

ANALISIS BUTIR SOAL DENGAN ITEMAN 4.0 PADA PENILAIAN AKHIR SEMESTER SEKOLAH DASAR

Zafrullah¹, Reza Kastara², Eko Sutrisno Riantoro³, Anugrah Arya Bakti⁴
^{1,2,3,4}Program Studi Penelitian dan Evaluasi Pendidikan, Sekolah Pascasarjana,
Universitas Negeri Yogyakarta
1zafrullah.2022@student.uny.ac.id

ABSTRACT

Daily assessment harian is an essential tool in education as it helps teachers evaluate students' learning progress. One of the commonly used evaluation formats is multiple-choice questions. The objective of this research is to analyze the items in the End-of-Semester Assessment in the form of multiple-choice questions. This study adopts a quantitative descriptive research design with statistical data analysis interpreted into descriptions. The assessment grid used covers the subject of Civic Education (PKn) for fifth-grade students. The research was conducted in the year 2023, and the test instrument was administered to 32 fifth-grade students in a primary school in Selat Panjang, Riau, Indonesia. Findings from the study indicate that the average score in the End-of-Semester Assessment is 23.875, with a standard deviation of 5.993. The minimum score recorded is 10, and the maximum score is 33. The Mean P value of 0.682 indicates that the difficulty level falls under the "Medium" category. The Mean Rpbis value of 0.359 suggests that the overall item discrimination is considered "Moderate." The obtained reliability coefficient is 0.854, indicating a sufficiently high level of reliability. Despite obtaining satisfactory interpretations for most aspects, there are some items that require revision, particularly in terms of their distractors, to improve the accuracy of interpretation. Well-crafted distractors play a crucial role in effectively assessing students' understanding and abilities.

Keywords: Item Analysis, Iteman, Classical Test Theory

ABSTRAK

Penilaian harian merupakan suatu alat penting dalam pendidikan karena membantu guru dalam mengevaluasi kemajuan pembelajaran peserta didik. Salah satu bentuk evaluasi yang sering digunakan adalah pertanyaan dengan pilihan ganda. Tujuan penelitian ini adalah untuk menganalisis butir soal Penilaian Akhir Semester dengan bentuk pilihan ganda. Penelitian ini menggunakan penelitian deskriptif kuantitatif, dengan analisis data statistik yang diinterpretasikan menjadi deskripsi. Kisi-kisi yang digunakan adalah materi Pendidikan Kewarganegaraan (PKn) kelas V. Penelitian ini dilaksanakan di tahun 2023, dengan instrumen tes diujikan kepada 32 siswa kelas V di salah satu sekolah dasar di Selat Panjang, Riau, Indonesia. Dari temuan peneliti menunjukkan bahwa rata-rata pada Penilaian Akhir Sekolah menunjukkan angka sebesar 23.875, Norma Baku 5.993,

Skor Terkecil sebesar 10, dan Skor Terbesar sebesar 33. Nilai *Mean P* menunjukkan angka sebesar 0.682 menunjukkan bahwa tingkat kesukaran berada pada kategori “Sedang”, nilai *Mean Rpbis* sebesar 0.359, hal ini menandakan bahwa nilai Daya Beda secara keseluruhan memiliki keterangan “Sedang”. Dan reliabilitas yang diperoleh reliabilitasnya yakni nilai *Alpha* sebesar 0.854 dengan nilai reliabilitas cukup tinggi. Meskipun mendapatkan hasil interpretasi yang cukup baik untuk sebagian besar sisi, terdapat beberapa soal yang perlu direvisi dalam hal pengecohannya, agar interpretasi menjadi lebih akurat. Pengecoh dalam suatu soal sangat penting untuk menguji pemahaman dan kemampuan siswa secara efektif.

Kata Kunci: Analisis Butir Soal, Itekan, Teori Tes Klasik

A. Pendahuluan

Semakin hari, dunia mengalami perkembangan yang signifikan, terkhusus dalam bidang kependidikan (Zafrullah et al., 2023; Zafrullah & Zetriuslita, 2021). Dalam era yang semakin berkembang ini, penilaian harian menjadi semakin penting sebagai bentuk evaluasi dalam pendidikan. Evaluasi yang tepat memberikan bantuan bagi guru dalam menilai pemahaman siswa dan meningkatkan kualitas pembelajaran di kelas. Dalam mengembangkan evaluasi pembelajaran, ada tiga aspek yang perlu dinilai, yaitu program pembelajaran, proses pembelajaran, dan hasil pembelajaran. Tujuan dari evaluasi ini adalah untuk mengumpulkan informasi yang menjadi pedoman dalam mengukur kemajuan, perkembangan, dan pencapaian belajar siswa, serta efektivitas

pengajaran yang dilakukan oleh guru (Anggis, 2023). Dalam mengikuti perkembangan zaman, evaluasi pembelajaran juga harus mengakomodasi perubahan dan kemajuan dalam bidang pendidikan. Dengan demikian, evaluasi pembelajaran tidak hanya berfokus pada penilaian kemampuan siswa, tetapi juga melibatkan komponen-komponen yang terkait dengan proses pembelajaran itu sendiri oleh para guru (Raibowo & Nopiyanto, 2020).

Guru memiliki peran penting dalam melaksanakan evaluasi pembelajaran siswa untuk memantau keberhasilan pembelajaran. Dalam hal ini, guru diharapkan melakukan penilaian terhadap hasil belajar siswa guna mengevaluasi sejauh mana pencapaian pembelajaran telah tercapai (Guangul et al., 2020; Supena et al., 2021). Proses evaluasi

pembelajaran di sekolah dapat menggunakan berbagai jenis pertanyaan, termasuk pertanyaan dengan format pilihan ganda (Daniel et al., 2019). Pilihan ganda merupakan salah satu bentuk tes yang umum digunakan dalam pendidikan karena kemampuannya untuk mengukur pemahaman siswa secara cepat dan efisien dalam waktu yang singkat (Kung et al., 2023). Di tingkat sekolah dasar, penilaian harian juga sering menggunakan pertanyaan dengan format pilihan ganda yang disusun oleh guru-guru pengajar. Meskipun demikian, belum ada penelitian yang secara khusus mengevaluasi validitas dan reliabilitas butir soal pada penilaian harian tersebut. Oleh karena itu, penting untuk diingatkan bahwa terdapat berbagai jenis penilaian yang dapat digunakan dalam mengukur kemajuan dan pemahaman siswa dalam pembelajaran.

Soal-soal yang berkualitas adalah soal-soal yang mampu mencerminkan kemampuan dan keterampilan siswa pada tingkat yang telah ditentukan. Untuk memastikan kualitasnya, instrumen pengukuran harus memenuhi kriteria validitas dan reliabilitas yang dapat diandalkan.

Validitas mengacu pada kemampuan instrumen untuk memberikan informasi yang akurat dan sesuai dengan tujuan pengukuran, sedangkan reliabilitas menunjukkan bahwa hasil pengukuran yang diperoleh konsisten setelah dilakukan pengukuran beberapa kali (Hannington et al., 2022; Muqorobin & Triana, 2022). Oleh karena itu, penting bagi guru untuk memilih soal-soal yang memenuhi standar validitas dan reliabilitas agar hasil pengukuran yang diperoleh dapat akurat dan dapat diandalkan (Matthews et al., 2022). Analisis butir soal juga perlu dilakukan untuk mengevaluasi tingkat kesulitan, daya pembeda, dan efektivitas distraktor, yang bertujuan untuk mengidentifikasi apakah soal-soal tersebut terlalu mudah atau terlalu sulit bagi siswa, serta untuk menilai kemampuan setiap butir soal dalam membedakan antara siswa yang telah memahami materi dengan siswa yang belum memahaminya.

Analisis butir soal pilihan ganda merupakan suatu prosedur yang terstruktur untuk memperoleh informasi khusus mengenai butir soal evaluasi (Cahyo et al., 2022). Terdapat beberapa faktor yang harus diperhatikan dalam analisis butir soal

agar dapat dikategorikan sebagai soal evaluasi yang baik. Pertama, penting untuk memeriksa validitas tes, yaitu sejauh mana soal evaluasi dapat mengukur hal-hal yang seharusnya diukur. Kedua, reliabilitas butir soal evaluasi perlu diperhatikan sebagai indikator ketepatan tes tersebut. Ketiga, tingkat kesulitan soal merupakan perbandingan antara jumlah siswa yang menjawab benar dengan total jumlah siswa. Keempat, daya beda soal digunakan untuk mengukur kemampuan tes dalam membedakan antara siswa yang telah memahami materi dengan siswa yang belum. Kelima, tingkat pengecoh soal mencerminkan pilihan jawaban siswa terhadap kemungkinan jawaban yang disediakan pada setiap butir soal evaluasi. Terakhir, efektivitas pengecoh soal mengukur sejauh mana pengecoh dapat membedakan antara siswa yang benar-benar memahami materi dengan siswa yang menebak.

Dari hasil observasi yang sudah dilakukan, didapatkan temuan bahwa beberapa peserta didik yang berprestasi tidak lebih baik daripada peserta didik dengan kemampuan standar. Hal ini dibuktikan dengan

bahwa nilai peserta didik pada penilaian akhir semester terdapat soal yang terlalu mudah, sehingga terdapat kesenjangan antara tingkat kesulitan soal dan kemampuan siswa, yang dapat mempengaruhi hasil evaluasi peserta didik. Hasil penelitian yang sejenis yakni Pamujo & Romadhoni (2022) menunjukkan bahwa kualitas soal berada pada angka sebesar 86%. Penelitian lain menunjukkan bahwa daya beda pada soal yang diujikan adalah 7 soal berkategori buruk, 13 soal berkategori cukup, 14 soal berkategori baik, dan 6 soal berkategori sangat baik (Pradani & Efendi, 2023). Dengan memperoleh informasi tentang tingkat kesulitan, daya pembeda, tingkat pengecoh pada soal, dan reliabilitas, guru dapat melakukan perbaikan pada penyusunan soal dan memberikan umpan balik yang lebih efektif kepada siswa.

Dari penjelasan yang sudah dijelaskan diatas, maka peneliti tertarik melakukan penelitian dengan menganalisis butir soal penilaian akhir semester pada sekolah dasar.

B. Metode Penelitian

Jenis penelitian yang dilakukan adalah penelitian deskriptif kuantitatif,

yang berfokus pada analisis data menggunakan metode perhitungan angka atau statistik (Sugiyono, 2017). Penelitian ini dilaksanakan pada tahun 2023. Subjek pada penelitian ini adalah 32 siswa kelas V di salah satu Sekolah Dasar di Selat Panjang, Provinsi Riau, Indonesia. Soal yang digunakan adalah soal Penilaian Akhir Semester (PAS) pada mata pelajaran Pendidikan Kewarganegaraan (PKn) di kelas V. Pengumpulan data dilakukan dengan melakukan wawancara dengan Guru pada sekolah dasar tersebut. Untuk melakukan tingkat kesukaran, daya pembeda, efektivitas distraktor, dan reliabilitas, peneliti menggunakan aplikasi *Iteman 4.0*.

Adapun kriteria tingkat kesukaran soal dapat dilihat pada Tabel 1 dibawah ini:

Tabel 1 Klasifikasi Tingkat Kesukaran Soal

Tingkat Kesukaran	Keterangan
0,00-0,30	Sukar
0,31-0,70	Sedang
0,71-1,00	Mudah

Sumber: Istiyono (2018)

Untuk indeks daya beda atau diskriminasi pada butir soal dapat dilihat pada Tabel 2 dibawah ini.

Tabel 2 Kriteria Daya Diskriminasi

Kriteria	Keterangan
$D \leq 0,199$	Buruk (ditolak)
0,200 – 0,299	Cukup Baik (Perlu Revisi)
0,300 – 0,399	Sedang (Tidak Perlu Direvisi)
$D \geq 0,400$	Bagus Sekali

Sumber: Istiyono (2018)

Untuk kriteria efektivitas distraktor dapat dilihat pada *proportional endorsing* yang ada pada *Iteman 4.0*. Sedangkan untuk mendapatkan reliabilitas yang baik adalah nilai reliabilitasnya berada pada diatas 0.70 (Cho & Kim, 2015).

C.Hasil Penelitian dan Pembahasan

Tujuan pada penelitian ini adalah untuk melihat kualitas butir soal pada pilihan ganda Penilaian Akhir Semester (PAS) pada mata pelajaran Pendidikan Kewarganegaraan. Dengan analisis menggunakan *Iteman 4.0*, didapatkan hasil analisisnya yang dapat dilihat pada Tabel 3 berikut ini.

Tabel 3 Hasil “Summary Statistics” Secara Keseluruhan

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
Scored Items	35	23.875	5.993	10	33	0.682	0.359

Dari tabel 3, dapat diinterpretasikan bahwa rata-rata

pada Penilaian Akhir Sekolah menunjukkan angka sebesar 23.875,

Norma Baku 5.993, Skor Terkecil yang diperoleh yakni siswa memperoleh jawaban yang benar sebesar 10, dan Skor Terbesar yang diperoleh adalah siswa memperoleh jawaban yang benar sebesar 33. Dari nilai *Mean P* menunjukkan angka sebesar 0.682 menunjukkan bahwa tingkat kesukaran soal pada Penilaian Akhir Semester ini berada pada kategori "Sedang". *Mean Rpbis* menunjukkan daya beda untuk keseluruhan soal, dengan nilai *Mean Rpbis* sebesar 0.359, hal ini menandakan bahwa nilai Daya Beda memiliki keterangan "Sedang".

Adapun hasil analisis tingkat kesukaran per item dapat dilihat pada Tabel 4 berikut ini.

Tabel 4 Hasil Analisis Tingkat Kesukaran Per Butir

Kategori	Butir	Jumlah
Mudah	9, 11, 13, 15, 16, 17,	15
	18, 19, 20, 21, 24, 28, 30, 32, 34	
Sedang	4, 5, 12, 22, 23, 25, 26, 27, 29, 31	10
Sulit	1, 2, 3, 6, 7, 8, 10, 14, 33, 35	10

Sumber: *Modifikasi Peneliti*

Berdasarkan tabel yang disebutkan sebelumnya, terdapat 10 soal yang termasuk dalam kategori "Sedang" dan "Sulit". Soal-soal ini dianggap berimbang karena jumlahnya cukup signifikan dan

mewakili tingkat kesulitan yang lebih tinggi. Namun, perlu diperhatikan bahwa soal-soal yang terlalu sulit dapat memiliki dampak negatif pada siswa. Soal-soal yang berlebihan dalam tingkat kesulitan dapat menyebabkan siswa kehilangan rasa percaya diri dan merasa tidak mampu menjawabnya. Hal ini dapat menghambat kemampuan mereka untuk menunjukkan pemahaman dan pengetahuan yang sebenarnya mereka miliki.

Penting bagi peserta didik untuk memperhatikan tingkat kesulitan yang sesuai dalam menyusun soal. Kesulitan soal yang tepat akan memungkinkan siswa untuk menunjukkan pemahaman mereka tanpa merasa terlalu ditekan atau terbebani. Jika soal-soal terlalu sulit, maka tujuan evaluasi menjadi terdistorsi, karena siswa mungkin tidak mampu menunjukkan pemahaman mereka dengan baik. Oleh karena itu, disarankan agar soal-soal yang terlalu sulit direvisi atau dibuang, dan kembali merujuk pada soal-soal dengan tingkat kesulitan "Mudah" yang sesuai dengan materi, sehingga dapat memberikan kesempatan yang adil bagi siswa untuk menunjukkan

kemampuan mereka dalam menguasai materi yang diajarkan.

Untuk hasil analisis daya beda per butir dapat dilihat pada Tabel 5 berikut ini.

Tabel 5 Hasil Analisis Daya Pembeda Per Butir

Kategori	Butir	Jumlah
Tidak Baik	1, 2, 3, 6, 7, 8, 14, 26, 35	9
Cukup Baik	17, 21, 28, 33	4
Baik	4, 5, 9, 10, 11, 12, 13, 15, 16, 18, 19, 20, 22, 23, 24, 25, 27, 29, 30, 31, 32, 34	22

Sumber: *Modifikasi Peneliti*

Berdasarkan hasil daya beda yang tertera dalam tabel, dapat disimpulkan bahwa terdapat 9 soal yang termasuk dalam kategori "Tidak Baik" dan 4 soal yang termasuk dalam kategori "Cukup Baik". Hal ini menunjukkan bahwa ada ketimpangan dalam tingkat kesulitan antara kedua kategori soal tersebut. Soal-soal yang dinyatakan "Tidak Baik" memiliki daya beda yang rendah, yang artinya mereka tidak efektif dalam membedakan antara siswa yang memiliki tingkat pemahaman yang berbeda. Sementara itu, soal-soal yang dinyatakan "Cukup Baik" memiliki daya beda yang lebih baik, yang

menunjukkan bahwa mereka mampu membedakan antara siswa yang memiliki kemampuan yang sedikit berbeda.

Daya beda mengacu pada kemampuan soal dalam membedakan antara siswa yang memiliki tingkat pemahaman yang berbeda. Soal dengan daya beda yang rendah cenderung memberikan hasil yang serupa di antara siswa, sehingga tidak memberikan informasi yang signifikan tentang kemampuan individu. Sebaliknya, soal dengan daya beda yang tinggi dapat memberikan hasil yang berbeda antara siswa yang memiliki tingkat pemahaman yang berbeda. Oleh karena itu, perlu diperhatikan untuk memperbaiki atau mengganti soal-soal yang memiliki daya beda rendah agar evaluasi dapat memberikan informasi yang lebih akurat tentang kemampuan siswa.

Untuk efektivitas distraktor, peneliti hanya mengambil tiga contoh, yakni distraktor dengan daya beda Tidak Baik, dan Cukup Baik.

Item information

Seq.	ID	Key	Scored	Num Options	Domain	Flags
8	no8	B	Yes	4	1	K, LR

Item statistics

N	P	Total Rpbis	Total Rbis	Alpha w/o
32	0.469	-0.182	-0.229	0.885

Option statistics

Option	N	Prop.	Rpbis	Rbis	Mean	SD	Color	
A	8	0.250	0.132	0.180	24.750	5.471	Maroon	
B	15	0.469	-0.182	-0.229	23.267	7.191	Green	**KEY**
C	2	0.063	0.136	0.288	26.500	9.192	Blue	
D	7	0.219	0.002	0.003	23.429	3.735	Olive	
Omit	0							
NotAdmin	0							

Quantile plot data

Option	N	0-20%	20-40%	40-60%	60-80%	80-100%	Color	
A	8	0.167	0.286	0.167	0.286	0.333	Maroon	
B	15	0.667	0.286	0.333	0.571	0.500	Green	**KEY**
C	2	0.000	0.143	0.000	0.000	0.167	Blue	
D	7	0.167	0.286	0.500	0.143	0.000	Olive	

Gambar 1 Hasil Efektivitas Distraktor Kategori “Tidak Baik”

Item nomor 8 memiliki nilai korelasi poin biserial (r_{pbis}) sebesar -0.182 untuk kunci jawaban, sedangkan nilai r_{pbis} untuk alternatif jawaban bernilai positif. Hal ini menandakan bahwa terdapat hubungan kebalikan antara jawaban yang benar dan respons siswa. Dalam konteks ini, semakin tinggi skor pada kunci jawaban, semakin rendah skor respons siswa. Hal ini menunjukkan bahwa siswa yang menjawab dengan benar pada item nomor 8 cenderung memiliki respons yang lebih rendah daripada siswa yang menjawab salah. Hal ini mengindikasikan bahwa item nomor 8 memiliki tingkat kesulitan yang lebih tinggi, di mana siswa kesulitan dalam menjawab dengan benar dan sebaliknya siswa cenderung memberikan jawaban yang salah. Faktor-faktor lain seperti kebingungan dalam pemahaman materi atau kekurangan pengetahuan mungkin berperan dalam memengaruhi hubungan ini. Oleh karena itu, item nomor 8 perlu diperhatikan dalam analisis lebih lanjut untuk

meningkatkan pemahaman dan pencapaian siswa.

Item information

Seq.	ID	Key	Scored	Num Options	Domain	Flags
17	no17	A	Yes	4	1	

Item statistics

N	P	Total Rpbis	Total Rbis	Alpha w/o
32	0.875	0.248	0.398	0.852

Option statistics

Option	N	Prop.	Rpbis	Rbis	Mean	SD	Color	
A	28	0.875	0.248	0.398	24.536	5.764	Maroon	**KEY**
B	3	0.094	-0.226	-0.392	19.000	8.544	Green	
C	1	0.031	-0.094	-0.233	20.000	0.000	Blue	
D	0	0.000	--	--	--	--	Olive	
Omit	0							
NotAdmin	0							

Quantile plot data

Option	N	0-20%	20-40%	40-60%	60-80%	80-100%	Color	
A	28	0.833	0.714	1.000	0.857	1.000	Maroon	**KEY**
B	3	0.167	0.143	0.000	0.143	0.000	Green	
C	1	0.000	0.143	0.000	0.000	0.000	Blue	
D	0	0.000	0.000	0.000	0.000	0.000	Olive	

Gambar 2 Hasil Efektivitas Distraktor Salah Satu Soal dengan Kategori “Cukup Baik”

Item nomor 17 memiliki nilai korelasi poin biserial (r_{pbis}) sebesar -0.248 untuk kunci jawaban, sedangkan nilai r_{pbis} untuk alternatif jawaban bernilai positif dan tidak bernilai. Hal ini menandakan bahwa adanya hubungan positif antara jawaban yang benar dan respons siswa. Dalam konteks ini, semakin tinggi skor pada kunci jawaban, semakin tinggi pula skor respons siswa. Hal ini menunjukkan bahwa siswa yang menjawab dengan benar pada item nomor 17 cenderung memiliki respons yang lebih tinggi daripada siswa yang menjawab salah atau tidak menjawab. Interpretasi ini mengindikasikan bahwa item nomor 17 memiliki tingkat kesulitan yang relatif rendah, di mana siswa cenderung mampu menjawab dengan

benar. Faktor-faktor seperti pemahaman materi yang baik atau pengetahuan yang memadai mungkin berperan dalam memengaruhi hubungan ini. Oleh karena itu, item nomor 17 dapat dianggap sebagai

indikator yang baik dalam mengukur pemahaman siswa terhadap materi yang diujikan.

Untuk melihat Reliabilitas secara keseluruhan, dapat dilihat pada Tabel 6 berikut ini.

Tabel 6 Hasil Reliabilitas Secara Keseluruhan

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0.854	2.292	0.889	0.796	0.756	0.941	0.886	0.861

Dari tabel 4, dapat dilihat reliabilitasnya yakni nilai *Alpha* sebesar 0.854. Hal ini menunjukkan bahwa nilai reliabilitas pada soal ini berada pada kategori "Sangat Reliabel" dikarenakan nilai tersebut berada di atas 0.7, dan sesuai dengan pernyataan dari Cho & Kim (2015).

E. Kesimpulan

Berdasarkan temuan dari peneliti, maka didapatkan kesimpulan yakni rata-rata pada Penilaian Akhir Sekolah menunjukkan angka sebesar 23.875, Norma Baku 5.993, Skor Terkecil sebesar 10, dan Skor Terbesar sebesar 33. Nilai *Mean P* menunjukkan angka sebesar 0.682 menunjukkan bahwa tingkat kesukaran berada pada kategori "Sedang", nilai *Mean Rpbis* sebesar 0.359, hal ini menandakan bahwa nilai Daya Beda secara keseluruhan

memiliki keterangan "Sedang". Dan reliabilitas yang diperoleh reliabilitasnya yakni nilai *Alpha* sebesar 0.854 dengan nilai reliabilitas cukup tinggi.

Meskipun mendapatkan hasil interpretasi yang cukup baik untuk sebagian besar sisi, terdapat beberapa soal yang perlu direvisi dalam hal pengecohannya, agar interpretasi menjadi lebih akurat. Pengecoh dalam suatu soal sangat penting untuk menguji pemahaman dan kemampuan siswa secara efektif. Jika pengecoh tidak memadai, hal ini dapat mengarah pada hasil yang bias atau kurang representatif. Oleh karena itu, penting untuk meninjau kembali pengecoh yang digunakan dalam beberapa soal dan memastikan bahwa mereka secara efektif membingungkan siswa yang tidak memahami materi yang diujikan. Dengan melakukan revisi pada

pengecoh yang tidak efektif, interpretasi hasil tes dapat menjadi lebih akurat dan memberikan informasi yang lebih bermakna tentang kemampuan siswa dalam memahami materi yang diujikan.

DAFTAR PUSTAKA

- Anggis, E. V. (2023). EVALUASI KRITERIA PENILAIAN PEMBELAJARAN BIOLOGI BERBASIS COUNTENANCE. *Jurnal Penjaminan Mutu*, 9(01), 93–99.
- Cahyo, A. N., Luriawati, D., & Wagiran, W. (2022). Analisis Butir Soal Penilaian Keterampilan Kebahasaan pada Pembelajaran Teks Eksplanasi Kelas XI. *JBSI: Jurnal Bahasa Dan Sastra Indonesia*, 2(01), 11–22.
- Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods*, 18(2), 207–230.
- Daniel, M., Rencic, J., Durning, S. J., Holmboe, E., Santen, S. A., Lang, V., Ratcliffe, T., Gordon, D., Heist, B., & Lubarsky, S. (2019). Clinical reasoning assessment methods: a scoping review and practical guidance. *Academic Medicine*, 94(6), 902–912.
- Guangul, F. M., Suhail, A. H., Khalit, M. I., & Khidhir, B. A. (2020). Challenges of remote assessment in higher education in the context of COVID-19: a case study of Middle East College. *Educational Assessment, Evaluation and Accountability*, 32, 519–535.
- Hannington, T., Asuman, B., & Mary, B. (2022). Savings and women involvement in business in Kasese District: A Case of Women Entrepreneurs in Hima Town Council. *IDOSR Journal of Current Issues in Social Sciences*, 8(1), 38–44.
- Istiyono, E. (2018). *Pengembangan instrumen penilaian dan analisis hasil belajar fisika dengan teori tes klasik dan modern*. UNY Press Yogyakarta.
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., & Maningo, J. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digital Health*, 2(2), e0000198.
- Matthews, R. A., Pineault, L., & Hong, Y.-H. (2022). Normalizing the use of single-item measures: Validation of the single-item compendium for organizational psychology. *Journal of Business and Psychology*, 37(4), 639–673.
- Muqorobin, M., & Triana, H. (2022). Validitas Instrumen Tingkat Kecemasan Siswa Terhadap Pembelajaran Matematika Dalam Penelitian Eksperimen. *JISIP (Jurnal Ilmu Sosial Dan Pendidikan)*, 6(2).
- Pamujo, P., & Romadhoni, D. R. M. (2022). ANALISIS BUTIR SOAL EVALUASI PEMBELAJARAN PKn KELAS IV SEKOLAH DASAR DENGAN ITEMAN DI

KECAMATAN
BANJARNEGARA.
PROCEEDING UMSURABAYA,
1(1).

assisted by Adobe Flash CS6.
*Math Didactic: Jurnal Pendidikan
Matematika*, 7(2), 114–123.
<https://doi.org/10.33654/math.v7i2.1272>

- Pradani, R. A., & Efendi, A. (2023). Analisis Butir Soal Ujian Sekolah Menggunakan Program Iteman (Analysis of School Exam Questions Using the Iteman Program). *Indonesian Language Education and Literature*, 8(2), 275–289.
- Raibowo, S., & Nopiyanto, Y. E. (2020). Evaluasi Pembelajaran Pendidikan Jasmani Olahraga & Kesehatan pada SMP Negeri Se-Kabupaten Mukomuko melalui Pendekatan Model Context, Input, Process & Product (CIPP). *Jurnal Pendidikan Kesehatan Rekreasi*, 6(2), 146–165.
- Sugiyono. (2017). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Supena, I., Darmuki, A., & Hariyadi, A. (2021). The Influence of 4C (Constructive, Critical, Creativity, Collaborative) Learning Model on Students' Learning Outcomes. *International Journal of Instruction*, 14(3), 873–892.
- Zafrullah, Z., Suyanto, S., Wahyuni, A., Ayuni, R. T., & Novilanti, F. R. E. (2023). Development of Android-based Learning to Improve Computational Thinking Skills in Junior High School. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 7(2), 1309–1320.
- Zafrullah, Z., & Zetriuslita, Z. (2021). Learning interest of seventh grade students towards mathematics learning media