

Analisis Butir Soal Ulangan Harian Siswa Mata Pelajaran PPKn Kelas IV Menggunakan Rach Model di Madrasah Ibtidaiyah Negeri

Fitria Rifana^{*1}, Syahrul Ramadhan², Khamim Zarkasih Putro³

¹ Universitas Sunan Kalijaga Yogyakarta

¹ Universitas Sunan Kalijaga Yogyakarta

¹ Universitas Sunan Kalijaga Yogyakarta

e-mail: 21204082031@student.uin-suka.ac.id, syahrul.ramadhan@uin-suka.ac.id,

khamim.putro@uin-suka.ac.id

Submitted: 20-12-2023

Revised : 02-04-2024

Accepted: 17-05-2024

ABSTRACT. The study aimed to analyze the items on students' daily test questions on PPKN subjects made by the teacher to measure learning outcomes during the previous lesson process. Conducted on fourth-grade students at the 01 West Pasaman State Madrasah Ibtidaiyah. The method used in this research is descriptive quantitative using Rasch model data analysis with Win Step software. The results of the data analysis carried out were on the reliability data analysis using Winstep Cronbach Alpha 0.49 software. From the existing data, it is clear that the level of reliability of the questions is still relatively weak. The results of the Rasch model of person reliability and item reliability are 0.47 and 0.52. The results of the analysis of the quality of the items in terms of the validity of the items using the Rasch model obtained 6 questions that were said to be valid and 24 questions that were said to be invalid because they did not meet the requirements for the MNSQ outfit, ZSTD outfit, and the Point Measure correlation (PT measure Corr). In using the Rasch model it is felt to be more accurate because it meets these three criteria, so it is considered valid. above the P value or the level of difficulty of the items below 0.25, there are 3 items, meaning the item is very difficult. Questions that are above 0.75 have 5 items meaning the question is too easy. And questions that are scored between 0.25-0.75 have 17 questions. That is, 18 questions have a good level of difficulty. Analysis of distracting items using the Winstep Rasch Model software for 25 daily test questions for students of Civics class IV at Madrasah Ibtidaiyah Negeri 01 Pasaman Barat shows that there are 23 questions with good functioning distractors. Items where the distractor does not function properly are 5 questions.

Keywords: *Ulangan Harian, Butir Soal Ulangan, Rach Model, Kewarganegaraan, Madrasah Ibtidaiyah*

 <https://doi.org/10.54069/attadrib.v7i1.424>

How to Cite Rifana, F., Ramadhan, S. ., & Putro, K. Z. . (2024). Analisis Butir Soal Ulangan Harian Siswa Mata Pelajaran PPKn Kelas IV Menggunakan Rach Model di Madrasah Ibtidaiyah Negeri. *Attadrib: Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, 7(1), 99–110. <https://doi.org/10.54069/attadrib.v7i1.424>

PENDAHULUAN

Pembelajaran dilakukan dengan adanya interaksi peserta didik dengan pendidik akan memiliki kaitan satu dengan lainnya, pendidik harus mampu melakukan perubahan terhadap siswa sehingga siswa memiliki pengetahuan yang baik terhadap sesuatu hal sehingga memberikan hasil baik terhadap siswa (Adimsyah et al., 2023; Fauzi & Masrupah, 2024; Komarodin & Rofiq, 2023; Mardapi, 2017). Dengan demikian pendidik yang menjadi tempat utama bagi siswa dalam pendidikan, tidak hanya memberikan pembelajaran namun pendidik dapat melakukan pengevaluasian terhadap proses belajar siswa (Apriliani et al., 2024; Arif et al., 2023; Widyaningsih & Yusuf, 2018).

Melihat hasil dari kualitas peserta didik dapat dilihat melalui *assessment* atau disebut juga dengan penilaian. *Assesment* menjadi sebuah pedoman pendidik dalam melihat capaian tujuan pembelajaran. Menurut Russel & Airasian menjelaskan tentang penilaian yang menjadi poin utama pada pelajaran (Sumarni, 2019). *Assesmen* adalah suatu komponen pada proses pelajaran yang memiliki keterkaitan pada pelajaran.

Asesmen dalam pembelajaran yang ada pada madrasah menjadi sebuah hal yang memiliki penilaian yang sebagai penilaian *formatif* dan *sumatif*. Penilaian yang di berikan pendidik terhadap siswa bertujuan untuk mendapatkan hasil yang lebih baik kedepannya disebut penilaian *formatif*. Penilaian yang dilakukan dengan memberikan kepada siswa pembelajaran yang mendapatkan hasil yang baik pada pelajaran yang memberikan sebuah pelajaran dengan adanya umpan balik bagi peserta didik. Pelajaran yang diberikan menjadikan peserta didik lebih memahami pelajaran yang sebelumnya (Azis et al., 2022; Pereira et al., 2016). Pendidik memberikan pengetahuan kepada peserta didik dengan pelajaran yang mendapatkan pengetahuan selanjutnya kepada peserta didik. Asesmen dilaksanakan dalam pembelajaran yang sudah dikuasai peserta didik selama pembelajaran yang telah di lakukan selama pelajaran berlangsung disebut dengan *sumatif* (Allur et al., 2018). Peserta didik di berikan penilaian yang dapat mengukur hasil belajar peserta didik seperti halnya dengan memberikan ulangan harian seperti halnya ulangan harian sebagai uji coba selama pembelajaran sebelumnya, ujian semester ganjil dan semester genap (Agus et al., 2023; Munir et al., 2022).

Penilaian yang dilaksanakan kepada peserta didik untuk mendapatkan hasil belajar peserta didik agar mengetahui pemahaman peserta didik terhadap pembelajaran yang telah ada selama pembelajaran atau semester, dimana peserta didik diberikan penilaian untuk mengetahui tingkat prestasi antar peserta didik. Dari hasil penilaian ulangan/ujian yang telah dilaksanakan gunakan oleh guru sebagai berikut: a) menjadikan penilaian antar siswa yang satu dengan yang lain dalam melihat prestasi siswa dengan ulangan/ujian yang sama dilaksanakan. b) untuk mengetahui pembelajaran yang telah dipelajari siswa pada waktu yang lama dipengetahuan dan keterampilannya. c) untuk mengetahui pemahaman peserta didik terhadap suatu materi didalam pembelajaran. d) untuk memahami potensi peserta didik yang dapat di ukur untuk masa selanjutnya (Suryani, 2017). Penilaian dapat menjadi sebuah acuan bagi pendidik dalam memahami kemampuan dan potensi peserta didik selama proses belajar di kelas berlangsung. Teknik dalam penilaian terbagi menjadi 2 yaitu tes dan non tes. Tes adalah alat menjadi tempat pengumpulan pengetahuan dalam pencapaian tujuan pelajaran dan pembelajaran (Azizah et al., 2023; Fitria et al., 2022; Wahyudi, 2010). dalam kata lain tes juga adalah langkah yang salah satu terdapat dalam penilaian yang dapat mengukur hasil belajar peserta didik dalam pembelajaran (Kadir, 2015).

Penilaian tes mempunyai tiga cakupan, sebagai berikut : isi teks memiliki kategori pada materi dengan diuji (*validitas isi*), *validitas konstruk* pada penulisan teks dengan baik, selanjutnya *reabilitas*. Tes mengatakan keribelian dapat menggunakan pengukuran pada tahap, pada siswa satu juga yang lain, namun mendapatkan hasil yang sama (Lafrarchi, 2020; Retnawati, 2014).

Penilaian non tes adalah adalah penilaian pada hasil belajar siswa dengan melaksanakan pengamat, observasi, dan lain-lain pada penilaian yang dilakukan terhadap siswa di sekolah (Mania, 2008). Tes berupa soal diujikan pendidik kepada peserta didik. Salah satu pembentukan soal ialah soal pilahan ganda yang dilakukan oleh pendidik untuk menilai ketercapaian tujuan pembelajaran di tingkat sekolah dasar. Soal pilihan ganda adalah digunakan dalam tes pada mengidentifikasi secara keseluruhan materi pelajaran (Fernanda & Hidayah, 2020). Suatu tes dikatakan baik atau tidaknya dapat dilihat dari segi *validitas* dan *reabilitas* (Wahyuningsih, 2015).

Soal-soal dapat dinyatakan valid ketika soal memiliki pengukuran yang dapat terukur pada hasil pembelajaran yang ditentukan seperti pengetahuan, keterampilan dan sikap sehingga soal yang ada meliputi tiga komponen tersebut (Sumintono & Widhiarso, 2015). Data dikatakan baik ketika kevalidatan dan *kreabilitas* pada instrumen sudah mencapai nilai yang tinggi dalam sebuah penelitian yang dilakukan (Hayati & Lailatussaadah, 2016).

Berdasarkan wawancara peneliti dengan pendidik di Madrasah Ibtidayah Negeri 01 Pasaman Barat bahwasannya tes yang dibuat pendidik belu melakukan pengujian secara empiri sehingga belum bisa menilai kemampuan peserta didik sebenarnya. Dalam melakukan analisis soal dapat dilakukan dalam dunia pendidikan. Pendekatan moderen dengan permodelan rasch. Menelaah butir soal pada jawaban siswa dari hasil menjawab soal yang telah diberikan pendidik dapat melihat tingkat kualitas butir soal dengan penggunaan teori tes klasik yang ada (Retnawati, 2014). Penilaian yang dilakukan pada teori tes klasik dapat berubah-ubah sesuai dengan kompetensi dan kualitas peserta didiknya dalam mengerjakan soal-soal yang diberikan.

Permodelan rasch merupakan salah satu pendekatan yang ada untuk mengetahui skor yang digunakan dalam data ujian yang belum secara penuh terselesaikan sepenuhnya yang masih dapat dilakukan perbaikan terhadap data yang telah ada (Erfan et al., 2020). Penerapan model ras model menjadikan sebuah hasil penilaian yang dilakukan memiliki nilai yang sangat baik. (Talib et al., 2018). Pemodelan rasch juga memiliki pemilihan bentuk butir soal dalam membuat tes yang menjadikan sebuah tes lebih bermakna juga mempunyai sebuah kualitas dalam keperluan pada sebuah tujuan pengujiannya (Sumintono & Widhiarso, 2015). Model rasch dalam penganalisis butir soal yang ada dapat menjadika siswa memiliki komponen dan kebutuhan yang sama sesuai potensi yang dimiliki siswa (Sumintono & Widhiarso, 2015).

Komponen kesukaran butir soal dan daya beda menjadi hal yang mendukung baik dan kurang baiknya sebuah kualitas soal yang ada (Wahyuningsih, 2015). Kesukaran butir soal adalah skala yang ada pada siswa dalam penentuan menjawab soal yang benar dengan penjumlahan di seluruh peserta tes yang ada (Rahayu & Djazari, 2016). Daya pembeda merupakan kecakapan pada butir soal dalam perbedaan komponen pada sudut pandang dalam mengukur kontras terdapat pada suatu komponen tersebut (Apriliani et al., 2024; Bagiyono, 2017; Fitriani et al., 2024). Soal yang mempunyai daya beda baik ketika dapat melihat sebuah perbedaan antara kompetensi yang tinggi dan rendah dari siswa tersebut (Sumintono & Widhiarso, 2015). Pemodelan rasch dapat dikatakan memiliki kualitas jika memenuhi sesuai dengan kebutuhan dan tujuan yang dilakukan pada pengujian. Salah satu cirri-cirinya memiliki potensi yang dapat mempengaruhi kemampuan yang ada pada tes (Sumintono & Widharso, 2015).

Analisis butir soal dilakukan dengan adanya programan di dalam computer untuk melihat seberapa memiliki peningkatan baik atau tidaknya dalam suatu penganalisisan didalam program tersebut yang dilakukan dengan menganalisis soal dan tes yang ada, sehingga dapat mengukur adanya potensi dan kualitas sebuah soal dan juga kualita siswa. Dimana guru juga dapat melihat dari menganalisis tersebut seberapa layaknya soal yang dibuat untuk diberikan kepada siswa. menganalisis soal yang digunakan tidak hanya dengan soal pilihan ganda namun juga dapat dilakukan pada soal essay yang telah dibuat guru maupun sebuah instansi atau sekolah. Dari hal tersebut seorang guru dan sekolah dapat melihat kualitas soal yang telah digunakan dalam penilaian hasil belajar siswa yang telah dilaksanakan di waktu belajar sebelumnya yang menjadi sebuah patokan bagi guru dalam melihat progres siswa pada pembelajaran (Sumintono & Widharso, 2015).

Berdasarkan pernyataan diatas dapat disimpulkan dalam pengujian rasch model dilakukan pengujian validitas, reliabilitas, tingkat kesukaran serta daya beda soal guna untuk melihat seberapa berkualitasnya sebuah instrumen tes yang telah dibuat untuk melihat kapasitas soal yang digunakan. Soal yang memiliki kategori sangat mudah akan menjadikan siswa tidak terlalu memiliki potensi tinggi dalam menjawab soal yang ada dikarenakan siswa akan merasakan soal yang diberikan sudah biasa dan tingkatan soalnya tidak terlalu menantang (Perdana, 2018). Sedangkan pada soal yang terlalu sulit membuat siswa menjadi tidak bersemangat dalam menjawab soal dikarenakan siswa akan merasakan soal yang dibuat terlalu susah dan menjadikan siswa putus asa karena merasa soal tersebut terlalu sangat susah untuk di pahami sesuai dengan pemahaman siswa (Iskandar & Rizal, 2018).

METODE PENELITIAN

Peneliti menggunakan penelitian deskriptif kuantitatif. Subyek penelitian yang dilakukan adalah siswa Madrasah Ibtidayah Negeri 01 Pasaman Barat dengan ulangan harian, siswa kelas IV pada mata pelajaran PPKn peserta didik. Analisis butir soal dilakukan pada soal guru yang telah dibuat dalam penilaian hasil belajar siswa di Madrasah Ibtidayah Negeri 01 Pasaman Barat yang mengajar di kelas IV pada saat ini. variabel yang diteliti adalah kemampuan kognitif peserta didik pada ulangan harian siswa instrument tes digunakan sebanyak 25 butir yang memiliki empat pilihan jawaban. Penggunaan soal berbentuk pilihan ganda akan menjadi mudah melihat respon peserta didik. Analisis instrument tes dilakukan dengan menggunakan software Winstep dengan rasch model. Aspek penilaian deskriptif kuantitatif diantaranya Validitas isi, reabilitas, daya beda soal, tingkat kesukaran, dan fungsi distractor (pengecoh).

Teknik pengambilan data digunakan pada tes jenis, pilihan ganda (*multiple choice*) pada empat pilihan jawaban. Menggunakan soal yang dibuat oleh pendidik mata pelajaran PPKn berjumlah 25 butir soal. Setelah diujikan kepada peserta didik butir soal dianalisis menggunakan teori klasik dengan analisis permodelan rasch digunakan *software winstep*. (Arikunto 2016)

HASIL DAN PEMBAHASAN

Hasil

Reabilitas

Reliabilitas merupakan kecakapan yang dapat memberikan hal dengan konsisten dan adanya kestabilan pada pengukuran yang dilakukan pada jangka waktu beda namun dengan materi yang sejenis. Pengukuran dengan reliable dapat memberikan hasil yang tetap dan juga adanya kestabilan dalam pengukuran yang dilakukan dengan materi yang sama namun pada waktu yang beda (Mansyur, 2015).

Mengukur konsistensinya dalam analisis data dapat dilakukan dengan tes satu uji coba. Dalam mengukur konsistensi internya dapat dilakukan dengan beberapa metode seperti, metode belah dua (rumus Spearman-Brown, rumus Flanagan, dan rumus Rulon), 2) metode Kuder-Richardson (KR-20 dan KR-21) dan metode Cronbach Alpha. Mengukur konsistensi internalnya menggunakan metode Cronbach Alpha salah satu metode kerap dipergunakan, dikarenakan dapat melakukan perhitungan dalam krelibilitas sebuah tes pada bentuk pilihan ganda ataupun uraian. (Saifuddin Azwar, 2012:98; Djemari Mardapi, 2008:122) Menjelaskan bahawa krelibilitas yang dikategorikan sudah cukup baik minimalnya 0,70. Sedangkan dibawah dan lebih dari 0,70 tersebut belum dapat dikatakan krelibilitas yang baik

Tabel 1. Reability analisis Winstep
SUMMARY OF 45 MEASURED PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	14.7	25.0	.65	.50	1.01	.03	.94	-.02
SEM	.4	.0	.11	.01	.04	.16	.06	.11
P.SD	2.9	.0	.73	.05	.24	1.04	.42	.71
S.SD	2.9	.0	.74	.05	.24	1.06	.42	.72
MAX.	21.0	25.0	2.56	.71	1.86	3.00	2.37	1.90
MIN.	8.0	25.0	-.94	.47	.64	-1.98	.46	-1.13
REAL RMSE	.53	TRUE SD	.50	SEPARATION	.94	PERSON RELIABILITY	.47	
MODEL RMSE	.51	TRUE SD	.52	SEPARATION	1.03	PERSON RELIABILITY	.52	
S.E. OF PERSON MEAN = .11								

Person raw score-to-measure correlation= .99, cronbach alpha (KR-20) person raw score “test” reliability = .49, sem= 2.06

Tabel 2. SUMMARY OF 25 MEASURED ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S. E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	27.6	45.0	.00	.41	1.00	.15	.94	.08
SEM	2.2	.0	.30	.03	.02	.19	.05	.21
P. SD	10.3	.0	1.46	.16	.12	.92	.25	1.03
S. SD	10.6	.0	1.49	.16	.12	.94	.26	1.05
MAX.	44.0	45.0	3.97	1.02	1.26	2.35	1.37	2.62
MIN.	2.0	45.0	-3.38	.32	.78	-1.36	.28	-1.41
REAL RMSE	.45	TRUE SD	1.39	SEPARATION	3.11	ITEM	RELIABILITY	.91
MODEL RMSE	.44	TRUE SD	1.39	SEPARATION	3.17	ITEM	RELIABILITY	.91
S. E. OF ITEM MEAN = .30								

Data analisis yang dilihat dari Tabel 1 menunjukkan data reabilitas analisis dengan menggunakan foware iteman scores alpha dengan score 0, 504. Sedangkan tabel 2 analisis data dari software winstep cronbach alpha 0, 49. Dari data yang sudah ada tersebut menjelaskan bahwa tingkat kreabilitas dari soal masih tergolong lemah .

Kuliatas butir soal dari aspek reliabilitas butir pada pemodelan rasch, mempunyai kategori pada penentuan nilai Item Reliabiliti dan person Reliability yang sesuai dengan kategori tersebut yaitu, (Sumintono & Widhiarso, 2015).

Tabel 3 . Kriteria reliabilitas pada model rasch

Nilai Reliability (person/item)	Interpretasi
>0,94	Istimewa
0,91-0,94	Bagus sekali
0.81-0,90	Bagus
0,67-0,80	Cukup
<0,67	Lemah

Tabel 4. Kategori nilai person reability dan item reliably butir soal

Nilai Alpha cronba	Interpretasi
>0,80	Bagus sekali
0,70-0,80	Bagus
0.60-0,70	cukup
0,50-0,60	jelek
<0,50	buruk

Hasil rasch model dari person rliability dan item reability butir soal 0.47 dan 0.52. sedangkan pada item reabilit. 0.91. hasil dari cronbach alpha (KR-20) person raw score “test” reliability = . 49 sem= 2.06 dari kriteria reliabilitas person reliability terdapat pada interpretasi jelek sedangkan item reliability bagus sekali pada nilai alpha cronbach terdapat pada interprestasi buruk.

Vadilitas

Analisis kualitas soal dapat menentukan perbedaan di antara rangkaian seri dan paralel dapat terlihat pada komponen validitas, reliabilitas, serta tingkat kesulitan soal. Validitas butir soal diuji pada metode korelasi product moment. Soal dapat dinyatakan valid ketika koefisien korelasi $r_{xy} > r_{tabel}$ dan jika $r_{xy} \leq r_{tabel}$ sehingga soal dinyatakan tidak valid (Syofian, 2015). Pemodelan Rasch, dapat dilihat dari mutu butir soal pada kategori validitas yang dapat memenuhi kriteria yaitu (Sumintono & Widhiarso, 2015).

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

Tabel 5. Hasil MISFIT ORDER Winstep

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL CORR.	EXP.	OBS%	MATCH EXP%	ITEM
					MNSQ	ZSTD	MNSQ	ZSTD					
3	19	45	.99	.32	1.26	2.35	1.37	2.62	A-.04	.32	53.3	64.8	3
24	35	45	-.74	.37	1.10	.56	1.33	1.14	B-.09	.27	80.0	78.0	24
11	19	45	.99	.32	1.18	1.68	1.29	2.13	C-.06	.32	57.8	64.8	11
23	39	45	-1.40	.45	1.06	.30	1.18	.54	D-.12	.22	86.7	86.7	23
12	23	45	.59	.32	1.16	1.69	1.17	1.45	E-.11	.32	51.1	63.4	12
8	27	45	.19	.32	1.15	1.41	1.13	.90	F-.13	.31	57.8	65.5	8
9	25	45	.39	.32	1.10	1.04	1.12	.99	G-.18	.32	57.8	64.1	9
16	2	45	3.97	.74	1.06	.30	1.01	.30	H-.08	.16	95.6	95.6	16
1	39	45	-1.40	.45	1.03	.20	1.03	.22	I-.17	.22	86.7	86.7	1
4	18	45	1.09	.32	1.02	.24	1.02	.22	J-.29	.32	68.9	65.6	4
13	18	45	1.09	.32	1.01	.16	1.02	.19	K-.30	.32	64.4	65.6	13
19	29	45	-.02	.33	.99	-.03	.95	-.27	L-.33	.31	68.9	67.5	19
15	23	45	.59	.32	.98	-.20	.98	-.16	M-.35	.32	64.4	63.4	15
14	29	45	-.02	.33	.95	-.42	.90	-.57	N-.39	.31	68.9	67.5	14
20	35	45	-.74	.37	.95	-.19	.94	-.12	O-.33	.27	80.0	78.0	20
21	41	45	-1.88	.53	.94	-.01	.77	-.24	P-.28	.18	91.1	91.1	21
25	35	45	-.74	.37	.92	-.34	.85	-.48	Q-.38	.27	80.0	78.0	25
2	44	45	-3.38	1.02	.91	.23	.28	-.42	R-.33	.10	97.8	97.8	2
5	30	45	-.13	.33	.90	-.79	.87	-.71	S-.44	.30	75.6	68.8	5
10	35	45	-.74	.37	.89	-.54	.75	-.88	T-.45	.27	75.6	78.0	10
7	16	45	1.30	.33	.87	-1.09	.80	-1.29	U-.51	.32	66.7	68.0	7
22	29	45	-.02	.33	.84	-1.36	.79	-1.28	V-.52	.31	73.3	67.5	22
18	41	45	-1.88	.53	.83	-.30	.46	-1.00	W-.48	.18	91.1	91.1	18
17	11	45	1.90	.36	.78	-1.20	.67	-1.41	X-.59	.30	82.2	76.7	17
MEAN	26.5	45.0	.24	.46	1.00	.15	.94	.08			74.0	74.8	
P.S.D	11.5	.0	1.84	.32	.12	.92	.25	1.03			13.2	11.0	

Dari pernyataan di atas pada hasil analisis kualitas soal dilihat pada validitas butir soal dengan menggunakan model rasch diperoleh 6 soal akan dinyatakan valid dan 24 soal dinyatakan tidak valid dikarenakan tidak memenuhi syarat outfit MNSQ, Outfit ZSTD, serta *Point Measure correlation*(PT measure Corr). Dalam menggunakan model rasch ini dirasa lebih akurat karena memenuhi ketiga kriteria ini, maka maka tersebut dianggap valid.

Tingkat Kesukaran

Tabel 6. Measure Order (Rasch Model)

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL CORR.	EXP.	OBS%	MATCH EXP%	ITEM
					MNSQ	ZSTD	MNSQ	ZSTD					
6	0	45	5.90	1.81	MAXIMUM MEASURE				.00	.00	100.0	100.0	6
16	2	45	3.97	.74	1.06	.30	1.01	.30	.08	.16	95.6	95.6	16
17	11	45	1.90	.36	.78	-1.20	.67	-1.41	.59	.30	82.2	76.7	17
7	16	45	1.30	.33	.87	-1.09	.80	-1.29	.51	.32	66.7	68.0	7
4	18	45	1.09	.32	1.02	.24	1.02	.22	.29	.32	68.9	65.6	4
13	18	45	1.09	.32	1.01	.16	1.02	.19	.30	.32	64.4	65.6	13
3	19	45	.99	.32	1.26	2.35	1.37	2.62	-.04	.32	53.3	64.8	3
11	19	45	.99	.32	1.18	1.68	1.29	2.13	.06	.32	57.8	64.8	11
12	23	45	.59	.32	1.16	1.69	1.17	1.45	.11	.32	51.1	63.4	12
15	23	45	.59	.32	.98	-.20	.98	-.16	.35	.32	64.4	63.4	15
9	25	45	.39	.32	1.10	1.04	1.12	.99	.18	.32	57.8	64.1	9
8	27	45	.19	.32	1.15	1.41	1.13	.90	.13	.31	57.8	65.5	8
14	29	45	-.02	.33	.95	-.42	.90	-.57	.39	.31	68.9	67.5	14
19	29	45	-.02	.33	.99	-.03	.95	-.27	.33	.31	68.9	67.5	19
22	29	45	-.02	.33	.84	-1.36	.79	-1.28	.52	.31	73.3	67.5	22
5	30	45	-.13	.33	.90	-.79	.87	-.71	.44	.30	75.6	68.8	5
10	35	45	-.74	.37	.89	-.54	.75	-.88	.45	.27	75.6	78.0	10
20	35	45	-.74	.37	.95	-.19	.94	-.12	.33	.27	80.0	78.0	20
24	35	45	-.74	.37	1.10	.56	1.33	1.14	.09	.27	80.0	78.0	24
25	35	45	-.74	.37	.92	-.34	.85	-.48	.38	.27	80.0	78.0	25
1	39	45	-1.40	.45	1.03	.20	1.03	.22	.17	.22	86.7	86.7	1
23	39	45	-1.40	.45	1.06	.30	1.18	.54	.12	.22	86.7	86.7	23
18	41	45	-1.88	.53	.83	-.30	.46	-1.00	.48	.18	91.1	91.1	18
21	41	45	-1.88	.53	.94	-.01	.77	-.24	.28	.18	91.1	91.1	21
2	44	45	-3.38	1.02	.91	.23	.28	-.42	.33	.10	97.8	97.8	2
MEAN	26.5	45.0	.24	.46	1.00	.15	.94	.08			74.0	74.8	
P.S.D	11.5	.0	1.84	.32	.12	.92	.25	1.03			13.2	11.0	

Hasil analisis di tabel 6 menggunakan rasch model. Tingkat kesukaran butir soal dinyatakan dengan melihat nilai logit tiap butir soal yang tertera pada kolom measure. Nilai logit tinggi menyatakan tingkat kesulitan soal yang paling tinggi. Penentuannya dilihat pada nilai standar deviasi dan rata-rata logit butir soal. Pada penelitian ini, untuk nilai rata-rata logit butir soalnya ialah 0.00 dan nilai standar deviasinya 0,98. Ada 3 kategori tingkat kesukaran butir soal menurut odell rasch yaitu:

Tabel 7. Kategori tingkat kesukaran butir soal

Nilai Logit	Kategori
>0,98	Sangat sukar
0,0 – 0,98	Kesukaran Yang baik
<0,98	Mudah

Kategori tingkat kesukaran butir soal yang dilihat pada nilai measure ada 7 soal yaitu soal nomor 17, 16, 13, 7, 6, dan 4 nilai butir soal yang di atas 0,98 artinya tingkat kesukaran butir soalnya sangat sukar. Pada nilai measure yang termasuk kategori mudah di bawah 0,98 terdapat 5 soal pada nomor 23, 21, 18, 2, 1. Selanjutnya pada butir soal dengan nilai measure 0,0 – 0,98 terdapat pada nomor 25, 24, 22, 20, 19, 15, 14, 12, 11, 10, 9, 8, 5, 3 yang berjumlah 14 soal.

Daya beda Rasch Model

Daya beda Rasch model dengan tingkat abilitas individu menjadi alat dalam membedakan kecakapan siswa dapat menjawab soal dan yang tidak dapat menjawab soal. Daya beda Rasch model dilihat pada:

- Nilai Outfit MNSQ (Mean Square) yang diterima adalah: $0,5 < \text{Outfit} - \text{MNSQ} < 1,5$
- Nilai Outfit ZSTD (Z –Standard) yang diterima adalah: $-2,0 < \text{ZSTD} < +2,0$
- Nilai Pt Measure Corr (Point Measure Correlation): $0,4 < \text{Point Measure Corr} < 0,85$ (Sumintono & Widhiarso, 2015).

Data daya beda dari soal, di rasch model dapat dilihat pada daya beda di sisi kecakapan siswa disaat menjawab butir soal. Dijelaskan dalam tabel 6. *Measure Order (Rasch Model)*. Siswa telah memenuhi ketiga syarat person fit. Berarti, pola respon siswa sudah ideal.

Pengecoh

Tabel 8. Item Category/Option/Distractor Frequencies: Measure Order

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA		ABILITY		S.E. MEAN	INFIT MNSQ	OUTF MNSQ	PTMA CORR.	ITEM
			COUNT	%	MEAN	P.SD					
6	C	0	45	100	.65	.73	.11	.0	.0	.00	6
16	D	0	6	13	.47	.70	.31	.7	.8	-.10	16
	B	0	22	49	.63	.92	.20	1.4	1.2	-.03	
	A	0	15	33	.72	.33	.09	.6	.9	-.06	
	C	1	2	4	.92	.25	.25	1.6	1.0	.08	
17	A	0	19	42	.19	.54	.13	.6	.6	-.54	17
	C	0	11	24	.59	.52	.16	.9	.9	-.05	
	D	0	4	9	.94	.40	.23	1.1	1.3	.12	
	B	1	11	24	1.41	.61	.19	.7	.6	.59	
7	C	0	4	9	.05	.43	.25	.6	.6	-.26	7
	D	0	11	24	.37	.49	.15	.8	.8	-.22	
	B	0	14	31	.47	.72	.20	1.0	1.0	-.17	
	A	1	16	36	1.15	.63	.16	.8	.8	.51	
4	C	0	8	18	.17	.66	.25	.8	.7	-.31	4
	D	0	5	11	.54	.31	.16	.9	.9	-.05	
	A	0	14	31	.63	.69	.19	1.2	1.2	-.02	
	B	1	18	40	.91	.75	.18	1.0	1.0	.29	
13	A	0	11	24	.34	.62	.20	.9	.9	-.24	13
	C	0	8	18	.46	.65	.25	1.0	.9	-.12	
	D	0	8	18	.67	.87	.33	1.4	1.6	.01	
	B	1	18	40	.92	.64	.16	1.0	1.0	.30	
3	A	0	8	18	.51	.43	.16	.9	.9	-.09	3
	C	0	8	18	.52	.60	.23	1.0	1.0	-.08	
	D	0	10	22	.93	.58	.19	1.4	1.5	.21	
	B	1	19	42	.61*	.89	.21	1.4	1.5	-.04	

11	B	0	12	27	.51	.78	.24	1.2	1.2	-.12	11
	A	0	13	29	.69	.68	.20	1.2	1.4	.03	
	D	0	1	2	.92	.00		1.2	1.3	.06	
	C	1	19	42	.70*	.73	.17	1.2	1.3	.06	
12	D	0	4	9	.28	.46	.27	.8	.8	-.16	12
	C	0	10	22	.58	.51	.17	1.1	1.1	-.05	
	B	0	8	18	.70	.59	.22	1.3	1.3	.03	
	A	1	23	51	.73	.85	.18	1.2	1.3	.11	
15	D	0	4	9	.05	.66	.38	.7	.7	-.26	15
	B	0	10	22	.35	.79	.26	1.0	1.1	-.22	
	A	0	8	18	.61	.57	.21	1.1	1.1	-.03	
	C	1	23	51	.90	.64	.14	.9	.9	-.35	
9	D	0	9	20	.39	.89	.31	1.1	1.2	-.18	9
	B	0	5	11	.41	.54	.27	1.0	.9	-.12	
	C	0	6	13	.74	.63	.28	1.4	1.3	.05	
	A	1	25	56	.77	.68	.14	1.0	1.0	.18	
8	C	0	6	13	.45	.44	.20	1.0	.9	-.11	8
	D	0	4	9	.46	.52	.30	1.0	1.0	-.08	
	A	0	8	18	.64	.50	.19	1.2	1.2	-.01	
	B	1	27	60	.73	.84	.16	1.2	1.2	.13	
14	C	0	4	9	.00	.42	.24	.7	.6	-.25	14
	D	0	3	7	.07	.38	.27	.7	.6	-.12	
	A	0	9	20	.46	.53	.19	1.1	1.0	-.13	
	B	1	29	64	.86	.74	.14	1.0	1.0	.39	
19	D	0	3	7	-.01	.32	.22	.7	.6	-.24	19
	A	0	8	18	.37	.55	.21	1.0	1.0	-.18	
	B	0	5	11	.45	.52	.26	1.1	1.0	-.10	
	C	1	29	64	.83	.76	.14	1.0	1.0	.33	
22	B	0	2	4	-.01	.22	.22	.6	.6	-.19	22
	A	0	10	22	.07	.65	.22	.8	.8	-.43	
	D	0	4	9	.39	.25	.14	.9	.9	-.11	
	C	1	29	64	.93	.66	.12	.9	.9	.52	
5	A	0	5	11	.04	.49	.24	.7	.7	-.30	5
	D	0	2	4	.12	1.05	1.05	1.1	1.0	-.16	
	C	0	8	18	.32	.54	.21	.9	.9	-.21	
	B	1	30	67	.88	.66	.12	.9	.9	.44	
10	A	0	6	13	-.04	.40	.18	.7	.6	-.37	10
	B	0	4	9	.17	.47	.27	.9	.8	-.21	
	C	1	35	78	.82	.70	.12	.9	.9	.45	
20	B	0	5	11	.04	.44	.22	.8	.7	-.30	20
	C	0	5	11	.37	.85	.43	1.1	1.2	-.13	
	A	1	35	78	.78	.69	.12	.9	.9	.33	
24	D	0	3	7	-.17	.58	.41	.7	.6	-.30	24
	A	0	5	11	.79	.77	.39	1.5	1.8	.07	
	B	0	2	4	.92	.25	.25	1.8	1.6	.08	
	C	1	35	78	.68*	.71	.12	1.0	1.1	.09	
25	B	0	5	11	-.01	.62	.31	.8	.7	-.32	25
	D	0	3	7	.15	.39	.27	.9	.7	-.19	
	C	0	2	4	.47	.70	.70	1.2	1.2	-.05	
	A	1	35	78	.80	.69	.12	.9	.9	.38	
1	A	0	3	7	.06	.71	.50	.9	.8	-.22	1
	B	0	2	4	.33	.11	.11	1.2	.9	-.09	
	C	0	1	2	1.17	.00		2.7	2.0	.11	
	D	1	39	87	.70*	.73	.12	1.0	1.0	.17	
23	D	0	6	13	.43	.64	.29	1.2	1.2	-.12	23
	C	1	39	87	.68	.73	.12	1.1	1.0	.12	
18	D	0	4	9	-.46	.29	.17	.6	.4	-.48	18
	C	1	41	91	.76	.66	.10	.8	.9	.48	
21	A	0	3	7	-.16	.66	.47	.7	.7	-.30	21
	C	0	1	2	.44	.00		1.4	1.0	-.04	
	D	1	41	91	.71	.70	.11	.9	1.0	.28	
2	C	0	1	2	-.94	.00		.4	.3	-.33	2
	D	1	44	98	.69	.69	.11	.9	.9	.33	

Berdasarkan tabel 8 diatas, analisis butir pengecoh menggunakan software Winstep Rasch Model terhadap 25 soal ulangan harian siswa kelas IV di Madrasah Ibtidaiyah menunjukkan bahwa terdapat 23 soal dengan keberfungsian pengecoh yang baiknya. Sedangkan Butir soal dengan pengecohnya tidak keberfungsian dengan baik yaitu 6 soal.

Pembahasan

Analisis pemodelan rasch berdasarkan Georg Rasch tahun 1960-an dan diperkenalkan Ben Wirght. Data mentah dengan menggunakan data dikotomi dalam memperlihatkan kecakapan siswa. Pada analisis model Rasch dengan satu model mengetahui hubungan pada soal dan siswa (Sumintono and Widhiarso 2014).

Data yang didapatkan dalam penggunaan winstep dilakukan pengolahan dengan cara mengaitkan antara data siswa dengan soal dalam skala sama. Perolehan nilai dapat memberikan kesempatan siswa dalam menjawab soal yang diberikan secara benar. Penggunaan model rach menjadikan guru dapat mengetahui seberapa pahamnya siswa dalam memahami soal yang memiliki tingkatan soal yang ada dan juga melihat seberapa pahamnya siswa yang memiliki kecakapan tinggi dapat menyelesaikan soal lebih mudah (Bond, T. G., & Fox 2011).

Analisis model Rasch memiliki tingkatan kesesuaian butir soal yang akan memberikan penjelasan terhadap butir soal yang sesuai dalam mengukur yang dilakukan pada soal. Dalam pembuatan soal yang telah diuji coba kepada siswa jika terdapat soal yang tidak fit, sehingga hal tersebut terdapat kesalahan dalam soal, maka pendidik dapat melakukan perbaikan terhadap soal yang telah dibuat tersebut dan juga melakukan perbaikan dalam pembelajaran yang diberikan (Sumintono, B., & Widhiarso 2015). Rasch model memiliki penyesuaian yang di singkatkan dengan penyesuaian butir (item fit). Menentukan item fit menurut Boone et al pada Sumintono (2014) dengan kategori yaitu: (1) Nilai Outfit mean square (MNSQ) yang diterima = $0,5 < MNSQ < 1,5$; (2) Nilai Outfit Z-standard (ZSTD) yang diterima = $-2,0 < ZSTD < +2,0$; (3) Nilai Measure Correlation (Pr Mean Corr): $0,4 < Pt Measure Corr < 0,85$. Analisis data dilakukan dengan software Winsteps melalui model Rasch jika nilai jumlah kuadrat tengah (mean square) adalah 1,0 pada nilai distandarkan (Z-standarized value) merupakan 0,0. Didalam analisis dilakukan dengan mengetahui soal terlalu mudah atau terlalu sulit dan responden yang outlier (Nuryanti, Masykuri, and Susilowati 2018).

Analisis *rasch model* memiliki tahap dalam penganalisisan butir soal pilihan ganda dan *essay* sebagai berikut:

Reabilitas

Analisis pemodelan rasch dapat memberikan informasi dengan keutuhan, memiliki mutu instrument dengan respon siswa dengan keutuhan dengan responden dan item soal. Reliabilitas butir soal dilihat dari nilai Alpha Cronbach apabila nilainya $< 0,50$ (buruk), $0,50 - 0,60$ (jelek), $0,60-0,70$ (cukup), $0,70-0,80$ (bagus), dan $>0,80$ (bagus sekali). (Chan, Ismail, & Sumintono, 2014). Person measure menunjukkan rata-rata skor nilai keseluruhan dalam pengerjaan soal. Sumintono menjelaskan dengan mengukur reliabilitas, ditunjukkan dengan nilai person reability dan item reability. Kriteria reliabilitas sebagai berikut: $< 0,67$ (lemah); $0,67-0,8$ (cukup); $0,8-0,9$ (bagus); $0,91-0,94$ (bagus sekali); $> 0,94$ (istimewa) (Sumintono, B., & Widhiarso 2015).

Validitas

Tingkat kesesuaian butir soal (validitas) yang dilakukan dengan memberikan penjelasan terhadap butir soal memiliki fungsi normal melaksanakan pengukuran atau tidak. Menurut Boone kategori dengan pemeriksaan yang disesuaikan pada item yang tidak sesuai (outlier atau misfits), yaitu: 1) Nilai Outfit mean square (MNSQ) yang diterima: $0,5 < MNSQ < 1,5$, 2) Nilai Outfit Z-Standard (Outfit ZSTD) diterima: $-2.0 < ZSTD < +2.0$, 3) Nilai Point Measure Correlation (PT-Measure Corr) nilai yang diterima: $0,4 < PT-Measure Corr$

Tingkat Kesukaran

Butir item soal dikatakan baik jika butir item memiliki tidak terlalu sukar dan tidak juga terlalu mudah bagi siswa, dapat dikatakan bahwa derajat kesukaran didalam kategori sedang atau cukup. Angka indeks kesukaran dapat dilakukan dengan rumus penjelasan dari oleh Robert L. Thorndike. Klasifikasi tingkat kesukaran soal dapat ditentukan yaitu: $0,00 - 0,30$ (sulit); $0,31- 0,70$ (sedang); $0,71-1,00$ (mudah) (Ngalimun 2018). Nana Sujana (2002:135) menjelaskan pada tingkat kesukaran mudah, sedang dan sukar. Tingkat kesukaran yang baik merupakan $0,25$ sampai $0,75$. Soal memiliki tingkat kesukaran di bawah $0,25$ sehingga soal itu sukar, maka sebaliknya soal yang mempunyai tingkat kesukaran di atas $0,75$ adalah soal dikatakan terlalu mudah (Hanifah 2014).

Daya Beda

Rasch model memiliki analisis dengan tingkatan abilitas individu dengan instrument dalam perbedaan kecakapan siswa dapat menjawab jawaban soal dan tidak menjawab soal. Pada pemodelan Rasch digunakan analisis pada tingkat abilitas individu sebagai alat untuk membedakan kemampuan peserta didik yang mampu menjawab soal dan yang tidak mampu menjawab soal. Menurut Boone et. al. kategori memiliki tingkatan dalam penyesuaian individu (person fit) memiliki

persamaan dalam item fit yang digunakan pada nilai outfit mean square (MNSQ) yang diterima: $0,5 < \text{MNSQ} < \text{ZSTD}$

Pengecoh

Pada analisis butir soal pada model rasch digunakan pemrograman winsteps dalam efektivitas pengecoh digunakan dengan nilai Data Count %, jika nilai memiliki lebih dari sama dengan 5 % maka pengecoh berfungsi. namun, efektivitas pengecoh soal dikatakan berfungsi dalam nilai Average Ability bernilai negatif, merupakan pengecoh ditentukan pada siswa pada kemampuan rendah (Purniasari, Masykuri, and Ariani 2021). Pengecoh juga dapat dilakukan fungsi jika semua pengecoh ada pemilihnya, dan dinyatakan tidak berfungsi dengan baik jika ada pengecoh yang tidak ada pemilihnya (Himawan and Nurgiyantoro 2022).

KESIMPULAN

Analisis butir soal yang dapat digunakan pada model Rasch dengan winstep dapat disimpulkan pada data reabilitas analisis dengan menggunakan software winstep cronbach alpha 0,49. Dari data yang sudah ada tersebut menjelaskan bahwa tingkat kreabilitas dari soal masih tergolong lemah. Hasil rasch model dari person reliability dan item reability butir soal 0,47 dan 0,52. hasil analisisnya kualitas tes dilihat pada validitas butir soal dengan menggunakan model rasch diperoleh 6 soal dinyatakan valid dan 24 soal dinyatakan tidak valid dikarenakan persyaratan outfit MNSQ, Outfit ZSTD, serta Point Measure correlation (PT measure Corr) yang tidak dipenuhi.

Dalam menggunakan model rasch ini dirasa lebih akurat karena memenuhi ketiga kriteria ini, maka tersebut dianggap valid. diatas P value atau tingkat kesukaran butir soalnya yang berada dibawah 0,25 ada 3 butir soal, berarti soal itu sangat sukar. Soal yang berada di atas 0,75 ada 5 butir soal berarti soal itu terlalu mudah. Dan soal yang berada di score diantara 0,25-0,75 ada 17 butir soal. Artinya, ada 18 soal yang mempunyai tingkat kesukaran yang baik. analisis butir pengecoh menggunakan software Winstep Rasch Model terhadap 25 soal ulangan harian kelas IV di Madrasah Ibtidaiyah Negeri 01 Pasaman Barat menunjukkan bahwa terdapat 23 soal pada fungsi pengecoh yang baik. 6 Butir soal yang pengecohnya tidak memiliki fungsi secara baik.

REFERENSI

- Adimsyah, F. A., Fauzi, A., & Rofiq, M. H. (2023). Pengaruh Penggunaan Media Pembelajaran Dakon Terhadap Peningkatan Hasil Belajar Peserta Didik. *Chalim Journal of Teaching and Learning (CJoTL)*, 3(1), Article 1.
- Agus, A., Juliadharma, M., & Djamaluddin, M. (2023). Application of the CIPP Model in Evaluation of The Inclusive Education Curriculum in Madrasah Aliyah. *Nidhomul Haq: Jurnal Manajemen Pendidikan Islam*, 8(1), Article 1. <https://doi.org/10.31538/ndh.v8i1.2705>
- Allur, E., Heras-Saizarbitoria, I., Boiral, O., & Testa, F. (2018). Quality and Environmental Management Linkage: A Review of the Literature. *Sustainability*, 10(11), Article 11. <https://doi.org/10.3390/su10114311>
- Apriliani, I., Pahrudin, A., Koderi, K., & Syafril, S. (2024). Management of Inclusive Education: An Implementation. *Munaddhomah: Jurnal Manajemen Pendidikan Islam*, 5(1), Article 1. <https://doi.org/10.31538/munaddhomah.v5i1.935>
- Arif, M., Aziz, M. K. N. bin A., Harun, M., & Ma'arif, M. A. (2023). Strengthening The Sense of Patriotism in Madrasah Ibtidaiyah, Indonesia Based on The Islamic Boarding School System. *Tafkir: Interdisciplinary Journal of Islamic Education*, 4(1), Article 1. <https://doi.org/10.31538/tijie.v4i1.226>

- Azis, A., Abou-Samra, R., & Aprilianto, A. (2022). Online Assessment of Islamic Religious Education Learning. *Tafkir: Interdisciplinary Journal of Islamic Education*, 3(1), Article 1. <https://doi.org/10.31538/tijie.v3i1.114>
- Azizah, H. N., Muchtar, N. E. P., & Putra, F. T. (2023). Pesantren as a Pillar of Islamic Civilization Development in Indonesia. *Academicus: Journal of Teaching and Learning*, 2(1), Article 1. <https://doi.org/10.59373/academicus.v2i1.19>
- Bagiyono, B. (2017). Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal Ujian Pelatihan Radiografi Tingkat 1 The Analysis of Difficulty Level and Discrimination Power of Test Items of Radiography Level 1 Examination. *Widyanuklida*, 16(1), 1–12.
- 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), 11–19.
- Fauzi, A., & Masrupah, S. (2024). Pengaruh Model Pembelajaran Kooperatif Tipe Team Games Tournament (TGT) Terhadap Hasil Belajar Siswa. *Ngaos: Jurnal Pendidikan Dan Pembelajaran*, 2(1), Article 1. <https://doi.org/10.59373/ngaos.v2i1.7>
- Fernanda, J. W., & Hidayah, N. (2020). Analisis Kualitas Soal Ujian Statistika Menggunakan Classical Test Theory dan Rasch Model. *Square: Journal of Mathematics and Mathematics Education*, 2(1), 49–60.
- Fitria, R. N., Alwasih, A., & Hakim, M. N. (2022). Strategi Kepala Madrasah Dalam Meningkatkan Prestasi Akademik Siswa. *Academicus: Journal of Teaching and Learning*, 1(1), Article 1.
- Fitriani, N., Anam, S., & Maulana, A. (2024). Building Literacy of Early Age Students' Language; Teacher Managerial Competence and Legal-Rational Authority of Boarding School Leaders. *Munaddhomah: Jurnal Manajemen Pendidikan Islam*, 5(1), Article 1. <https://doi.org/10.31538/munaddhomah.v5i1.707>
- Hayati, S., & Lailatussaadah, L. (2016). Validitas dan Reliabilitas Instrumen Pengetahuan Pembelajaran Aktif, Kreatif dan Menyenangkan (pakem) Menggunakan Model Rasch. *Jurnal Ilmiah Didaktika: Media Ilmiah Pendidikan Dan Pengajaran*, 16(2), 169–179.
- Iskandar, A., & Rizal, M. (2018). Analisis Kualitas Soal di Perguruan Tinggi Berbasis Aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12–23.
- Kadir, A. (2015). Menyusun dan Menganalisis Tes Hasil Belajar. *Al-TA'DIB: Jurnal Kajian Ilmu Kependidikan*, 8(2), 70–81.
- Komarodin, M., & Rofiq, A. (2023). Islamic Boarding School Management in Forming The Religious Character of Students. *Chalim Journal of Teaching and Learning*, 3(1), Article 1. <https://doi.org/10.31538/cjotl.v3i1.428>
- Lafrarchi, N. (2020). Assessing Islamic Religious Education Curriculum in Flemish Public Secondary Schools. *Religions*, 11(3), Article 3. <https://doi.org/10.3390/rel11030110>
- Mania, S. (2008). Teknik Non Tes: Telaah Atas Fungsi Wawancara dan Kuesioner Dalam Evaluasi Pendidikan. *Lentera Pendidikan: Jurnal Ilmu Tarbiyah Dan Keguruan*, 11(1), 45–54.
- Mardapi, D. (2017). Pengukuran Penilaian dan Evaluasi Pendidikan Edisi 2. *Yogyakarta: Parama Publishing*.
- Munir, M. S., Muhtadin, M. A., & Septiansyah, M. A. L. (2022). Semester Credit System Program Evaluation at MTsN 2, Kediri City. *Munaddhomah: Jurnal Manajemen Pendidikan Islam*, 3(3), Article 3. <https://doi.org/10.31538/munaddhomah.v3i3.279>

- Perdana, S. A. (2018). Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes lasik Dan Rasch Model. *Jurnal Kiprah*, 6(1), 41–48.
- Pereira, D., Flores, M. A., & Niklasson, L. (2016). Assessment revisited: A review of research in Assessment and Evaluation in Higher Education. *Assessment & Evaluation in Higher Education*, 41(7), 1008–1032. <https://doi.org/10.1080/02602938.2015.1055233>
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1).
- Retnawati, H. (2014). Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana. *Yogyakarta: Nuha Medika*.
- Sumarni, S. (2019). Designing Ict Competences-Integrated Assessment Instruments Of Practical Key Teaching Competences For English Language Education Study Program. *Ijlecr-International Journal Of Language Education And Culture Review*, 5(1), 47–55.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Trim komunikata.
- Suryani, Y. E. (2017). Pemetaan Kualitas Empirik Soal Ujian Akhir Semester Pada Mata Pelajaran Bahasa Indonesia SMA di Kabupaten Klaten. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 21(2), 142–152.
- Syofian. (2015). *Metode Penelitian Kuantitatif Dilengkapi Perbandingan Perhitungan Manual & SPSS*. Prenamedia Group.
- Talib, A. M., Alomary, F. O., & Alwadi, H. F. (2018). Assessment Of Student Performance For Course Examination Using Rasch Measurement Model: A Case Study Of Information Technology Fundamentals course. *Education Research International*, 2018.
- Wahyudi, W. (2010). Assesment Pembelajaran Berbasis Portofolio di Sekolah. *Jurnal Visi Ilmu Pendidikan*, 2(1).
- Wahyuningsih, E. T. (2015). Analisis Butir Soal Tes Objektif Buatan Guru Ulangan Semester Ganjil Mata Pelajaran Ekonomi Kelas X di SMA Negeri 1 Mlati tahun ajaran 2013/2014. *Universitas Negeri Yogyakarta*.
- 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).