

ANALISIS BUTIR SOAL PENILAIAN TENGAH SEMESTER MENGGUNAKAN MODEL RASCH DI SMK NEGERI 5 SURABAYA

Ikhwa Nudin

S1 Pendidikan Teknik Mesin, Fakultas Teknik, Universitas Negeri Surabaya

Email : ikhwa.19017@mhs.unesa.ac.id

Rachmad Syarifudin Hidayatullah

Jurusan Teknik Mesin, Fakultas Teknik, Universitas Negeri Surabaya

Email : rachmadhidayatullah@unesa.ac.id

Abstrak

Pendidikan merupakan unsur yang sangat penting dalam mengembangkan potensi diri setiap manusia. dalam dunia pendidikan, pembelajaran terdiri atas tiga tahap, yakni tahap perencanaan dan tahap pelaksanaan serta tahap evaluasi. Sesuai dengan pembelajaran abad 21, yaitu berpikir kritis, berpikir kreatif dan inovatif, dan keterampilan komunikasi serta keterampilan berkolaborasi. Maka, guru harus mengembangkan keterampilan siswa yang relevan dengan abad 21, salah satunya termasuk keterampilan berpikir kritis. Jadi dalam penerapannya dengan melaksanakan metode pembelajaran *Problem Based Learning* (PBL) dan pertanyaan-pertanyaan yang digunakan dalam ujian harus megandung pertanyaan HOTS (*High Order Thinking Skill*). Penelitian ini termasuk penelitian berjenis *expost facto* dengan pendekatan deskriptif kuantitatif menggunakan model Rasch. Tujuan penelitian ini untuk mendeteksi kualitas dari butir soal pada Penilaian Tengah Semester yang menganalisis indeks kesulitan dan indeks daya beda, validitas dan reliabilitas dari butir soal yang diujikan yang bermanfaat untuk meningkatkan pengelolaan pelaksanaan evaluasi pembelajaran dan meningkatkan informasi tentang kemampuan siswa serta kualitas pertanyaan yang diberikan dengan model rasch. Dari analisis yang dilakukan dengan perangkat lunak winsteps 3.08.1 didapatkan bahwa instrumen butir soal pada Penilaian Tengah Semester Pemeliharaan Sasis dan Pindah Tenaga Kendaraan Ringan nilai logit *person reliability* sebesar 0.49 sedangkan *item reliability* nilai logit senilai 0.94 dan nilai logit *alpha cronbach* senilai 0.81. Sehingga instrumen butir soal dapat digunakan untuk mengevaluasi pembelajaran pada pengujian berikutnya.

Kata Kunci: analisis butir soal, evaluasi pembelajaran, model rasch.

Abstract

Education is a very important element in developing the potential of every human being. in the world of education, learning consists of three stages, namely the planning stage and the implementation stage and the evaluation stage. In accordance with 21st century learning, namely critical thinking skills, creative and innovative thinking skills, and communication skills and collaboration skills. Thus, teachers must develop student skills that are relevant to the 21st century, one of which includes critical thinking skills. So in its application by implementing the Problem Based Learning (PBL) learning method and the questions used in the exam must contain HOTS (High Order Thinking Skill) questions. This type of research an ex post facto type of research with a quantitative descriptive approach using the Rasch model. The purpose of this study was to detect the quality of the items in the Mid-Semester Assessment which analyzed the index of difficulty and index of discrimination, validity and reliability of the items tested which were useful for improving the management of implementation the learning evaluation and increasing information about students' abilities and the quality of the questions given with rasch model. From the analysis carried out with the winsteps 3.08.1 software, item instrument found that on the mid-semester assessment of light vehicle chassis and power shifter maintenance had a logit person reliability value of 0.49, while item reliability had a logit value of 0.94 and a logit alpha cronbach value of 0.81. So that the instrument items can be used for evaluate learning in the next test.

Keywords: item analysis, learning evaluation, Rasch model.

PENDAHULUAN

Pendidikan merupakan unsur yang sangat penting dalam mengembangkan dan membangun potensi diri setiap orang. Sebagaimana yang disebutkan dalam Undang-undang Republik Indonesia Nomor 20 tahun 2003 perihal Sistem Pendidikan Nasional pada pasal 1 mengemukakan bahwasanya pendidikan adalah usaha sadar dan terencana

untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara.

Dalam dunia pendidikan, pembelajaran memiliki tiga tahapan yakni tahap perencanaan, tahap pelaksanaan dan

tahap evaluasi. Tahap evaluasi mengukur sejauh mana tujuan pembelajaran tercapai (Erfan et al., 2020). Menurut (Arikunto, 2011) evaluasi merupakan kegiatan penilaian yang memiliki tujuan untuk mengukur keberhasilan pembelajaran. Pengukuran melibatkan aktivitas membandingkan suatu prediksi dengan suatu standar. Hasil dari pengukuran kemudian diinterpretasikan dan diperhitungkan dalam bentuk penilaian (Zainal, 2020). Proses penilaian melibatkan pengambilan keputusan untuk menentukan tujuan, hasil pembelajaran, umpan balik dari peserta didik (Ibarra-Sáiz et al., 2021). Sedangkan penilaian memiliki tujuan untuk melihat tingkat kemahiran dan pemahaman para peserta didik dalam menguasai materi yang telah diajarkan (Kalahatu, 2021). Evaluasi pendidikan adalah proses dari kegiatan pembelajaran yang tidak dapat dipisahkan (Li et al., 2021), karena kegiatan pembelajaran harus diikuti kegiatan penilaian (Mauliandri et al., 2021). Evaluasi bertujuan untuk mengidentifikasi upaya yang telah dilakukan dalam proses pembelajaran yang dilakukan. Evaluasi dilakukan untuk mengidentifikasi berhasil atau tidaknya pendidikan dalam mencapai tujuannya (Devi et al., 2022). Untuk mencapai tujuan pembelajaran, evaluasi harus dilakukan dengan sistematis dan berkesinambungan untuk menemukan proses pembelajaran yang baik (Muhammad et al., 2021). Dalam pembelajaran diperlukan evaluasi untuk mengetahui seberapa efektif kegiatan pembelajaran yang telah diimplementasikan pada peserta didik (UTAMI, 2021).

Tes merupakan teknik penilaian dimana siswa diminta untuk menyelesaikan berbagai butir atau tugas yang harus dikerjakan untuk menghasilkan nilai tentang perilaku peserta didik (Sawaluddin & Muhammad, 2020). Pada jenjang Sekolah Menengah Kejuruan, evaluasi pembelajaran dilakukan per semester yang diadakan dua kali, yaitu Penilaian Akhir Semester dan Penilaian Akhir Semester. Sesuai pembelajaran abad 21 yang mencakup keterampilan berpikir kritis, keterampilan berpikir kreatif dan inovatif, dan keterampilan komunikasi serta keterampilan berkolaborasi (Rosnaeni, 2021). Berpikir kritis ialah kemampuan untuk memprioritaskan pemikiran memproses dan mengolah informasi sehingga bisa dipertanggungjawabkan keabsahannya (Muhammad Santoso & Arif, 2021). Maka, guru harus mengembangkan keterampilan siswa yang relevan dengan pembelajaran abad 21, diantaranya adalah keterampilan berpikir kritis. Jadi dalam penerapan pembelajaran dengan menggunakan metode *Problem Based Learning* (PBL) dan pertanyaan-pertanyaan yang digunakan dalam ujian harus mengandung pertanyaan HOTS (*High Order Thinking Skill*). Dari kegiatan Pengenalan Lingkungan Persekolahan (PLP) butir-butir soal tidak dilakukan uji kualitas soal secara tertulis, ini menghasilkan penilaian

semu yang berarti bahwa kemampuan belajar yang sebenarnya tidak terukur.

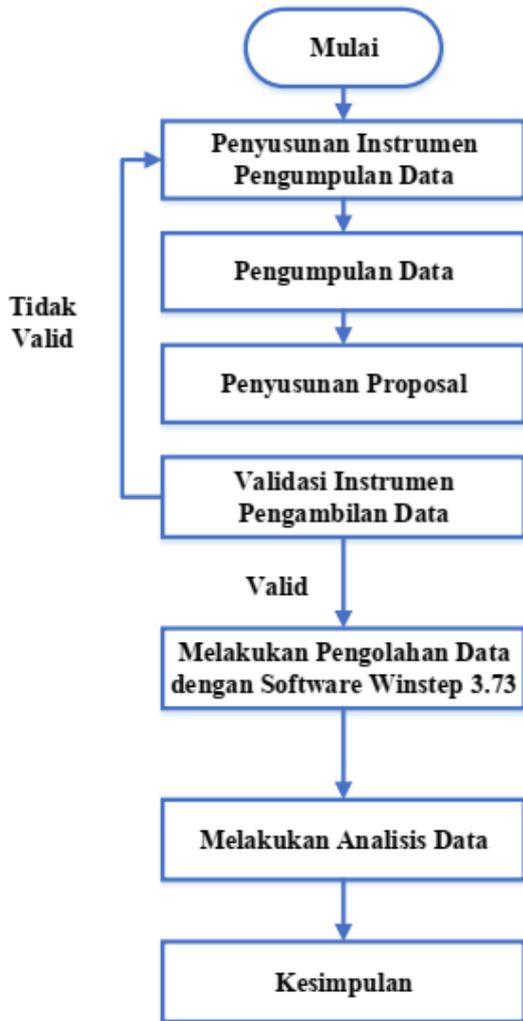
Analisis butir soal dapat digunakan sebagai informasi diagnostik untuk menentukan apakah siswa memahami apa yang telah dipelajari dan untuk meningkatkan kualitas soal dengan memperbaiki atau menghilangkan pertanyaan yang tidak valid (Fauziana & Dessy Wulansari, 2021). Penelitian ini memiliki tujuan untuk mendeteksi kualitas dari butir soal Penilaian Tengah Semester yang menganalisis validitas, reliabilitas, dan indeks kesulitan serta indeks daya beda dari butir soal yang diujikan yang bermanfaat untuk meningkatkan pengelolaan pelaksanaan evaluasi pembelajaran dan meningkatkan informasi tentang kemampuan siswa serta kualitas pertanyaan yang diberikan dengan model Rasch. Pemodelan Rasch memiliki kemiripan dengan model IPL yaitu pengukuran yang sama-sama menekankan pada tingkat kesulitan (Widhiarso, 2015).

Pendekatan model Rasch mempunyai pendekatan yang tidak sama dalam penggunaan nilai atau data mentah dari hasil ujian dalam ranah evaluasi pendidikan. Penggunaan pengukuran model Rasch pada data mentah dari hasil ujian memiliki tujuan supaya mendapatkan suatu skala penilaian menggunakan interval yang sama sehingga dapat memperlihatkan data yang akurat, baik kemampuan dari peserta didik maupun kualitas dari soal yang diberikan kepada peserta didik (Erfan et al., 2020).

METODE

Penelitian yang dilakukan termasuk dalam jenis penelitian *ex post facto* menggunakan pendekatan deskriptif kuantitatif dengan model Rasch untuk menganalisis validitas, reliabilitas, dan indeks kesulitan serta indeks daya beda dari butir soal yang diujikan. Penelitian kuantitatif memerlukan angka, dimulai dari pengumpulan data, dan penafsiran data, serta menampilkan hasil (Arikunto, 2002) untuk mendeskripsikan fakta-fakta yang ada, baik yang bersifat alamiah maupun perancangan manusia (Syaodih, 2010). Data penelitian diperoleh dari hasil Penilaian Tengah Semester kelas XI Teknik Manajemen dan Perawatan Otomotif (TMPO) SMKN 5 Surabaya pada mata pelajaran Pemeliharaan Sasis dan Pemindah Tenaga Kendaraan Ringan yang terdiri dari 61 siswa dengan metode *purposive sampling*. Teknik *sample purposive*, disebut juga pengambilan sampel penilaian, yaitu peserta dipilih dengan sengaja karena kualitas yang dimiliki (Etikan, 2016). Instrumen pengumpulan data berupa soal yang diberikan kepada peserta didik melalui lembar jawaban. Dalam soal terdapat 15 butir soal dengan kategori pilihan ganda. Dari hasil jawaban peserta didik, jika benar dalam menjawab soal maka diberi nilai 1. Dan apabila salah menjawab soal atau soal tidak dijawab maka mendapat nilai 0. Hal ini

memudahkan untuk di proses ke dalam analisis perangkat lunak, yaitu winsteps versi 3.08.1.



Bagan 1. Desain Penelitian

HASIL DAN PEMBAHASAN

Hasil dari penelitian yang dilakukan di SMK Negeri 5 Surabaya terdiri dari dua kelas XI dengan jumlah 61 siswa dengan mata pelajaran Pemeliharaan Sasis dan Pemindah Tenaga Kendaraan Ringan. Agar validitas instrument diketahui, maka dilakukan uji validitas pada tiga bidang, yaitu ahli konstruk, ahli materi, dan ahli Bahasa yang terdiri dari lima validator dari guru di jurusan otomotif pada masing-masing bidang dan beberapa guru bahasa di SMK Negeri 5 Surabaya. Dalam penelitian kualitatif, bagian hasil berisi bagian-bagian detail dalam wujud beberapa sub pokok yang terkait dengan fokus penelitian dan kategori-kategori.

Tabel 1. Hasil Uji Validator

Butir Soal	Validasi					
	Materi		Konstruk		Bahasa	
	Nilai	%	Nilai	%	Nilai	%
1	3,19	80%	3,54	89%	3,36	84%
2	3,21	80%	3,66	92%	3,42	86%
3	2,92	73%	3,64	91%	3,41	85%
4	3,23	81%	3,47	87%	3,47	87%
5	3,33	83%	3,41	85%	3,51	88%
6	3,14	79%	3,45	86%	3,43	86%
7	3,13	78%	3,39	85%	3,43	86%
8	3,19	80%	3,53	88%	3,43	86%
9	3,19	80%	3,46	87%	3,37	84%
10	3,14	79%	3,62	91%	3,53	88%
11	3,16	79%	3,43	86%	3,46	87%
12	3,13	78%	3,50	88%	3,50	88%
13	3,09	77%	3,58	90%	3,54	89%
14	2,67	67%	3,54	89%	3,46	87%
15	3,05	76%	3,54	89%	3,45	86%

Dari uji validasi ahli diambil dari nilai rata-rata setiap bidang dapat disimpulkan bahwa butir soal pada mata pelajaran Pemeliharaan Sasis dan Pemindah Tenaga Kendaraan Ringan dikategorikan valid dan bisa digunakan untuk pengujian selanjutnya. Kategori tersebut dapat dilihat dengan acuan tabel berikut.

Tabel 2. Skala Likert

Persentase	Interpretasi
0%-20%	Sangat Tidak Valid
21%-40%	Tidak Valid
41%-60%	Cukup Valid
61%-80%	Valid
81%-100%	Sangat Valid

(Riduwan, 2010)

Hasil jawaban Penilaian Tengah Semester (PTS) peserta didik diperoleh melalui lembar jawaban siswa melalui soal yang diberikan. Siswa yang mengerjakan soal akan dicatat dan diketahui apakah siswa menjawab dengan benar atau salah. Data respon siswa untuk menjawab Penilaian Tengah Semester (PTS) dianalisis menggunakan rash model dengan software winstep. Hasil analisis ditunjukkan pada gambar berikut.

TABLE 3.1 DATA Nilai.xlsx 201970WS.TXT Mar 28 2023 7:17
 INPUT: 61 PERSON 15 ITEM REPORTED: 61 PERSON 15 ITEM 2 CATS MINISTEP 5.4.1.0

SUMMARY OF 61 MEASURED PERSON

TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	7.3	15.0	-.13	.64	1.01	-.05	1.03
SEM	3.8	0	-.12	.01	-.05	-.16	-.10
P. SD	2.4	0	-.97	.07	.37	1.27	.79
S. SD	2.4	0	-.97	.07	.37	1.28	.79
MAX.	13.0	15.0	2.44	1.07	1.98	2.97	5.74
MIN.	1.0	15.0	-3.25	.60	-.56	-1.97	-.32

REAL RMSE .69 TRUE SD .67 SEPARATION .97 PERSON RELIABILITY .49
 MODEL RMSE .64 TRUE SD .72 SEPARATION 1.13 PERSON RELIABILITY .56
 S.E. OF PERSON MEAN = .12

PERSON RAW SCORE-TO-MEASURE CORRELATION = .99 (approximate due to missing data)
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .51 SEM = 1.66
 (approximate due to missing data)
 STANDARDIZED (50 ITEM) RELIABILITY = .81

SUMMARY OF 15 MEASURED ITEM

TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	29.5	61.0	0.00	.32	-.98	.10	1.03
SEM	3.8	0	1.35	.01	-.04	-.31	-.09
P. SD	14.2	0	1.31	.05	-.14	1.17	.33
S. SD	14.7	0	1.35	.05	-.15	1.22	.35
MAX.	50.0	61.0	2.40	.45	1.25	2.49	1.76
MIN.	6.0	61.0	-1.89	.28	-.69	-2.47	-.59

REAL RMSE .33 TRUE SD 1.26 SEPARATION 3.81 ITEM RELIABILITY .94
 MODEL RMSE .33 TRUE SD 1.27 SEPARATION 3.89 ITEM RELIABILITY .94
 S.E. OF ITEM MEAN = .35

ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00 (approximate due to missing data)
 Global statistics: please see table 44.
 UNMEAN=.0000 USCALE=1.0000

Gambar 1. Summary Statistics Person

Berdasarkan analisis data menggunakan software winstep, diperoleh nilai penting mengenai *person reliability*, *item reliability*, *nilai cronbach alpha*, dan *peso measure*. Nilai yang diperoleh ditunjukkan dari tabel dibawah ini.

Tabel 3. Summary Statistics

	Keterangan	Nilai
Logit	Person	0.13
	Item	0
Reliabilitas	Person Reliability	0.49
	Item Reliability	0.94
	Alpha Cronbach	0.81
Outfit MNSQ	Peson	1.03
	Item	1.03
Outfit ZSTD	Person	0.09
	Item	0.14

Berdasarkan tabel 3. diatas diperlihatkan nilai logit dari *person* atau *measure* senilai 0.13 dan nilai *item measure* senilai 0 yang berarti nilai *person measure* lebih tinggi daripada nilai *item measure*. Dapat disimpulkan bahwasanya abilitias peserta didik lebih tinggi daripada level kesulitan soal. Dengan kata lain, adanya potensi siswa menjawab dengan benar pada semua soal. Sehingga peserta didik yang mempunyai abilitas yang tinggi mampu menjawab dengan benar pada soal yang paling sulit dan paling mudah. Disisi lain, *item reliability* senilai 0.94 dan *person reliability* senilai 0.49 serta *Alpha Cronbach* senilai 0.81. Dilihat dari nilai di atas dapat disimpulkan bahwasanya tingkat konsistensi dari jawaban peserta didik cukup tinggi serta kualitas instrumen butir soal memiliki nilai reliabilitas senilai 0.94.

Selain itu, pada tabel diatas dipaparkan nilai *Outfit Mean Squared (Outfit MNSQ)* senilai 1.03 baik pada

kolom *person* maupun pada kolom *item*. Nilai 1.03 tergolong pada kriteria fit yaitu berada diantara $0,5 < MNSQ < 1,5$. Menunjukkan instrumen tes yang digunakan pengujian sesuai dengan model untuk mengukur kompetensi peserta didik. Selanjutnya, nilai *Outfit Z Standarized (Outfit ZSTD)* sebesar 0.09 pada kolom *person* dan 0.14 pada kolom *item*. Nilai 0.09 pada kolom *person* dan 0.14 pada kolom *item* berada diantara $-0,2 < ZSTD < 0,2$ yang berarti data mempunyai potensi nilai yang rasional. Dapat diambil bahwasanya keseluruhan butir soal dapat digunakan sebagai instrumen tes pada pengujian selanjutnya karena telah sesuai dengan model rasch.

Distribusi butir soal dikatakan tidak fit dengan model bisa dilihat pada gambar 2. Di bawah, batas butir soal dikatakan fit dengan model apabila mencakup baik salah satu ataupun kedua syarat berikut ini. Pertama, nilai Outfit MNSQ berada diantara 0,5 sampai 1,5; nilai Outfit ZSTD berada diantara -2,0 sampai 2,0; dan nilai kolerasi butir dengan total skor terletak diantara 0,4 sampai 0,85(Sumintono, 2018).

	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.
	1.02	.16	1.76	1.27	A .16	.24
	1.11	1.24	1.44	2.66	B .23	.38
	1.25	2.49	1.35	1.93	C .11	.37
	1.03	.22	1.33	.83	D .19	.28
	1.07	.65	1.22	1.03	E .25	.35
	1.17	1.81	1.17	1.12	F .22	.38
	1.07	.77	1.05	.31	G .30	.37
	1.02	.20	1.02	.15	H .36	.37
	.96	-.09	.80	-.49	g .41	.36
	.96	-.44	.90	-.66	f .44	.38
	.87	-.87	.78	-.98	e .52	.38
	.86	-.79	.77	-.88	d .51	.38
	.85	-.65	.63	-1.13	c .53	.36
	.81	-.77	.59	-.96	b .48	.28
	.69	-2.47	.62	-2.06	a .67	.38

Gambar 2. Misfit Order

Dari gambar diatas dapat dilihat dari syarat dinyatakan fit dengan model, sehingga instrumen dapat digunakan dengan catatan diganti atau direvisi pada beberapa soal yang belum memenuhi ketiga syarat. Beberapa butir soal yang tidak memenuhi beberapa syarat diatas dapat dimungkinkan adanya butir soal yang mudah ditebak oleh peserta didik karena ada kesamaan karakteristik butir-butir soal. Untuk tingkat kesulitan butir soal bisa dilihat pada tabel di bawah ini:

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.
2	6	61	2.40	.45
6	9	61	1.89	.38
14	9	61	1.89	.38
5	20	61	.72	.29
12	24	61	.39	.28
4	25	61	.32	.28
3	26	61	.24	.28
9	30	61	-.07	.28
15	30	61	-.07	.28
11	32	61	-.23	.28
8	42	61	-1.04	.30
13	44	61	-1.23	.31
7	46	61	-1.43	.32
1	50	61	-1.89	.36
10	50	61	-1.89	.36
MEAN	29.5	61.0	.00	.32
P. SD	14.2	.0	1.31	.05

Gambar 3. Measure Order

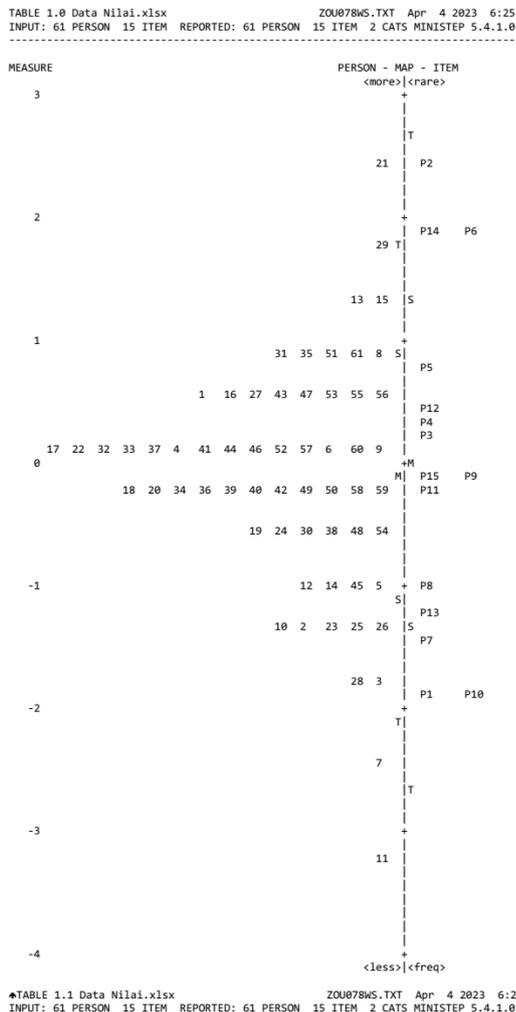
Dari tabel diatas menunjukkan nilai dari setiap butir soal dari yang tertinggi sampai terendah. Untuk butir soal nomor 2 memiliki nilai tertinggi sebesar 2.40 yang mengartikan bahwa butir soal nomor 2 termasuk soal yang paling sulit. Sedangkan butir soal nomor 10 mempunyai nilai terendah sebesar -1.89 yang mengartikan bahwa butir soal nomor 10 termasuk soal paling mudah.

GUTTMAN SCALOGRAM OF RESPONSES:

PERSON	ITEM	
	1 1 1 1 1 1	
	107381953425642	
21	+111111110110111	21
29	+111111011101110	29
13	+111111101110010	13
15	+111111101011010	15
8	+111111110110000	8
31	+111111101011000	31
35	+111111101011000	35
51	+111111101011000	51
61	+111111101011000	61
1	+110111101010010	1
16	+111111100000011	16
27	+111001001101110	27
43	+111110110101000	43
47	+111111100110000	47
53	+111111110100000	53
55	+111110011001001	55
56	+111111111000000	56
4	+111111110000000	4
6	+110111111010000	6
9	+101111111000000	9
17	+011011000101110	17
22	+110101110001100	22
32	+111110011000100	32
33	+111110011001000	33
37	+111111010100000	37
41	+111111010100000	41
44	+111111010100000	44
46	+111110001010001	46
52	+111110011001000	52
57	+101010110101010	57
60	+111110011000100	60
18	+110001110001100	18
20	+101110110100000	20
34	+111111010000000	34
36	+111110001010000	36
39	+111000010111000	39
40	+111110001010000	40
42	+111110001010000	42
49	+111010010101000	49
50	+111110011000000	50
58	+111110001010000	58
59	+111111100000000	59
19	+100100101010100	19
24	+000100101110100	24
30	+111110001000000	30
38	+101010011010000	38
48	+011010001110000	48
54	+111110000010000	54
5	+010001110100000	5
12	+010001110001000	12
14	+110101100000000	14
45	+111001000100000	45
2	+100100100100000	2
10	+011100010000000	10
23	+010001000110000	23
25	+001101000000100	25
26	+000000110101000	26
3	+101000001000000	3
28	+010000000001001	28
7	+110000000000000	7
11	+000000100000000	11
	1 1 1 1 1 1	
	107381953425642	

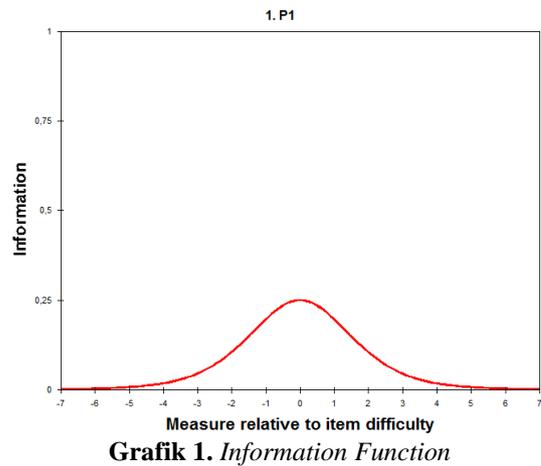
Gambar 4. Guttman Scalogram

Dari gambar diatas dapat dilihat *respon error* yang ditunjukkan subjek. Dari hasil jawaban peserta didik dengan kode 11 memiliki kemampuan rendah, namun mampu menjawab soal dengan kode 9 yang sedang, sedangkan untuk soal dengan kode 10 yang termasuk mudah justru salah dalam menjawab. Dapat diidentifikasi terdapat tebakan dari peserta didik kode 11 pada soal kode 9, sehingga dapat disimpulkan peserta didik kode 11 menjawab dengan benar soal kode 9 hanya dengan menebak. Disisi lain, berlandaskan gambar diatas maka dapat dilihat siapa yang lebih unggul diantara peserta didik kode 13 dan 15. Jika hanya melihat jawaban yang benar maka peserta didik kode 13 dan 15 memiliki kemampuan yang sama. Tapi jika dilihat peserta didik dengan kode 15 benar daalam mengerjakan soal yang lebih sulit, sehingga dapat disimpulkan peserta didik kode 15 dapat dikatakan lebih pandai dari peserta didik kode 13. Jadi, kemampuan seorang diri tidak hanya diketahui dari nilai mentahannya, namun juga mempertimbangkan tingkat kesulitan butir soal yang dipandah sulit, sedang atau mudah dalam mengerjakan(Rasmuin & Luddin, 2022).



Gambar 5. Wright Map

Gambar di atas menunjukkan daya pembeda setiap soal dan peserta didik yang diurutkan mulai dari soal paling sulit ke soal paling mudah dan siswa yang bisa menjawab soal paling sulit sampai yang bisa menjawab soal yang paling mudah. Dari nilai analisis diatas dapat diketahui bahwasanya peserta didik dengan kode 21 mempunyai level kemampuan yang tinggi, sedangkan peserta didik dengan kode 11 memiliki level kemampuan yang rendah. Selain itu dapat dilihat soal yang mempunyai tingkat kesulitan dalam kategori sulit adalah soal dengan kode P2, P14, dan P6. Sedangkan soal yang mempunyai tingkat kesulitan dalam kategori sedang adalah soal dengan kode P5, P12, P4, P3, P15, P9, dan P11. Kemudian soal yang mempunyai tingkat kesulitan dalam kategori mudah adalah soal dengan kode P8, P13, P7, P1, dan P10.



Dari gambar diatas menunjukkan informasi pengukuran yang diperoleh dari Penilaian Tengan Semester (PTS) mata pelajaran Pemeliharaan Sasis dan Pemindah Tenaga Kendaraan Ringan. Sumbu X menampilkan tingkat kemampuan peserta didik dalam melakukan tes. Sedangkan sumbu Y menampilkan angka fungsi informasi. Mengacu pada grafik diatas, diperoleh informasi pengukuran dengan sangat tinggi pada tingkat kemampuan sedang. Jadi instrumen Penilaian Tengah Semester untuk mata pelajaran Pemeliharaan Sasis dan Pemindah Tenaga Kendaraan Ringan cocok apabila diaplikasikan pada peserta didik dengan kemampuan sedang.

PENUTUP

Simpulan

Instrumen butir soal mata pelajaran Pemeliharaan Sasis dan Pemindah Tenaga Kendaraan Ringan fit dengan model rasch. Hal tersebut ditunjukkan dengan nilai logit dari person atau measure senilai 0.13 dan nilai item measure senilai 0. Dapat disimpulkan bahwasanya

abilitas siswa cenderung lebih tinggi daripada tingkat kesulitan soal. Sementara itu *item reliability* senilai 0.94 dan *person reliability* senilai 0.49 dan *Alpha Cronbach* senilai 0.81. dari nilai tersebut dapat diartikan bahwasanya tingkat konsistensi jawaban peserta didik cukup tinggi dan kualitas instrumen butir soal memiliki nilai reliabilitas senilai 0.94. Angka lain yang ditampilkan pada tabel diatas yaitu nilai *Outfit Mean Squared (Outfit MNSQ)* senilai 1.03 baik pada kolom *person* ataupun *item*. Dapat diartikan bahwasanya instrumen tes yang digunakan sesuai dengan model untuk mengukur kompetensi peserta didik. Selanjutnya, nilai *Outfit Z Standarized (Outfit ZSTD)* senilai 0.09 pada kolom *person* dan 0.14 pada kolom *item* yang berarti data mempunyai kemungkinan nilai yang rasional. Hal ini dapat disimpulkan bahwa butir soal sudah sesuai dengan model rasch dan dapat digunakan instrumen tes pada mata pelajaran Pemeliharaan Chasis dan Pemindah Tenaga Kendaraan Ringan. Dilihat dari syarat dinyatakan fit dengan model, maka soal nomor 9 adalah tidak fit atau misfit, sehingga pertanyaan butir soal nomor 9 dapat di ganti atau di revisi. Untuk butir soal nomor 2 memiliki nilai tertinggi sebesar 2.40 yang mengartikan bahwa butir soal nomor 2 termasuk soal yang paling sulit. Sedangkan butir soal nomor 10 memiliki nilai terendah sebesar -1.89 yang mengartikan bahwa butir soal nomor 10 termasuk soal yang paling mudah.

Saran

- Kepada guru sebagai pengajar supaya terbiasa untuk menganalisis butir soal sebelum dan sesudah soal diujikan kepada peserta didik, supaya guru dapat mengevaluasi apakah soal yang diujikan sudah sesuai berdasarkan tujuan pembelajaran dengan baik atau tidak.
- Butir soal yang diteliti dinyatakan sangat valid, maka dari itu, butir soal tersebut dapat digunakan untuk pengujian Penilaian Tengah Semester (PTS) selanjutnya. Dengan beberapa catatan soal yang tidak valid supaya diganti atau dibetulkan.

DAFTAR PUSTAKA

- Arikunto, S. (2002). *Prosedur Penelitian suatu pendekatan praktik*. PT Asdi Mahasatya.
- Arikunto, S. (2011). *Prosedur Penelitian : Suatu Pendekatan Praktik*. Rineka Cipta.
- Devi, M. Y., Hidayanthi, R., & Fitria, Y. (2022). Model-Model Evaluasi Pendidikan dan Model Sepuluh Langkah dalam Penilaian. *Jurnal Basicedu*, 6(1), 675–683. <https://doi.org/10.31004/basicedu.v6i1.1934>
- Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). ANALISIS KUALITAS SOAL KEMAMPUAN MEMBEDAKAN RANGKAIAN SERI DAN PARALEL MELALUI TEORI TES KLASIK DAN MODEL RASCH. *Indonesian Journal Of Educational Research and Review*, 3(1), 11. <https://doi.org/10.23887/ijerr.v3i1.24080>
- Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fauziana, A., & Dessy Wulansari, A. (2021). Analisis Kualitas Butir Soal Ulangan Harian di Sekolah Dasar dengan Model Rasch. *Ibriez : Jurnal Kependidikan Dasar Islam Berbasis Sains*, 6, 10–19. <https://doi.org/10.21154/ibriez.v6i1.112>
- Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., & Boud, D. (2021). The quality of assessment tasks as a determinant of learning. *Assessment & Evaluation in Higher Education*, 46(6), 943–955. <https://doi.org/10.1080/02602938.2020.1828268>
- Kalahatu, M. F. (2021). PERSEPSI PESERTA PELATIHAN DASAR TERHADAP PENGGUNAAN QUIZZZ SEBAGAI METODE EVALUASI PEMBELAJARAN. *Akademika*, 10(01), 163–178. <https://doi.org/10.34005/akademika.v10i01.1228>
- Li, T., Yeung, M., Li, E., & Leung, B. (2021). How formative are assessments for learning activities towards summative assessment? *International Journal of Teaching and Education*, 9(2), 42–57. <https://doi.org/10.52950/TE.2021.9.2.004>
- Mauliandri, R., Maimunah, M., & Roza, Y. (2021). Kesesuaian Alat Evaluasi Dengan Indikator Pencapaian Kompetensi Dan Kompetensi Dasar Pada RPP Matematika. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, 5(1), 803–811. <https://doi.org/10.31004/cendekia.v5i1.436>
- Muhammad, M., Widyaningrum, H. K., Masjid, A. Al, Komariah, K., & Sumarwati, S. (2021). Pelaksanaan Prosedur Evaluasi Pembelajaran Bahasa Indonesia di SMK Pekanbaru pada Masa Pandemi. *Stilistika: Jurnal Pendidikan Bahasa Dan Sastra*, 14(2), 109. <https://doi.org/10.30651/st.v14i2.8262>
- Muhammad Santoso, A., & Arif, S. (2021). Efektivitas Model Inquiry dengan Pendekatan STEM Education terhadap Kemampuan Berfikir Kritis Peserta Didik. *Jurnal Tadris IPA Indonesia*, 1(2), 73–86. <https://doi.org/10.21154/jtii.v1i2.123>
- Rasmuin, R., & Luddin, S. (2022). Tingkat Kesulitan Soal Buatan Guru Bidang Studi Matematika Menurut Teori Tes Klasik pada Tingkat SMP di Kota Baubau. *Jurnal Akademik Pendidikan Matematika*, 33–40. <https://doi.org/10.55340/japm.v8i1.699>

- Riduwan. (2010). *METODE DAN TEKNIK MENYUSUN PROPOSAL PENELITIAN*. CV ALFABETA.
- Rosnaeni, R. (2021). Karakteristik dan Asesmen Pembelajaran Abad 21. *Jurnal Basicedu*, 5(5), 4341–4350.
<https://doi.org/10.31004/basicedu.v5i5.1548>
- Sawaluddin, S., & Muhammad, S. (2020). Langkah-Langkah dan Teknik Evaluasi Hasil Belajar Pendidikan Agama Islam. *Jurnal PTK Dan Pendidikan*, 6(1).
<https://doi.org/10.18592/ptk.v6i1.3793>
- Sumintono, B. (2018). Rasch Model Measurements as Tools in Assesment for Learning. *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017)*.
<https://doi.org/10.2991/icei-17.2018.11>
- Syaodih, N. (2010). *METODE PENELITIAN PENDIDIKAN*. PT REMAJA ROSDAKARYA.
- UTAMI, L. W. S. (2021). PENGGUNAAN GOOGLE FORM DALAM EVALUASI HASIL BELAJAR PESERTA DIDIK DI MASA PANDEMI COVID-19. *TEACHING: Jurnal Inovasi Keguruan Dan Ilmu Pendidikan*, 1(3), 150–156.
<https://doi.org/10.51878/teaching.v1i3.453>
- Widhiarso, B. S. & W. (2015). *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Trim Komunikata.
- Zainal, N. F. (2020). Pengukuran, Assesment dan Evaluasi dalam Pembelajaran Matematika. *Laplace: Jurnal Pendidikan Matematika*, 3(1), 8–26. <https://doi.org/10.31537/laplace.v3i1.310>

