



Analisis pengukuran hasil ulangan menggunakan teori tes klasik dan model Rasch pada siswa SMA

Annisa El Hasna ^{a,1,*}

^a Program Studi Pendidikan Biologi, Universitas Jember, Jember, Jawa Timur, 68121, Indonesia.

¹ annisaelhasna25@gmail.com

* Corresponding author.

INFORMASI ARTIKEL

Lini Masa Artikel

Draft diterima : 2024-06-16

Revisi diterima : 2024-12-28

Diterbitkan : 2025-04-24

Keywords

Biology test

Classical test theory

Rasch Model;

ABSTRAK

Evaluasi adalah bagian dan tahapan penting yang harus dilalui pendidik untuk menjamin pembelajaran efektif. Hasil evaluasi dapat dijadikan bahan refleksi guru dalam mengembangkan dan menyelenggarakan program dan kegiatan pembelajaran. Penelitian bertujuan untuk membandingkan analisis penilaian atau evaluasi menggunakan Teori Tes Klasik dengan Model Rasch dan dapat mengetahui perbedaan analisis dengan menggunakan kedua model tersebut. Jenis penelitian pada penelitian ini yakni metode kualitatif secara pendekatan deskriptif. Sampel yang digunakan yakni hasil ulangan harian BAB Virus dengan jumlah butir soal sejumlah 15 soal. Analisis penilaian berdasarkan teori tes klasik menggunakan aplikasi Microsoft Excel dan analisis penilaian berdasarkan Model Rasch menggunakan aplikasi Winsteps. Analisis validitas pada teori tes klasik serta Model Rasch berjumlah sembilan soal valid serta tidak valid enam soal, dan valid lima soal serta tidak valid 10 soal. Analisis Uji Reliabilitas melalui teori tes klasik diperoleh nilai sebesar 0,672 dengan kategori sedang serta dari Item Reliability diperoleh nilai 0,87 dengan kategori bagus. Rerata tingkat kesulitan soal pada teori tes sebesar 0,463 pada tingkat sedang, menurut model Rasch terdapat tiga soal sangat mudah, mudah lima soal, sulit tiga soal serta sangat empat soal. Analisis daya pembeda berdasarkan teori tes klasik yang berada pada kategori buruk satu item, enam butir cukup, lima butir tinggi, dan tiga item sangat tinggi. Daya pembeda butir soal berdasarkan Model Rasch diperoleh indeks separasi butir soal $H = 3,81$.

ABSTRACT

Evaluation is an important part and stage that educators must go through to ensure effective learning. Evaluation results can be used as material for teacher reflection in developing and organizing learning programs and activities. The research aims to compare the analysis of assessment or evaluation using Classical Test Theory with the Rasch model and can find out the differences in analysis using the two models. The type of research in this study is a qualitative method with a descriptive approach. The sample used is the results of the Virus Chapter daily test with a total of 15 questions. Assessment analysis based on classical test theory using Microsoft Excel application and assessment analysis based on Rasch Model using Winsteps Rasch application. Validity analysis on classical test theory and Model Rasch amounted to nine valid questions and six invalid questions, and five valid questions and 10 invalid questions. Reliability test analysis through classical test theory obtained a value of 0.672 with a moderate category and from Item Reliability obtained a value of 0.87 with a good category. The average difficulty level of questions on the test theory is 0.463 at a moderate level, according to the Rasch Model there are three very easy questions, five easy questions, three difficult questions, and four very difficult questions. Analysis of distinguishing power based on classical test theory which is in the bad category of one item, six items are sufficient, five items are high, and three items are very high. The distinguishing power of items based on the Rasch Model obtained the item separation index $H = 3.81$.

Cara Sitasi Artikel Ini (APA Style):

Hasna, A. E. (2025). Analisis pengukuran hasil ulangan menggunakan teori tes klasik dan model Rasch pada siswa SMA. *Bio-Pedagogi*. 14(1), 1-9. <https://dx.doi.org/10.20961/bio-pedagogi.v14i1.88425>.

Artikel ini dapat diakses secara bebas dengan lisensi [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/).



PENDAHULUAN

Dalam sistem pendidikan Indonesia terdapat kegiatan evaluasi untuk memantau, memvalidasi, dan menentukan kualitas pendidikan ([Datuk & Arifin, 2023](#)). Evaluasi dilakukan untuk memantau perkembangan, kemajuan dan peningkatan hasil belajar siswa ([Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia, 2013](#)). Pada setiap akhir pelajaran, siswa dinilai untuk mengetahui penguasaan materi yang dipelajari. Hasil dari penilaian berguna untuk menjadi dasar dalam merevisi dan melengkapi program serta aktivitas pembelajaran ([Wardani et al., 2022](#)). Sebutan pada sistem penilaian yakni pengukuran, penilaian serta evaluasi. Penilaian adalah tahapan menaksir hasil capaian belajar siswa melalui pengerjaan serta penggarapan informasi.

Analisis pada penilaian dapat dilakukan dengan banyak cara, misalnya dengan menggunakan teori tes klasik atau menggunakan pemodelan Rasch ([Gunasti et al., 2023](#)). Teori tes klasik digunakan untuk mengukur keterampilan intelektual dengan mendeskripsikan hubungan antara nilai tes yang diamati dan skor sebenarnya yang tak teramati ([Solichin, 2017](#)). Parameter utama yang dianalisis dengan teori tes klasik meliputi: kesulitan soal, perbedaan soal, dan reliabilitas. Analisis tes klasik juga berguna mengetahui fungsi pengecoh pada soal pilihan ganda ([Sumaryanta, 2021](#)). Namun analisis tes klasik mempunyai kelemahan yaitu karakteristik tes mempengaruhi hasil pengukuran dan kemampuan peserta tes mempengaruhi nilai parameter soal ([Triana, 2024](#)).

Model *Rasch* merupakan metode pengujian lanjutan yang dapat mengatasi beberapa keterbatasan dalam teori tes klasik, seperti kompleksitas tes dan kurangnya kontrol terhadap klasifikasi hasil tes ([Widyaningsih & Yusuf, 2018](#)). Evaluasi dengan menggunakan model Rasch melibatkan pemberian kesempatan kepada siswa untuk menyelesaikan masalah dengan benar, tergantung pada hubungan antara keterampilan siswa dengan kesulitan masalah yang dipelajari ([Bond, 2015](#); [Sekarningtias et al., 2023](#)). Model Rasch muncul menjadi model pengukuran karena pengukuran pendidikan unggul dalam objektivitas yang spesifik serta estimasi parameter butir yang tinggi dan stabil. Selain itu Model Rasch dapat menyebabkan nilai standar error pengukuran yang mampu membuat akurasi perhitungan meningkat, sehingga direkomendasikan untuk digunakan oleh para pendidik dalam mengukur dan menilai hasil belajar siswa untuk mengerti kemampuan siswa yang sebenarnya. ([Darmana et al., 2021](#); [Muhammad et al., 2023](#)). Berdasarkan uraian diatas, penelitian ini bertujuan untuk membandingkan analisis penilaian atau evaluasi menggunakan Teori Tes Klasik dengan Model Rasch dan dapat mengetahui perbedaan hasil analisis dengan menggunakan kedua model tersebut.

METODE

Penelitian ini dilaksanakan pada Bulan Agustus – September 2022. Penelitian dilaksanakan di satu SMAN di Balung. Metode yang digunakan dalam penelitian ini yaitu metode kualitatif berdasarkan pendekatan deskriptif. Sampel yang digunakan yakni hasil ulangan harian 36 siswa.

Soal ulangan diberikan dan diujikan pada siswa melalui *website Nearpod*. Soal ulangan harian berjumlah 15 soal serta berbentuk pilihan ganda. Hasil ulangan harian siswa kemudian dianalisis menggunakan teori tes klasik menggunakan aplikasi Microsoft Excel serta analisis evaluasi berdasarkan model Rasch menggunakan aplikasi Winsteps. Parameter yang diuji meliputi: validitas, reliabilitas, efektivitas distraktor, tingkat kesulitan, dan daya pembeda.

HASIL DAN PEMBAHASAN

Hasil ulangan pada siswa SMA kelas X.2 dianalisis dengan teori tes klasik menggunakan aplikasi Microsoft Excel serta berdasarkan Pemodelan Rasch menggunakan *Software Winsteps*. Parameter yang digunakan yakni validitas, reliabilitas, efektivitas distraktor, tingkat kesulitan, dan daya pembeda.

Hasil Uji Analisis Validitas

Tabel 1. Perbandingan Hasil Analisis Validitas Butir Soal Melalui Pendekatan Teori Tes Klasik dan Model Rasch

Hasil	Nomor Soal	
	Teori Tes Klasik	Permodelan Rasch
Valid	3, 4, 6, 7, 9, 11, 13, 14, 15	3, 7, 11, 13, 15
Tidak Valid	1, 2, 5, 8, 10, 12	1, 2, 4, 5, 6, 8, 9, 10, 12, 14

Instrumen dinyatakan valid bila berguna untuk mengukur sesuatu yang perlu diukur. Fungsi uji validitas yakni melihat valid atau tidaknya suatu alat ukur ([Alfiatunnisa et al., 2022](#) ; [Souliisa & Manuputty, 2021](#)). Berdasarkan teori pengujian klasik, soal valid jika nilai $r_{xy} \geq r$ tabel, sedangkan soal dikatakan tidak valid jika nilai r_{xy} adalah $\leq r$ tabel. Hasil analisis validitas Rasch dapat dilihat dengan memeriksa hold MNSQ, ZSTD hold, dan Pt Measure Corr. Rentang nilai pada Outfit MNSQ berkisar $0,50 < x < 1,50$. Rentang nilai pada Outfit ZSTD berkisar $-2,0 < x < 2,0$. Rentang nilai pada PT MEGA CORR berkisar $0,4 < x < 0,85$. Dengan demikian dapat diketahui bahwa suatu pertanyaan tergolong valid jika nilainya berada dalam rentang nilai ketiga kriteria tersebut, sedangkan suatu pertanyaan tergolong tidak valid jika nilainya tidak berada dalam rentang nilai ketiga kriteria tersebut. Soal dinyatakan valid dapat dipakai pada tes atau disimpan dalam bank soal. Sedangkan soal dinyatakan tidak valid akan disesuaikan serta dikoreksi dengan standar kelulusan menggunakan teknik persiapan ([Sari, 2022](#)).

Reliabilitas Butir Soal

Uji reliabilitas adalah uji yang mengukur hasil suatu pengukuran bisa dipercaya dan berfungsi dalam mengidentifikasi konsistensi alat ukur ([Nazura et al., 2021](#)). Uji ini memakai rumus Cronbach's Alpha karena hasilnya lebih akurat ([Janna & Herianto, 2021](#)). Soal dikatakan reliabel apabila sesuai kriteria koefisien korelasi reliabilitas instrumen yang disajikan pada **Tabel 2**. Kriteria menentukan nilai Item Reliability dalam model Rasch menurut [Sumintono et al. \(2015\)](#), disajikan pada **Tabel 3**. Perbandingan hasil analisis reliabilitas berdasarkan teori tes klasik serta pemodelan Rasch terlihat pada **Tabel 4**.

Tabel 2. Kriteria Koefisien Korelasi Reliabilitas Instrumen

Koefisien Korelasi	Interpretasi
$0,90 \leq r \leq 1,00$	Sangat Tinggi
$0,70 \leq r < 0,90$	Tinggi
$0,40 \leq r < 0,70$	Sedang
$0,20 \leq r < 0,40$	Rendah
$R < 0,20$	Sangat Rendah

Tabel 3. Kriteria Reliabilitas pada pemodelan Rasch

Nilai Reliability (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. Perbandingan hasil analisis reliabilitas soal menggunakan pendekatan teori tes klasik dan pemodelan Rasch

Teori Tes Klasik			Permodelan Rasch		
Alpha Cronbach	Kategori	Person Reliability	Kategori	Item Reliability	Kategori
0,672	Sedang	0,63	Lemah	0,87	Bagus

Tabel 4 menunjukkan berdasarkan teori tes klasik soal tergolong memiliki reliabilitas sedang. Terlihat juga perbedaan antara nilai *person reliability* 0,63 yang tergolong rendah dengan nilai *item reliability* 0,87 yang tergolong bagus. Selisih skor yang diperoleh menunjukkan bahwa konsistensi jawaban siswa saat menjawab soal ulangan harian pada bab virus sangat rendah. Nilai *item reliability* sebesar 0,87 artinya kualitas item instrumen mempunyai aspek reliabilitas yang baik. Rendahnya nilai reliabilitas ini disebabkan beberapa faktor seperti lama dan singkatnya ujian, kualitas soal, soal yang relevan bagi siswa, petunjuk menjawab soal sebelum ujian dimulai, kedisiplinan penguji, pengawasan, suasana lingkungan serta tempat tes ([Arikunto, 2016](#)).

Efektivitas Distraktor

Setiap tes objektif harus menggunakan alternatif jawaban yang telah disiapkan, termasuk jawaban yang dapat menyesatkan atau mengganggu ([Wahyudi, 2010](#)). Ini tentang melihat seberapa dalam seorang siswa memahami konsep suatu topik tertentu ([Herrmann-abel & Deboer, 2016](#)). Kandidat tertarik atau terdorong untuk memilih distraksi sebagai jawaban yang benar. Distraksi dikatakan efektif jika banyak kandidat yang tertipu dengan distraksi yang dipasang pada item tersebut dan sebaliknya ([Madaniyah et al., 2015](#); [Subari et al., 2021](#)). Berdasarkan teori tes klasik, Efektivitas Distraktor pada butir soal nomor 4;5;12;13;15 pada salah satu opsi pilihannya efektivitas distraktor nya tak bekerja dengan baik dikarenakan tidak terdapat siswa yang mengambil salah satu opsi tersebut. Hasil analisis efektivitas distractor pada *Model Rasch* dapat diketahui pada *average ability*-nya yang nilainya semakin naik pada tiap butir pilihan jawabannya. Dapat diketahui pada soal nomor 1, 2, 7, dan 12 distraktor nya tidak bekerja dengan baik dikarenakan *average ability*-nya menurun pada butir pilihan jawabannya, sedangkan soal nomor 3,4,5,6,8,9,10,11,13,14, dan 15 distraktor nya berkerja dengan baik dikarenakan *average ability*-nya semakin naik pada tiap butir pilihan jawabannya. Hasil analisis efektivitas distractor menggunakan *Model Rasch* (lampiran).

Tingkat Kesulitan

Tabel 5. Tingkat Kesulitan Butir Soal dengan Pendekatan Teori Tes Klasik

Nilai Indeks Kesulitan Soal	Interpretasi Indeks Kesulitan	Butir Soal
IK = 0,00	Sangat Sulit	-
$0,00 < IK \leq 0,30$	Sulit	6,9,10,14
$0,30 < IK \leq 0,70$	Sedang	1, 2, 3, 4, 5, 7, 8, 11, 15
$0,70 < IK \leq 1,00$	Mudah	12, 13
IK = 1	Terlalu Mudah	-

Tabel 6. Tingkat Kesulitan Butir Soal dengan Pemodelan Rasch

Nilai Measure (logit)	Interpretasi Indeks Kesulitan	Butir Soal
Measure logit < -1	Sangat Mudah	1, 12, 13
$-1 \leq \text{Measure logit} \leq 0$	Mudah	2, 4, 5, 11, 15
$0 \leq \text{Measure logit} \leq 1$	Sulit	3, 7, 8
Measure logit > 1	Sangat Sulit	6, 9, 10, 14

Berdasarkan **Tabel 5**, dengan menggunakan teori tes klasik diketahui sebanyak empat soal terkategori sulit, sembilan soal terkategori sedang, dan dua soal terkategori mudah. Hasil yang didapat menggunakan Model Rasch (Tabel 6) menunjukkan hasil yang berbeda. Tingkat kesulitan soal model Rasch tercermin dari nilai pengukuran satuan logit setiap item soal. Berdasar pada kriteria pada Tabel 6 terlihat nilai pengukuran (logit), butir soal dikelompokkan menjadi empat kategori sesuai dengan tingkat kesulitan butir soal, yakni butir soal sangat mudah, butir soal mudah, dan butir soal mudah. soal sulit serta soal sangat sulit dengan sebaran hampir merata pada tiap kriteria.

Soal yang baik ialah soal tak terlalu sulit namun tak terlalu mudah ([Tyas et al., 2020](#)). Soal yang terlalu mudah tak akan menggugah siswa untuk mencoba menyelesaikannya, sedangkan soal yang terlalu sulit akan memberikan harapan kepada siswa karena berada di luar kemampuannya untuk menyelesaikan soal tersebut ([Halik et al., 2019](#); [Putri et al., 2023](#)). Soal dengan tingkat kesulitan sedang dapat digunakan kembali dengan mengimpor ke bank soal Jika tingkat kesulitannya berkisar dari sulit hingga sangat sulit, soal dapat disimpan dan digunakan dalam tes bersifat ketat. Soal juga dapat dikaitkan dengan penyebab soal yang sulit dijawab ([Istika et al., 2019](#)).

Daya Pembeda Soal

Daya pembeda suatu pertanyaan ialah kapabilitas pertanyaan untuk mengklasifikasikan siswa mampu menjawab pertanyaan maupun siswa yang punya pengetahuan tinggi dengan siswa yang rendah pengetahuan dalam menjawab pertanyaan ([Erfan et al., 2020](#)). Hasil perhitungan indeks daya diskriminasi butir soal berdasar pada teori tes klasik dikelompokkan menjadi tiga kategori yang dijelaskan pada **Tabel 7**.

Tabel 7. Tingkat Kesulitan Butir Soal dengan Pemodelan Rasch

Daya Pembeda (DP)	Interpretasi Indeks Kesulitan	Butir Soal
$DP \geq 0,70$	Sangat Tinggi	3, 4, 11
$0,40 \leq DP < 0,70$	Tinggi	6, 7, 9, 13, 15
$0,20 \leq DP < 0,40$	Cukup	1, 5, 8, 10, 12, 14
$DP < 0,20$	Rendah	2

Dari Tabel 7, dapat diketahui satu soal terkategori rendah, sebanyak enam soal terkategori cukup, sebanyak lima soal terkategori tinggi, dan tiga soal terkategori sangat tinggi. Pada pendekatan Model Rasch penggolongan responden berlandaskan indeks separasi responden. Makin besar nilai separasi item maka bobot instrumen pada keseluruhan responden serta butir soal makin bagus, disebabkan mampu menganalisis golongan responden dan kelompok butir ([Sumintono, 2016](#)). Persamaan lainnya dalam mengidentifikasi pengelompokan lebih teliti menggunakan **persamaan strata (H)**:

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

SUMMARY OF 36 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	6.9	15.0	-.23	.64	1.00	-.09	1.09	.02
SEM	.5	.0	.19	.01	.05	.19	.12	.18
P.SD	2.9	.0	1.14	.09	.31	1.11	.73	1.06
S.SD	2.9	.0	1.16	.09	.31	1.12	.74	1.08
MAX.	12.0	15.0	1.76	1.08	1.88	2.63	3.94	2.44
MIN.	1.0	15.0	-3.19	.59	.53	-2.27	.45	-1.79
REAL RMSE	.69	TRUE SD			1.32	Person	RELIABILITY	.63
MODEL RMSE	.65	TRUE SD	.94	SEPARATION	1.44	Person	RELIABILITY	.68
S.E. OF Person MEAN = .19								
Person RAW SCORE-TO-MEASURE CORRELATION = .99								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .67 SEM = 1.65								
SUMMARY OF 15 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	16.7	36.0	.00	.42	.97	-.12	1.09	.18
SEM	2.0	.0	.32	.01	.05	.32	.11	.33
P.SD	7.5	.0	1.20	.05	.19	1.19	.42	1.22
S.SD	7.7	.0	1.24	.05	.19	1.23	.43	1.27
MAX.	30.0	36.0	2.29	.56	1.28	1.83	1.95	2.07
MIN.	4.0	36.0	-2.20	.37	.62	-2.63	.52	-2.10
REAL RMSE	.43	TRUE SD	1.12	SEPARATION	2.61	Item	RELIABILITY	.87
MODEL RMSE	.42	TRUE SD	1.13	SEPARATION	2.68	Item	RELIABILITY	.88
S.E. OF Item MEAN = .32								

Gambar 1. Hasil *Summary Statistic* untuk 40 butir soal dengan pemodelan Rasch

Dari **Gambar 1** diketahui nilai separasi item sebesar 2,61 maka nilai $H = 3,81$ diterapkan menjadi 4 yakni terdapat 4 kelompok item serta untuk responden nilai pemisahannya adalah 1,32 dengan $H = 2,09$ diterapkan menjadi dua menyatakan kelompok responden bisa dibagi menjadi dua kelompok berlandaskan nilai keterpisahan respondennya. Soal dengan daya diferensial tinggi dan sangat tinggi akan ditempatkan di bank soal serta berguna untuk tes selanjutnya. Elemen dengan daya diferensial yang cukup dan rendah dapat dihilangkan dan tidak dihilangkan pada pengujian mendatang atau dapat dilacak dan ditingkatkan sehingga dapat digunakan dalam pengujian mendatang (Tilaar & Hasriyanti, 2019).

KESIMPULAN

Hasil penelitian menunjukkan kualitas soal Ulangan Harian bab virus berdasar pada metode pengujian teori klasik secara umum dikatakan baik. Sedangkan berdasar metode model Rasch masih belum bisa dikatakan baik. Analisis validitas pada teori tes klasik adalah sembilan soal dianggap valid dan enam soal tidak valid, sedangkan melalui model Rasch terdapat lima soal yang valid serta sepuluh soal yang tidak valid. Analisis pengujian reliabilitas dengan menggunakan teori tes klasik didapat nilai Cronbach's Alpha untuk reliabilitas yakni 0,672 dengan pengkategorian sedang dan dari reliabilitas soal diperoleh nilai reliabilitas sebesar 0,87 dengan kategori baik. Rerata indeks kesulitan soal menurut pendekatan teori tes klasik rata-rata sebesar 0,463, sedangkan hasil analisis kesukaran menurut model Rasch mengindikasikan empat tingkat kesukaran soal, khususnya tiga soal sangat sulit, mudah, lima soal mudah, tiga soal sulit, dan empat soal sangat sulit. Analisis daya diskriminan didasarkan pada teori tes klasik, dengan satu butir soal berkategori rendah, enam butir berkategori sedang, lima butir berkategori tinggi, dan tiga butir berkategori sangat tinggi. Daya diskriminasi item berbasis model Rasch dicapai oleh delapan kelompok item berdasarkan indeks pemisahan item $H = 3,81$.

DAFTAR PUSTAKA

- Alfiatunnisa, E., Zulfah Khairunnisa, H., Hayati, S., & Listya Maulida, V. (2022). Uji Validitas Dan Reliabilitas Terhadap Kemandirian Siswa Sekolah Dasar Kelas 1. *JURNAL HURRIAH: Jurnal Evaluasi Pendidikan Dan Penelitian*, 3(2), 29–36. <https://doi.org/10.56806/jh.v3i2.81>
- Arikunto, S. (2016). *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara.
- Bond, T. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Sciences, Third Edition* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315814698>
- Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa*, I., & Nurhaslinda, N. (2021). Analysis of Rasch Model for the Validation of Chemistry National Exam Instruments. *Jurnal Pendidikan Sains Indonesia*, 9(3), 329–345. <https://doi.org/10.24815/jpsi.v9i3.19618>
- Datuk, A., & Arifin, A. (2023). Urgensi Akreditasi Sekolah dalam Upaya Pemetaan Mutu Pendidikan di Provinsi Nusa Tenggara Timur. *Attractive: Innovative Education Journal*, 5(1), 54–64. <https://doi.org/10.51278/aj.v5i1.557>
- Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Tes Klasik Dan Model Rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19. <https://doi.org/10.23887/ijerr.v3i1.24080>
- Halik, A. S., Mania, S., & Nur, F. (2019). Analisis Butir Soal Ujian Akhir Sekolah (Uas) Mata Pelajaran Matematika Pada Tahun Ajaran 2015/2016 Smp Negeri 36 Makassar. *Al Asma : Journal of Islamic Education*, 1(1), 11. <https://doi.org/10.24252/asma.v1i1.11249>
- Herrmann-abell, C. F., & Deboer, G. E. (2016). *modeling to investigate hierarchies of chemistry misconceptions and detect Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with in*. February. <https://doi.org/10.1039/C1RP90023D>
- Gunasti, A., Sanosra, A., Muhtar, M., Mufarida, N. A., & Satoto, E. B. (2023). Pemanfaatan rasch model untuk mengukur kemampuan peserta pengkaderan formal. *JMM (Jurnal Masyarakat Mandiri)*, 7(2), 1544–1557. <https://doi.org/10.31764/jmm.v7i2.13826>
- Istika, N., Zamsir, & Prajono, R. (2019). Quality Of Repeated Test Subject Semester Final Grade Math Vii Semester Ii Junior State 12 Kendari. *Jurnal Penelitian Pendidikan Matematika*, 7(2), 127–140. <https://doi.org/10.30605/pedagogy.v10i1.5672>
- Janna, N. M., & Herianto. (2021). Artikel Statistik yang Benar. *Jurnal Darul Dakwah Wal-Irsyad (DDI)*, 18210047, 1–12. <https://doi.org/10.31219/osf.io/v9j52>
- Madaniyah, J., Instrumen, K., & Hasil, T. (2015). KUALITAS INSTRUMEN TES HASIL BELAJAR Khaerudin 1. 2, 212–235. <https://doi.org/10.55210/al-fikru.v5i1.1576>
- Muhammad, I., Marina Angraini, L., Darmayanti, R., & Sugianto, R. (2023). Students' Interest in Learning Mathematics Using Augmented Reality: Rasch Model Analysis. *Edutechnium Journal of Educational Technology*, 1(1), 89–99. <https://doi.org/10.71365/edujet.v1i2.34>
- Nazura, Saputri, D. F., & Angraeni, L. (2021). *Pengembangan Tes Diagnostik Three Tier Test pada Materi*

Pesawat Sederhana untuk Peserta Didik Kelas VIII SMP. 04(02), 54–60.
<https://doi.org/10.31571/jpsa.v4i2.2425>

Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia. (2013). *Standar Penilaian Pendidikan*.

Putri, S., Zulyusri, & Violita. (2023). Analisis Kualitas Soal Ujian Tengah Semester Biologi Kelas XII SMA Negeri 6 Padang Tahun Pelajaran 2021/2022. *Jurnal Metaedukasi: Jurnal Ilmiah Pendidikan*, 4(1), 9–15. <https://doi.org/10.37058/metaedukasi.v4i1.4090>

Sari, R. P. (2022). ANALISIS BUTIR SOAL UJIAN AKHIR SEMESTER GENAP MATA PELAJARAN MATEMATIKA KELAS XI SMK NEGERI 1 PERCUT SEI TUAN T . A . 2019 / 2020 *Jurnal Penelitian Pendidikan MIPA*, 07(01), 91–99. <http://dx.doi.org/10.32696/jp2mipa.v7i1.1360>

Sekarningtias, F. O., Rusilowati, A., & Darsono, T. (2023). *Tinjauan Literatur Sistematis :Tes Diagnostik untuk Mengidentifikasi Miskonsepsi Sains dengan Model Rasch*. *Unnes Physics Education Journal*: 13(1), 15–27. <https://doi.org/10.15294/upej.v13i1.72044>

Solichin, M. (2017). Analisis daya beda soal, taraf kesukaran, validitas butir tes, interpretasi hasil tes dan validitas ramalan dalam evaluasi pendidikan. *Dirasat: Jurnal Manajemen dan Pendidikan Islam*, 2(2), 192–213. <https://doi.org/10.26594/dirasat.v2i2.879>

Soulisa, I., & Manuputty, P. (2021). Pengembangan Alat Penilaian Pembelajaran Bahasa Indonesia Berbasis Kompetensi Menggunakan Pendekatan Kontekstual Di Sma Papua Kota Sorong. *KREDO : Jurnal Ilmiah Bahasa Dan Sastra*, 4(2), 579–593. <https://doi.org/10.24176/kredo.v4i2.4996>

Subari, A., Lufri, L., & Syamsurizal, S. (2021). Analisis Butir Soal Ujian Akhir Semester Ganjil Biologi Kelas XI MAN 2 Kota Jambi. *Bioscientist: Jurnal Ilmiah Biologi*, 9(1), 45. <https://doi.org/10.33394/bjib.v9i1.3612>

Sumaryanta. (2021). Teori Tes Klasik dan Teori Respon Butir: Konsep dan Contoh Penerapannya. In *Cetakan Pertama* (Vol. 15, Issue 2).

Sumintono, B. (2016). Seminar Nasional Pendidikan IPA Prosiding Seminar Nasional Pendidikan IPA “Mengembangkan Keterampilan Berpikir Tingkat Tinggi Melalui Pembelajaran IPA” Penerbit: S2 IPA UNLAM PRESS PENILAIAN KETERAMPILAN BERPIKIR TINGKAT TINGGI: APLIKASI PEMODELAN RASCH P. *Prosiding Seminar Nasional Pendidikan IPA*, 978–602.

Tilaar, A. L. F., & Hasriyanti. (2019). Analisis Butir Soal Semester Ganjil Mata Pelajaran Matematika pada Sekolah Menengah Pertama. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 8(1), 57–68. <https://doi.org/10.15408/jp3i.v8i1.13068>

Triana, A. (2024). Analisis Model Rasch pada Tes Diagnostik Kesulitan Belajar Matematika Siswa SMA. *Jurnal Tarbiyah Dan Ilmu Keguruan Borneo*, 5(1), 119–132. <https://doi.org/10.21093/jtikborneo.v5i3.7135>

Tyas, E. H., Hamdu, G., & Pranata, O. H. (2020). Analisis Soal Pilihan Ganda dengan Menggunakan Pemodelan RASCH untuk Mengukur Kemampuan Siswa dalam Mengurutkan Bilangan

-
- Pecahan di Sekolah Dasar. *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, 7(2), 1-12. <https://doi.org/10.17509/pedadidaktika.v7i2.24773>
- Wahyudi, W. (2010). Assesment Pembelajaran Berbasis Portofolio di Sekolah. *Jurnal Visi Ilmu Pendidikan*, 2(1), 288-296. <https://doi.org/10.26418/jvip.v2i1.370>
- Wardani, H. K., Darusuprapti, F., & Hajaroh, M. (2022). Model-Model Evaluasi Pendidikan Dasar (Scriven Model, Tyler Model, dan Goal Free Evaluation). *Jurnal Pendidikan: Riset dan Konseptual*, 6(1), 36-49. https://doi.org/10.28926/riset_konseptual.v6i1.446
- 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). <https://doi.org/10.30870/gravity.v4i1.3116>
-

LAMPIRAN

EFEKTIVITAS DISTRAKTOR

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: MEASURE ORDER												
ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	Item	
14	A	0	2	6	-1.14	.61	.61	.2	.3	-.19	S14	
	C	0	15	42	-.73	1.00	.27	.6	.6	-.37		
	E	0	10	28	-.24	.97	.32	1.0	1.0	.00		
	D	0	5	14	.69	.51	.26	1.1	1.8	.32		
	B	1	4	11	.98	1.10	.63	.8	1.0	.37		
9	A	0	4	11	-1.79	.87	.50	.2	.2	-.48	S9	
	B	0	3	8	-.90	.00	.00	.2	.4	-.18		
	C	0	5	14	-.64	1.05	.52	1.1	.9	-.14		
	E	0	17	47	.04	.85	.21	1.0	1.3	.22		
	D	1	7	19	.59	1.10	.45	1.1	1.1	.35		
6	A	0	3	8	-1.19	1.53	1.08	.8	.7	-.25	S6	
	C	0	13	36	-.80	.99	.29	.9	.8	-.38		
	D	0	8	22	-.03	.84	.32	1.2	1.3	.09		
	E	0	3	8	.41	.34	.24	1.1	1.5	.17		
	B	1	9	25	.53	.91	.32	1.0	.9	.38		
10	A	0	7	19	-.87	.49	.20	.4	.4	-.28	S10	
	B	0	12	33	-.43	1.28	.39	1.1	1.1	-.12		
	C	0	4	11	-.18	.44	.25	.6	.9	.01		
	E	0	4	11	.13	1.02	.59	1.7	1.9	.11		
	D	1	9	25	.36	1.24	.44	1.2	1.6	.30		
3	A	0	5	14	-1.15	.89	.44	.6	.5	-.32	S3	
	B	0	5	14	-1.08	.42	.21	.3	.4	-.30		
	E	0	3	8	-.65	.17	.12	.5	.6	-.11		
	D	0	11	31	-.59	.97	.31	.8	.8	-.21		
	C	1	12	33	.94	.69	.21	.6	.5	.72		
8	D	0	1	3	-1.76	.00		.1	.2	-.23	S8	
	B	0	6	17	-1.03	.19	.08	.3	.4	-.31		
	A	0	4	11	-.09	.46	.26	.9	1.1	.04		
	E	0	13	36	-.05	1.29	.37	1.7	2.0	.12		
	C	1	12	33	.06	1.17	.35	1.3	2.0	.18		
7	C	0	7	19	-1.27	1.02	.42	.6	.5	-.45	S7	
	A	0	9	25	-.70	.81	.29	.8	.8	-.24		
	B	0	2	6	.17	.35	.35	1.4	1.5	.09		
	E	0	2	6	.38	.91	.91	2.1	2.5	.13		
	D	1	16	44	.36*	.99	.25	1.0	.9	.46		
5	D	0	2	6	-.71	.18	.18	.6	.6	-.10	S5	
	B	0	13	36	-.67	.96	.28	.9	.9	-.29		
	A	0	4	11	-.12	1.42	.82	2.4	2.4	.04		
	E	1	17	47	.14	1.13	.28	1.2	1.4	.31		

5	D	0	2	6	-.71	.18	.18	.6	.6	-.10	S5
	B	0	13	36	-.67	.96	.28	.9	.9	-.29	
	A	0	4	11	-.12	1.42	.82	2.4	2.4	.04	
	E	1	17	47	.14	1.13	.28	1.2	1.4	.31	
15	D	0	2	6	-2.47	.72	.72	.2	.1	-.48	S15
	A	0	10	28	-.68	.66	.22	.9	.8	-.24	
	C	0	5	14	-.57	1.07	.53	1.2	1.4	-.12	
	B	1	19	53	.33	.95	.22	.9	.8	.52	
11	A	0	8	22	-1.29	1.01	.38	.7	.5	-.50	S11
	D	0	6	17	-.81	.58	.26	.8	.7	-.23	
	B	0	2	6	.00	.90	.90	1.8	1.9	.05	
	C	1	20	56	.35	.94	.21	.9	.8	.56	
2	B	0	4	11	-.94	.62	.36	.7	.6	-.22	S2
	E	0	1	3	-.90	.00		.6	.6	-.10	
	C	0	4	11	-.85	1.45	.84	1.4	1.1	-.19	
	D	0	6	17	.01	1.25	.56	2.2	2.4	.09	
	A	1	21	58	-.01*	1.03	.23	1.2	1.1	.22	
4	C	0	3	8	-2.14	.95	.67	.4	.3	-.50	S4
	A	0	9	25	-1.10	.44	.15	.6	.5	-.44	
	E	0	2	6	-.71	.18	.18	.8	.7	-.10	
	B	1	22	61	.43	.82	.18	.7	.6	.72	
1	D	0	2	6	-1.53	.23	.23	.4	.3	-.28	S1
	E	0	4	11	-1.26	.67	.39	.7	.5	-.32	
	C	0	4	11	-.36	1.30	.75	1.4	2.6	-.04	
	B	0	2	6	.53	.36	.36	3.1	2.7	.16	
	A	1	24	67	.01*	1.08	.22	1.1	1.4	.30	
13	B	0	6	17	-1.54	1.01	.45	.7	.6	-.51	S13
	C	0	2	6	-1.10	.20	.20	.8	.6	-.18	
	D	1	28	78	.11	.96	.18	.8	.8	.56	
12	B	0	2	6	-2.04	1.15	1.15	.6	.4	-.39	S12
	C	0	3	8	-1.38	.68	.48	.8	.6	-.30	
	E	0	1	3	1.76	.00		10.0	10.0	.29	
	A	1	30	83	-.06*	.96	.18	.8	.9	.33	

5	D	0	2	6	-.71	.18	.18	.6	.6	-.10	S5
	B	0	13	36	-.67	.96	.28	.9	.9	-.29	
	A	0	4	11	-.12	1.42	.82	2.4	2.4	.04	
	E	1	17	47	.14	1.13	.28	1.2	1.4	.31	
15	D	0	2	6	-2.47	.72	.72	.2	.1	-.48	S15
	A	0	10	28	-.68	.66	.22	.9	.8	-.24	
	C	0	5	14	-.57	1.07	.53	1.2	1.4	-.12	
	B	1	19	53	.33	.95	.22	.9	.8	.52	
11	A	0	8	22	-1.29	1.01	.38	.7	.5	-.50	S11
	D	0	6	17	-.81	.58	.26	.8	.7	-.23	
	B	0	2	6	.00	.90	.90	1.8	1.9	.05	
	C	1	20	56	.35	.94	.21	.9	.8	.56	
2	B	0	4	11	-.94	.62	.36	.7	.6	-.22	S2
	E	0	1	3	-.90	.00		.6	.6	-.10	
	C	0	4	11	-.85	1.45	.84	1.4	1.1	-.19	
	D	0	6	17	.01	1.25	.56	2.2	2.4	.09	
	A	1	21	58	-.01*	1.03	.23	1.2	1.1	.22	
4	C	0	3	8	-2.14	.95	.67	.4	.3	-.50	S4
	A	0	9	25	-1.10	.44	.15	.6	.5	-.44	
	E	0	2	6	-.71	.18	.18	.8	.7	-.10	
	B	1	22	61	.43	.82	.18	.7	.6	.72	
1	D	0	2	6	-1.53	.23	.23	.4	.3	-.28	S1
	E	0	4	11	-1.26	.67	.39	.7	.5	-.32	
	C	0	4	11	-.36	1.30	.75	1.4	2.6	-.04	
	B	0	2	6	.53	.36	.36	3.1	2.7	.16	
	A	1	24	67	.01*	1.08	.22	1.1	1.4	.30	
13	B	0	6	17	-1.54	1.01	.45	.7	.6	-.51	S13
	C	0	2	6	-1.10	.20	.20	.8	.6	-.18	
	D	1	28	78	.11	.96	.18	.8	.8	.56	
12	B	0	2	6	-2.04	1.15	1.15	.6	.4	-.39	S12
	C	0	3	8	-1.38	.68	.48	.8	.6	-.30	
	E	0	1	3	1.76	.00		10.0	10.0	.29	
	A	1	30	83	-.06*	.96	.18	.8	.9	.33	