 5^{th}$, kriteria yang digunakan yaitu: (>15%: buruk, 10-15%: cukup, 5-10%: baik, < 5%: sangat baik).

Analisis reliabilitas yang didapatkan dari program *Winstep*® yaitu ditunjukkan pada *output tables 3.1 Summary statistic*, tabel ini memberikan informasi tentang bagaimana bagaimana hasil analisis reliabilitas antara peserta didik dengan item. Menurut Sumintono dan Widhiarso, (2015) tabel ini memberikan hasil gambaran yaitu sebagai berikut:

- a. *Person* dengan item *measure* menyatakan bahwa nilai rerata peserta didik dan itemnya, yaitu nilai terkecil memperlihatkan tingkat keahlian peserta didik terhadap soalnya yang rendah dengan nilai logit 0,0.
- b. *Nilai person reliability* dengan item *reliability* yaitu nilai reliabilitas peserta didik dengan nilai reliabilitas item soalnya, kriteria yang dipakai yaitu: (< 0,67: lemah, 0,67-0,80: cukup, 0,81-0,90: bagus, 0,91-0,94: bagus sekali, > 0,94: istimewa).
- c. *Nilai alpa crombach*, dipakai dalam mengukur reliabilitasnya yang meliputi interaksi antara soal dan itemnya, kriteria yang dipakai yaitu: (<0,50: buruk, 0,50-0,60: jelek, 0,60-0,70: cukup, 0,70-0,80: bagus, >0,80: bagus sekali).

Separation (pemisahan) adalah penggabungan peserta didik dan itemnya. Jika bertambah tinggi nilai *separation*-nya, maka semakin baik kualitas instrumennya. Perumusan yang dipakai untuk membuktikan penggabungan peserta didik dengan item terdapat pada persamaan (1).

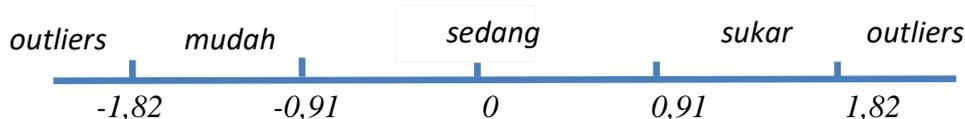
$$H = \frac{[(4 \times Separation) + 1]}{3} \dots\dots\dots (1)$$

Keterangan:

H : Nilai strata (penggabungan)

Separation: Nilai dari *Separation* terhadap informan yang dijadikan

Hasil analisis tingkat kesukaran butir soal (*item measure*) yang didapatkan dari program *Winstep*® yaitu memberikan tampilan nilai *logit* dari setiap butir soal secara terperinci. Hasil *output* 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. Pengelompokkan 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* $0,00 + 0,91 \text{ SD} = 0,91$, sehingga akan diperoleh kelompok soal sebagaimana pada Gambar 1. (Sumintono & Widhiarso, 2015).



Gambar 1. Tingkat Kesukaran Butir Soal

Hasil analisis daya pembeda yang didapatkan dari program *Winstep*® yaitu ditunjukkan pada *output tables 1. Variabel maps*. Daya pembeda suatu soal merupakan kemampuan dalam membedakan antara peserta didik yang pandai atau berkemampuan tinggi dan peserta didik yang kurang pandai atau berkemampuan rendah (Arikunto, 2013). Rumus yang dapat digunakan untuk menghitung daya pembeda soal menggunakan program *Microsoft Excel* versi 2016 terdapat pada persamaan 2.

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} \dots\dots\dots (2)$$

Keterangan:

D = daya pembeda

J_A = banyaknya peserta kelompok atas

J_B = banyaknya peserta kelompok bawah

B_A = banyaknya peserta kelompok atas yang menjawab dengan benar

B_B = banyaknya peserta kelompok bawah yang menjawab dengan benar

Sumber: dimodifikasi dari (Arikunto, 2013)

Hasil analisis efektivitas pengecoh soal yang didapatkan dari program *Winstep*® yaitu ditunjukkan pada *output tables 13. Item Measure*. Pengecoh dikatakan dapat menjalankan fungsinya dengan baik apabila pengecoh tersebut dipilih sekurang-kurangnya 5% dari seluruh peserta tes. Pengecoh yang telah menjalankan fungsinya dengan baik dapat digunakan kembali pada tes-tes berikutnya (Marthunis et al., 2015).

HASIL DAN PEMBAHASAN

1. Validitas

Hasil analisis validitas soal pilihan ganda yang dihasilkan berupa validitas konstruk dan validitas isi (konten). Hasil analisis yang dilakukan dengan model Rasch didapatkan hasil pada *output tables item dimensionality*, hasil analisis dapat dilihat pada Gambar 2.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units			
	Eigenvalue	Observed	Expected
Total raw variance in observations =	31.3450	100.0%	100.0%
Raw variance explained by measures =	16.3450	52.1%	51.2%
Raw variance explained by persons =	12.1877	38.9%	38.1%
Raw Variance explained by items =	4.1573	13.3%	13.0%
Raw unexplained variance (total) =	15.0000	47.9%	100.0%
Unexplned variance in 1st contrast =	4.9751	15.9%	33.2%
Unexplned variance in 2nd contrast =	3.5040	11.2%	23.4%
Unexplned variance in 3rd contrast =	1.8063	5.8%	12.0%
Unexplned variance in 4th contrast =	1.3538	4.3%	9.0%
Unexplned variance in 5th contrast =	.9818	3.1%	6.5%

Gambar 2. Hasil Analisis Validitas Konstruk dengan *Winsteps*

Gambar 2 menunjukkan hasil validitas konstruk pada garis warna biru terdapat nilai *Raw variance explained by measures* diperoleh 52,1%, hampir sama dengan nilai yang diprediksi oleh model Rasch yaitu 51,2%, dalam hal ini validitas konstruk secara observasi hampir sama dengan nilai yang di prediksi oleh model Rasch. Penilaian validitas konstruk tersebut dimaknai 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 et al., 2016). Model Rasch juga lebih mudah digunakan karena telah disertai dengan analisis secara langsung menggunakan aplikasi komputer (Mair & Hatzinger, 2007).

Hasil analisis validitas isi (konten) dapat dilihat pada tingkat kesesuaian butir soal ditunjukkan pada *output tables item fit order* sebagaimana pada Tabel 1.

Tabel 1. Hasil Analisis Validitas pada Soal Pilihan Ganda

Kriteria	Indeks Validitas			No. Butir	Jumlah	Persentase
	MNSQ	ZSTD	<i>PT-Mean Corr</i>			
Valid	3	12	8	S1, S2, S3, S4, S11, S12, S14 dan S15	8	60%
Tidak Valid	12	0	7	S5, S6, S7, S8, S9, S10 dan S13	7	40%

Tabel 1 menunjukkan hasil analisis validitas isi (konten) pada tingkat kesesuaian butir soal pilihan ganda yaitu dari 15 butir soal PAS diperoleh sebanyak 8 soal yang dinyatakan valid dan 7 soal yang tidak valid karena tidak memenuhi kriteria *Outfit* MNSQ dan *Pt-Mean corr*. Apabila soal telah dinyatakan valid artinya soal tersebut 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 *output tables summary statistic* yang dapat dilihat pada Tabel 2.

Tabel 2. Hasil Analisis Reliabilitas pada Soal Pilihan Ganda

Variabel	Rata-rata Logit (SD)	<i>Separation</i>	<i>Reliability</i>	α <i>Crombach</i>
Peserta didik	2,94	0,85	0,42	0,94
Item soal	0,00	0,11	0,01	

Tabel 2 menunjukkan hasil nilai reliabilitas peserta didik yaitu 0,42, dan nilai reliabilitas butir soal didapatkan nilai sebesar 0,01 yang berarti bahwa reliabilitas peserta didik dan butir soal tersebut bermakna lemah karena nilai tersebut $< 0,67$. Sedangkan nilai *alpha crombach* yaitu 0,94 yang artinya nilai reliabilitas antara peserta didik dan butir soal berkriteria bagus sekali. Penilaian tersebut didasarkan pada nilai *alpha crombach* yang dapat mengukur interaksi antara peserta didik dan butir-butir soal secara keseluruhan ditentukan pada nilai *separation* dan nilai *alpha crombach*. Pengelompokkan peserta didik dan butir soal dapat ditentukan sebagaimana pada persamaan (1).

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

$$H_{\text{peserta didik}} = \frac{[(4 \times 0,85) + 1]}{3} = \frac{[4,4]}{3} = 1,467$$

$$H_{\text{butir soal}} = \frac{[(4 \times 0,11) + 1]}{3} = \frac{[1,44]}{3} = 0,48$$

Didapatkan nilai $H_{\text{Peserta didik}}$ sebanyak 1,467, yang bermakna *separation* peserta didik kurang baik karena kemampuan kognitif peserta didik hanya didominasi sebatas 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 *seperation* peserta didik kurang baik dibutuhkan hasil analisis item dan person pada Tabel 5. Sedangkan untuk nilai $H_{\text{Butir Soal}}$ sebanyak 0,48, yang bermakna *seperation* soal bernilai kurang baik karena hanya bernilai 0.

3. Tingkat Kesukaran Butir Soal

Berdasarkan hasil analisis data didapatkan hasil tingkat kesukaran 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 (seperation)*. Dari hasil ini didapatkan nilai rata-rata *logit measure* yaitu 0,00 dan nilai SD sebesar 1,04.

Tabel 3. Hasil *item measure* analisis kesukaran butir soal

Item statistics	Measure	Keterangan
Mean	0,0	Berkriteria sukar
S.D	1,04	

Berdasarkan hasil analisis pada Tabel 3 didapatkan pengelompokkan tingkat kesukaran butir soal yang meliputi *outliers*, soal sukar, soal sedang dan soal mudah, dapat dilihat pada tabel 4.

Tabel 4. Deskripsi Hasil Analisis Tingkat Kesukaran Soal Pilihan Ganda

Kategori	No. Butir	Jumlah	Presentase
Outliers	1 dan 4	2	14%
Sukar	12, 14 dan 15	3	20%
Sedang	3, 7, 10 dan 13	4	26%
Mudah	2, 5, 6, 8, 9 dan 11	6	40%
Total		15	100%

Tabel 4 menunjukkan hasil analisis tingkat kesukaran soal didominasi oleh soal yang berkriteria mudah, karena berdasarkan model Rasch butir soal tersebut termasuk mempunyai tingkat kesukaran yang kurang baik karena soal tersebut lebih didominasi oleh soal yang berkriteria mudah yang artinya soal tersebut terlalu mudah untuk dikerjakan oleh peserta didik. Sedangkan kategori *outlier* maksudnya butir soal tersebut tidak dapat digunakan sebaiknya diganti dengan soal yang baru. Pemodelan Rasch dapat memberikan informasi yang detail tentang tingkat kesulitan instrumen yang diberikan (Curtis & Boman, 2007).

4. Daya Pembeda Soal

Berdasarkan hasil analisis data didapatkan hasil analisis daya pembeda soal yang dilakukan untuk menganalisis kemampuan butir soal pilihan ganda agar dapat membedakan antara peserta didik yang menguasai materi dan peserta didik yang belum mampu dalam menguasai materi. Hasil analisis daya pembeda yang dilakukan dengan *Winsteps* dapat ditunjukkan pada *output tables Variabel maps*. Hasil analisis berdasarkan pengelompokkan tingkat kemampuan peserta didik dapat dilihat pada Tabel 5.

Tabel 5. Hasil Analisis Tingkat Kemampuan Peserta Didik

Tingkat Kemampuan	Peserta didik	Jumlah	Persentase
Di atas Rara-rata	02P, 03P, 04P, 06P, 07L, 10P, 13P, 16L, 21P, 22P, 26P, 27P, 28P	13	46%
Rata-rata	01L, 05L, 08P, 09L, 12P, 19P, 20P, 23P, 24L, 25L	10	36%
Di bawah Rata-rata	17P, 14P, 11P, 15P, 18L	5	18%

Total	28	100%
--------------	----	------

Hasil analisis *item* dan *person* menunjukkan kemampuan peserta didik yang didominasi pada tingkat kemampuan di atas rata-rata sebesar 46% artinya tingkat kemampuan peserta didik untuk menjawab soal dengan benar sudah baik, akan tetapi untuk butir soalnya kurang baik karena soal tersebut lebih didominasi oleh soal yang berkriteria mudah yang artinya soal tersebut perlu dilakukan perbaikan agar dapat digunakan pada tes berikutnya. Hasil analisis *item* dan *person* bersesuaian dengan hasil *HPeserta didik* yang hanya didominasi satu tingkatan kemampuan peserta didik. Penggunaan model Rasch dapat mengukur tingkat perbedaan kemampuan peserta didik dalam menjawab item soal yang diberikan (Andersen, 1973). Selain itu untuk dapat melihat daya pembeda dari butir soal dilakukan dengan perhitungan manual menggunakan program *Microsoft Excel* versi 2016. Deskripsi dari hasil analisis daya pembeda soal menggunakan program *Microsoft Excel* versi 2016 dapat dilihat pada Tabel 6.

Tabel 6. Deskripsi Hasil Analisis Daya Pembeda Soal Pilihan Ganda

Kriteria	No. Butir Soal	Jumlah	Presentase
Baik Sekali	-	-	0%
Baik	14	1	7%
Cukup	2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13 dan 15	12	80%
Kurang	1 dan 4	2	13%
Total		15	100%

Tabel 6 menunjukkan hasil analisis daya pembeda pada butir soal yang digunakan pada Penilaian Akhir Semester gasal tersebut didominasi oleh butir soal dengan daya pembeda yang cukup, karena terdapat 12 butir soal atau sekitar 80% soal yang mempunyai daya pembeda cukup yang berarti bahwa soal-soal tersebut sudah layak digunakan.

5. Efektivitas Pengecoh Soal

Berdasarkan hasil analisis data didapatkan hasil analisis efektivitas pengecoh pada soal pilihan ganda yang ditunjukkan pada *output tabel Item measure*. Option pengecoh dikatakan berfungsi apabila telah dipilih oleh sekurang-kurangnya 5% orang. Terdapat 75 pilihan jawaban yang digunakan, dimana 15 pilihan jawaban lainnya merupakan kunci jawaban dari soal. Hasil analisis dapat dilihat pada Tabel 7.

Tabel 7. Hasil Analisis Keberfungsian Pengecoh Soal Pilihan Ganda

Kriteria yang Dipilih	Jumlah Option	Persentase
Berfungsi > 5%	9	12%
Tidak Berfungsi <5%	31	41%
Tidak Dipilih	20	27%
Kunci Jawaban	15	20%
Jumlah	75	100%

Tabel 8. Deskripsi Hasil Analisis Efektivitas Pengecoh Soal

Kriteria yang Dipilih	Jumlah Butir (dengan option yang dipilih)	Persentase	Keterangan
Berfungsi > 5%	8	60%	Baik
Tidak Berfungsi <5%	7	40%	Kurang baik

Tabel 7 dan 8 menunjukkan hasil analisis efektivitas pengecoh soal yang digunakan pada Penilaian Akhir Semester gasal yang didominasi oleh butir soal dengan pengecoh/*distractor* yang baik, karena terdapat 8 butir soal atau sekitar 60% soal yang *option* pengecohnya berfungsi dengan baik (jumlah pemilihnya >5%). Melalui pemodelan Rasch dapat diperoleh instrument soal yang baik untuk mengukur tingkat pemahaman peserta didik dalam penelitian sehingga sangat direkomendasikan penggunaannya terutama dalam bidang pendidikan (Ismail, dkk, 2010).

KESIMPULAN DAN SARAN

Analisis kualitas butir soal PAS menggunakan model Rasch pada materi fisika kelas X IPA SMA Oikoumene Manokwari dapat disimpulkan bahwa:

1. Validitas tingkat kesesuaian butir soal yaitu sebanyak 8 butir soal (60%) yang valid dan 7 butir soal (40%) yang tidak valid.
2. Reliabilitas peserta didik sebesar 0,42 bermakna lemah, reliabilitas butir soal sebesar 0,01 bermakna lemah dan reliabilitas antara peserta didik dengan butir soal 0,94 bermakna bagus sekali.
3. Tingkat kesukaran soal secara umum dalam kategori yang kurang baik dengan jumlah 6 butir soal (40%), karena soal tersebut lebih didominasi oleh soal yang berkriteria mudah yang artinya soal tersebut terlalu mudah untuk dikerjakan oleh peserta didik.
4. Daya pembeda pada soal dikatakan mempunyai daya pembeda yang cukup, karena terdapat 12 butir soal (80%) masuk dalam kriteria tersebut.
5. Efektivitas pengecoh soal dikatakan mempunyai kriteria yang baik, karena terdapat 8 butir soal (60%) yang *option* pengecohnya sudah berfungsi dengan baik.

Kesimpulannya bahwa analisis kualitas butir soal PAS menggunakan model Rasch dinyatakan valid, reliabel, tingkat kesukaran butir soalnya kurang baik, daya pembeda yang cukup, dan efektivitas pengecohnya sudah baik.

SARAN

Saran dalam penelitian ini yaitu butir soal yang sudah dibuat sebaiknya diujikan terlebih dahulu dan dianalisis untuk melihat kualitas dari soal tersebut sebelum digunakan. Sebaiknya pada saat pembagian soal, pendidik hendaknya membagi peserta didik secara acak supaya tidak saling bekerja sama dalam mengerjakan soal. Butir soal yang kualitasnya baik sebaiknya dimasukkan ke dalam bank soal supaya dapat digunakan pada tes-tes berikutnya. Sedangkan untuk butir soal yang kualitasnya belum baik dapat diperbaiki/direvisi dan diujikan kembali untuk melihat kualitasnya setelah diperbaiki. Jika kualitasnya menjadi baik maka butir soal tersebut dapat dimasukkan ke dalam bank soal.

DAFTAR PUSTAKA

- Andersen, E. B. (1973). A Goodness of Fit Test for the Rasch Model. *Psychometrika*, 38(1), 123–140.
- Arikunto, S. (2013). *Dasar-dasar Evaluasi Pendidikan*. PT Bumi Aksara.
- Peraturan Pemerintah Republik Indonesia N0. 19 Tahun 2005 tentang Standar Nasional Pendidikan, 1 (2005).
- Curtis, D. D., & Boman, P. (2007). X-ray your Data with Rasch. *International Education*

Journal, 8(2), 249–259.

- Liesfi, N. F. R. (2016). Analisis Butir Soal Ulangan Akhir Semester Ganjil Mata Pelajaran Teori Kejuruan Akuntansi. *Kajian Pendidikan Akuntansi Indonesia*, 5(6), 1–11. <https://journal.student.uny.ac.id/index.php/kpai/article/view/5757>
- 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. <https://doi.org/10.18637/jss.v020.i09>
- Marthunis, M., Khaldun, I., & Zulfadli. (2015). Analisis Kualitas Butir Soal Ujian Semester Genap Mata Pelajaran Kimia Kelas X MAN Model Banda Aceh Tahun Pelajaran 2014 / 2015 Menggunakan Program Proanaltes. *Jurnal Ilmiah Mahasiswa Pendidikan Kimia*, 1(4), 70–78.
- Prayetno, E. (2021). *Analisis Butir Soal Penilaian Tengah Semester pada Mata Pelajaran IPA Kelas VII SMP Yapis Manokwari Tahun Ajaran 2019/2020*. Unipa Papua.
- Rahayu, T. D., Purnomo, B. H., & 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), 39–43.
- Ratumanan, T. G., & Laurens, T. (2011). *Penilaian Hasil Belajar pada Tingkat Satuan Pendidikan*. 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. *Jurnal Sistem Kesehatan*, 2(1), 47–55. <https://doi.org/10.24198/jsk.v2i1.10419>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Bandung. Trim Komunikata.
- Widyaningsih, S. wahyu, & Yusuf, I. (2018). Analisis Soal Modul Laboratorium Fisika Sekolah I Menggunakan Racsh Model. *Gravity: Jurnal Ilmiah Penelitian Dan Pembelajaran Fisika*, 4(1), 13–46.