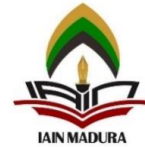




**GHÂNCARAN: JURNAL PENDIDIKAN
BAHASA DAN SASTRA INDONESIA**

<http://ejournal.iainmadura.ac.id/ghancaran>
E-ISSN: 2715-9132; P-ISSN: 2714-8955
DOI 10.19105/ghancaran.v7i2.21239



Analisis Kualitas Butir Soal Asesmen Sumatif Mata Pelajaran Bahasa Indonesia di Sekolah

Siti Kasiyani* & Siti Isnaniah*

*Tadris Bahasa Indonesia, UIN Raden Mas Said Surakarta, Indonesia
Alamat surel: sitikasiyaniiii@gmail.com; sisnaniah7@gmail.com

Abstract

Keywords:
Item analysis;
Summative
assessment;
Anates.

Exam questions are often designed without prior comprehensive analysis of the quality of each item. As a result, many questions fail to optimally measure student abilities. This study aimed to assess the quality of summative assessment items in the Indonesian language subject for eighth-grade students at Batik Special Program Junior High School, Surakarta, in the 2024/2025 academic year. This study employed a quantitative descriptive approach with the aid of Anates version 4 software to analyze validity, reliability, difficulty level, discriminating power, and distractor effectiveness. The subjects were all 25 eighth-grade students. The data analyzed consisted of 30 multiple-choice questions. Based on the analysis results, several conclusions were obtained: 1) as many as 53% of the questions were included in the valid category, 2) the reliability coefficient showed a value of 0.72 which placed the questions in the high reliability category, 3) the difficulty level of most of the questions was at the easy to very easy level, 4) the discriminating power analysis showed that 37% of the questions did not have the ability to differentiate students well, 5) the effectiveness of distractors indicated that most of the answer options did not function optimally. These findings indicate that the quality of the questions still needs to be improved, especially in the aspects of discriminating power and distractor effectiveness, so that it can be used as a basis for improving the preparation of assessment instruments in the future.

Abstrak

Kata Kunci:
Analisis soal;
Asesmen sumatif;
Anates.

Penyusunan soal ujian kerap dilakukan tanpa didahului oleh analisis yang komprehensif terhadap kualitas setiap butir soal. Akibatnya, banyak soal yang tidak mampu mengukur kemampuan siswa secara optimal. Dilaksanakannya penelitian ini memiliki tujuan untuk mengkaji kualitas butir soal asesmen sumatif pada mata pelajaran Bahasa Indonesia kelas VIII di SMP Batik Program Khusus Surakarta Tahun Ajaran 2024/2025. Penelitian ini menerapkan pendekatan deskriptif kuantitatif dengan bantuan perangkat lunak Anates versi 4 untuk menganalisis aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh. Subjek penelitian adalah seluruh siswa kelas VIII yang berjumlah 25 orang. Data yang dianalisis berupa 30 butir soal pilihan ganda. Berdasarkan hasil analisis, diperoleh beberapa kesimpulan: 1) sebanyak 53% butir soal termasuk dalam kategori valid, 2) koefisien reliabilitas menunjukkan nilai sebesar 0,72 yang menempatkan soal dalam kategori reliabel tinggi, 3) tingkat kesukaran sebagian besar soal berada pada level mudah hingga sangat mudah, 4) analisis daya pembeda menunjukkan bahwa 37% butir soal belum memiliki kemampuan membedakan peserta didik secara baik, 5) efektivitas

pengecoh mengindikasikan bahwa sebagian besar opsi jawaban tidak berfungsi secara maksimal. Temuan tersebut menunjukkan bahwa kualitas soal masih perlu ditingkatkan, khususnya pada aspek daya pembeda dan efektivitas pengecoh, sehingga dapat dijadikan dasar dalam perbaikan penyusunan instrumen asesmen pada masa mendatang.

Dikirim: 18 Juli 2025; Revisi: 16 November 2025; Diterbitkan: 7 Januari 2026

©Ghâncaran: Jurnal Pendidikan Bahasa dan Sastra Indonesia

Tadris Bahasa Indonesia

Institut Agama Islam Negeri Madura, Indonesia

PENDAHULUAN

Evaluasi hasil belajar merupakan bagian penting dalam proses pendidikan untuk mengetahui seberapa jauh tujuan pembelajaran telah tercapai (Zainuri, dkk., 2021). Setiap proses pembelajaran tidak dapat dipisahkan dari kegiatan evaluasi pembelajaran (Himawan & Suyata, 2021). Tujuan dari pelaksanaan evaluasi adalah untuk menilai seberapa jauh efektivitas metode-metode yang telah diterapkan ketika proses belajar mengajar dalam kurun waktu tertentu (Nafs, dkk., 2023). Menurut Riinawati (2021), evaluasi pendidikan dimaknai sebagai suatu proses atau tindakan untuk mengukur nilai atau keberhasilan dari berbagai aspek dalam dunia pendidikan yang bertujuan untuk mengukur seberapa efektif strategi yang diterapkan dalam mencapai tujuan pembelajaran. Evaluasi merupakan konsep yang bersifat menyeluruh, mencakup serangkaian kegiatan sistematis dalam mengumpulkan, mengolah, menganalisis, dan menafsirkan berbagai informasi yang relevan guna menentukan tingkat efektivitas pelaksanaan proses pembelajaran serta ketercapaian tujuan dan hasil belajar peserta didik.

Evaluasi hasil belajar siswa umumnya dilakukan melalui metode tes dan non-tes. Di antara kedua metode tersebut, tes merupakan instrumen yang paling sering digunakan dalam praktik evaluasi pendidikan karena dirancang secara khusus untuk mengukur dan menilai tingkat pencapaian hasil belajar peserta didik. Penggunaan tes terutama difokuskan pada ranah kognitif, yakni aspek yang berkaitan dengan pemahaman, penguasaan, dan kemampuan siswa dalam mengolah materi pembelajaran yang telah diajarkan, sehingga hasilnya dapat dijadikan dasar pengambilan keputusan pembelajaran secara objektif dan terukur (Rahayu, 2024). Dalam dunia pendidikan, tes berfungsi sebagai instrumen untuk menggali kemampuan tertentu serta membedakan tingkat kompetensi antar siswa. Oleh sebab itu, perencanaan dan penyusunan soal tes harus dilakukan dengan sebaik mungkin.

Dalam evaluasi terdapat berbagai jenis asesmen, salah satunya adalah asesmen sumatif. Menurut Depdiknas (2007), asesmen sumatif adalah bentuk penilaian yang dilakukan setelah siswa menyelesaikan suatu program pembelajaran. Tujuan dari

penilaian ini untuk mengukur seberapa jauh pencapaian siswa terhadap target kompetensi. Pelaksanaannya dapat berupa ulangan akhir semester, ujian tingkat akhir pada satuan pendidikan, atau sejenisnya sebagai bentuk penilaian akhir dari pencapaian kompetensi siswa. Kualitas soal-soal dalam asesmen sumatif sangat menentukan validitas dan reliabilitas hasil evaluasi tersebut. Agar dapat mengetahui kualitas suatu soal baik itu dari segi validitasnya, reliabilitasnya, tingkat kesukarannya, daya pembedanya, maupun efektivitas pengecohnya maka perlu dilakukan analisis butir soal.

Mengkaji atau menganalisis kualitas soal menjadi langkah yang penting bagi guru dalam upaya memperbaiki kualitas soal yang telah disusun (Magdalena, dkk., 2021). Kegiatan ini dilakukan untuk menilai dan memperbaiki mutu soal agar instrumen evaluasi yang digunakan dapat mengukur kemampuan siswa secara tepat dan efektif. Pada praktiknya, soal ujian sering disusun tanpa melalui proses analisis yang komprehensif terhadap kualitas butir soal. Akibatnya, banyak soal yang tidak mampu mengukur kemampuan siswa secara optimal, baik karena tingkat kesulitan yang tidak sesuai, daya pembeda yang rendah, maupun adanya distraktor yang tidak berfungsi secara efektif. Keberfungsian suatu soal tes dapat diketahui melalui kegiatan analisis soal (Mustaqim & Sulisti, 2024). Oleh karena itu, diperlukan analisis kuantitatif terhadap soal-soal asesmen untuk memastikan fungsi dan kualitasnya.

Arikunto (2018) membagi beberapa cara menilai tes atau menganalisis soal menjadi empat. Empat cara tersebut meliputi meneliti secara jujur, mengadakan analisis soal, mengecek validitas, dan yang terakhir mengecek reliabilitas. Terdapat tiga permasalahan yang erat kaitannya dengan analisis butir soal, yaitu taraf kesukaran, daya pembeda, pola atau bentuk jawaban soal, dan efektivitas pengecoh. Analisis validitas dilakukan untuk mengetahui sejauh mana suatu instrumen tes mampu mengukur secara cermat, tepat, dan sesuai dengan tujuan pengukuran yang telah ditetapkan. Analisis reliabilitas bertujuan untuk menilai tingkat konsistensi atau keandalan hasil suatu pengukuran sehingga dapat dipercaya dan digunakan secara berulang (Marambaawang, dkk., 2023). Dengan analisis-analisis tersebut dapat didapatkan informasi mengenai kekurangan sebuah soal sekaligus sebagai petunjuk untuk melakukan sebuah perbaikan.

Proses pembelajaran tentu melibatkan berbagai upaya strategis yang dirancang untuk mencapai hasil yang optimal (Wachidah, dkk., 2021), salah satunya adalah dengan mengkaji atau menganalisis kualitas soal. Analisis kualitas butir soal memiliki peran krusial dalam menghasilkan soal-soal yang berkualitas untuk digunakan sebagai perangkat evaluasi (Kaka, dkk., 2024). Analisis kuantitatif butir soal dapat dilakukan dengan memakai perangkat lunak khusus, satu di antaranya adalah *software* Anates

versi 4.0. *Software* Anates merupakan sebuah perangkat lunak yang didesain guna menganalisis efektivitas soal sehingga sangat membantu guru dalam mengevaluasi hasil belajar siswa. *Software* tersebut dikembangkan oleh Karnoto yang merupakan seorang dosen Psikologi dari Universitas Pendidikan Indonesia (UPI), bekerja sama dengan Yudi Wibisono seorang konsultan di bidang komputer (Karnoto & Wibisono, 2004). Penggunaan Anates memungkinkan proses analisis dilakukan secara cepat, sistematis, dan akurat.

Penelitian ini dilaksanakan dengan tujuan untuk mengkaji kualitas soal asesmen sumatif mata pelajaran Bahasa Indonesia kelas VIII melalui studi kasus di SMP Batik Program Khusus Surakarta Tahun Ajaran 2024/2025 menggunakan pendekatan kuantitatif dengan bantuan *software* Anates versi 4.0. Kajian difokuskan pada aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh setiap butir soal. Temuan penelitian diharapkan mampu menyajikan gambaran objektif mengenai kualitas soal serta menjadi bahan evaluasi dalam penyempurnaan soal di masa yang akan datang.

Analisis butir soal dengan memanfaatkan alat bantu *software* Anates menjadi pendekatan yang efektif untuk mengevaluasi kualitas instrumen asesmen. Dalam penelitian oleh Syafiriya, dkk. (2024) ditemukan bahwa reliabilitas soal tergolong cukup (0,65) tetapi sebagian besar soal tergolong sangat mudah (80%) dan hanya 3 dari 15 soal yang layak digunakan tanpa revisi. Penelitian ini memiliki kesamaan dengan penelitian penulis dalam penggunaan metode kuantitatif dan instrumen Anates untuk menganalisis validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta pengecoh. Namun, penelitian ini berbeda dari segi konteks karena berfokus pada soal HOTS di mata pelajaran Akuntansi SMK, serta tidak menggunakan teori Arikunto sebagai dasar analisis.

Penelitian oleh Jatmoko, dkk. (2025) didapatkan hasil bahwa 95% soal valid, reliabilitas tinggi (0,78), dan sebagian besar soal tergolong sukar. Penelitian ini sejalan dengan penelitian penulis dalam hal metode dan pemanfaatan Anates tetapi pendekatannya lebih menekankan pada aspek diagnostik kognitif dan diterapkan pada mata pelajaran Teknik Dasar Otomotif. Penelitian ini memperlihatkan pentingnya evaluasi menyeluruh terhadap soal sebagai bagian dari pengukuran kompetensi siswa tetapi berbeda dengan penelitian penulis yang fokus pada asesmen sumatif Bahasa Indonesia.

Sementara itu, penelitian oleh Fiska, dkk. (2021) didapatkan hasil bahwa validitas soal sangat rendah (hanya 5% soal yang valid) serta reliabilitas soal belum memenuhi kategori baik (0,54). Penelitian ini memiliki kesamaan dengan penelitian penulis dalam

penggunaan *software* Anates dan pendekatan klasik, serta konteks satuan pendidikan SMP. Namun, fokusnya lebih kepada ulangan harian IPA, sedangkan penelitian penulis terfokus pada asesmen sumatif mata pelajaran Bahasa Indonesia.

Dengan demikian, ketiga penelitian ini memperkuat pentingnya analisis butir soal dalam meningkatkan kualitas evaluasi pembelajaran. Persamaan utamanya terletak pada pendekatan kuantitatif dan penggunaan *software* Anates, sementara perbedaannya terletak pada konteks mata pelajaran, tingkat pendidikan, dan dasar teori yang digunakan. Penelitian penulis menambahkan kontribusi penting dengan mengintegrasikan teori Arikunto dalam menganalisis kualitas butir soal asesmen sumatif Bahasa Indonesia di SMP sehingga memberikan dasar teoretis yang kuat untuk interpretasi hasil analisis.

METODE

Penelitian ini tergolong sebagai penelitian deskriptif kuantitatif yang dilaksanakan bersamaan dengan pelaksanaan asesmen sumatif semester genap di SMP Batik Program Khusus Surakarta tahun ajaran 2024/2025 tepatnya mata pelajaran Bahasa Indonesia kelas VIII, yaitu dalam rentang waktu bulan Juni 2025. Populasi penelitian ini mencakup seluruh siswa kelas VIII di SMP Batik Program Khusus Surakarta dengan jumlah sebanyak 25 siswa, meliputi kelas VIII A dan B. Lokasi penelitian berada di SMP Batik Program Khusus Surakarta yang beralamat di Jalan Agus Salim nomor 17, Sondakan, Laweyan, Surakarta. Data yang dianalisis merupakan soal pilihan ganda yang berjumlah 30 butir dari jumlah keseluruhan 35 butir soal asesmen sumatif. Lima butir soal lainnya berbentuk uraian sehingga tidak memungkinkan untuk dianalisis dari segi efektivitas pengecohnya sehingga penelitian ini hanya memfokuskan analisis pada 30 soal pilihan ganda. Dalam proses analisisnya memanfaatkan perangkat lunak Anates versi 4.0, yang dikenal efektif dalam mengukur aspek-aspek penting, seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta pengecoh pada soal.

Penelitian kuantitatif merupakan penelitian yang digunakan untuk memahami suatu fenomena secara menyeluruh dalam konteks aslinya dengan mengandalkan data yang dikumpulkan dari situasi nyata (Hardani, dkk., 2020). Dalam proses ini, peneliti berperan langsung dalam mengumpulkan dan menganalisis data. Tahapan tersebut meliputi (1) peneliti melakukan pengambilan data dokumen berupa soal, lembar jawaban, dan kunci jawaban siswa; (2) seluruh jawaban siswa direkap dan dimasukkan ke dalam perangkat lunak Anates; (3) hasil anates diinterpretasikan untuk menentukan kualitas butir soal berdasarkan kriteria analisis menurut Arikunto (2018); (4) hasil interpretasi divalidasi

melalui diskusi dengan guru mata pelajaran. Menurut Sugiyono (2013), pendekatan kuantitatif deskriptif diterapkan guna mendeskripsikan atau menggambarkan data yang diperoleh dari objek penelitian tanpa melakukan manipulasi terhadap variabel. Sebagai informasi pendukung, peneliti melakukan wawancara dengan pihak sekolah yakni guru mata pelajaran Bahasa Indonesia untuk memperoleh informasi kontekstual, serta pengumpulan data utama berupa dokumen soal, kunci jawaban, dan lembar jawaban siswa melalui teknik dokumentasi. Sebagaimana disarankan oleh Arikunto bahwa dokumentasi merupakan teknik pengumpulan data yang dilakukan dengan mengkaji dokumen sebagai sumber informasi utama (Arikunto, 2018).

HASIL DAN PEMBAHASAN

Langkah penting untuk menjamin validitas dan reliabilitas alat evaluasi yang diterapkan dalam pembelajaran dengan melaksanakan analisis kualitas butir soal. Hasil dari analisis ini dirancang untuk dapat memberikan pandangan secara objektif mengenai kualitas soal yang digunakan serta menjadi dasar perbaikan instrumen evaluasi pada asesmen berikutnya. Berikut analisis validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh soal asesmen sumatif semester genap yang dilaksanakan di SMP Batik Program Khusus (PK) Surakarta tahun ajaran 2024/2025 untuk mata pelajaran Bahasa Indonesia kelas VIII. Penelitian ini mendapatkan hasil sebagai berikut.

Validitas

Validitas merupakan derajat ketepatan sebuah soal atau tes dalam mengukur yang memang idealnya diukur sesuai dengan tujuan penyusunannya (Musyarofah & Farida, 2021). Pengujian validitas soal dilaksanakan dengan menggunakan alat bantu *software* Anates versi 4.0 seperti yang telah tertera pada Gambar 1.

KORELASI SKOR BUTIR DG SKOR TOTAL
=====

Jumlah Subyek= 25
Butir Soal= 30
Nama berkas: D:\TUGAS KULIAH\REFERENSI TUGAS AKHIR\BISMILLAH S.PD..ANA

No Butir	Korelasi	Signifikansi
1	0,216	-
2	NAN	NAN
3	NAN	NAN
4	0,382	Signifikan
5	NAN	NAN
6	0,443	Signifikan
7	0,700	Sangat Signifikan
8	-0,172	-
9	0,030	-
10	0,397	Signifikan
11	0,348	-
12	0,421	Signifikan
13	0,484	Sangat Signifikan
14	0,030	-
15	0,082	-
16	NAN	NAN
17	0,013	-
18	0,555	Sangat Signifikan
19	0,076	-
20	0,428	Signifikan
21	0,625	Sangat Signifikan
22	0,400	Signifikan
23	-0,175	-
24	0,523	Sangat Signifikan
25	0,147	-
26	0,402	Signifikan
27	0,319	-
28	0,387	Signifikan
29	0,469	Sangat Signifikan
30	-0,128	-

Gambar 1. Hasil Uji Validitas

Hasil perhitungan validitas setiap butir soal kemudian dibandingkan dengan nilai r tabel pada taraf signifikansi 5%. Dengan jumlah subjek penelitian sebanyak 25 siswa, diperoleh nilai r tabel sebesar 0,381. Kriteria penentuan validitas adalah bahwa butir soal dinyatakan valid apabila nilai r hitung lebih besar daripada r tabel ($r_{hitung} > r_{tabel}$). Sebaliknya, apabila nilai r hitung lebih kecil daripada r tabel ($r_{hitung} < r_{tabel}$), maka butir soal dinyatakan tidak valid. Distribusi hasil validitas soal selanjutnya disajikan dalam tabel berikut.

Validitas	Nomor Soal	Jumlah	Persentase
Valid	1, 4, 6, 7, 10, 12, 13, 16, 28, 20, 21, 22, 24, 26, 28, 29	16	53%
Tidak Valid	2, 3, 5, 8, 9, 11, 14, 15, 17, 19, 23, 25, 27, 30	14	47%

Tabel 1. Distribusi Validitas

Tabel 1 menyajikan hasil distribusi valid dan tidak validnya suatu soal yang diperoleh melalui pengolahan data menggunakan perangkat lunak Anates. Berdasarkan

analisis tersebut diperoleh hasil bahwa dari total 30 butir soal, 16 (53%) di antaranya tergolong valid sedangkan 14 butir soal (47%) dikategorikan tidak valid. Oleh karena itu, dapat ditarik kesimpulan bahwa terdapat 16 butir soal yang memenuhi kriteria validitas sedangkan 14 butir soal lainnya memerlukan perbaikan. Penggunaan soal yang tidak valid berpotensi menghasilkan penilaian yang kurang akurat dalam mencerminkan kemampuan sebenarnya dari siswa. Butir soal yang memenuhi kriteria validitas dapat digunakan kembali serta diarsipkan dalam bank soal, sementara butir soal yang belum memenuhi validitas perlu direvisi dengan menyesuaikan pada indikator pembelajaran serta meningkatkan pemahaman dan keterampilan dalam penyusunan soal yang baik dan sesuai kaidah.

Reliabilitas

Reliabilitas mengacu pada tingkat kepercayaan suatu instrumen dalam mengumpulkan data, yang mengindikasikan bahwa instrumen atau perangkat tersebut telah memenuhi kriteria sebagai alat ukur yang baik serta memiliki tingkat konsistensi yang memadai. (Saputri, dkk., 2023). Untuk mengetahui tingkat reliabilitas suatu soal dapat ditentukan menggunakan klasifikasi kriteria sebagai berikut.

Indeks Reliabilitas	Klasifikasi
$r_{11} < 0,20$	Sangat rendah
$0,20 \leq r_{11} < 0,40$	Rendah
$0,40 \leq r_{11} < 0,70$	Cukup
$0,70 \leq r_{11} < 0,90$	Tinggi
$0,90 \leq r_{11} \leq 1,00$	Sangat tinggi

Tabel 2. Klasifikasi Reliabilitas

(Sumber: Azis 2016)

Semakin tinggi koefisien reliabilitas suatu tes (mendekati angka 1) maka semakin besar pula tingkat konsistensi hasil yang dihasilkan oleh tes tersebut. Dengan demikian, reliabilitas dianggap dapat dipercaya apabila tes mampu memberikan hasil yang relatif stabil setiap kali digunakan. Hasil uji reliabilitas soal disajikan dalam gambar berikut.

RELIABILITAS TES

=====

Rata2= 23,32

Simpang Baku= 3,24

KorelasiXY= 0,57

Reliabilitas Tes= 0,72

Nama berkas: D:\TUGAS KULIAH\REFERENSI TUGAS AKHIR\BISMILLAH S.PD..ANA

No.Urut	Kode/Nama Subyek	Skor Ganjil	Skor Genap	Skor Total
1	Afifah Nur Majid	11	14	25
2	Alief Zain Ihsan	11	14	25
3	Bimo Purno K.	14	15	29
4	Davian Riza	9	10	19
5	Melvin Setya W.	9	13	22
6	M. Hafidz H.	11	10	21
7	Muhammad Bint...	11	11	22
8	M. Fariz Al F...	9	10	19
9	Mu'iza Indra ...	9	11	20
10	Raka Herdians...	11	12	23
11	Zaidan Faiz A.	9	11	20
12	Adita Putri K...	9	9	18
13	Agustina Rahm...	11	10	21
14	Aine Maulida M.	12	11	23
15	Amalina Fausti	12	10	22
16	Avionna Clari...	6	8	14
17	Azhifa Ramadh...	10	10	20
18	Khirana Nalya W.	11	11	22
19	Najla Dzaleiyyah	12	12	24
20	Naysha Adelia P.	14	13	27
21	Salsabila Ali...	13	10	23
22	Selena Anjani...	14	12	26
23	Sheza Adzkiya...	13	12	25
24	Tabriz Kumari...	14	12	26
25	Tasya Galleneya	13	13	26

Gambar 2. Hasil Uji Reliabilitas

Perhitungan reliabilitas soal asesmen sumatif semester genap mata pelajaran Bahasa Indonesia kelas VIII SMP Batik PK Surakarta Tahun Ajaran 2024/2025 dilaksanakan dengan alat bantu berupa *software* Anates versi 4.0 tersebut kemudian diinterpretasikan berdasarkan kriteria bahwa jika nilai r hitung $\geq 0,70$ maka soal dikategorikan reliabel atau mempunyai tingkat reliabilitas yang tinggi (Qirom, dkk., 2020). Berdasarkan hasil perhitungan reliabilitas, diperoleh bahwa koefisien reliabilitas untuk soal dengan besaran 0,72. Dari hasil tersebut dapat diambil kesimpulan bahwa soal Asesmen Sumatif Semester Genap Mata Pelajaran Bahasa Indonesia kelas VIII SMP Batik PK Surakarta Tahun Ajaran 2024/2025 tersebut mempunyai tingkatan reliabilitas yang tinggi. Temuan ini sekaligus mengindikasikan bahwa soal tersebut memiliki reliabilitas yang memadai dikarenakan nilai koefisien reliabilitasnya berada pada angka $\geq 0,70$.

Hasil ini menjadi relevan dengan penelitian Basri (2021) yang memperoleh nilai koefisien reliabilitas sebesar 0,80. Hal ini juga memperkuat pandangan bahwa suatu instrumen atau perangkat tes yang berada dalam tingkat validitas tinggi pada setiap butirnya cenderung mempunyai tingkat reliabilitas yang baik. Pendapat ini didukung oleh pernyataan Siyoto dan Sodik (2015) yang menyatakan bahwa soal yang valid pasti

bersifat reliabel tetapi soal yang bersifat reliabel belum dapat diastikan kevalidannya. Artinya, jika sebuah soal telah valid secara teoretis, maka soal tersebut juga dapat dianggap telah reliabel secara teoretis. Oleh karena itu, dapat ditarik kesimpulan bahwa soal asesmen sumatif semester genap Mata Pelajaran Bahasa Indonesia kelas VIII SMP Batik PK Surakarta Tahun Ajaran 2024/2025 tergolong baik dari aspek reliabilitas.

Tingkat Kesukaran

Tingkat kesukaran butir soal mencerminkan persentase peserta tes yang dapat menjawab soal dengan tepat (Qadir, dkk., 2024). Menurut Magdalena (2022), ketika penyusunan naskah ujian, disarankan untuk menggunakan butir soal dengan tingkat kesukaran yang seimbang, yaitu 25% soal berada pada kategori sukar, 50% termasuk kategori sedang, dan 25% lainnya tergolong mudah. Untuk mengetahui tingkat kesukaran suatu soal dapat ditentukan menggunakan klasifikasi kriteria sebagai berikut.

Indeks Kesukaran (P)	Klasifikasi
0,00 – 0,19	Sangat sukar
0,20 – 0,39	Sukar
0,40 – 0,59	Sedang
0,60 – 0,79	Mudah
0,80 – 1,00	Sangat mudah

Tabel 3. Klasifikasi Tingkat Kesukaran

(Sumber: Qadir, Huda, & Hermina, 2024)

Soal yang baik ditandai oleh tingkat kesukaran yang berada pada kategori sedang, artinya tidak terlalu mudah dan tidak terlalu sukar. Nilai indeks kesukaran yang semakin mendekati angka 1 menunjukkan bahwa soal tersebut memiliki tingkat kesukaran yang semakin rendah. Semakin rendah nilai indeks kesukaran (mendekati 0) maka semakin tinggi tingkat kesukaran soal tersebut. Hasil uji tingkat kesukaran disajikan dalam gambar berikut.

TINGKAT KESUKARAN

=====

Jumlah Subyek= 25

Butir Soal= 30

Nama berkas: D:\TUGAS KULIAH\REFERENSI TUGAS AKHIR\BISMILLAH S.PD..ANA

No Butir	Jml Betul	Tkt. Kesukaran(%)	Tafsiran
1	21	84,00	Mudah
2	25	100,00	Sangat Mudah
3	25	100,00	Sangat Mudah
4	19	76,00	Mudah
5	25	100,00	Sangat Mudah
6	15	60,00	Sedang
7	15	60,00	Sedang
8	24	96,00	Sangat Mudah
9	23	92,00	Sangat Mudah
10	20	80,00	Mudah
11	22	88,00	Sangat Mudah
12	17	68,00	Sedang
13	18	72,00	Mudah
14	23	92,00	Sangat Mudah
15	20	80,00	Mudah
16	25	100,00	Sangat Mudah
17	5	20,00	Sukar
18	9	36,00	Sedang
19	23	92,00	Sangat Mudah
20	18	72,00	Mudah
21	18	72,00	Mudah
22	18	72,00	Mudah
23	15	60,00	Sedang
24	20	80,00	Mudah
25	18	72,00	Mudah
26	23	92,00	Sangat Mudah
27	21	84,00	Mudah
28	22	88,00	Sangat Mudah
29	15	60,00	Sedang
30	21	84,00	Mudah

Gambar 3. Hasil Uji Tingkat Kesukaran

Hasil pengujian tingkat kesukaran menunjukkan bahwa tidak ada butir soal yang berada pada kategori sangat sukar (0%). Terdapat 1 butir soal (3%) yang termasuk kategori sukar, 6 butir soal (20%) berkategori sedang, 12 butir soal (40%) tergolong mudah, dan 11 butir soal (37%) berada pada kategori sangat mudah. Rincian distribusi tingkat kesukaran soal tersebut dapat dilihat pada Tabel 4.

Tingkat Kesukaran	Nomor Soal	Jumlah	Persentase
Sangat Sukar	-	-	0%
Sukar	17	1	3%
Sedang	6, 7, 12, 18, 23, 29	6	20%
Mudah	1, 4, 10, 13, 15, 20, 21, 22, 24, 25, 27, 30	12	40%
Sangat mudah	2, 3, 5, 8, 9, 11, 14, 16, 19, 26, 28	11	37%

Tabel 4. Distribusi Tingkat Kesukaran

Hasil distribusi tingkat kesukaran menunjukkan bahwa soal Asesmen Sumatif Semester Genap Mata Pelajaran Bahasa Indonesia kelas VIII SMP Batik PK Surakarta Tahun Ajaran 2024/2025 memiliki tingkat kesukaran yang kurang seimbang sehingga kualitasnya belum optimal. Sebagai tindak lanjut hasil analisis, butir soal yang memiliki tingkat kesukaran sedang dinilai paling ideal sehingga layak untuk dipertahankan, digunakan kembali. Selanjutnya, disimpan dalam bank soal sebagai instrumen evaluasi pada pelaksanaan asesmen. Hal ini dianggap telah memenuhi standar kualitas yang baik sedangkan soal-soal yang tergolong terlalu mudah sebaiknya diperbaiki terlebih dahulu agar sesuai dengan standar yang diharapkan.

Daya Pembeda

Daya pembeda merupakan indikator sejauh mana sebuah butir soal dapat membedakan antara siswa dari kelompok atas (berkemampuan tinggi) dan siswa dari kelompok bawah (berkemampuan rendah), (Qomariyah, 2022). Daya pembeda pada suatu tes digunakan untuk menilai kemampuan soal dalam membedakan siswa berdasarkan tingkat kemampuannya (Nurhalimah, dkk., 2022). Daya pembeda merupakan salah satu indikator penting yang menggambarkan kemampuan suatu butir soal dalam membedakan siswa yang memiliki tingkat kemampuan tinggi dengan siswa yang berkemampuan rendah, sehingga dapat diketahui sejauh mana butir soal tersebut efektif dalam mengukur variasi kemampuan peserta didik. Indeks ini membantu menilai seberapa baik soal membedakan tingkat kemampuan siswa.

Untuk mengetahui daya pembeda suatu soal dapat ditentukan menggunakan klasifikasi kriteria sebagai berikut.

Indeks Diskriminasi (D)	Klasifikasi
$D < 0,00$	Tidak baik
$0,00 < D \leq 0,20$	Kurang baik
$0,20 < D \leq 0,40$	Cukup
$0,40 < D \leq 0,70$	Baik
$0,70 < D \leq 1,00$	Sangat baik

Tabel 5. Klasifikasi Daya Pembeda

(Sumber: Zalukhu et al. 2023)

Butir soal dengan indeks daya pembeda yang tinggi menunjukkan bahwa soal tersebut mampu berfungsi secara efektif dalam membedakan siswa yang memiliki kemampuan tinggi dan siswa yang berkemampuan rendah. Sebaliknya, nilai indeks daya pembeda yang rendah menandakan bahwa butir soal kurang optimal dalam

mengidentifikasi perbedaan tingkat kemampuan siswa, sehingga daya ukur soal terhadap variasi kemampuan peserta didik menjadi lemah. Hasil uji daya pembeda disajikan dalam gambar berikut.

DAYA PEMBEDA
=====

Jumlah Subyek= 25
Klp atas/bawah(n)= 7
Butir Soal= 30
Nama berkas: D:\TUGAS KULIAH\REFERENSI TUGAS AKHIR\BISMILLAH S.PD..ANA

No Butir	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1	6	4	2	28,57
2	7	7	0	0,00
3	7	7	0	0,00
4	7	3	4	57,14
5	7	7	0	0,00
6	5	2	3	42,86
7	7	1	6	85,71
8	6	7	-1	-14,29
9	7	7	0	0,00
10	7	3	4	57,14
11	7	6	1	14,29
12	7	4	3	42,86
13	7	3	4	57,14
14	7	7	0	0,00
15	5	6	-1	-14,29
16	7	7	0	0,00
17	2	1	1	14,29
18	6	1	5	71,43
19	7	7	0	0,00
20	7	2	5	71,43
21	7	3	4	57,14
22	6	3	3	42,86
23	3	5	-2	-28,57
24	7	4	3	42,86
25	6	4	2	28,57
26	7	6	1	14,29
27	7	6	1	14,29
28	7	5	2	28,57
29	6	2	4	57,14
30	6	7	-1	-14,29

Gambar 4. Hasil Uji Daya Pembeda

Berdasarkan hasil pengujian daya pembeda, diperoleh variasi kualitas butir soal dalam membedakan kemampuan siswa. Sebanyak 11 butir soal (37%) berada pada kategori tidak baik, 4 butir soal (13%) termasuk kategori kurang baik, 3 butir soal (10%) tergolong cukup baik, 9 butir soal (30%) berkategori baik, dan 3 butir soal (10%) menunjukkan daya pembeda yang sangat baik. Hasil distribusi daya pembeda soal tersebut dapat dilihat pada Tabel 6 berikut.

Daya Pembeda	Nomor Soal	Jumlah	Persentase
Tidak baik	2, 3, 5, 8, 9, 14, 15, 16, 19, 23, 30	11	37%
Kurang baik	11, 17, 26, 27	4	13%
Cukup baik	1, 25, 28	3	10%
Baik	4, 6, 10, 12, 12, 21, 22, 24, 29	9	30%
Sangat baik	7, 18, 20	3	10%

Tabel 6. Distribusi Daya Pembeda

Apabila suatu butir soal dapat dijawab dengan mudah oleh siswa berkemampuan tinggi maupun rendah maka butir soal tersebut tidak mampu membedakan tingkat kemampuan siswa sehingga dikategorikan tidak memiliki daya pembeda. Begitu pula jika soal tidak dapat dijawab oleh keduanya maka soal tersebut dianggap tidak baik karena gagal membedakan tingkat kemampuan siswa. Berdasarkan hasil analisis, soal Asesmen Sumatif Semester Genap Mata Pelajaran Bahasa Indonesia kelas VIII SMP Batik PK Surakarta Tahun Ajaran 2024/2025 menunjukkan bahwa sebagian besar soal masih perlu diperbaiki atau diganti. Terutama yang berada pada kategori tidak baik dan kurang baik, agar kualitas evaluasi pembelajaran dapat lebih optimal dalam membedakan kemampuan siswa. Di sisi lain, butir soal yang memiliki daya pembeda baik hingga sangat baik layak dipertahankan dan dapat langsung diarsipkan dalam bank soal. Adapun butir soal dengan daya pembeda cukup disarankan untuk direvisi terlebih dahulu agar kualitasnya meningkat sebelum dimasukkan ke dalam bank soal.

Efektivitas Pengecoh

Efektivitas pengecoh atau distraktor dalam soal tes mengacu pada sejauh mana distraktor tersebut mampu menarik perhatian siswa yang kurang cakap, sehingga memilih alternatif jawaban yang salah tersebut (Bano, dkk., 2022). Dengan kata lain, distraktor yang efektif adalah pilihan jawaban yang mampu menggiring siswa dengan pemahaman rendah untuk memilihnya sehingga berperan dalam membedakan kemampuan siswa. Penilaian terhadap efektivitas pengecoh dalam setiap butir soal dilakukan dengan menggunakan kriteria seperti yang terdapat pada Tabel 5 tentang klasifikasi efektivitas pengecoh.

Pengecoh yang Berfungsi	Klasifikasi
3	Sangat baik
2	Baik
1	Kurang baik
0	Tidak baik

Tabel 7. Klasifikasi Efektivitas Pengecoh

(Sumber: Nafs, dkk., 2023)

Efektivitas pengecoh dapat dinilai melalui pola sebaran jawaban siswa. Suatu pengecoh dianggap efektif jika dipilih oleh minimal 5% dari peserta tes. Artinya, jika sebuah pilihan jawaban yang salah menarik perhatian lebih dari 5% siswa maka pengecoh tersebut berfungsi dengan baik.

KUALITAS PENGECOH

=====

Jumlah Subyek= 25

Butir Soal= 30

Nama berkas: D:\TUGAS KULIAH\REFERENSI TUGAS AKHIR\BISMILLAH S.PD..ANA

No Butir	a	b	c	d	*
1	2+	21**	0--	2+	0
2	0	0	0	25**	0
3	0	0	25**	0	0
4	0--	19**	3+	3+	0
5	0	0	25**	0	0
6	2+	7---	15**	1-	0
7	6--	1-	3++	15**	0
8	24**	1---	0--	0--	0
9	1+	23**	1+	0--	0
10	20**	1+	3--	1+	0
11	2--	22**	0--	1++	0
12	3++	3++	17**	2+	0
13	18**	2++	5---	0--	0
14	2---	0--	0--	23**	0
15	2++	20**	1+	2++	0
16	25**	0	0	0	0
17	12--	8++	5**	0--	0
18	9**	13---	1--	2-	0
19	2---	23**	0--	0--	0
20	7---	0--	0--	18**	0
21	18**	3+	1-	3+	0
22	5---	1-	18**	1-	0
23	7---	15**	3++	0--	0
24	1+	2++	2++	20**	0
25	2++	4-	18**	1-	0
26	23**	2---	0--	0--	0
27	21**	1+	0--	3---	0
28	2--	1++	0--	22**	0
29	8---	0--	15**	2+	0
30	3---	21**	1+	0--	0

Gambar 5. Hasil Uji Efektivitas Pengecoh

Setiap pilihan jawaban pada butir soal dinyatakan berfungsi apabila diberi tanda (-) untuk kurang baik, (+) untuk baik, dan (++) untuk sangat baik. Sementara itu, pilihan yang tidak berfungsi ditandai dengan (--) untuk buruk, (---) untuk sangat buruk, serta dapat ditandai dengan blok warna tertentu yang menunjukkan bahwa opsi tersebut tidak berfungsi. Hasil uji efektivitas pengecoh menandakan bahwa terdapat 5 soal yang tergolong sangat baik (17%), terdapat 7 butir soal (23%) yang tergolong baik, 9 butir soal (30%) termasuk dalam kategori kurang baik, 9 butir soal (30%) masuk dalam kategori tidak baik. Hasil distribusi efektivitas pengecoh soal tersebut dapat dilihat pada Tabel 8 berikut.

Efektivitas Pengecoh	Nomor Soal	Jumlah	Persentase
Sangat baik	12, 15, 21, 24, 25	5	17%
Baik	1, 4, 6, 7, 9, 10, 22	7	23%
Kurang baik	11, 13, 17, 18, 23, 27, 28, 29, 30	9	30%
Tidak baik	2, 3, 5, 8, 14, 16, 19, 20, 26	9	30%

Tabel 8. Distribusi Efektivitas Pengecoh

Hasil efektivitas pengecoh menunjukkan bahwa soal Asesmen Sumatif Semester Genap Mata Pelajaran Bahasa Indonesia kelas VIII SMP Batik PK Surakarta Tahun Ajaran 2024/2025 menunjukkan bahwa sebagian besar distraktor belum berfungsi secara optimal dalam mengecoh siswa, sehingga diperlukan perbaikan atau penggantian distraktor pada soal-soal yang termasuk dalam kategori kurang baik dan tidak baik untuk meningkatkan kualitas asesmen di masa mendatang. Sebagai tindak lanjut, butir soal yang termasuk dalam kategori sedang dapat disimpan dalam bank soal karena dianggap memenuhi kriteria yang bermutu, yaitu relevan, representatif, dan berfungsi membedakan kemampuan siswa secara memadai (Brookhart, 2024). Sementara itu, soal-soal yang tergolong terlalu mudah sebaiknya diperbaiki terlebih dahulu dengan cara memperkaya konteks bacaan, meningkatkan kompleksitas stimulus, atau bisa juga dengan cara meningkatkan kualitas pengecoh agar sesuai dengan standar yang diharapkan.

SIMPULAN

Berdasar pada hasil analisis yang telah dilakukan terhadap butir soal asesmen sumatif mata pelajaran Bahasa Indonesia kelas VIII di SMP Batik Program Khusus Surakarta tahun ajaran 2024/2025, dapat disimpulkan bahwa kualitas soal secara umum masih perlu ditingkatkan. Dari 30 soal yang dianalisis, sebanyak 53% tergolong valid, sedangkan sisanya belum memenuhi kriteria validitas. Soal memiliki reliabilitas tinggi dengan koefisien sebesar 0,72, menunjukkan bahwa instrumen ini cukup konsisten. Namun, tingkat kesukaran menunjukkan dominasi soal yang terlalu mudah, sementara dari segi daya pembeda, 37% soal termasuk kategori tidak baik dan tidak mampu membedakan kemampuan siswa secara optimal. Selain itu, efektivitas pengecoh juga masih kurang memadai, di mana 60% distraktor kurang dan bahkan tidak berfungsi sebagaimana mestinya. Oleh karena itu, perlu dilakukan revisi terhadap butir soal yang belum memenuhi standar agar instrumen penilaian menjadi lebih representatif dalam mengukur kompetensi siswa. Adapun rekomendasi untuk penelitian selanjutnya adalah agar dilakukan pengembangan soal berbasis indikator kompetensi secara lebih menyeluruh, disertai uji coba lapangan dengan jumlah responden yang lebih besar serta pendekatan analisis yang tidak hanya kuantitatif, tetapi juga dikombinasikan dengan analisis kualitatif guna memahami konteks pemahaman siswa terhadap soal.

Ucapan Terima Kasih

Ucapan terima kasih penulis ucapkan kepada tim redaksi *Ghancaran: Jurnal Pendidikan Bahasa dan Sastra Indonesia* yang telah memberikan ruang publikasi bagi karya ini. Tidak lupa, penulis menyampaikan terima kasih kepada Universitas Islam Negeri (UIN) Raden Mas Said Surakarta yang telah memberikan dukungan akademik dalam pelaksanaan penelitian yang penulis lakukan, serta kepada dosen pembimbing yang dengan penuh kesabaran dan ketelitian telah membimbing penulis selama proses penyusunan artikel. Ucapan terima kasih juga disampaikan kepada guru dan seluruh staf SMP Batik Program Khusus Surakarta yang telah memberikan izin, dukungan, dan bantuan dalam proses pengumpulan data. Semoga artikel ini dapat memberikan kontribusi yang positif bagi pengembangan kualitas evaluasi pembelajaran di sekolah.

DAFTAR RUJUKAN

- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan Edisi 3*. Bumi Aksara.
- Azis, A. (2016). Analisis Tes Buatan Guru Bidang Studi Matematika Kelas V SD 1 Katobengke. *Edumatica : Jurnal Pendidikan Matematika*, 6(1), 15–24.
- Bano, V. O., Marambaawang, D. N., & Njoeroemana, Y. (2022). Analisis Kriteria Butir Soal Ujian Sekolah Mata Pelajaran IPA di SMP Negeri 1 Waingapu. *Ideas: Jurnal Pendidikan, Sosial, dan Budaya*, 8(1), 145-152.
- Basri, K., Baidowi, B., Junaidi, J., & Turmuzi, M. (2021). Analisis Butir Soal Ulangan Semester Ganjil Mata Pelajaran Matematika Kelas VIII SMP. *Griya Journal of Mathematics Education and Application*, 1(4), 682–694.
- Brookhart, S. M. (2024). *Education Sciences Teachers Revisited*. 14(7), 751-765.
- Depdiknas. (2007). *Standar penilaian pendidikan*. Badan Standar Nasional Pendidikan.
- Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2021). Analisis Butir Soal Ulangan Harian IPA Menggunakan Software Anates pada Pendekatan Teori Tes Klasik. *Natural Science Education Research*, 4(1), 65–76.
- Hardani, Andriani, H., Ustiawaty, J., Utami, E. F., Istiqomah, R. R., Fardani, R. A., Sukmana, D. J., & Auliya, N. H. (2020). *Metode Penelitian Kualitatif*. CV. Pustaka Ilmu Grup.
- Himawan, R., & Suyata, P. (2021). Analisis Kebutuhan Pengembangan Soal HOTS dalam Pembelajaran Teks Pidato Persuasif di MGMP SMP Wilayah Kabupaten Bantul. *Ghancaran: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 117–129.
- Jatmoko, D., Pambudi, A., Setiyono, K. A., Nugroho, B., Jl, A., & Dahlan, K. H. A. (2025). Analisis Butir Soal Teknik Dasar Otomotif Berbasis Diagnostik Kognitif dengan Anates Versi 4.02. *Jurnal Riset Rumpun Ilmu Pendidikan*, 4(1), 461-471.
- Kaka, L., Bano, V. O., & Njoeroemana, Y. (2024). Efektivitas Analisis Butir Soal Pilihan Ganda Menggunakan Aplikasi Anates di SMPN 2 Kanatang. *Jurnal Inovasi Penelitian*, 4(9), 1441-1450.
- Karnoto, & Wibisono, Y. (2004). *Anates Program Khusus Analisis Tes Pilihan Ganda dan Uraian versi 4.0 untuk Window*. Universitas Pendidikan Indonesia.
- Magdalena, I. (2022). *Teori dan Praktik Evaluasi Pembelajaran SD*. CV. Jejak.
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *BINTANG: Jurnal Pendidikan dan Sains*, 3(2), 198–214.
- Marambaawang, D., Bano, V. O., & Hada Enda, R. R. (2023). Analisis Kualitas Butir Soal

- Penilaian Akhir Semester Gasal Tahun 2021/2022 Menggunakan Iteman Di SMP Negeri 1 Kambera. *DE_Journal: Dharmas Education Journal*, 4(1), 233–243.
- Mustaqim, M., & Sulisti, H. (2024). Analisis Butir Soal Pas Matematika Peminatan: Daya Pembeda, Tingkat Kesukaran, dan Kualitas Pengecoh. *Al-'Adad: Jurnal Tadris Matematika*, 3(1), 44–56.
- Musyarofah, A., & Farida. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu Arrib Journal Of Arabic Education*, 1(1), 34–44.
- Nafs, H., Sridana, N., Hikmah, N., & Soeprianto, H. (2023). Analisis Kualitas Butir Soal Ulangan Akhir Semester Genap Mata Pelajaran Matematika Kelas VII SMPN 6 Mataram Tahun Ajaran 2022/2023. *Jurnal Ilmiah Profesi Pendidikan*, 8(4), 2324–2331.
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan Antara Validitas Item Dengan Daya Pembeda dan Tingkat Kesukaran Soal Pilihan Ganda Pas. *Natural Science Education Research*, 4(3), 249–257.
- Qadir, A., Huda, N., & Hermina, D. (2024). Analisis Butir Tes: Tingkat Kesukaran, Daya Pembeda dan Efektivitas Pengecoh. *Al Furqan: Jurnal Agama, Sosial dan Budaya*, 3(3), 1450–1467.
- Qirom, M. S., Sridana, N., & Prayitno, S. (2020). Pengembangan Soal Matematika Berbasis Higher Order Thinking Skills pada Lingkup Materi Ujian Nasional untuk Tingkatan Sekolah Menengah Pertama. *Jurnal Pijar MIPA*, 15(5), 466–472.
- Qomariyah, L. (2022). Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal TOAFL Universitas Hasyim Asy'ari Tebuireng Jombang. *Lisanan Arabiya: Jurnal Pendidikan Bahasa Arab*, 6(1), 1–18.
- Rahayu, L. P. (2024). Kualitas Soal Bahasa Indonesia Kelas XI SMAN 2 Bangko Pusako : Analisis Butir Soal. *Jurnal Onoma: Pendidikan, Bahasa dan Sastra*, 10(4), 3755–3762.
- Riinawati. (2021). *Pengantar Evaluasi Pendidikan*. Thema Publishing.
- Saputri, H. A., Zuhijrah, Larasati, N. J., & Shaleh. (2023). Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran, dan Daya Beda Butir Soal. *Didaktik: Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 9(5), 2986–2995.
- Siyuto, S., & Sodik, M. A. (2015). *Dasar Metodologi Penelitian*. Literasi Media Publishing.
- Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. CV Alfabeta.
- Syafiriya, D. M., & Hakim, L. (2024). Analisis Butir Soal HOTS untuk Menguji Pemahaman Siswa SMK Jurusan Akuntansi Menggunakan Software Anates. 4(1), 149-171.
- Wachidah, L. R., Laila, Y., Irmawati, A., & Amin, S. (2021). Implementasi Penggunaan Tes Essay dalam Evaluasi Pembelajaran Daring pada Siswa Kelas VII SMP Negeri 1 Tlanakan. *GHANCARAN: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 16–26.
- Zainuri, A., Aquami, & AnNur, S. (2021). *Evaluasi Pendidikan (Kajian Teoritik)*. Penerbit Qiara Media.
- Zalukhu, K. O., Khairuddin, K., Syamsuar, S., & Sari, D. N. (2023). Evaluasi Instrumen Pembelajaran Pendidikan Jasmani Olahraga dan Kesehatan Kelas VII Sekolah Menengah Pertama Negeri 30. *Jurnal JPDO*, 6(6), 89–96.