

ANALISIS BUTIR SOAL BAHASA INDONESIA KELAS VI SD ISLAM TERPADU AL-AHYA TAHUN PELAJARAN 2024/2025

Hera Hani Lia^{1*}

¹Universitas Bina Darma, Jendral Ahmad Yani No.3, 9/10 Ulu, Palembang, 30264, Indonesia

*Alamat email penulis koresponden: herahanilia8@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengevaluasi kualitas butir soal Penilaian Tengah Semester (PTS) mata pelajaran Bahasa Indonesia kelas VI SD Islam Terpadu Al-Ahya tahun pelajaran 2024/2025. Kualitas yang dianalisis meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Metode yang digunakan adalah metode campuran (mixed methods) dengan pendekatan kuantitatif dan kualitatif. Hasil penelitian menunjukkan bahwa dari 20 butir soal, hanya 6 butir yang valid secara empiris. Nilai reliabilitas soal sebesar 0,519, menunjukkan bahwa soal tidak reliabel. Tingkat kesukaran menunjukkan 50% soal tergolong sedang, 20% mudah, dan 30% sukar. Sebanyak 50% butir soal memiliki daya beda buruk. Efektivitas pengecoh menunjukkan bahwa 50% soal memiliki distraktor yang tidak berfungsi. Penelitian ini merekomendasikan revisi soal untuk meningkatkan kualitas evaluasi pembelajaran.

Kata kunci analisis butir soal, bahasa Indonesia

Abstract

This study aims to evaluate the quality of mid-semester exam (PTS) items in the Indonesian language subject for 6th grade at SD Islam Terpadu Al-Ahya during the 2024/2025 academic year. The analysis covers item validity, reliability, difficulty level, discrimination power, and distractor effectiveness. The research employed a mixed-methods approach with both quantitative and qualitative techniques. The findings show that only 6 out of 20 items were empirically valid. The reliability score was 0.519, indicating low reliability. Regarding difficulty, 50% of items were moderate, 20% easy, and 30% difficult. Moreover, 50% of items had poor discrimination power. Half of the distractors were ineffective. The study recommends revising the test items to improve assessment quality.

Keywords: Item Analysis, Indonesian Language

PENDAHULUAN

Evaluasi pembelajaran merupakan proses penting untuk mengetahui pencapaian tujuan pendidikan. Dalam konteks pendidikan dasar, kualitas soal ujian harus dikaji secara menyeluruh agar dapat mencerminkan kemampuan siswa secara objektif. Penelitian ini dilakukan untuk menganalisis kualitas butir soal PTS Bahasa Indonesia kelas VI di SD Islam Terpadu Al-Ahya. Fokus penelitian mencakup validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Menurut Popham (2021), Evaluasi memiliki peran penting dalam proses pendidikan karena membantu guru,

sekolah, dan pembuat kebijakan untuk memahami tingkat pencapaian tujuan pendidikan dan melakukan perbaikan yang diperlukan. Evaluasi tidak hanya terbatas pada penilaian akhir (sumatif), tetapi juga mencakup penilaian formatif yang dilaksanakan selama proses pembelajaran. Penelitian relevan dilakukan oleh Anggita Fitria Prameswari dkk (2021) dengan judul *Analisis Soal Ulangan Harian Mata Pelajaran Bahasa Indonesia Siswa Kelas X SMA N 1 Jetik Bantul*. Tujuan dari penelitian ini adalah menganalisis soal ulangan harian pada mata pelajaran Bahasa Indonesia siswa kelas X di SMA N 1 Jetis Bantul. Hasil analisis menunjukkan terdapat 2 soal yang terlalu sukar (8%), 20 soal yang sedang (80%) dan soal yang terlalu mudah sebanyak (12%). Hasil analisis menunjukkan daya pembeda soal untuk kriteria diterima sebanyak 5 soal (20%), kriteria diterima dengan perbaikan ada 8 soal (32%), kriteria diperbaiki ada 9 soal (36%) dan kategori dibuang 3 soal (12%). Hasil analisis secara keseluruhan, yakni pada 25 soal pilihan ganda hanya ada 13 soal yang dapat diterima yaitu 52%, sedangkan soal yang tidak baik untuk digunakan ada 12 soal yaitu 48%. Hasil analisis yang dilakukan oleh peneliti menunjukkan bahwa tingkat kesukarannya sedang. Hal tersebut menunjukkan 25 soal pilihan ganda ada 80% presentase dari soal yang ber kriteria sedang. Dari hasil keseluruhan soal ada 2 soal yang terlalu sukar dengan presentase 8% dan ada 3 soal yang terlalu mudah dengan presentase 12%. Daya pembeda menunjukkan daya pembeda soal dengan kriteria diterima sebanyak 5 soal dengan presentase sebesar 20%, untuk soal dengan kriteria diterima dengan perbaikan sebanyak 8 soal dengan presentase sebesar 32%, soal dengan kriteria diperbaiki ada sebanyak 9 soal dengan presentase 36% sedangkan untuk soal dengan kriteria dibuang, ada 3 soal dengan presentase 12%. Penelitian ini diharapkan dapat memberikan kontribusi baru dalam pemahaman tentang bagaimana kualitas soal-soal Bahasa Indonesia dapat memengaruhi evaluasi dan pencapaian hasil belajar siswa. Dengan demikian, penelitian ini akan mengidentifikasi dan menganalisis tingkat kesukaran dan daya beda butir soal Bahasa Indonesia yang digunakan di kelas IV SD islam terpadu Al-Ahya.

METODE PENELITIAN

Penelitian ini menggunakan metode campuran (*mixed methods*), yaitu perpaduan antara pendekatan kuantitatif dan kualitatif. Data kualitatif dikumpulkan melalui wawancara dan dokumen, sedangkan data kuantitatif diperoleh melalui analisis hasil jawaban peserta didik terhadap 20 butir soal pilihan ganda. Subjek yang digunakan

dalam penelitian ini adalah peserta didik kelas VI SD islam terpadu Al- ahyia Banyuasin 1 dengan jumlah total sebanyak 20 peserta didik. Dalam penelitian ini, teknik pengumpulan data dilakukan dengan menggunakan tiga metode utama, yaitu dokumentasi, observasi, dan wawancara. Analisis data dilakukan dengan bantuan perangkat lunak SPSS versi 27.

HASIL DAN PEMBAHASAN

Validitas Soal Ujian Penilaian Tengah Semester

Berdasarkan hasil analisis validitas isi yang dilihat dengan mencocokkan kisi-kisi soal menunjukkan bahwa soal ujian PTS Mata Pelajaran Bahasa Indonesia Kelas VI Tahun Pelajaran 2024/2025 telah sesuai dengan indikator tujuan pembelajaran. Pengujian validitas butir soal dilakukan dengan menganalisis data primer berupa butir soal yang telah diolah berdasarkan benar atau salah jawaban setiap butir soal dengan responden 20 peserta didik, lalu data tersebut diinput ke dalam program komputer SPSS versi 27 untuk memperoleh nilai r hitung atau *pearson correlation* pada setiap butir soal. Langkah selanjutnya adalah membandingkan nilai r hitung (*pearson correlation*) dan nilai r tabel untuk $N=20$ dengan distribusi signifikansi uji dua arah. Diketahui bahwa r tabel untuk $N = 20$ adalah $df = N-2$ ($20-2$) = 18 dengan distribusi signifikansi 5% adalah 0,443. Dasar keputusan yang diambil adalah sebagai berikut:

1. Jika nilai r hitung > dari nilai r tabel, maka dinyatakan valid
2. Jika nilai r hitung < dari nilai r tabel, maka dapat dinyatakan tidak valid.

Selajutnya mencari tahu nilai *pearson correlation* untuk menganalisis 20 butir soal dengan perbandingan nilai r tabel. Perhatikan gambar dibawah ini.

Item	Item Label	Pearson Correlation
Item 1	Item 1	0.45
Item 2	Item 2	0.48
Item 3	Item 3	0.42
Item 4	Item 4	0.46
Item 5	Item 5	0.44
Item 6	Item 6	0.47
Item 7	Item 7	0.43
Item 8	Item 8	0.49
Item 9	Item 9	0.41
Item 10	Item 10	0.46
Item 11	Item 11	0.44
Item 12	Item 12	0.47
Item 13	Item 13	0.43
Item 14	Item 14	0.48
Item 15	Item 15	0.42
Item 16	Item 16	0.46
Item 17	Item 17	0.44
Item 18	Item 18	0.47
Item 19	Item 19	0.43
Item 20	Item 20	0.49

Gambar 1. Hasil correlation pearson validitas spss 27

Gambar diatas merupakan hasil output spss *correlation pearson* validitas butir soal yang hasilnya dianalisa dengan nilai r tabel seperti tabel dibawah ini.

No. Soal	Pearson Correlation	Nilai Sig.	R Tabel	Kesimpulan
Soal 1	0,512	0,21	0,443	Valid
Soal 2	0,458	0,42	0,443	Valid
Soal 3	0,343	0,138	0,443	Tidak Valid
Soal 4	0,052	0,829	0,443	Tidak Valid
Soal 5	0,244	0,301	0,443	Tidak Valid
Soal 6	0,574	0,008	0,443	Valid
Soal 7	0,325	0,162	0,443	Tidak Valid
Soal 8	-0,068	0,775	0,443	Tidak Valid
Soal 9	0,434	0,056	0,443	Tidak Valid
Soal 10	0,322	0,167	0,443	Tidak Valid
Soal 11	0,260	0,269	0,443	Tidak Valid
Soal 12	-0,129	0,589	0,443	Tidak Valid
Soal 13	0,052	0,829	0,443	Tidak Valid
Soal 14	0,624	0,003	0,443	Valid
Soal 15	0,576	0,008	0,443	Valid
Soal 16	0,722	<0,001	0,443	Valid
Soal 17	-0,036	0,880	0,443	Tidak Valid
Soal 18	0,134	0,573	0,443	Tidak Valid
Soal 19	0,325	0,161	0,443	Tidak Valid
Soal 20	0,307	0,187	0,443	Tidak Valid

Tabel 1. Hasil Analisis Validitas Butir soal Ujian PTS Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025

Berdasarkan tabel analisis diatas diketahui bahwa butir soal yang dinyatakan valid sebanyak 6 butir soal dan 14 butir soal dinyatakan tidak valid. Jadi, dapat disimpulkan bahwa penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025 memiliki validitas butir soal yang tidak valid karena 70% butir soal dinyatakan tidak valid.

Reliabilitas

Reliability Statistics	
Cronbach's Alpha	N of Items
.519	20

Gambar 2. Hasil Uji Realibilitas

Hasil uji reliabilitas diperoleh nilai *Cronbach's Alpha* 0,519.

Nilai $0,519 < 0,60$ = tidak reliabel. Hal ini berdasarkan pada dasar pengambilan uji reliabilitas *Cronbach Alpha* menurut Wiratna Sujarweni (2014:85), yaitu jika nilai *Alpha* $> 0,60$ maka reliabel, jika nilai *Alpha* $< 0,60$ maka tidak reliabel. Artinya 20 butir soal penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025 yang telah dinyatakan tidak reliabel atau tidak konsisten.

Tingkat Kesukaran

Soal	Mean	Interpretasi
1	0,75	Mudah
2	0,40	Sedang
3	0,70	Sedang
4	0,40	Sedang
5	0,60	Sedang
6	0,25	Sukar
7	0,45	Sedang
8	0,15	Sukar
9	0,45	Sedang
10	0,85	Mudah
11	0,35	Sedang
12	0,95	Mudah
13	0,40	Sedang
14	0,80	Mudah
15	0,60	Sedang
16	0,30	Sukar
17	0,20	Sukar
18	0,15	Sukar
19	0,20	Sukar
20	0,50	Sedang

Tabel 2. Hasil Analisa Tingkat Kesukaran

Tingkat kesukaran butir soal penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya menunjukkan butir soal dengan kategori mudah sebanyak 20%, butir soal dengan kategori sedang sebanyak 50% dan butir soal dengan kategori sukar sebanyak 30%.

Daya Beda

Berdasarkan hasil output SPSS daya beda yang telah dianalisis berdasarkan indeks daya beda diketahui bahwa butir soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran

2024/2025 yang berkategori baik sekali sebanyak 1 butir soal, berkategori baik sebanyak 5 butir soal, berkategori cukup sebanyak 4 butir soal, dan berkategori buruk sebanyak 10 butir soal. Hasil output SPSS dan hasil analisis dapat dilihat pada gambar dan tabel berikut ini.

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Soal_1	8.70	6.958	.382	.466
Soal_2	9.05	6.997	.301	.478
Soal_3	8.75	7.355	.186	.502
Soal_4	9.05	8.155	-.125	.560
Soal_5	8.85	7.608	.068	.524
Soal_6	9.20	6.800	.454	.453
Soal_7	9.00	7.368	.152	.508
Soal_8	9.30	8.326	-.194	.555
Soal_9	9.00	7.053	.272	.484
Soal_10	8.60	7.516	.199	.501
Soal_11	9.10	7.568	.090	.520
Soal_12	8.50	8.263	-.205	.541
Soal_13	9.05	8.155	-.125	.560
Soal_14	8.65	6.766	.523	.445
Soal_15	8.85	6.661	.438	.449
Soal_16	9.15	6.345	.627	.413
Soal_17	9.25	8.303	-.178	.558
Soal_18	9.30	7.905	.005	.528
Soal_19	9.25	7.461	.188	.502
Soal_20	8.95	7.418	.132	.512

Gambar 3. Output Daya Beda

Soal	Corrected Item-Total Correlation	Interpretasi
1	0.382	Cukup
2	0.301	Cukup
3	0.186	Buruk
4	-0.125	Buruk
5	0.068	Baik
6	0.454	Baik
7	0.152	Buruk
8	-0.194	Buruk
9	0.272	Cukup
10	0.199	Buruk
11	0.090	Baik Sekali
12	-0.205	Cukup
13	-0.125	Buruk
14	0.523	Baik
15	0.438	Baik
16	0.627	Baik
17	-0.178	Buruk
18	0.005	Buruk
19	0.188	Buruk
20	0.132	Buruk

Tabel 3. Hasil Analisis Daya Beda

Persentase daya beda soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-Ahya Banyuasin 1 Tahun Pelajaran 2024/2025 menunjukkan persentase butir soal dengan kategori baik sekali sebanyak 5%, butir soal dengan kategori baik sebanyak 25%, butir soal dengan kategori cukup sebanyak 20% dan butir soal dengan kategori buruk sebanyak 50%.

Efektivitas pengecoh (*distractor effectiveness*)

Menunjukkan hasil analisis efektivitas pengecoh terhadap 20 butir soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025 memiliki 10 butir soal dengan distraktor yang berfungsi, yaitu pada butir soal nomor 6,8,9,10,11,15,17,18,19,20 dan 10 butir soal memiliki distraktor yang tidak berfungsi, yaitu pada butir soal nomor 1,2,3,4,5,7,12,13,14,16. Suatu butir soal dinyatakan memiliki pengecoh soal dengan baik, jika semua atau 4 pengecoh soalnya berfungsi atau minimal dipilih 5% oleh seluruh peserta tes, apabila salah satu pengecoh soal tidak berfungsi atau tidak dipilih oleh minimal 5% oleh seluruh peserta tes, maka butir soal tersebut dinyatakan memiliki pengecoh soal yang tidak baik atau tidak berfungsi (Himawan & Nurgiyantoro, 2022).

Pembahasan

Penelitian menganalisis 20 butir soal PTS Bahasa Indonesia kelas VI di SD IT Al-Ahya.

Yang mana hasilnya :

1. Validitas soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025 dinyatakan valid atau sahih pada kategori validitas isi sedangkan pada kategori validitas empiris atau validitas butir soal dinyatakan tidak valid atau tidak sahih dengan persentase 70%.
2. Reliabilitas, analisis reliabilitas terhadap soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025 dilakukan secara menyeluruh dengan menggunakan SPSS serta dengan model iAlpha. Hasil perhitungan koefisien reliabilitas tes diperoleh 0,519 kemudian dikonsultasikan dengan patokan menurut teori Wiratna Sujarweni (2014), yaitu jika hasil tes sama dengan atau lebih besar daripada 0,60 maka dapat dinyatakan reliabilitas yang tinggi (=reliabel). Jika hasil tes lebih kecil atau kurang dari 0,60 maka dapat dinyatakan belum memiliki reliabilitas yang tinggi (un-reliabel). Nilai 0,519 yang diperoleh lebih kecil daripada 0,60 maka dapat dinyatakan tidak reliabel.
3. Tingkat Kesukaran, hasil analisis tingkat kesukaran terhadap soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025 menunjukkan bahwa soal yang

berkategori mudah sebanyak 3 butir soal, berkategori sedang sebanyak 10 butir soal, dan berkategori sukar sebanyak 6 butir soal. Banyaknya butir soal yang berkategori sedang menunjukkan soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025 memiliki tingkat kesukaran yang baik. Hal ini sesuai dengan teori evaluasi oleh Arifin (2012) bahwa Suatu tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah. Oleh sebab itu, soal yang berkategori sedang dapat dimasukkan dalam bank soal untuk dapat digunakan kembali oleh peserta didik Angkatan selanjutnya. Sedangkan untuk soal yang berkategori sukar dan mudah digugurkan atau dapat dilakukan analisis apa penyebab peserta didik terlalu sulit dan terlalu mudah dalam menjawab soal tersebut. Hasil penelitian ini didukung oleh teori evaluasi Arikunto (2018) yang menggolongkan indeks tingkat kesukaran yaitu soal dengan kategori sukar memiliki indeks 0,00 – 0,30, soal dengan kategori sedang memiliki indeks 0,31 – 0,70, dan soal dengan kategori mudah memiliki indeks 0,71 – 1,00.

4. Daya Beda, hasil analisis daya pembeda soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025 menunjukkan soal dengan kategori baik sekali sebanyak 5% dari 20 soal, soal dengan kategori baik sebanyak 25% dari 20 soal, soal dengan kategori cukup sebanyak 20% dari 20 soal, dan soal dengan kategori buruk sebanyak 50% dari 20 soal. Hasil tersebut didukung oleh teori evaluasi dari Anas Sudijono (2012) yang menyatakan bahwa kegiatan daya beda dilakukan untuk menjawab pertanyaan: “Apakah peserta didik yang dianggap pandai pada umumnya akan menjawab benar, dan apakah peserta didik yang dianggap kurang pandai pada umumnya akan menjawab salah?”. Jika yang didapat adalah peserta didik yang dianggap pandai lebih banyak menjawab benar daripada peserta didik yang dianggap kurang pandai, maka daya pembeda pada butir soal berfungsi dengan baik. Sebaliknya jika peserta didik yang dianggap kurang pandai lebih banyak menjawab soal dengan benar dibandingkan peserta didik yang dianggap pandai, maka daya pembeda soal dikatakan tidak berfungsi dengan baik karena dianggap bertentangan dengan tujuan tes itu sendiri.
5. Efektivitas pengecoh, hasil analisis efektivitas pengecoh soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya

Banyuasin 1 Tahun Pelajaran 2024/2025 memiliki pengecoh soal yang berfungsi dengan persentase 50% dari 20 soal dan pengecoh soal yang tidak berfungsi dengan persentase 50% dari 20 soal. Samanya persentase pengecoh soal yang berfungsi dengan pengecoh soal yang tidak berfungsi dapat dinyatakan bahwa soal tersebut mampu mengecoh peserta didik dan dianggap berhasil. Hasil penelitian ini didukung oleh teori evaluasi Anas Sudijono (2012) yang mengatakan "Distraktor yang dianggap berhasil jika sekurang-kurang berhasil dipilih oleh 5% dari seluruh peserta tes". Berdasarkan pembahasan di atas dapat dinyatakan keberfungsian efektivitas pengecoh soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Kelas VI SD islam terpadu Al-ahya Banyuasin 1 Tahun Pelajaran 2024/2025 telah menjalankan fungsinya dengan baik.

KESIMPULAN

Berdasarkan hasil penelitian dan analisis terhadap butir soal Penilaian Tengah Semester (PTS) mata pelajaran Bahasa Indonesia kelas VI SD Islam Terpadu Al-Ahya Tahun Pelajaran 2024/2025, maka dapat disimpulkan validitas isi menunjukkan bahwa seluruh butir soal telah sesuai dengan indikator dalam kisi-kisi dan tujuan pembelajaran yang telah ditetapkan. Hal ini menunjukkan bahwa dari sisi materi, soal telah sesuai dengan kompetensi dasar dan materi ajar yang berlaku. Validitas empiris (statistik) menunjukkan bahwa hanya 6 dari 20 butir soal (30%) yang memenuhi kriteria valid berdasarkan perhitungan korelasi Pearson dengan SPSS. Sedangkan 14 soal (70%) tidak valid. Hasil uji reliabilitas menggunakan metode Cronbach's Alpha memperoleh nilai sebesar 0,519. Nilai ini $< 0,60$, menunjukkan bahwa soal-soal PTS Bahasa Indonesia kelas VI tidak reliabel. Artinya, soal tidak menunjukkan konsistensi dalam mengukur kemampuan siswa, sehingga perlu perbaikan dari segi penyusunan dan pengujian soal. Dari hasil analisis tingkat kesukaran diperoleh bahwa 4 soal (20%) berada dalam kategori mudah, 10 soal (50%) berada dalam kategori sedang, dan 6 soal (30%) kategori sukar.

Analisis terhadap daya beda menunjukkan bahwa 1 soal (5%) memiliki daya beda sangat baik, 5 soal (25%) baik, 4 soal (20%) cukup, dan 10 soal (50%) buruk. Hasil analisis terhadap pengecoh (distraktor) menunjukkan bahwa dari 20 soal yang dianalisis 10 soal (50%) memiliki pengecoh yang berfungsi, 10 soal (50%) memiliki pengecoh yang tidak berfungsi. Ini berarti bahwa hanya separuh soal yang distraktornya

mampu mengecoh siswa dengan baik. Distraktor yang tidak berfungsi dapat menyebabkan siswa langsung mengeliminasi pilihan, sehingga mengurangi kualitas soal sebagai alat evaluasi.

DAFTAR RUJUKAN

- Anastasi, A., & Urbina, S. (2020). *Psychological testing* (7th ed.). Pearson Education.
- Arikunto, S. (2022). *Dasar-dasar evaluasi pendidikan*. Bumi Aksara.
- Ismail, F., & Astuti, H. (Eds.). (2020). *Evaluasi pendidikan*. Universitas Islam Negeri Raden Fatah Palembang. <http://repository.radenfatah.ac.id>
- Sari, D., & Nugroho, I. (2023). Peningkatan kualitas soal ujian dalam mengukur keterampilan berpikir kritis siswa. *Jurnal Pendidikan dan Pembelajaran*, 16(4), 98–110.
- Syamsul, A. (2021). Evaluasi kualitas soal ujian dalam menilai kemampuan siswa: Perspektif kurikulum dan pembelajaran. *Jurnal Pendidikan Indonesia*, 18(2), 124–136.
- Yamin, M. (2021). *Evaluasi pembelajaran: Teori, konsep, dan aplikasinya dalam pembelajaran abad 21*. Penerbit Kencana.