

Penyusunan Butir Soal Unsur Bahasa Arab (Bunyi, Kosakata, dan Struktur Kalimat) dan Uji Instrumen Butir Soal di MTs S Aluppang

Muh. Naufal Fitra¹, Dewi Chamidah², Sitti Najmianti Rahma Basit³, Sardiyannah⁴

^{1,2}Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia

³Universitas Islam Negeri Alauddin, Makassar, Indonesia

⁴Universitas Islam Ahmad Dahlan, Sinjai, Indonesia

E-mail: muh.naufalfitra@gmail.com

DOI: 10.47435/naskhi.v6i1.2964



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

Abstract

Based on previous research related to the evaluation of items that have been prepared, there are two main problems. First, there is no clear description of the items given. Second, the variety of difficulty levels and forms of questions is not sufficient. In addition, many questions given to students have not been tested for validity and reliability. This study aims to develop items of Arabic language elements (sound, vocabulary, and sentence structure) and test the question instrument at MTs S Aluppang. This research is expected to produce valid and reliable items so that they can be used to measure the level of mastery of Arabic language elements in MTs S Aluppang students more accurately and objectively. This research is a type of research and development. The development model used in this study is based on the stages of assessment planning. These steps are compiling test specifications, writing test questions, reviewing test questions, conducting test trials, and analyzing test items. The test of the test instrument was carried out for class VIII students of MTs S Aluppang, totaling 13 students. The results showed that the preparation of items had gone through the appropriate stages and produced 10 items for each aspect of language elements. In general, the test of the item instrument shows that there are 16 items or 53% of all items that are suitable for reuse and there are 14 items or 47% of all items that must be corrected or discarded.

Keywords: *Preparation of Question Items; Arabic Language Element Questions; Test Item Instrument Questions*

Abstrak

Berdasarkan penelitian terdahulu terkait evaluasi terhadap butir soal yang telah disusun menunjukkan adanya dua permasalahan utama. *Pertama*, tidak ada deskripsi yang jelas pada butir soal yang diberikan. *Kedua*, variasi tingkat kesulitan dan bentuk soal belum memadai. Selain itu, soal yang diberikan kepada peserta didik masih banyak yang belum diuji validitas dan reliabilitasnya. Penelitian ini bertujuan untuk menyusun butir soal unsur bahasa Arab (bunyi, kosakata, dan struktur kalimat) dan menguji instrumen soal di MTs S Aluppang. Penelitian ini diharapkan dapat menghasilkan butir soal yang valid dan reliabel sehingga dapat digunakan untuk mengukur tingkat penguasaan unsur-unsur bahasa Arab pada peserta didik MTs S Aluppang

dengan lebih akurat dan objektif. Penelitian ini merupakan jenis penelitian dan pengembangan. Model pengembangan yang digunakan dalam penelitian ini berdasarkan tahapan perencanaan penilaian. Langkah-langkah tersebut adalah menyusun spesifikasi tes, menulis soal tes, menelaah soal tes, melakukan uji coba tes, dan menganalisis butir soal. Uji coba instrumen tes dilakukan bagi peserta didik kelas VIII MTs S Aluppang yang berjumlah 13 peserta didik. Hasil penelitian menunjukkan bahwa penyusunan butir soal telah melalui tahapan yang sesuai dan menghasilkan 10 butir soal untuk masing-masing aspek unsur bahasa. Secara umum, uji instrumen butir soal menunjukkan terdapat 16 butir soal atau 53% dari keseluruhan butir soal yang layak untuk digunakan kembali dan terdapat 14 butir soal atau 47% dari keseluruhan butir soal yang harus diperbaiki atau dibuang.

Kata Kunci: Penyusunan Butir Soal; Soal Unsur Bahasa Arab; Uji Instrumen Butir Soal

1. Pendahuluan

Bahasa Arab merupakan komponen penting dalam kurikulum madrasah, dipelajari secara menyeluruh di semua tingkatan pendidikan madrasah, termasuk Madrasah Tsanawiyah (Nurkholis, 2019, p. 233). Di beberapa madrasah, mata pelajaran Bahasa Arab merupakan mata pelajaran yang masih menggunakan Kurikulum 2013. Berdasarkan Keputusan Menteri Agama Nomor 183 Tahun 2019, mata pelajaran Bahasa Arab di Madrasah Tsanawiyah bertujuan untuk (1) Mengembangkan kemampuan berkomunikasi, baik secara lisan maupun tulisan dalam bahasa Arab. Hal ini mencakup empat kecakapan berbahasa, yaitu *istima'* (kemampuan menyimak dan memahami pembicaraan dalam bahasa Arab), *kalam* (kemampuan berbicara dan mengungkapkan diri dalam bahasa Arab), *qira'ah* (kemampuan membaca dan memahami teks bahasa Arab), *kitabah* (kemampuan menulis dan menghasilkan karya tulis dalam bahasa Arab), (2) Menumbuhkan kesadaran tentang pentingnya bahasa Arab sebagai salah satu bahasa asing yang bermanfaat untuk proses pembelajaran dan pengkajian sumber-sumber Islam dan sebagai alat komunikasi, dan (3) Mengembangkan pemahaman tentang hubungan erat antara bahasa dan budaya dan memperluas cakrawala budaya. Sehingga, hal ini dapat membekali peserta didik dengan kemampuan bahasa yang mumpuni, meliputi gramatika, wacana, strategi, sosiologis, dan budaya. Kemampuan berbahasa Arab ini sangat diperlukan bagi peserta didik MTs untuk melanjutkan pendidikan ke jenjang yang lebih tinggi, khususnya di bidang keagamaan.

Salah satu aspek krusial dalam pembelajaran bahasa Arab adalah penguasaan unsur-unsur bahasanya. Penguasaan unsur bahasa Arab, meliputi bunyi (*ashwat*), kosakata (*mufradat*), dan struktur kalimat (*tarkib*), menjadi fondasi bagi peserta didik untuk memahami dan menggunakan bahasa Arab dengan baik dan benar (Suib et al., 2022, p. 149). Penguasaan unsur-unsur bahasa ini ibarat fondasi bangunan. Fondasi yang kokoh akan menopang bangunan yang kokoh pula. Demikian pula dengan penguasaan unsur-unsur bahasa Arab. Semakin baik penguasaan unsur-unsur bahasa Arab, semakin baik pula kemampuan peserta didik dalam memahami dan menggunakan bahasa Arab.

Menilai tingkat penguasaan unsur-unsur bahasa Arab pada peserta didik MTs memerlukan alat ukur yang sah dan teruji. Tes menjadi salah satu instrumen yang tepat. Kualitas suatu tes ditentukan oleh tingkat validitasnya. Tes yang valid adalah tes yang mampu mengukur dengan tepat apa yang ingin diukur. Untuk mencapai validitas tes, diperlukan serangkaian langkah-langkah yang terencana dan sistematis. (Ainin, 2016, p. 293). Kualitas soal tes menjadi kunci utama dalam mendapatkan hasil pengukuran yang valid dan reliabel. Oleh karena itu, penyusunan soal tes harus dilakukan secara cermat dan sistematis, dengan memperhatikan kaidah-kaidah yang berlaku.

Salah satu tahapan dalam menyusun tes adalah tahapan perencanaan yang dipaparkan oleh Munip (2017, p. 131). Langkah-langkah tersebut antara lain: *Pertama*, penyusunan spesifikasi tes yang meliputi menentukan tujuan tes dibuat, menyusun kisi-kisi tes, memilih bentuk tes, dan menentukan panjang tes. *Kedua*, penulisan soal tes yang merupakan langkah mengubah indikator menjadi pertanyaan-pertanyaan yang sesuai dengan kisi-kisi yang telah dirumuskan sebelumnya. *Ketiga*, telaah soal tes yang bertujuan untuk memeriksa kembali soal-soal yang telah dibuat untuk

memastikan tidak ada kesalahan atau kekurangan dan mendapatkan masukan dari para ahli atau praktisi untuk meningkatkan kualitas soal. *Keempat*, uji coba tes dengan melakukan tes pada sekelompok kecil peserta didik untuk mendapatkan data empiris tentang kualitas soal. Melalui uji coba dapat diperoleh data tentang reliabilitas, validitas, tingkat kesukaran, dan lain-lain. *Kelima*, analisis butir soal terhadap masing-masing butir soal yang telah disusun. Melalui analisis butir ini dapat diketahui antara lain: validitas butir, reliabilitas butir, tingkat kesukaran butir soal, dan daya pembeda butir soal. *Keenam*, perbaikan tes terutama pada bagian-bagian soal yang masih belum sesuai dengan yang diharapkan berdasarkan hasil analisis butir soal. *Ketujuh*, Perakitan butir-butir soal telah dianalisis dan diperbaiki menjadi satu kesatuan tes, menyusun tes dengan memperhatikan urutan soal yang logis dan sistematis, dan memastikan format tes rapi, jelas, dan mudah dibaca.

Berdasarkan kajian literatur, ditemukan beberapa permasalahan terkait dengan penyusunan butir soal unsur bahasa Arab di MTs. Permasalahan tersebut antara lain: *Pertama*, butir soal yang diberikan tidak dideskripsikan dengan jelas dan tegas (Laili, 2020, p. 111). *Kedua*, banyak butir soal yang tidak valid daripada yang valid (Tanjung et al., 2024, p. 347). *Ketiga*, tingkat kesulitan dan bentuk tes yang diberikan masih belum bervariasi (Fauzi, 2023, p. 635). Permasalahan pertama dan ketiga juga didapatkan peneliti di MTs S Aluppang. Sedangkan, berkaitan dengan validitas dan reliabilitas butir soal, guru Bahasa Arab MTs S Aluppang mengakui bahwa soal yang diberikan kepada peserta didik masih belum diuji validitas dan reliabilitasnya. Permasalahan-permasalahan tersebut menyebabkan hasil tes yang diperoleh peserta didik tidak dapat mencerminkan kemampuan mereka yang sebenarnya.

Oleh karena itu, perlu dilakukan penelitian untuk menyusun butir soal unsur bahasa Arab (bunyi, kosakata, dan struktur kalimat) dan menguji instrumen soal di MTs S Aluppang. Penelitian ini diharapkan dapat menghasilkan butir soal yang valid dan reliabel sehingga dapat digunakan untuk mengukur tingkat penguasaan unsur-unsur bahasa Arab pada peserta didik MTs S Aluppang dengan lebih akurat dan objektif.

2. Metode

Jenis penelitian yang digunakan adalah Penelitian dan Pengembangan (*Research and Development*). Sedangkan model pengembangan yang digunakan berdasarkan tahapan perencanaan penilaian menurut Munip (2017, p. 131). Langkah-langkah tersebut adalah menyusun spesifikasi tes, menulis soal tes, menelaah soal tes, melakukan usji coba tes, menganalisis butir soal, memperbaiki tes, dan merakit tes. Namun, dalam penelitian ini, penyusunan butir soal terbatas hingga tahapan analisis butir soal.

Uji coba instrumen tes dilakukan bagi peserta didik kelas VIII MTs S Aluppang Kabupaten Bone yang berjumlah 13 peserta didik. Teknik sampling yang digunakan adalah sampel jenuh, di mana seluruh populasi dalam penelitian digunakan sebagai sampel penelitian (Suriani et al., 2023, p. 30). Uji coba instrumen tes dilakukan dengan menguji validitas dan realibitas tes. Kemudian, soal dianalisis menggunakan analisis tingkat kesukaran butir soal dan daya pembeda butir soal.

3. Hasil dan Pembahasan

3.1. Menyusun Spesifikasi Tes

Langkah awal dalam mengembangkan tes yang berkualitas adalah dengan menetapkan spesifikasi tes. Spesifikasi ini memuat uraian lengkap mengenai karakteristik yang harus dimiliki tes tersebut. Memiliki spesifikasi tes yang jelas bagaikan memiliki peta jalan yang terperinci. Hal ini akan memudahkan proses penulisan soal, dan memastikan bahwa semua soal memiliki tingkat kesulitan yang relatif sama, tanpa terpengaruh oleh siapapun yang menulisnya. Penyusunan spesifikasi tes ini mencakup beberapa kegiatan yang penting, di antaranya (1) menentukan tujuan tes, (2) menyusun kisi-kisi, (3) memilih bentuk tes, dan (4) menentukan panjang tes.

Tujuan tes dalam penelitian ini disusun dengan tujuan untuk tes formatif. Tes formatif bagaikan cermin bagi proses pembelajaran. Melalui tes ini, guru dapat memperoleh data tentang sejauh mana peserta didik memahami materi yang telah diajarkan. Data ini kemudian dimaknai

dan diubah menjadi umpan balik yang berharga untuk meningkatkan kualitas pembelajaran (Firani Putri & Supratman Zakir, 2023, p. 176). Dengan kata lain, tes formatif adalah evaluasi yang dilakukan secara berkala di akhir setiap pembahasan materi atau topik. Tujuannya adalah untuk mengetahui apakah proses pembelajaran telah berjalan sesuai dengan rencana.

Tes dalam penelitian ini disusun untuk buku Bahasa Arab MTs Kelas VIII terbitan Direktorat KSKK Madrasah (Masrukhin, 2020) pada bab 6 dengan topik *عيادة المرضى* (Menjenguk Orang Sakit). Kisi-kisi soal dapat dilihat pada Tabel 1 berikut.

Tabel 1. Kisi-Kisi Soal

No.	KD*	Materi	Indikator Soal	Bentuk Tes	No. Soal
1	3.11 Memahami fungsi sosial, struktur teks, dan unsur kebahasaan (bunyi, kata, dan makna) dari teks sederhana yang berkaitan dengan tema: <i>عيادة المرضى</i> yang melibatkan tindak tutur mendoakan orang sakit dengan memperhatikan gramatikal susunan <i>الفعل الماضي والجملة الفعلية</i>	<i>عيادة المرضى</i>	Peserta didik diminta membedakan bunyi setiap kata atau kalimat yang mirip	Pilihan Ganda	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
			Peserta didik diminta menjodohkan kosakata dengan gambar yang sesuai	Menjodohkan	11, 12, 13, 14, 15
			Peserta didik diminta mengisi isian yang kosong dengan kata yang telah disediakan	Menjodohkan	16, 17, 18, 19, 20
			Peserta didik diminta menunjukkan struktur kalimat yang salah	Benar-Salah	21, 22, 23, 24, 25, 26, 27, 28, 29, 30

*sumber: Masrukhin (2020:88)

Sebagaimana dalam Tabel 1 di atas, tes yang disusun adalah tes objektif dengan beragam bentuk soal. Peneliti menggunakan bentuk soal pilihan ganda untuk soal unsur suara, bentuk soal menjodohkan untuk soal unsur kosakata, dan bentuk soal benar-salah untuk soal unsur struktur kalimat (Munip, 2017). Bentuk tes pilihan ganda merupakan jenis tes di mana para peserta didik dihadapkan dengan beberapa pilihan jawaban untuk satu pertanyaan atau pernyataan. Jawaban yang benar hanya satu, dan peserta didik harus memilih jawaban yang paling tepat dari beberapa alternatif yang disediakan. Struktur tes pilihan ganda umumnya terdiri dari dua bagian, yaitu pernyataan atau pokok soal dan alternatif jawaban, yang mencakup kunci jawaban dan pengecoh (distraktor).

Bentuk tes menjodohkan terdiri dari dua kelompok pertanyaan, yaitu pernyataan yang berisi beberapa pernyataan yang harus dicocokkan dan jawaban yang berisi beberapa pilihan jawaban yang harus dihubungkan dengan pernyataan yang sesuai. Tugas peserta didik adalah menemukan pasangan yang paling tepat antara pernyataan dan jawaban yang tersedia. Bentuk tes benar-salah terdiri dari sejumlah pernyataan yang dinilai berdasarkan kebenarannya. Peserta didik harus menentukan apakah setiap pernyataan benar atau salah. Jenis tes benar-salah yang digunakan dalam penelitian ini adalah tanpa pembetulan. Artinya, peserta didik tidak perlu membetulkan apabila jawaban yang dipilih salah.

Tes yang disusun adalah tes objektif yang relatif lebih mudah sehingga pengerjaannya relatif lebih cepat berkisar 30 detik hingga 60 detik (Faisal, 2022, p. 6). Jumlah soal yang disusun

sebanyak 30 soal dan setiap soal dapat diselesaikan selama satu menit. Sehingga waktu yang diperlukan dalam menyelesaikan semua soal adalah 30 menit.

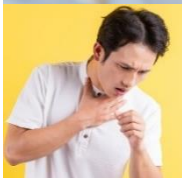
3.2. Menulis Soal Tes

Berikut soal tes yang disusun berdasarkan kisi-kisi yang telah dirumuskan pada tahapan sebelumnya.

Dengarkanlah kata atau kalimat yang diucapkan oleh guru dan pilihlah jawaban yang sesuai!

د. سُدَاعٌ	ج. طُدَاعٌ	ب. صُدَاعٌ	أ. تُدَاعٌ	١
د. مُصْطَشَقِي	ج. مُسْطَشَقِي	ب. مُصْتَشَقِي	أ. مُسْتَشَقِي	٢
د. طَبِيبٌ	ج. طَبِيبٌ	ب. تَابِيبٌ	أ. تَبِيبٌ	٣
د. ذَهَبِنٌ	ج. ذَهَبِنَا	ب. سَحَابِنٌ	أ. سَحَابِنَا	٤
د. مَرِيذٌ	ج. مَرِيضٌ	ب. مَرِيذٌ	أ. مَرِيضٌ	٥
د. صَرَ مَحْمُودٌ نَاجِحًا	ج. صَارَ مَحْمُودٌ نَاجِحًا	ب. صَرَ مَحْمُودٌ نَجِحًا	أ. صَارَ مَحْمُودٌ نَجِيفًا	٦
د. شُفِيْطٌ فَاطِمَةٌ	ج. شُفِيْثٌ فَاطِمَةٌ	ب. شُوفِيْطٌ فَاطِمَةٌ	أ. شُوفِيْثٌ فَاطِمَةٌ	٧
د. تَنَاوَلْتُ الدَّوَاءَ	ج. تَنَاوَلْتُ الدَّوَاءَ	ب. تَنَاوَلْتُ الدَّوَاءَ	أ. تَنَاوَلْتُ الدَّوَاءَ	٨
د. أَشْعُرُ بِأَلَمٍ	ج. أَشْعُرُ بِأَلَمٍ	ب. أَشْعُرُ بِعَلَمٍ	أ. أَشْعُرُ بِأَلَمٍ	٩
د. عِبَادَةُ المُرْضَى	ج. عِبَادَةُ المُرْذَى	ب. إِيَادَةُ المُرْذَى	أ. إِيَادَةُ المُرْضَى	١٠

Hubungkan kata dengan gambar yang benar!



•

•

الدَّوَاءُ ١١

•

•

التَّعَبُ ١٢

•

•

حَرَارَةٌ ١٣

•

•

رِيَاضَةُ المَسْنِي ١٤

•

•

التَّهَابُ فِي الحَلْقِ ١٥

Isi bagian yang kosong dengan kata yang sesuai!

غَابَثَ

اسْتَأْذَنَتْ

تَحَسَّنَ

دَعَتْ

يَحْتَاجُ

حَالُ الْمَرِيضِ بَعْدَ أَنْ تَنَاوَلَ الدَّوَاءَ _____	١٦
الْمَرِيضُ _____ إِلَى الْهُدُوءِ	١٧
فَاطِمَةُ عَنِ الْمُدْرَسَةِ لِأَنَّهَا مَرِيضَةٌ _____	١٨
رِيَانَةُ وَالِدَيْهَا فِي زِيَارَةِ فَاطِمَةَ _____	١٩
لَمَّا وَصَلَتْ رِيَانَةُ إِلَى صَدِيقَتِهَا الْمَرِيضَةِ _____ لَهَا الشِّفَاءَ	٢٠

Berikan tanda X pada (ص) jika kalimat benar dan tanda X pada (خ) jika salah!

أَنَا : وَصَلْتُ إِلَى صَدِيقِي	(ص خ)	٢١
نَحْنُ : جَلَسَ عَلَى الْكُرْسِيِّ	(ص خ)	٢٢
هُوَ : فَهَمَّتِ الدِّرَاسَةَ	(ص خ)	٢٣
هِيَ : تَنَاوَلَتْ فَاطِمَةُ الدَّوَاءَ	(ص خ)	٢٤
أَنْتِ : شَعَرْتُ بِالْأَلَمِ	(ص خ)	٢٥
أَنْتِ : شَعَرْتُ بِالْأَلَمِ	(ص خ)	٢٦
أَنَا : أَتَنَاوَلُ الدَّوَاءَ	(ص خ)	٢٧
نَحْنُ : تَفَهَّمُ الدِّرَاسَةَ	(ص خ)	٢٨
أَنْتِ : تَشْعُرُ بِالْأَلَمِ	(ص خ)	٢٩
أَنْتِ : تَشْعُرِينَ بِالْأَلَمِ	(ص خ)	٣٠

3.3. Menelaah Soal Tes

Soal tes yang telah ditulis kemudian diberikan kepada guru mata pelajaran Bahasa Arab di MTs S Alupang untuk ditelaah berdasarkan instrumen telaah butir soal objektif dalam Munip (2017, p. 140). Hasil telaah butir soal dapat dilihat dalam Tabel 2 berikut.

Tabel 2. Telaah Butir Soal

No	Komponen Telaah Butir Soal	Skor			
		1 (TS)	2 (KS)	3 (S)	4 (SS)
A. Aspek Materi					
1	Soal sesuai dengan indikator				✓
2	Distraktor berfungsi dengan baik			✓	
3	Hanya ada satu kunci jawaban yang paling tepat				✓
B. Aspek Konstruksi					
4	Pokok soal dirumuskan dengan singkat, jelas dan tegas				✓
5	Rumusan pokok soal dan pilihan jawaban merupakan pertanyaan yang diperlukan				✓
6	Pokok soal tidak memberi petunjuk ke kunci jawaban				✓
7	Pokok soal bebas dari pernyataan yang bersifat negatif ganda				✓
8	Gambar, grafik, tabel, diagram, wacana, dan sejenisnya yang terdapat pada soal jelas dan berfungsi			✓	
9	Panjang pilihan jawaban relatif sama				✓

No	Komponen Telaah Butir Soal	Skor			
		1 (TS)	2 (KS)	3 (S)	4 (SS)
10	Pilihan jawaban tidak menggunakan pernyataan “Semua jawaban di atas salah” atau “Semua pilihan jawaban di atas benar” dan sejenisnya				✓
11	Pilihan jawaban yang berbentuk angka atau waktu harus disusun berdasarkan urutan besar kecilnya angka tersebut atau kronologis;				✓
12	Butir-butir soal tidak bergantung pada jawaban soal sebelumnya			✓	
C. Aspek Bahasa/ Budaya					
13	Menggunakan bahasa yang sesuai dengan kaidah Bahasa Indonesia/ Bahasa Arab;				✓
14	Menggunakan bahasa yang komunikatif				✓
15	Tidak menggunakan bahasa yang berlaku setempat (bias budaya)				✓
16	Pilihan jawaban tidak mengulang kata/kelompok kata yang sama				✓
Total		0	0	3	13

sumber: Guru Bahasa Arab MTs S Aluppong

Berdasarkan Tabel 2 di atas, diperoleh informasi bahwa tiga indikator mendapatkan skor tiga dan 13 indikator mendapatkan skor 4. Berdasarkan skor tersebut, maka diperoleh skor total 61. Selanjutnya, kualitas soal dapat dikategorikan berdasarkan kategorisasi pada Tabel 3 berikut.

Tabel 3. Kategorisasi Kualitas Soal

No	Rentang Skor	Kualitas
1	49-64	Baik
2	33-48	Sedang
3	17-32	Kurang

sumber: Munip (2017, p. 141)

Berdasarkan Tabel 3 di atas, maka kualitas soal yang telah disusun dapat dikategorikan sebagai soal yang baik. Namun, perlu diingat bahwa telaah butir soal ini mengacu pada satu teori tertentu sehingga perlu dilakukan penilaian yang lebih komprehensif melalui uji coba instrumen tes di MTs S Aluppong untuk mengetahui uji validitas, uji reliabilitas, tingkat kesukaran, dan tingkat daya pembeda butir soal..

3.4. Menguji Coba Tes



Gambar 1. Dokumentasi Uji Coba Tes

Tes diujicobakan pada peserta didik kelas VIII MTs S Aluppang yang berjumlah 13 peserta didik. Berikut hasil uji coba yang diperoleh peneliti.

Tabel 4. Hasil Uji Coba Tes

Peserta Didik	Skor	Nilai	Keterangan
Peserta Didik A	9	30.00	Tidak Lulus
Peserta Didik B	21	70.00	Lulus
Peserta Didik C	12	40.00	Tidak Lulus
Peserta Didik D	24	80.00	Lulus
Peserta Didik E	7	23.33	Tidak Lulus
Peserta Didik F	24	80.00	Lulus
Peserta Didik G	23	76.67	Lulus
Peserta Didik H	20	66.67	Tidak Lulus
Peserta Didik I	23	76.67	Lulus
Peserta Didik J	12	40.00	Tidak Lulus
Peserta Didik K	8	26.67	Tidak Lulus
Peserta Didik L	24	80.00	Lulus
Peserta Didik J	11	36.67	Tidak Lulus

KKM yang diterapkan oleh guru Bahasa Arab MTs S Aluppang adalah 70. Berdasarkan Tabel 4 di atas, sebanyak 7 (54%) peserta didik yang tidak memenuhi KKM dan 6 (46%) peserta didik yang memenuhi KKM. Mayoritas peserta didik (54%) belum mencapai KKM. Hal ini menunjukkan bahwa soal-soal tes mungkin terlalu sulit atau peserta didik belum menguasai materi dengan baik. Untuk mengetahui tingkat kesulitan butir soal, peneliti melakukan analisis butir soal berdasarkan hasil uji coba di MTs S Aluppang.

3.5. Menganalisis Butir Soal

3.5.1. Uji Validitas Butir Soal

Validitas tes adalah tingkat sejauh mana tes tersebut dapat mengukur apa yang ingin diukur. Tes yang valid akan menghasilkan data yang akurat dan dapat dipercaya, sehingga dapat digunakan sebagai dasar untuk pengambilan keputusan yang tepat (Munip, 2017, p. 244). Instrumen tes dikatakan valid jika mampu mengukur secara tepat apa yang hendak diukur, menilai secara akurat apa yang hendak dinilai, dan mengevaluasi secara tepat apa yang hendak dievaluasi. Untuk mengetahui tingkat validitas tes, dapat dilakukan uji validitas. Uji validitas ini menggunakan statistik tertentu untuk menganalisis data yang diperoleh dari tes.

Dalam penelitian ini, peneliti menggunakan *software IBM SPSS Statistics 25* untuk melakukan uji validitas tes. Nilai Sig. (2-tailed) yang lebih kecil dari 0.05 menunjukkan bahwa tes tersebut valid. Berikut adalah hasil penghitungan validitas.

Tabel 5. Uji Validitas Butir Soal

Nomor Soal	Pearson Correlation	Sig. (2-tailed)	Keterangan
Soal 1	0.765	0.002	Valid
Soal 2	0.723	0.005	Valid
Soal 3	0.704	0.007	Valid
Soal 4	0.125	0.684	Tidak Valid
Soal 5	0.639	0.019	Valid
Soal 6	0.640	0.018	Valid
Soal 7	0.324	0.28	Tidak Valid
Soal 8	0.679	0.011	Valid
Soal 9	-0.084	0.785	Tidak Valid
Soal 10	0.241	0.428	Tidak Valid
Soal 11	0.625	0.022	Valid
Soal 12	0.636	0.02	Valid
Soal 13	0.625	0.022	Valid
Soal 14	0.13	0.671	Tidak Valid
Soal 15	0.700	0.008	Valid

Nomor Soal	Pearson Correlation	Sig. (2-tailed)	Keterangan
Soal 16	0.704	0.007	Valid
Soal 17	0.201	0.511	Tidak Valid
Soal 18	0.704	0.007	Valid
Soal 19	0.274	0.365	Tidak Valid
Soal 20	0.641	0.018	Valid
Soal 21	0.195	0.524	Tidak Valid
Soal 22	0.640	0.018	Valid
Soal 23	0.625	0.022	Valid
Soal 24	-0.595	0.032	Valid
Soal 25	0.729	0.005	Valid
Soal 26	0.779	0.002	Valid
Soal 27	-0.048	0.876	Tidak Valid
Soal 28	0.723	0.005	Valid
Soal 29	0.729	0.005	Valid
Soal 30	0.704	0.007	Valid

Berdasarkan Tabel 5 di atas, terdapat 21 butir soal atau 70% dari keseluruhan soal yang tergolong valid, yaitu pada nomor 1, 2, 3, 5, 6, 8, 11, 12, 13, 15, 16, 18, 20, 22, 23, 24, 25, 26, 28, 29, dan 30. Diketahui pula 9 butir soal atau 30% dari keseluruhan soal yang tergolong tidak valid, yaitu pada nomor 4, 7, 9, 10, 14, 17, 19, 21, dan 27.

Suatu tes dikatakan memiliki validitas tinggi jika hasil ukurnya sesuai dengan maksud dan tujuan diadakannya tes. Hal ini berarti bahwa tes tersebut mampu memberikan informasi yang akurat tentang apa yang ingin diketahui oleh penguji. Sebaliknya, tes yang menghasilkan data tidak relevan dengan tujuan pengukuran dikatakan sebagai tes yang memiliki validitas rendah. Tes seperti ini tidak akan memberikan informasi yang bermanfaat dan tidak dapat digunakan untuk pengambilan keputusan yang tepat (Fahmi et al., 2022, p. 99). Hal ini dapat diinterpretasikan bahwa mayoritas soal yang telah disusun memiliki validitas dan dapat digunakan sebagai alat ukur yang tepat dalam pembelajaran unsur bahasa Arab.

Meskipun tes secara keseluruhan telah teruji validitasnya, penemuan 9 butir soal yang tidak valid patut menjadi perhatian. Hal ini menunjukkan adanya potensi untuk meningkatkan kualitas tes di masa depan. Langkah penting berikutnya adalah melakukan analisis mendalam terhadap butir-butir soal yang tidak valid. Analisis ini bertujuan untuk mengevaluasi sejauh mana butir soal tersebut memenuhi kriteria validitas. Hal ini dapat dilakukan dengan meninjau kembali spesifikasi tes, kisi-kisi soal, dan butir soal itu sendiri dan mengidentifikasi faktor-faktor yang mungkin menyebabkan ketidakvalidan. Apakah karena kesalahan dalam penulisan soal, ketidaksesuaian dengan materi pelajaran, atau ambiguitas jawaban? (Anshari et al., 2024, p. 974).

3.5.2. Uji Reliabilitas Butir

Reliabilitas tes merupakan salah satu syarat penting bagi sebuah tes untuk menjadi instrumen penilaian yang handal. Tes yang reliabel akan menghasilkan hasil yang konsisten dan stabil, sehingga dapat dipercaya sebagai dasar untuk pengambilan keputusan (Munip, 2017, p. 260). Kejegan dan ketetapan adalah kunci utama dalam reliabilitas tes. Artinya, tes yang reliabel akan memberikan hasil yang sama meskipun dilakukan beberapa kali dengan kondisi yang sama. Hal ini menunjukkan bahwa hasil tes tidak dipengaruhi oleh faktor-faktor yang tidak relevan, seperti kebetulan, kelelahan, atau suasana hati peserta tes. Kepercayaan terhadap hasil tes sangatlah penting dalam proses penilaian. Jika hasil tes berubah-ubah, maka informasi yang dihasilkan tidak dapat dipercaya dan tidak dapat digunakan untuk mengambil keputusan yang tepat.

Dalam penelitian ini, peneliti menggunakan *software IBM SPSS Statistics 25* untuk melakukan uji reliabilitas tes dengan metode Cronbach's Alpha. Metode ini digunakan untuk menghitung tingkat konsistensi internal butir-butir soal dalam tes. Instrumen tes dinyatakan reliabel apabila nilai *Cronbach's Alpha* lebih besar dari nilai R tabel. Berikut adalah hasil penghitungan reliabilitas butir soal:

Tabel 6. Uji Reliabilitas Butir Soal

Case Processing Summary

		N	%
Cases	Valid	13	100.0
	Excluded ^a	0	.0
	Total	13	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's	
Alpha	N of Items
.881	30

Berdasarkan Tabel 6 di atas, tes yang telah disusun dikategorikan reliabel. Hal ini sesuai dengan keterangan bahwa nilai Cronbach's Alpha > R tabel. Di mana R tabel adalah 0.55 dengan tingkat signifikansi 5% dan nilai N=13. Hal ini dapat diinterpretasikan bahwa soal yang disusun dapat dipercaya karena cenderung memberikan hasil yang tetap.

3.5.3. Tingkat Kesukaran Butir Soal

Soal yang baik bagaikan jembatan yang menghubungkan kemampuan dan tantangan. Soal yang terlalu mudah ibarat jembatan yang terlalu pendek, tidak memberikan kesempatan bagi peserta didik untuk berkembang dan menunjukkan kemampuan terbaiknya. Sebaliknya, soal yang terlalu sulit bagaikan jurang yang curam, membuat peserta didik putus asa dan kehilangan semangat (Asrul et al., 2014, p. 149). Soal yang ideal adalah soal yang memiliki tingkat kesulitan yang tepat. Soal ini tidak terlalu mudah sehingga membosankan, dan tidak terlalu sulit sehingga membuat frustrasi. Soal yang tepat akan merangsang peserta didik untuk berpikir kritis, berusaha keras, dan mencapai potensinya. Tingkat kesukaran ini dinyatakan dalam bentuk Indeks Kesukaran Soal (IKS) yang berkisar 0.00-1.00 dengan kriteria sebagai berikut: 0.00-0.30 (sukar); 0.31-0,70 (sedang); 0.71-1 (mudah). Nilai Indeks Kesukaran Soal dapat diperoleh dengan rumus sebagai berikut.

$$P = \frac{B}{JS}$$

Keterangan:

P = Nilai Indeks Kesukaran Soal

B = Jumlah peserta didik yang menjawab benar

JS = Jumlah peserta didik keseluruhan (Asrul et al., 2014, p. 149)

Peneliti menggunakan bantuan *software Microsoft Excel* untuk mengetahui tingkat kesukaran butir soal dan kategorisasi setiap butir soal. Berikut adalah hasil yang diperoleh.

Tabel 7. Tingkat Kesukaran Butir Soal

Nomor Soal	Jumlah Jawaban Benar	Tingkat Kesukaran	Kriteria
Soal 1	5	0.38	Sedang
Soal 2	10	0.77	Mudah
Soal 3	9	0.69	Sedang
Soal 4	6	0.46	Sedang
Soal 5	8	0.62	Sedang
Soal 6	10	0.77	Mudah
Soal 7	4	0.31	Sedang
Soal 8	9	0.69	Sedang
Soal 9	6	0.46	Sedang
Soal 10	6	0.46	Sedang

Nomor Soal	Jumlah Jawaban Benar	Tingkat Kesukaran	Kriteria
Soal 11	4	0.31	Sedang
Soal 12	6	0.46	Sedang
Soal 13	4	0.31	Sedang
Soal 14	7	0.54	Sedang
Soal 15	4	0.31	Sedang
Soal 16	9	0.69	Sedang
Soal 17	10	0.77	Mudah
Soal 18	9	0.69	Sedang
Soal 19	4	0.31	Sedang
Soal 20	7	0.54	Sedang
Soal 21	6	0.46	Sedang
Soal 22	10	0.77	Mudah
Soal 23	4	0.31	Sedang
Soal 24	6	0.46	Sedang
Soal 25	9	0.69	Sedang
Soal 26	9	0.69	Sedang
Soal 27	9	0.69	Sedang
Soal 28	10	0.77	Mudah
Soal 29	9	0.69	Sedang
Soal 30	9	0.69	Sedang

Berdasarkan Tabel 7 di atas, diketahui bahwa terdapat 25 soal atau 83% dari keseluruhan soal yang dikategorikan sedang dan 5 soal atau 17% dari keseluruhan soal yang dikategorikan mudah. Sehingga dapat diinterpretasikan bahwa soal yang disusun telah baik karena soal dengan IKS sedang jauh lebih banyak dibandingkan soal dengan IKS mudah atau sukar.

Butir-butir soal yang tergolong "sedang" merupakan aset berharga dalam proses pembelajaran. Sebaiknya catat dan simpan soal-soal ini dalam buku bank soal untuk digunakan kembali di masa depan. Soal-soal ini memiliki tingkat kesulitan yang tepat untuk mengukur pemahaman peserta didik tanpa membuat mereka frustrasi (Susanto et al., 2015, p. 213). Untuk butir-butir soal yang sukar dan mudah, ada beberapa langkah yang bisa diambil. *Pertama*, lebih baik dibuang dan digantikan dengan soal yang lebih sesuai. *Kedua*, analisis soal untuk mencari tahu apa yang membuatnya sukar atau mudah. Apakah karena materi yang terlalu kompleks, materi yang terlalu sederhana, instruksi yang tidak jelas, atau kesalahan penulisan soal? Setelah diperbaiki, soal dapat digunakan kembali di masa depan. *Ketiga*, gunakan soal yang sukar untuk seleksi ketat atau soal yang mudah untuk seleksi yang longgar.

3.5.4. Daya Pembeda Butir Soal

Daya pembeda soal adalah ibarat alat ukur kemampuan. Sebuah soal yang baik haruslah mampu membedakan dengan jelas antara peserta didik yang berkemampuan tinggi dengan yang berkemampuan rendah. Indeks Diskriminasi (D) adalah angka yang menunjukkan seberapa besar daya pembeda suatu soal. Semakin tinggi nilai D, semakin baik daya pembeda soal tersebut (Asrul et al., 2014, p. 151). Indeks Diskriminasi (daya pembeda) ini berkisar antara -1.00 sampai 1.00, dengan kriteria sebagai berikut: D < 0 (Sangat Buruk); 0.00-0.20 (Buruk); 0.21-0.40 (Cukup); 0.41-0.70 (Baik); 0.71-1.00 (Sangat Baik). Nilai Indeks Diskriminasi dapat diperoleh dengan rumus berikut:

$$D = \frac{BA}{JA} - \frac{BB}{JB}$$

Keterangan:

D = Indeks Diskriminasi

BA = Jumlah peserta kelompok atas yang menjawab benar

- BB = Jumlah peserta kelompok bawah yang menjawab benar
 JA = Jumlah keseluruhan peserta kelompok atas
 JB = Jumlah keseluruhan peserta kelompok bawah (Asrul et al., 2014, p. 153)

Peneliti menggunakan bantuan *software Microsoft Excel* untuk mengetahui tingkat daya pembeda butir soal dan kategorisasi setiap butir soal. Berikut adalah hasil yang diperoleh.

Tabel 8. Daya Pembeda Butir Soal

Nomor Soal	Indeks Diskriminasi	Kriteria
Soal 1	0.83	Baik Sekali
Soal 2	0.50	Baik
Soal 3	0.67	Baik
Soal 4	0.17	Buruk
Soal 5	0.50	Baik
Soal 6	0.50	Baik
Soal 7	0.17	Buruk
Soal 8	0.67	Baik
Soal 9	0.00	Buruk
Soal 10	0.17	Buruk
Soal 11	0.67	Baik
Soal 12	0.67	Baik
Soal 13	0.67	Baik
Soal 14	0.00	Buruk
Soal 15	0.67	Baik
Soal 16	0.67	Baik
Soal 17	0.33	Cukup
Soal 18	0.67	Baik
Soal 19	0.33	Cukup
Soal 20	0.67	Baik
Soal 21	0.17	Buruk
Soal 22	0.50	Baik
Soal 23	0.67	Baik
Soal 24	-0.67	Buruk Sekali
Soal 25	0.67	Baik
Soal 26	0.67	Baik
Soal 27	0.00	Buruk
Soal 28	0.50	Baik
Soal 29	0.67	Baik
Soal 30	0.67	Baik

Berdasarkan Tabel 8 di atas, diketahui bahwa terdapat 1 butir soal atau 3% soal yang dikategorikan memiliki daya pembeda yang baik sekali, yaitu pada nomor 1; terdapat 19 butir soal atau 63% soal yang dikategorikan memiliki daya pembeda baik, yaitu pada nomor 2, 3, 5, 6, 8, 11, 12, 13, 15, 16, 18, 20, 22, 23, 25, 26, 28, 29, dan 30; terdapat 2 butir soal atau 7% soal yang dikategorikan memiliki daya pembeda yang cukup, yaitu pada nomor 17 dan 19; terdapat 7 butir soal atau 23% soal yang dikategorikan memiliki daya pembeda buruk, yaitu pada nomor 4, 7, 9, 10, 14, 21, dan 27; dan terdapat 1 butir soal atau 3% soal yang dikategorikan memiliki daya pembeda buruk sekali, yaitu pada nomor 24. Sehingga dapat diinterpretasikan bahwa soal yang disusun secara keseluruhan memiliki daya pembeda yang rata-rata baik.

Berdasarkan hasil analisis daya pembeda item tes hasil belajar, berikut adalah tindak lanjut yang direkomendasikan: *Pertama*, catat dan simpan soal dengan daya pembeda cukup, baik, dan

baik sekali dalam Buku Bank Soal untuk digunakan pada tes selanjutnya. Soal-soal dengan kategori ini menunjukkan kualitas yang memadai dan layak digunakan kembali di masa depan. *Kedua*, perbaiki dan gunakan kembali soal dengan daya pembeda buruk atau buruk sekali. Soal-soal ini memiliki potensi untuk diperbaiki. Lakukan analisis mendalam untuk mengidentifikasi kekurangannya dan lakukan revisi agar daya pembedanya meningkat. Setelah diperbaiki, soal-soal ini dapat digunakan kembali dalam tes selanjutnya. Atau buang dan lupakan jika setelah dianalisis dan diperbaiki, daya pembeda soal masih belum menunjukkan hasil yang memuaskan, sebaiknya buang dan tidak gunakan kembali dalam tes. *Ketiga*, hindari soal dengan indeks diskriminasi negatif. Jangan gunakan dalam tes selanjutnya dan lakukan revisi menyeluruh untuk memperbaikinya (Susanto et al., 2015, p. 215).

Berdasarkan hasil dan diskusi di atas, setiap butir soal dapat diinterpretasikan sebagaimana pada tabel berikut.

Tabel 9. Keterangan Tindak Lanjut Setiap Butir Soal

Nomor Soal	Validitas	Kesukaran	Daya Pembeda	Keterangan
Soal 1	Valid	Sedang	Baik Sekali	Layak
Soal 2	Valid	Mudah	Baik	Revisi/Penghapusan
Soal 3	Valid	Sedang	Baik	Layak
Soal 4	Tidak Valid	Sedang	Buruk	Revisi/Penghapusan
Soal 5	Valid	Sedang	Baik	Layak
Soal 6	Valid	Mudah	Baik	Revisi/Penghapusan
Soal 7	Tidak Valid	Sedang	Buruk	Revisi/Penghapusan
Soal 8	Valid	Sedang	Baik	Layak
Soal 9	Tidak Valid	Sedang	Buruk	Revisi/Penghapusan
Soal 10	Tidak Valid	Sedang	Buruk	Revisi/Penghapusan
Soal 11	Valid	Sedang	Baik	Layak
Soal 12	Valid	Sedang	Baik	Layak
Soal 13	Valid	Sedang	Baik	Layak
Soal 14	Tidak Valid	Sedang	Buruk	Revisi/Penghapusan
Soal 15	Valid	Sedang	Baik	Layak
Soal 16	Valid	Sedang	Baik	Layak
Soal 17	Tidak Valid	Mudah	Cukup	Revisi/Penghapusan
Soal 18	Valid	Sedang	Baik	Layak
Soal 19	Tidak Valid	Sedang	Cukup	Revisi/Penghapusan
Soal 20	Valid	Sedang	Baik	Layak
Soal 21	Tidak Valid	Sedang	Buruk	Revisi/Penghapusan
Soal 22	Valid	Mudah	Baik	Revisi/Penghapusan
Soal 23	Valid	Sedang	Baik	Layak
Soal 24	Valid	Sedang	Buruk Sekali	Revisi/Penghapusan
Soal 25	Valid	Sedang	Baik	Layak
Soal 26	Valid	Sedang	Baik	Layak
Soal 27	Tidak Valid	Sedang	Buruk	Revisi/Penghapusan
Soal 28	Valid	Mudah	Baik	Revisi/Penghapusan
Soal 29	Valid	Sedang	Baik	Layak
Soal 30	Valid	Sedang	Baik	Layak

Berdasarkan Tabel 9 di atas, soal yang layak untuk digunakan kembali berjumlah 16 butir soal atau 53% dari keseluruhan butir soal, yaitu pada nomor 1, 3, 5, 8, 11, 12, 13, 15, 16, 18, 20, 23, 25, 26, 29, dan 30; dan soal yang harus diperbaiki atau dibuang berjumlah 14 butir soal atau 47% dari keseluruhan butir soal, yaitu pada nomor 2, 4, 6, 7, 9, 10, 14, 17, 19, 21, 22, 24, 27, dan 28.

Peneliti mengategorikan soal yang layak untuk digunakan kembali yaitu soal yang memenuhi unsur validitas, memiliki tingkat kesukaran yang sedang, dan memiliki daya pembeda

yang cukup atau baik atau baik sekali. Apabila butir soal tidak memenuhi salah satu aspek dari ketiga aspek tersebut, maka soal tersebut harus diperbaiki atau dibuang dan tidak digunakan lagi pada tes berikutnya.

Soal-soal yang tidak memenuhi ketiga unsur tersebut perlu diperbaiki atau dibuang karena dapat menurunkan kualitas tes artinya soal-soal yang tidak valid, terlalu mudah, atau tidak memiliki daya pembeda yang cukup dapat mendistorsi hasil tes dan memberikan informasi yang tidak akurat tentang kemampuan peserta didik. Dengan menggunakan kembali soal-soal yang berkualitas tinggi dan membuang atau memperbaiki soal-soal yang tidak berkualitas, diharapkan tes dapat memberikan informasi yang akurat tentang kemampuan peserta didik dan membantu meningkatkan kualitas pembelajaran.

4. Simpulan

Berdasarkan hasil dan diskusi di atas, peneliti telah melaksanakan tahapan penyusunan butir soal unsur bahasa Arab yang terdiri atas soal bunyi, soal kosakata, dan soal struktur kalimat sesuai dengan tahapan persiapan penilaian dalam Munip (2017). Tahapan yang dilakukan oleh peneliti meliputi spesifikasi tes, menulis soal tes, menelaah soal tes, melakukan uji coba tes, dan menganalisis butir soal. Berdasarkan hasil analisis butir soal, didapati bahwa (1) terdapat 70% butir soal yang tergolong valid dan 30% butir soal yang tergolong tidak valid, (2) tes yang disusun dikategorikan reliabel, (3) terdapat 17% butir soal yang dikategorikan mudah dan 83% butir soal yang dikategorikan sedang, dan (4) terdapat 3% butir soal yang dikategorikan memiliki daya pembeda yang baik sekali, terdapat 63% butir soal yang dikategorikan memiliki daya pembeda baik, terdapat 7% butir soal yang dikategorikan memiliki daya pembeda yang cukup, terdapat 23% butir soal yang dikategorikan memiliki daya pembeda buruk; dan terdapat 3% butir soal yang dikategorikan memiliki daya pembeda buruk sekali. Secara umum, terdapat 16 butir soal atau 53% dari keseluruhan butir soal yang layak untuk digunakan kembali dan terdapat 14 butir soal atau 47% dari keseluruhan butir soal yang harus diperbaiki atau dibuang. Peneliti menyadari keterbatasan penelitian ini. *Pertama*, pada tahapan menelaah soal tes, peneliti tidak mendapatkan penilaian dan pendapat dari pakar. Sehingga, peneliti merekomendasikan adanya penilaian dan pendapat dari pakar untuk penelitian selanjutnya. *Kedua*, penelitian ini terbatas hingga tahapan analisis butir soal, sehingga peneliti merekomendasikan penelitian lanjutan yang melengkapi keseluruhan tahapan termasuk tahapan perbaikan tes dan perakitan soal tes.

Daftar Pustaka

- Ainin, M. (2016). Kesahihan Dalam Penyusunan Tes Bahasa Arab Di Madrasah / Sekolah. *Prosiding Nasional Bahasa Arab 11*, 293–296. <http://prosiding.arab-um.com/index.php/konasbara/article/view/75/68>
- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI. *Edukatif: Jurnal Ilmu Pendidikan*, 6(1), 964–975. <https://doi.org/10.31004/edukatif.v6i1.5931>
- Asrul, Ananda, R., & Rosinta. (2014). Evaluasi Pembelajaran. In *Ciptapustaka Media*. Citapustaka Media.
- Fahmi, B., Rizqi, S., & Harmeilinda, N. E. (2022). Analisis Butir Soal Bahasa Arab MAS Pondok Pesantren Assalam Kampar Riau. *Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, 6(1), 95–105. <https://doi.org/10.15575/jpba.v6i1.16193>
- Faisal, A. (2022). Administrasi Tes dalam Evaluasi Pembelajaran. *Darussalam*, 23(1), 1–12. <https://doi.org/dx.doi.org/10.58791/drs.v23i1.175>
- Fauzi, S. (2023). Analisis Soal Tes Kompetensi Kebahasaan Pada PAS Bahasa Arab di Madrasah Tsanawiyah. *Ukazh: Journal of Arabic Studies*, 4(2), 635–848. <https://doi.org/10.37274/ukazh.v4i2.873>
- Firani Putri, & Supratman Zakir. (2023). Mengukur Keberhasilan Evaluasi Pembelajaran: Telaah Evaluasi Formatif Dan Sumatif Dalam Kurikulum Merdeka. *Dewantara: Jurnal Pendidikan Sosial Humaniora*, 2(4), 172–180.

- <https://doi.org/10.30640/dewantara.v2i4.1783>
Keputusan Menteri Agama Nomor 183 Tahun 2019 Tentang Kurikulum Pendidikan Agama Islam Dan Bahasa Arab (2019). <https://dki.kemenag.go.id/storage/files/3-191031102640-5dba54703adf2.pdf>
- Laili, M. (2020). Ketepatan Kontruksi Butir Pilihan Ganda Bahasa Arab. *ALSUNIYAT: Jurnal Penelitian Bahasa, Sastra, Dan Budaya Arab*, 3(2), 111–124. <https://doi.org/10.17509/alsuniyat.v3i2.25272>
- Masrukhin. (2020). *BAHASA ARAB MTs KELAS VIII*. Direktorat KSKK Madrasah.
- Munip, A. (2017). *Penilaian pembelajaran bahasa Arab*. FITK UIN Sunan Kalijaga.
- Nurkholis, N. (2019). Evaluasi Implementasi Kurikulum 2013 Mata Pelajaran Bahasa Arab di Madrasah Tsanawiyah (MTs). *Al-Fathin: Jurnal Bahasa Dan Sastra Arab*, 2(02), 233–258. <https://doi.org/10.32332/al-fathin.v2i02.1907>
- Suib, M., Saputra, D., Fidri, M., & Nurhayati. (2022). Strategi Pembelajaran Berdasarkan Unsur-Unsur Bahasa Arab. *Jurnal As-Said*, 2(1), 149–161. <http://e-journal.institutabdullahsaid.ac.id/index.php/AS-SAID/article/view/88>
- Suriani, N., Risnita, & Jailani, M. S. (2023). Konsep Populasi dan Sampling Serta Pemilihan Partisipan Ditinjau Dari Penelitian Ilmiah Pendidikan. *Jurnal IHSAN : Jurnal Pendidikan Islam*, 1(2), 24–36. <https://doi.org/10.61104/ihsan.v1i2.55>
- Susanto, H., Rinaldi, A., & Novalia, N. (2015). Analisis Validitas Reliabilitas Tingkat Kesukaran dan Daya Beda Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika Kelas XII IPS di SMA Negeri 12 Bandar Lampung Tahun Ajaran 2014/2015. *Al-Jabar : Jurnal Pendidikan Matematika*, 6(2), 203–218. <https://doi.org/10.24042/ajpm.v6i2.50>
- Tanjung, M. A. H. R., Fahmi, A. A., Rahmanita, F., Habibah, I. F., & Qomari, N. (2024). Analisis Butir Soal Penilaian Akhir Tahun Pelajaran Bahasa Arab Kelas VII MTs Al-Ma'arif Rakit Banjarnegara Jawa Tengah. *Mantiqutayr: Journal of Arabic Language*, 4(1), 347–367. <https://doi.org/10.25217/mantiqutayr.v4i1.4038>