MASALIQ
Jurnal Pendidikan dan Sains

e-ISSN: 2808-8115
p-ISSN: 2809-1051

Terindeks: Sinta 5, Copernicus, Dimensions, Scilit, Lens, Crossref, Garuda, Google Scholar, Base, etc

https://doi.org/10.58578/masaliq.v6i3.9548

KUALITAS SOAL EVALUASI MATA PELAJARAN BIOLOGI SMA YANG DIKEMBANGKAN MENGGUNAKAN GEMINI AI DENGAN ANALISIS *RASCH MODEL*

Quality of High School Biology Evaluation Items Developed Using Gemini AI with Rasch Model Analysis

Mutiara Salsabila Warman, Relsas Yogica, Syamsurizal, Rahmawati D
Universitas Negeri Padang
relsasyo@fmipa.unp.ac.id

Article Info:

Submitted:	Revised:	Accepted:	Published:
Mar 2, 2026	Mar 30, 2026	Apr 11, 2026	Apr 16, 2026

Abstract

The limited variety of items in the item bank remains an obstacle for Biology teachers in developing quality learning evaluations, even though the items used should be measured empirically for their quality and validity in order to provide comprehensive evaluation results. This study aims to examine the quality of senior high school Biology test items developed through Gemini AI. This study employed a quantitative descriptive approach by utilizing primary data from one class at SMA Pembangunan Laboratorium UNP. The research stages included item development using AI, content validation by experts, item tryout, answer collection and scoring, and item quality analysis. The results showed that the senior high school Biology evaluation items generated by Gemini AI and analyzed using the Rasch Model had fairly good quality, although improvements were still needed in several specific item quality indicators. These findings indicate that the use of Gemini AI can support teachers in developing and evaluating Biology test items more systematically. This study affirms that the use of Gemini AI has the potential to become an effective alternative in item quality analysis, while also providing practical

Masaliq Journal is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

implications for teachers to utilize similar technology in developing evaluation items in various subjects in order to obtain more reliable and higher-quality instruments.

Keywords: Biology; Gemini AI; Item Quality; Rasch Model; Learning Evaluation

Abstrak: Keterbatasan ragam soal dalam *bank soal* masih menjadi kendala bagi guru Biologi dalam menyusun evaluasi pembelajaran yang berkualitas, padahal butir soal yang digunakan seharusnya diukur kualitas dan validitasnya secara empiris agar mampu memberikan hasil evaluasi yang komprehensif. Penelitian ini bertujuan untuk menguji kualitas soal Biologi SMA yang dikembangkan melalui *Gemini AI*. Penelitian ini menggunakan pendekatan deskriptif kuantitatif dengan memanfaatkan data primer dari satu kelas di SMA Pembangunan Laboratorium UNP. Tahapan penelitian meliputi pengembangan soal menggunakan *AI*, validasi isi oleh ahli, uji coba soal, pengumpulan dan koreksi jawaban, serta analisis kualitas soal. Hasil penelitian menunjukkan bahwa soal evaluasi Biologi SMA yang dihasilkan oleh *Gemini AI* dan dianalisis menggunakan *Rasch Model* memiliki kualitas yang cukup baik, meskipun masih memerlukan perbaikan pada beberapa indikator kualitas soal tertentu. Temuan ini menunjukkan bahwa pemanfaatan *Gemini AI* dapat mendukung guru dalam mengembangkan dan mengevaluasi butir soal Biologi secara lebih sistematis. Penelitian ini menegaskan bahwa penggunaan *Gemini AI* berpotensi menjadi alternatif yang efektif dalam analisis kualitas soal, serta memberikan implikasi praktis bagi guru untuk memanfaatkan teknologi serupa dalam pengembangan soal evaluasi pada berbagai mata pelajaran agar diperoleh instrumen yang lebih andal dan berkualitas.

Kata Kunci: Biologi; *Gemini AI*; Kualitas Soal; *Rasch Model*; Evaluasi Pembelajaran

PENDAHULUAN

Pembelajaran biologi di tingkat SMA berperan penting dalam membentuk pemahaman ilmiah peserta didik serta mengembangkan kemampuan berpikir kritis, logis, dan analitis. Selain menyampaikan konsep dan prinsip dasar, biologi juga melatih keterampilan siswa dalam mencari informasi dan menerapkan metode ilmiah untuk memecahkan masalah (Jayanti, 2021). Kurikulum Merdeka mengarahkan pembelajaran biologi agar lebih fleksibel, kontekstual, dan sesuai dengan kebutuhan peserta didik dalam menghadapi tantangan abad ke-21 (Rochim et al., 2024). Guru diberi keleluasaan untuk menyusun pembelajaran berdasarkan potensi siswa dan kondisi sekolah, sehingga proses belajar menjadi lebih adaptif (Sandy & Nugrahaningsih, 2024). Oleh karena itu, evaluasi yang relevan diperlukan untuk mendukung pembelajaran yang efektif, karena berfungsi sebagai alat ukur keberhasilan dalam mencapai tujuan pembelajaran biologi secara optimal. Evaluasi yang dilaksanakan secara sistematis, berkelanjutan, dan berbasis pendekatan autentik mampu memberikan gambaran utuh mengenai kemampuan peserta didik. Hal ini penting sebagai acuan bagi guru dalam menentukan langkah tindak lanjut pembelajaran (Wisman et al., 2021); serta dalam

mengevaluasi dan menyempurnakan proses pembelajaran yang telah berlangsung (Rosyadi, 2024). Dengan demikian, evaluasi memiliki peran krusial dalam memastikan efektivitas pembelajaran dan mendukung pencapaian tujuan pendidikan secara holistik.

Soal biologi yang baik harus memenuhi standar kualitas yang meliputi aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Soal yang baik harus sesuai dengan aspek isi menggunakan bahasa yang komunikatif dan sesuai dengan kaidah Bahasa Indonesia, serta tidak menimbulkan ambiguitas bagi siswa (Sapitri et al. , 2022). Soal juga harus mengacu pada indikator yang telah ditetapkan, serta mencakup berbagai level kognitif, mulai dari mengingat hingga menganalisis (Jannah et al. , 2023). Selain itu, proporsi tingkat kesulitan soal harus seimbang agar dapat mengukur kemampuan siswa secara 4 menyeluruh; misalnya, soal mudah, sedang, dan sulit harus tersebar secara proporsional (Rajagukguk & Naibaho, 2023). Efektivitas pengecoh dalam soal pilihan ganda juga menjadi faktor penting untuk menghindari jawaban yang mudah ditebak dan meningkatkan kualitas soal secara keseluruhan (Oktanin, 2023). Dengan memenuhi standar-standar tersebut, soal biologi dapat memberikan gambaran yang akurat tentang penguasaan materi siswa dan mendukung proses pembelajaran yang bermakna.

Berdasarkan hasil wawancara yang dilakukan peneliti bersama Ibu Dean Roslaini, selaku guru mata pelajaran Biologi di SMA Pembangunan Laboratorium UNP Kota Padang pada tanggal 10 April 2025, ditemukan beberapa kendala yang dihadapi guru dalam menyusun soal evaluasi pembelajaran. Salah satu tantangan utama dalam pelaksanaan evaluasi pembelajaran adalah terbatasnya ragam soal yang tersedia dalam bank soal SMA Pembangunan Laboratorium UNP, sehingga materi evaluasi yang digunakan cenderung bersifat repetitif dan kurang mampu menggambarkan cakupan kompetensi secara menyeluruh. Di samping itu, alokasi waktu yang dimiliki guru untuk merancang instrumen evaluasi sering kali tidak mencukupi, sehingga menghambat proses pengembangan soal yang inovatif dan sesuai dengan kebutuhan peserta didik. Upaya pemanfaatan teknologi sebenarnya telah mulai diterapkan dalam penyusunan soal, namun penerapannya masih belum berjalan secara maksimal karena guru mengalami kesulitan dalam mengakses dan mengimplementasikan teknologi secara berkelanjutan dalam kegiatan evaluasi.

Pemanfaatan teknologi kecerdasan buatan, seperti Gemini AI, dalam dunia pendidikan menjadi salah satu terobosan yang potensial untuk meningkatkan efektivitas penyusunan materi ajar serta instrumen evaluasi pembelajaran yang inovatif dan adaptif

terhadap kebutuhan zaman. Penelitian yang dilakukan oleh Suganda (2023) menunjukkan bahwa Gemini AI merupakan alternatif yang menjanjikan untuk diintegrasikan ke dalam dunia pendidikan, khususnya dalam pengembangan materi ajar dan soal evaluasi. Kemampuannya dalam menghasilkan 6 teks kreatif dalam berbagai format memberikan peluang bagi guru untuk menciptakan bahan ajar yang lebih menarik serta soal-soal yang bervariasi. Beberapa penelitian terdahulu menunjukkan bahwa ketertarikan terhadap pemanfaatan kecerdasan buatan dalam pengembangan soal evaluasi didasari oleh keterbatasan penelitian yang secara langsung menguji kualitas butir soal yang dibuat oleh AI. Penelitian oleh Kasprianto et al. (2025) secara eksplisit menyebutkan bahwa latar belakang dilakukannya penelitian mereka adalah belum adanya pengujian kualitas soal berbasis High Order Thinking Skills (HOTS) yang disusun menggunakan teknologi AI di lingkungan sekolah, sehingga mereka terdorong untuk melakukan kajian terhadap soal yang dikembangkan dengan bantuan AI. Sejalan dengan temuan Zuhri et al., (2024)) yang menyatakan bahwa meskipun kecerdasan buatan menawarkan berbagai kemudahan dan keunggulan dalam dunia pendidikan, penggunaannya tetap perlu disikapi secara cermat, khususnya dalam memastikan mutu butir soal yang digunakan dalam evaluasi. Penelitian ini berangkat dari asumsi serupa: bahwa belum ada studi yang secara sistematis menguji kualitas butir soal yang dihasilkan oleh kecerdasan buatan. Ketiadaan validasi empiris ini membatasi pemahaman kita tentang keandalan dan efektivitas butir soal AI, sehingga mendorong dilakukannya penelitian ini untuk menjembatani kesenjangan tersebut. Senada dengan itu, penelitian lain yang dilakukan dalam konteks pendidikan biologi juga memperlihatkan bahwa penggunaan AI masih berada dalam tahap eksploratif, dan belum banyak menyentuh aspek validasi kualitas instrumen evaluasi secara komprehensif. Dengan demikian, penelitian yang menitikberatkan pada pengujian kualitas butir soal Biologi yang dikembangkan melalui platform kecerdasan buatan, khususnya Web Gemini, masih memiliki relevansi yang tinggi untuk dilakukan, karena dapat mengisi kekosongan kajian yang hingga saat ini belum banyak ditelaah secara mendalam oleh penelitian sebelumnya.

METODE

Penelitian ini merupakan penelitian berjenis deskriptif kuantitatif. Pendekatan ini lebih menekankan pada penyajian fakta sebagaimana adanya berdasarkan hasil pengukuran atau perhitungan yang dilakukan secara sistematis. Penelitian dilakukan di SMA Pembangunan Laboratorium UNP dan di Departemen Biologi Fakultas Matematika dan Ilmu Pengetahuan Alam UNP. Penelitian ini dilaksanakan mulai dari semester I (gasal) tahun

ajaran 2025/2026. Subjek pada penelitian ini kelas X Fase E 1, di SMA Pembangunan Laboratorium UNP semester gasal tahun ajaran 2025/2026. Objek dalam penelitian ini adalah soal evaluasi mata pelajaran biologi yang dikembangkan melalui Gemini AI. Instrumen penelitian ini berupa lembar wawancara guru dan lembah validasi ahli. Pengumpulan data penelitian dilakukan dengan data primer yang didapatkan melalui hasil pengerjaan soal evaluasi oleh siswa dari butir tes/angket observasi dan data sekunder diperoleh dari dokumen pendukung seperti capaian pembelajaran dan kisi-kisi soal Biologi SMA yang didapatkan dari dokumentasi sekolah untuk panduan penyusunan soal evaluasi yang dikembangkan oleh Gemini AI. Teknik analisis data pada butir-butir soal evaluasi mata pelajaran biologi SMA yang dikembangkan oleh Gemini AI dengan mencari validitas dan kualitas item dengan indikator Outfit MNSQ, Outfit ZSTD, dan PT-Mea Corr. Analisis dilakukan dengan metode RASCH Model.

HASIL

Data dalam penelitian yang didapatkan dari subjek penelitian berdasarkan hasil pengerjaan soal oleh siswa kemudian dianalisis untuk mengidentifikasi kualitas dari butir soal Biologi yang diuji. Uji validitas menjadi aspek analisis dari indikator kualitas butir soal biologi. Temuan penelitian akan menyajikan tabel *misfit order* yang menunjukkan kualitas butir soal berdasarkan kesesuaian dengan model Rasch. Tabel misfit order meletakkan item soal yang paling fit berada di tengah-tengah. Evaluasi dilakukan melalui nilai Outfit MNSQ, Outfit ZSTD, dan PT-Measure Correlation

Tabel 1. Misfit Order Kelas X

Item	Outfit MNSQ	Outfit ZSTD	PT-MEA Corr.	Keterangan
S7	9.90	3.0	0.16	Misfit
S10	4.94	3.0	0.38	Misfit
S13	2.44	1.3	0.59	Misfit
S21	2.36	2.3	0.24	Misfit
S11	1.63	1.2	0.45	Misfit
S12	1.57	1.2	0.34	Misfit
S17	1.49	0.9	0.32	Fit meski prediktif
S1	1.21	0.5	0.33	Fit meski prediktif
S8	0.91	0.1	0.44	Fit meski prediktif
S3	0.88	0.0	0.47	Fit
S15	0.93	0.0	0.48	Fit
S2	0.86	0.1	0.37	Fit
S6	0.82	-0.2	0.52	Fit
S22	0.49	-0.3	0.41	Fit
S9	0.52	-0.4	0.45	Fit
S19	0.59	-0.4	0.53	Fit
S14	0.59	-0.3	0.49	Fit
S24	0.43	-0.5	0.50	Misfit
S4	0.33	-0.5	0.44	Misfit
S18	0.71	-0.3	0.58	Fit
S25	0.36	-0.6	0.49	Misfit
S16	0.42	-0.6	0.53	Misfit
S23	0.40	-0.7	0.56	Misfit
S20	0.41	-0.9	0.65	Misfit
S5	0.10	-0.7	0.75	Misfit

Berdasarkan tabel 1, dapat dilihat bahwa terdapat beberapa item yang tergolong misfit, yang ditandai dengan nilai MNSQ yang terlalu tinggi atau terlalu rendah serta nilai korelasi yang rendah, sehingga menunjukkan bahwa butir tersebut kurang mampu mengukur konstruk secara konsisten; S7, S10, S13, S21, S11, S12, S24, S4, S25, S16, S23, ,S20, S5. Sementara itu, butir yang termasuk kategori fit menunjukkan kesesuaian yang baik dengan model dan memiliki kemampuan yang memadai dalam membedakan responden; S3, S15, S2, S6, S22, S9, S19, S14, S18. Selain itu, terdapat pula item yang fit meskipun prediktif, yang berarti masih dapat digunakan karena memenuhi kriteria utama, meskipun memiliki keterbatasan dalam aspek tertentu; S17, S1, S8. Dengan demikian, butir-butir yang tergolong misfit perlu ditinjau lebih lanjut untuk diperbaiki atau dieliminasi agar kualitas instrumen menjadi lebih optimal.

Kualitas asesmen tidak hanya ditentukan oleh kesesuaian butir terhadap model (fit dan misfit), tetapi juga oleh tingkat reliabilitasnya. Reliabilitas menunjukkan konsistensi suatu instrumen dalam mengukur variabel laten serta kemampuannya menghasilkan skor yang stabil pada kondisi yang berbeda. Instrumen dinyatakan memiliki reliabilitas yang dapat

diterima apabila koefisiennya ≥ 0,70, nilai 0,80 tergolong baik, dan ≥ 0,90 termasuk sangat tinggi. Hasil analisis dikelas X menunjukkan nilai item reliability sebesar 0,90 yang berarti instrumen memiliki konsistensi yang sangat tinggi. Sedangkan hasil penilaian dari perhitungan Cronbach's Alpha pada analisis instrumen di kelas X adalah 0.86. Yang mana artinya nilai Cronbach's Alpha dari instrumen ini menunjukkan konsistensi yang sangat bagus.

Analisis Rasch Model dilakukan untuk melihat informasi umum yang bersifat mendasar pada *summary statistic*. Data yang tersaji pada Tabel 2 dibawah menunjukkan hasil rangkuman statistik dari kelas X yang menjadi inti dalam melakukan analisis lanjutan terhadap kualitas butir soal.

Tabel 2. Summary Statistic Kelas X

	Total Score	Count	Measure	Model Error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	20.1	31.0	0.00	0.54	0.95	-0.1	1.41	0.3
S.D.	7.6	0.0	1.85	0.12	0.34	1.1	1.99	1.1
Max.	28.0	31.0	4.45	0.85	1.91	2.7	9.90	3.0
Min.	3.0	31.0	-2.12	0.43	0.39	-2.2	0.10	-0.9
Real RMSE = 0.59			TRUE SD = 1.75		Separation = 2.96		Item Reliability = 0.90	
Model RMSE = 0.59			TRUE SD = 1.76		Separation = 3.17		Item Reliability = 0.91	
S.E. of Item Mean = 0.38								

Berdasarkan tabel 2, nilai mean measure sebesar 0.00 menunjukkan bahwa rata-rata kemampuan responden berada pada titik tengah skala logit. Nilai infit dan outfit MNSQ yang mendekati 1.00 mengindikasikan bahwa secara umum data telah sesuai dengan pemodelan rasch. Standar deviasi (S.D.) pada measure sebesar 1,85 menunjukkan adanya variasi kemampuan responden. Selain itu, nilai person separation sebesar 2.96–3.17 menunjukkan bahwa instrumen mampu membedakan responden ke dalam beberapa tingkat kemampuan. Sementara itu, item reliability sebesar 0.90–0.91 mengindikasikan bahwa butir soal memiliki konsistensi yang sangat baik dalam mengukur konstruk yang diteliti. Dengan demikian, secara keseluruhan instrumen yang digunakan pada kelas X dan data yang diperoleh dapat dinyatakan memiliki kualitas pengukuran yang baik.

PEMBAHASAN

Penelitian ini diawali dengan tahap observasi yang dilakukan di SMA Pembangunan Laboratorium UNP melalui pengisian angket observasi bersama salah satu guru biologi, yaitu Ibu Dean Roslaeni. Hasil observasi menunjukkan bahwa penggunaan soal evaluasi masih kurang bervariasi karena cenderung mengandalkan bank soal sekolah serta buku cetak yang dimiliki oleh guru. Selain itu, terdapat kendala lain berupa keterbatasan waktu dalam

memodifikasi soal-soal yang bersumber dari bank soal maupun buku cetak. Sebagai alternatif, guru telah mencoba memanfaatkan internet dan teknologi berbasis kecerdasan buatan (AI). Namun demikian, kualitas soal yang dihasilkan masih belum dapat dipastikan, mengingat belum banyak penelitian yang secara khusus mengkaji hal tersebut.

Berdasarkan temuan tersebut, peneliti kemudian mengembangkan instrumen soal evaluasi dengan bantuan Gemini AI untuk kelas X yang disesuaikan dengan kebutuhan pembelajaran. Instrumen yang telah disusun selanjutnya melalui tahap validasi logis oleh pakar (dosen dan guru) untuk menilai kesesuaian aspek materi, konstruksi, dan bahasa. Hasil validasi logis menunjukkan bahwa instrumen yang dikembangkan telah memenuhi kriteria kelayakan yaitu kelas X persentase kelayakan 78,75% yang dikategorikan tinggi. Tentunya disertai beberapa perbaikan pada bagian tertentu sesuai saran validator. Setelah dinyatakan layak, instrumen diujicobakan kepada peserta didik untuk memperoleh data respon. Setelah diujicobakan, data yang didapatkan dianalisis dengan bantuan aplikasi Winstep. Berdasarkan hasil analisis data menggunakan pemodelan Rasch, kualitas instrumen yang dikembangkan menunjukkan karakteristik yang beragam ditinjau dari aspek validitas, reliabilitas, serta fungsi distraktor. Secara umum, sebagian besar butir soal telah memenuhi kriteria kesesuaian model (fit) berdasarkan indikator outfit MNSQ, outfit ZSTD, dan PT-Measure Corr. Nilai-nilai tersebut menunjukkan bahwa item mampu berfungsi secara konsisten dalam mengukur konstruk yang sama.

Dalam konteks pemodelan Rasch, kesesuaian item terhadap model menjadi indikator utama bahwa instrumen memiliki validitas konstruk yang baik (Han, Boone, Staver, & Yale, 2014; Bond & Fox, 2015). Hasil ini sejalan dengan penelitian terbaru yang menyatakan bahwa analisis Rasch mampu memberikan informasi yang lebih komprehensif dalam mengevaluasi kualitas butir soal, khususnya dalam mengidentifikasi item yang tidak sesuai (misfit) dan berpotensi menurunkan kualitas instrumen (Sumintono & Widhiarso, 2015; Retnawati, 2019). Selain itu, nilai item reliability yang tinggi menunjukkan bahwa tingkat kesulitan butir soal tersusun secara stabil dan mampu membedakan tingkat kemampuan peserta didik dengan baik. Namun, nilai person reliability yang cenderung lebih rendah mengindikasikan bahwa konsistensi jawaban peserta didik masih bervariasi. Hal ini dapat disebabkan oleh heterogenitas kemampuan responden atau faktor lain seperti kurangnya pemahaman terhadap materi (Boone et al., 2014). Temuan ini didukung oleh penelitian dalam beberapa tahun terakhir yang menyebutkan bahwa dalam analisis Rasch, perbedaan antara item reliability dan person reliability merupakan hal yang umum terjadi, terutama jika distribusi

kemampuan peserta tidak merata (Andrich, 2018; Sumintono & Widhiarso, 2015). Dengan demikian, meskipun kualitas item sudah sangat baik, peningkatan kualitas pembelajaran tetap diperlukan untuk meningkatkan konsistensi respon peserta didik (Azizah & Wahyuningsih, 2020).

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa instrumen yang dikembangkan telah memiliki kualitas yang cukup baik berdasarkan analisis Rasch, baik dari segi validitas dan reliabilitas. Namun demikian, masih diperlukan perbaikan pada beberapa aspek, terutama peningkatan konsistensi respon peserta didik. Dengan melakukan revisi pada butir-butir yang belum optimal, diharapkan instrumen ini dapat digunakan secara lebih efektif dalam mengukur kemampuan peserta didik secara akurat dan objektif.

KESIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa evaluasi mata pelajaran Biologi SMA yang dihasilkan oleh Gemini AI yang sudah dianalisis dengan Rasch Model memiliki kualitas yang cukup baik dengan catatan perbaikan pada indikator kualitas soal tertentu. Hasil temuan ini diharapkan berkontribusi terhadap bidang pendidikan yaitu guru dapat memanfaatkan analisis pemodelan Rasch untuk mengevaluasi kualitas soal karena metode ini mampu memberikan informasi yang lebih komprehensif terkait karakteristik butir dan kemampuan peserta didik dibandingkan analisis klasik. Penelitian selanjutnya diharapkan dapat mengkaji lebih dalam terkait analisis soal dengan jumlah sampel yang lebih besar serta mengembangkan instrumen pada materi atau jenjang yang berbeda sehingga diperoleh hasil yang lebih general dan komprehensif. Penelitian selanjutnya juga disarankan untuk mengkaji lebih lanjut kualitas soal yang dihasilkan oleh teknologi AI dengan membandingkannya dengan soal buatan guru, sehingga dapat diketahui tingkat kelayakan penggunaanya dalam evaluasi pembelajaran.

DAFTAR PUSTAKA

Andrich, D. (2018). Controlling response dependence in the measurement of change using the Rasch model. *Statistical Methods in Medical Research, 27*(12), 3709–3725. https://doi.org/10.1177/0962280217710834

Azizah, A., & Wahyuningsih, S. (2020). Penggunaan Model Rasch untuk Analisis Instrumen Tes pada Mata Kuliah Matematika Aktuaria. *Jurnal Pendidikan Matematika (JUPITEK), 3*(1), 45–50. https://doi.org/10.30598/jupitekvol3iss1pp45-50

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. https://doi.org/10.4324/9781315814698

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education, 15*(4), rm4. https://doi.org/10.1187/cbe.16-04-0148

Han, C. (2019). William J. Boone, John R. Staver and Melissa S. Yale. *Rasch analysis in the human sciences. Journal of Research Design and Statistics in Linguistics and Communication Science, 5*(1–2), 208–211. https://doi.org/10.1558/jrds.37535

Jannah, I. K., Mahanal, S., & Mashfufah, A. (2023). Analisis Tingkat Kognitif Soal Asesmen Sumatif Akhir Semester I (ASAS I) IPA Berbasis Jenis Soal AKM berdasarkan Taksonomi Bloom di Kelas V SD Swasta Kota Malang. *JIIP - Jurnal Ilmiah Ilmu Pendidikan, 6*(2), 806–810. https://doi.org/10.54371/jiip.v6i2.1633

Jayanti, U. N. A. D., & Mahidin. (2021). *Perencanaan Pembelajaran Biologi: Tinjauan Teori, Praktik, dan Paradigma Wahdatul Ulum*. Perdana Publishing. https://repository.uinsu.ac.id/17030/

Kasprianto, R., Munawar, W., & Sriyono. (2025). Analisis Butir Soal High Order Thinking Skills (HOTS) Berbantuan Artificial Intelligence (AI) untuk Pembelajaran Perawatan dan Perbaikan Sasis Sepeda Motor di SMK. *ATIKANOTO: Journal of Automotive Engineering Education, 2*(1). https://doi.org/10.17509/atikanoto.v2i1.87360

Nurpitasari, D. (2022). Kualitas Butir Soal Biologi Kelas X IPA MAN 1 Merangin Tahun Ajaran 2021/2022. *EDU-BIO: Jurnal Pendidikan Biologi, 6*(2), 13–23. https://doi.org/10.30631/edubio.v6i2.17

Rajagukguk, M. J. T., & Naibaho, D. (2023). Mampu Memilih Soal Berdasarkan Tingkat Kesukaran. *Pediaqu: Jurnal Pendidikan Sosial dan Humaniora, 2*(4), 12736–12747. https://publisherqu.com/index.php/pediaqu/article/view/701

Rochim, A. A., Baharung, S., & Isnaini, I. (2024). Perencanaan Pembelajaran Biologi Berbasis Project Based Learning pada Kurikulum Merdeka di SMAN 1 Bungku Tengah. *Jurnal Pendidikan Tambusai, 8*(2). https://doi.org/10.31004/jptam.v8i2.15015

Sandy, D. A., & Nugrahaningsih, W. H. (2024). Implementasi Kurikulum Merdeka pada Mata Pelajaran Biologi di Sekolah Daerah Rural. *Prosiding Semnas Biologi XII Tahun 2024 FMIPA Universitas Negeri Semarang*, 69–74. https://proceeding.unnes.ac.id/semnasbiologi/article/view/3937

Sapitri, A., Kurniati, T., & Yuliawati, A. (2022). Analisis Kualitas Soal UAS Biologi SMA Kelas X dan XI MIA. *Bioeduca: Journal of Biology Education, 4*(1), 45–56. https://doi.org/10.21580/bioeduca.v4i1.8433

Sudianto, Amrillah Rosyadi, & Yusuf. (2024). Evaluasi Tingkat Pemahaman Konsep Siswa pada Materi Komponen Ekosistem dan Interaksi Antar Komponen Kelas X SMA Negeri 2 Bayan Kabupaten Lombok Utara. *Otus Education: Jurnal Biologi dan Pendidikan Biologi, 2*(2), 89–102. https://doi.org/10.62588/otusedu.2024.v2i2.0111

Suganda, A. (2023). Memilih AI yang Tepat untuk Guru: Perbandingan Fitur Gemini, ChatGPT, dan Claude AI. *Jurnal Inovasi Teknologi Dan Edukasi Teknik, 3*(11), 1–10. https://doi.org/10.17977/um068.v3.i11.2023.2

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

Wisman, Y., Effrata, E., & Tutesa, T. (2021). Penerapan Konsep Instrumen Evaluasi Hasil Belajar. *Jurnal Ilmiah Kanderang Tingang, 12*(1), 1–9. https://doi.org/10.37304/jikt.v12i1.105

Zuhri, N. Z., Syihabuddin, & Tatang. (2024). Analisis Validitas, Reliabilitas, dan Tingkat Kesukaran Soal Bahasa Arab Tingkat SMP Berbasis Artificial Intelligence (AI) melalui Platform QuestionWell. *Jurnal Pendidikan dan Pembelajaran Indonesia (JPPI), 4*(2), 693–704. https://doi.org/10.53299/jppi.v4i2.576