

## Analisis Butir Soal Bahasa Arab kelas X SMA Muhammadiyah 2 Sidoarjo Berbasis Aplikasi Anates

### Analysis of Arabic Language Question for 10<sup>th</sup> class SMA Muhammadiyah 2 Sidoarjo Based on the Anates Application

Alfa Izzul Haq<sup>1)</sup>, Najih Anwar.\*<sup>2)</sup>

<sup>1)</sup>Program Studi Pendidikan Bahasa Arab, Universitas Muhammadiyah Sidoarjo, Indonesia

<sup>2)</sup>Program Studi Pendidikan Bahasa Arab, Universitas Muhammadiyah Sidoarjo, Indonesia.

\*Email Penulis Korespondensi: [najihanwar@umsida.ac.id](mailto:najihanwar@umsida.ac.id)

**Abstract.** *This research aims to analyze the quality of Arabic Language questions based on validity, reliability, level of difficulty and distinguishing power using the Anates V4.0.9 program. The method used in this research is descriptive quantitative with documentary data collection techniques. The document is in the form of multiple choice questions for the Mid-Semester Summative Assessment with a total of 20 questions with 30 students from class 10 at SMA Muhammadiyah 2 Sidoarjo as respondents. The data analysis technique uses Anates with the results obtained showing that the validity of the questions is 80%, this is meaningful (significant), the reliability of the questions is 0.87, the level of difficulty of the questions is medium and the discriminating power of the questions is 80% or as many as 16 items are capable of differentiate students who have high and low understanding of the material.*

**Keywords** - Learning evaluation, analysis of test items, Anates

**Abstrak.** *Penelitian ini bertujuan untuk menganalisis kualitas butir soal Bahasa Arab berdasarkan validitas, reliabilitas, tingkat kesulitan dan daya pembeda menggunakan program Anates V4.0.9. Metode yang digunakan dalam penelitian ini adalah kuantitatif deskriptif dengan teknik pengumpulan data secara dokumentatif. Dokumen berupa soal pilihan ganda penilaian Sumatif Tengah Semester sejumlah 20 butir soal dengan responden 30 siswa dari kelas 10 SMA Muhammadiyah 2 Sidoarjo. Teknik analisis data menggunakan Anates dengan hasil yang di peroleh menunjukkan bahwa validitas soal adalah 80%, hal ini bermakna (signifikan), reliabilitas soal yakni 0,87, taraf kesukaran soal sedang dan daya pembeda soal sebesar 80% atau sebanyak 16 butir soal yang mampu membedakan siswa yang memiliki pemahaman terhadap materi tinggi dan rendah.*

**Kata Kunci** - Evaluasi pembelajaran, Analisis butir penilaian, Anates

## I. PENDAHULUAN

Proses pembelajaran bahasa Arab yang baik dapat dilihat dari seberapa sering diadakannya evaluasi untuk meninjau sejauh mana pencapaian tujuan pembelajaran bahasa Arab tersebut, dengan demikian guru dapat mengetahui dengan cepat apa yang harus diperbaiki. Evaluasi yang dimaksud adalah pengukuran pembelajaran (measurement).[1] Dalam konteks ini, pengukuran pembelajaran bahasa Arab dapat menggunakan dua metode, yaitu pendekatan tes dan pendekatan non-tes. Dalam mu'jam ikhtiba>ra>t tes (ikhtiba>r) diartikan sebagai alat ukur yang yang mencakup rangsang yang disiapkan secara ilmiah dengan tujuan menentukan besar kecilnya sesuatu yang hendak diukur.[2] Tes mencakup ujian tertulis yang disebut *ikhtiba>ra>t kita>biyyah*, dan ujian lisan yang dikenal sebagai *ikhtibar at syafawiyah*. Ujian lisan dimanfaatkan untuk mengukur kemampuan siswa dalam berbicara, seperti tingkat kefasihan berbicara, pemilihan kata yang tepat, pemahaman dialek, dan respons siswa terhadap informasi yang diberikan. Sementara itu, ujian tertulis dimanfaatkan untuk mengevaluasi keterampilan bahasa lainnya, seperti kemampuan mendengarkan, membaca, menulis, dan pemahaman tata bahasa.[3]

Tes tulis menurut syekh Ali al-huli dalam kitabnya *asa>li>b tadri>s al-lughah* adalah tes yang di berikan kepada siswa dengan pertanyaan dan jawaban berupa tulisan.[4] Tes tulis merupakan metode yang secara tidak langsung menilai kemampuan siswa melalui tanggapan mereka terhadap berbagai pertanyaan yang disusun. Oleh karena itu, sangat penting untuk merencanakan butir soal dengan cermat untuk menghindari kesalahan, sehingga instrumen tersebut memiliki kualitas yang baik. Dalam pembelajaran bahasa Arab, tes bisa diklasifikasikan menjadi empat kategori berdasarkan fungsinya, yakni: 1) Tes penempatan, 2) Tes diagnostik, 3) Tes formatif, dan 4) Tes sumatif.[5] Dari keempat pembagian tes tersebut terdapat tes sumatif, tes suamtif merupakan tes yang dilakukan setiap berakhirnya pemberian suatu program pembelajaran.[6] Tes ini dilakukan untuk monitoring perkembangan siswa setiap pertengahan maupun akhir semester guna mengetahui sejauh mana pencapaian siswa dalam pelajaran tertentu. [7]

Pembuatan butir soal bahasa Arab yang baik harus memperhatikan beberapa hal terkait mutu butir soal tersebut, yakni: *pertama* adalah validitas (Al-Si>dqu), yaitu seberapa tepatkah instrument penilaian dapat mengukur

apa yang dikehendaki untuk di ukur. Yang *kedua* adalah reliabilitas (Tsaba>t Al-as'ilah), yang mengacu pada seberapa handal butir soal tersebut ketika digunakan berulang kali, suatu butir soal dianggap memiliki reliabilitas jika hasilnya konsisten dan tidak berfluktuasi secara signifikan ketika diberikan kepada populasi yang sama. Reliabilitas sangat penting dalam penilaian karena membantu memastikan bahwa hasil tes adalah representasi yang konsisten dan dapat diandalkan dari karakteristik atau kemampuan yang sedang diukur, maksudnya apabila butir soal memiliki reliabilitas yang tinggi maka hasil butir soal apabila di ujikan dua kali dalam waktu berdekatan maka tidak menunjukkan perbedaan hasil yang signifikan.[8] *Ketiga*, tingkat kesulitan (Mustawa> as-su'u>bah) digunakan untuk membedakan antara butir soal yang mudah dan sulit. Jika sebuah pertanyaan menunjukkan tingkat kesulitan yang seimbang atau proporsional, itu mengindikasikan bahwa kualitas dari pertanyaan tersebut baik. *Empat*, daya beda (Mustawa< at-tamyi<z) Kemampuan daya pembeda suatu pertanyaan merujuk pada kapasitasnya untuk memisahkan siswa yang menunjukkan tingkat pemahaman yang berbeda. Kemampuan daya pembeda pertanyaan ini berkaitan dengan sejauh mana pertanyaan tersebut mampu membedakan siswa yang memiliki pemahaman tinggi dan yang memiliki pemahaman rendah. Dengan kata lain, pertanyaan yang memiliki kemampuan daya pembeda yang baik akan dengan tegas memisahkan antara siswa-siswa yang sepenuhnya memahami materi atau memiliki kinerja yang unggul dengan mereka yang memiliki pemahaman yang terbatas atau kinerja yang rendah.[9] Keempat indikator diatas digunakan untuk menganalisis butir soal, manfaatnya adalah untuk mengetahui apakah butir soal yang digunakan dalam instrumen tes sesuai atau tidak, serta sejauh mana butir soal tersebut dapat mengukur kemampuan siswa dengan akurat.[10]

Terdapat beberapa program komputer yang dapat digunakan menganalisis butir soal antara lain Anates. Anates merupakan sebuah perogram komputer yang dirancang khusus untuk melakukan evaluasi pada butir soal yang terdiri dari penilaian pilihan ganda dan penilaian uraian. Aplikasi ini berguna dalam membantu guru menilai kualitas butir soal yang mereka buat. Dengan penggunaan Anates, guru dapat mengevaluasi aspek-aspek seperti validitas, reliabilitas, tingkat kesulitan, kemampuan untuk membedakan antara peserta yang mampu dan tidak, serta dampak dari masing-masing penilaian.[11] Kelebihan utama dari Anates adalah dari segi tampilannya yang simple dan penggunaannya yang mudah memungkinkan guru yang kurang. Selain itu, data yang dikumpulkan oleh Anates dapat segera diolah, memungkinkan guru untuk dengan cepat memperoleh hasil seperti reliabilitas tes, pengelompokan penilaian yang baik dan buruk, analisis tingkat kesulitan, korelasi antara hasil pada masing-masing penilaian dengan hasil keseluruhan, serta kualitas pilihan jawaban yang diberikan. Anates memudahkan pengguna dalam melaksanakan tugas analisis ini. Manfaat Anates mirip dengan perangkat analisis penilaian lainnya, tetapi lebih dikenal dalam hal kemudahan pengoprasian dan keakuratannya. Perangkat ini sangat berguna dalam mengkaji data pada tes pilihan ganda yang diujikan, dan versi 4.0.9 Anates bahkan memungkinkan guru untuk secara praktis dan cepat memeriksa jawaban yang benar dan salah pada setiap penilaian. Kesimpulannya, Anates merupakan alat yang sangat membantu bagi guru dalam mengevaluasi kualitas tes yang mereka susun. Namun, masih banyak guru yang melakukan analisis butir soal secara manual, yang mengakibatkan proses analisis tersebut tidak efisien waktu dan memiliki tingkat akurasi yang rendah. Dengan menggunakan program komputer seperti Anates, guru dapat memperpendek waktu yang diperlukan dalam proses analisis dan meningkatkan keakuratan hasil analisis mereka.[12]

Beberapa penelitian yang relevan dengan topik ini, seperti yang dilakukan oleh Raudhatul Husna pada tahun 2022 dalam studinya berjudul "Tahli>lu Bunu>di as'ilah Al-lughah Al-arabiah fi> madrasah Aliyah bi Ma'had Jeumala Amal Pidie Jaya". Dalam hasil penelitian yang menggunakan Anates, didapati bahwa validitas butir soal mencapai (56%), reliabilitas mencapai (0,90), tingkat kesukaran butir soal terbagi menjadi tingkat mudah (36%), tingkat sedang (32%), dan tingkat kesulitan tinggi (36%). Butir soal juga dinilai cukup baik dalam membedakan antara kemampuan siswa. Dengan demikian, dapat disarankan bahwa penilaian-penilaian tersebut layak digunakan dalam konteks evaluasi pembelajaran. [13]

Penelitian kedua, yang dilaksanakan oleh Elviana pada tahun 2020 dalam studi yang berjudul "Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates," mengungkapkan bahwa tingkat daya pembeda butir soal sangat rendah, hanya sekitar 20% dari seluruh butir penilaian. Kesulitan penilaian cenderung berada pada tingkat sedang, dengan sekitar 64% butir penilaian termasuk dalam kategori tersebut. Validitas tes hanya mencapai 0,24, menunjukkan tingkat validitas yang rendah sekitar 24%, yang signifikan. Reliabilitas tes juga dinilai rendah, dan hanya sekitar 32% butir soal yang berhasil dengan baik dalam menguji kemampuan siswa.[14]

Penelitian ketiga, yang dilakukan oleh Deni Maulana dan Anwar Sanusi pada tahun 2020 dalam studi berjudul "Evaluasi Butir Penilaian Bahasa Arab Ujian Akhir Madrasah Bersama Daerah (Uambd) Madrasah Ibtidaiyah Tahun 2017-2018," memanfaatkan aplikasi Anates sebagai alat analisis. Hasil penelitian ini mengungkapkan temuan yang signifikan. Validitas butir soal sangat tinggi, mencapai 100%, menunjukkan bahwa butir penilaian tersebut sangat efektif dalam mengukur kompetensi yang dikehendaki. Reliabilitas butir soal juga mencapai tingkat konsistensi yang tinggi, dengan nilai tes sebesar 0,68, berada dalam rentang yang baik, yaitu antara 0,60 hingga 0,80. Tingkat kesulitan butir soal berkisar sekitar 63,23%, dapat diartikan bahwa sebagian besar penilaian mempunyai tingkat kesulitan yang moderat. Kemampuan butir soal untuk membedakan antara siswa juga ditemukan pada tingkat yang dapat diterima, sekitar 31,4%, meskipun terdapat potensi perbaikan sebesar 25,7%. Selain itu, dalam hal kualitas

pilihan jawaban, ditemukan bahwa ada 11 butir penilaian dengan penyebaran jawaban yang sangat baik, 5 butir soal dengan kualitas baik, 6 butir penilaian dengan kualitas kurang baik, dan empat butir soal dengan kualitas buruk.[15]

Berdasarkan tinjauan penelitian terdahulu yang telah diuraikan sebelumnya, dapat disimpulkan bahwa penelitian ini memiliki fokus yang serupa, yaitu tentang penggunaan alat evaluasi instrumen penilaian dan pemanfaatan aplikasi Anates. Meskipun demikian, perbedaan antara penelitian ini dan penelitian sebelumnya terletak pada subjek dan objek penelitiannya. Subjek dalam penelitian ini adalah siswa kelas X SMA Muhammadiyah 2 Sidoarjo, sementara objek penelitiannya adalah jawaban siswa pada soal bahasa Arab. Hasil pengamatan awal menunjukkan bahwa guru bahasa Arab di SMA Muhammadiyah 2 Sidoarjo telah melakukan analisis instrumen penilaian secara rutin, namun belum menerapkan aplikasi praktis seperti Anates dalam proses analisis tersebut. Dengan mengadopsi aplikasi seperti Anates, proses analisis instrumen penilaian dapat menjadi lebih efektif, efisien, dan tentu saja lebih akurat.

Rumusan masalah dalam penelitian ini adalah, bagaimana kualitas butir soal penilaian bahasa Arab untuk siswa kelas X di SMA Muhammadiyah 2 Sidoarjo dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dengan menggunakan program Anates?.[16] Tujuan penelitian ini adalah untuk mengetahui kualitas butir soal berdasarkan validitas, reliabilitas, tingkat kesulitan dan daya pembeda dengan program Anates versi 4.0.9.

## II. METODE

Penelitian ini menggunakan pendekatan deskriptif kuantitatif, pendekatan deskriptif digunakan untuk merincikan ciri-ciri suatu peristiwa yang sedang berlangsung selama penelitian berjalan, serta untuk menyelidiki faktor-faktor yang mempengaruhi fenomena tertentu tersebut. Dalam esensinya, metode deskriptif bertujuan untuk mengumpulkan informasi yang mencerminkan situasi yang sebenarnya saat penelitian dilakukan.[17] Di sisi lain, metode penelitian kuantitatif merupakan jenis penelitian yang berfokus pada penggunaan data dalam bentuk angka, mulai dari proses pengumpulan data, analisis data, hingga penyajian data. Pendekatan kuantitatif menekankan pada analisis data yang didasarkan pada angka-angka, yang selanjutnya dianalisis menggunakan metode statistik yang sesuai.[18]

Penelitian ini dilakukan di SMA Muhammadiyah 2 Sidoarjo, lokasi penelitian ini dipilih karena relevan dengan tujuan penelitian. Subjek penelitian ini adalah 30 siswa kelas X yang diambil secara acak dari 3 kelas yang berbeda. Objek penelitian mencakup jawaban siswa pada soal mata pelajaran Bahasa Arab tahun ajaran 2023/2024, termasuk penilaian dan kunci jawaban dari seluruh penilaian tersebut. Teknik dokumentasi dalam kajian ini dimanfaatkan sebagai metode pengumpulan data, memungkinkan pengumpulan data berupa penilaian dan kunci jawaban, serta hasil ujian dari seluruh siswa kelas X. Data dalam penelitian ini dianalisis secara kuantitatif dengan menggunakan perangkat lunak Anates versi 4.0.9.[19]

## JEES (JOURNAL OF ENGLISH EDUCATORS SOCIETY)

HOME ABOUT USER HOME CATEGORIES SEARCH CURRENT ARCHIVES ANNOUNCEMENTS PUBLICATION ETHICS STATEMENT EDITORIAL BOARD  
COPYRIGHT NOTICE INDEXING SERVICE FOCUS AND SCOPE AUTHOR GUIDELINES

Home > Vol 1, No 2 (2016) > Megenawati

### TERTIARY LEVEL EXCHANGE STUDENTS' PERSPECTIVES ON SELF-EFFICACY: TOWARD EFL WRITING

Fita Megenawati

#### ABSTRACT

This study aims to describe the students' self-efficacy on their writing competence. Descriptive study was implemented by distributing closed-ended questionnaires in addition to interview and the result of writing task. The subjects of this study were three students from Thailand. The students' responses in questionnaire were analyzed through frequency distributions and percentage. For the result of interview, it was transcribed in verbatim form and used coding techniques to classify the relevant points. The result of writing task became the supplementary data to confirm the findings and support conclusion. In a nutshell, the subjects of this study have moderate level of writing self-efficacy. Each student showed diverse selection in writing stage. The first student had moderate self-efficacy, but he relatively could cope with the writing problems. In the second student, the writing self-efficacy was the highest one, and it was proven from her better writing result. For the last student, similar to the writing quality, he considered himself weak in this skill.

#### KEYWORDS

EFL writing, self-efficacy, tertiary level

#### FULL TEXT:

PDF

#### REFERENCES

- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191.
- Brewer-Sandera, K. E., Swanson, E., & Tarry, R. M. (2002). Preliminary proficiency guidelines—Writing revised 2001. 1X(1), 9–15.
- Heath, P., Sullivan, J. R., & Guerra, N. S. (2007). Self-Efficacy and Goal Orientation. *Journal of Advanced Academics*, 18(3), 454–476. Retrieved from <http://files.eric.ed.gov/fulltext/EJ773183.pdf>
- Kalasin, B. (2009). Teaching EFL to the Generation: A survey of using YouTube as supplementary material with college EFL students in Thailand. *Call-EJ Online*, 10(2).
- Kerpelman, J. L., & Mosier, L. S. (2004). Rural African American Adolescents' Future Orientation: The Importance of Self-Efficacy, Control and Responsibility, and Identity Development. *Adolescence*, 40(2), 187–208. [https://doi.org/10.1207/s1532706xaj4002\\_5](https://doi.org/10.1207/s1532706xaj4002_5)
- Lenden, B., Chalmers, L. R., & Lenden, B. (2009). Language and experience: Evidence from the blind child. 6.
- Lisambriah, E. A., & Piatnick, P. R. (2005). THE ROLE OF SELF-EFFICACY BELIEFS IN STUDENT ENGAGEMENT AND LEARNING IN THE CLASSROOM. *Reading & Writing Quarterly*, 19(2), 119–137. <https://doi.org/10.1080/1077560308222>
- Mavrou, C. A. (2009). Using YouTube to Encourage Authentic Writing in EFL Classrooms. 1–12. *TESOL Reporter*, 42(11), 1–12.

ARTICLE TOOLS
Print this article
Indexing metadata
How to cite item
Finding References
Review policy
Email this article
Email the author
Post a Comment
ABOUT THE AUTHOR
Fita Megenawati Universitas Muhammadiyah Sidoarjo Indonesia
RELATED ITEMS
Show all
Journal Help
USER
You are logged in as... transilmalassar My Journals My Profile Log Out Log Out as User
NOTIFICATIONS
View (1 new) Manage
JOURNAL CONTENT

**Gambar 1.** Contoh gambar atau ilustrasi [1]**Tabel 1.** Contoh tabel [2]

No	Nama Jurnal	Fakultas	Prodi
1	JEES	FKIP	PBI
2	SEJ	FKIP	FKIP
3	Pedagogia	FKIP	FKIP
4	Rechtsidee	HUKUM	HUKUM

Jika penulis lebih dari satu, semua nama penulis dituliskan dengan dipisahkan oleh koma (.). Jika nama penulis terdiri dari dua kata, kata pertama penulis (*first name*) sebaiknya tidak disingkat. Jika nama penulis hanya terdiri dari satu kata, nama sebenarnya dituliskan dalam satu kata. Namun, di versi online (HTML) nama penulis yang hanya satu kata perlu dituliskan dalam dua kata yang berisi nama yang sama (berulang) untuk keperluan indeksasi dan metadata.

Jika penulis mempunyai lebih dari satu afiliasi, afiliasi tersebut dituliskan secara berurutan. Tanda *superscript* berupa nomor yang diikuti tanda tutup kurung, misalnya <sup>1)</sup>, diberikan di belakang nama penulis (lihat contoh). Jika semua penulis berasal dari satu afiliasi, tanda ini tidak perlu diberikan.

### III. HASIL DAN PEMBAHASAN

Pada bagian ini berisi pemaparan hasil penelitian yaitu terkait validitas, reliabilitas, tingkat kesukaran dan daya pembeda pada butir soal bentuk pilihan ganda pada soal bahasa Arab SMA Muhammadiyah 2 Sidoarjo dengan menggunakan Anates.

#### A. Validitas

Naskah manuskrip yang sudah memenuhi petunjuk penulisan UMSIDA Preprints Server (dalam format MS Word/Openoffice Writer) harus dikirimkan melalui *Online Submission System* di portal archive UMSIDA Preprints Server (<https://archive.umsida.ac.id>) setelah mendaftarkan sebagai Penulis di bagian "*Register*". Penulis diharapkan menggunakan *template* yang telah disediakan. Petunjuk pengiriman manuskrip secara daring dapat dilihat di bagian Petunjuk Submit Online di dokumen ini dan dari situs UMSIDA Preprints Server. Naskah manuskrip yang tidak sesuai petunjuk penulisan UMSIDA Preprints Server akan dikembalikan ke Penulis terlebih.

Penilaian butir soal dilakukan dengan membandingkan skor setiap butir dengan skor total. Proses ini melibatkan penggunaan teknik korelasi product moment. Bila terdapat korelasi yang positif dan signifikan antara skor butir dan skor total, hal ini menunjukkan bahwa butir tersebut dianggap valid.[20]

Catatan: Batas signifikansi koefisien korelasi sebagai berikut:

df (N-2)	P=0,05	P=0,01	df (N-2)	P=0,05	P=0,01
10	0,576	0,708	60	0,250	0,325
15	0,482	0,606	70	0,233	0,302
20	0,423	0,549	80	0,217	0,283
25	0,381	0,496	90	0,205	0,267
30	0,349	0,449	100	0,195	0,254
40	0,304	0,393	125	0,174	0,228
50	0,273	0,354	>150	0,159	0,208

Bila koefisien = 0,000 berarti tidak dapat dihitung.

**Gambar 1.** Batas signifikansi koefisien korelasi penilaian

Pada soal bahasa Arab terdapat sebanyak 20 butir soal pilihan ganda, sehingga dapat diketahui batas signifikansi koefisiennya berada direntang 0,423 sampai 0,549 maka hasilnya signifikan, apabila nilai lebih dari 0,549 maka hasilnya sangat signifikan sesuai dengan gambar 1. Dapat disimpulkan apabila nilai korelasinya berada dibawah 0,423 maka penilaian dinyatakan tidak signifikan atau tidak valid. Hasil analisa validitas butir soal pada aplikasi Anates adalah sebagai berikut;

**Tabel 1.** Hasil uji validitas dengan Anates

No	Nilai korelasi	Tingkat signifikansi (validitas)	catatan
1.	0.715	Sangat signifikan	Terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang dhomir muttashil.
2.	0.703	Sangat signifikan	Terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang dhomir muttashil.
3.	0.629	Sangat signifikan	Terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang dhomir muttashil.
4.	0.475	Signifikan	Terdapat korelasi positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang dhomir muttashil.
5.	0.578	Sangat signifikan	Terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang macam-macam dhomir.
6.	0.611	Sangat signifikan	Terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang dhomir munfashil.
7.	0.675	Sangat signifikan	terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang menerjemahkan kalimat bahasa Arab kedalam bahasa indonesia.
8.	0.138	Tidak signifikan	Tidak adanya korelasi positif dan signifikan antara skor item dengan skor total pada soal yang

9.	0.597	Sangat signifikan	berkaitan tentang menerjemahkan kalimat bahasa Arab kedalam bahasa Indonesia. Terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang dhomir munfashil.
10.	0.590	Sangat signifikan	Terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang dhomir munfashil.
11.	0.684	Sangat signifikan	terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang menyusun sebuah kalimat dari kata yang terurai.
12.	0.683	Sangat signifikan	terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang menyusun sebuah kalimat dari kata yang terurai
13.	0.556	Sangat signifikan	terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang menyusun sebuah kalimat dari kata yang terurai.
14.	0.454	Signifikan	terdapat korelasi positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang menyusun sebuah kalimat dari kata yang terurai.
15.	0.359	Tidak signifikan	Tidak adanya korelasi positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang menyusun sebuah kalimat dari kata yang terurai.
16.	0.593	Sangat signifikan	terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total yang berkaitan tentang mengartikan mufrodat.
17.	0.482	signifikan	Terdapat korelasi positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang mengartikan mufrodat.
18.	0.673	Sangat signifikan	terdapat korelasi yang sangat positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang dhomir muthasil.
19.	0.342	Tidak signifikan	Tidak adanya korelasi positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang dhomir muthasil.
20.	0.216	Tidak signifikan	1. Tidak adanya korelasi positif dan signifikan antara skor item dengan skor total pada soal yang berkaitan tentang benar salah. 2. Untuk soal berbentuk benar salah sebaiknya menggunakan bahasa Arab صحيح dan خطأ agar murid terbiasa dengan bahasa Arab.

**Catatan:** Hasil validitas soal dapat berbeda-beda karena berbagai faktor yang mempengaruhi kualitas soal seperti reliabilitas, tingkat kesukaran dan daya pembedanya. Semakin baik kualitas empat aspek tersebut dalam suatu soal maka semakin tinggi pula nilai validitasnya.[21]

Berdasarkan tabel 1, pada hasil analisis butir soal bahasa Arab didapati 16 butir soal atau apabila dijadikan bentuk presentase yaitu sebesar 80% dari total soal dinyatakan sebagai soal yang valid dengan koefisien diatas 0,423, dan 4 butir soal atau sebesar 20% dinyatakan tidak valid dengan koefisien dibawah 0,423. 4 butir soal yang dinyatakan tidak valid yaitu nomor 8 yang hanya memiliki validitas sebesar 0,138, nomor 15 memiliki 0,359, nomor 19 memiliki validitas 0,342 dan butir soal nomor 20 hanya memiliki validitas 0,216. Rendahnya validitas pada soal tersebut

menunjukkan bahwa soal tersebut sangat kurang dalam mengukur kemampuan siswa yang seharusnya ingin diukur oleh guru melalui soal tersebut. Dengan kata lain keempat butir soal tersebut harus dikaji ulang oleh guru atau perlu diganti sebab memiliki validitas yang rendah. [22]

## B. Reliabilitas

Reliabilitas butir soal mencerminkan sejauh mana butir soal tersebut konsisten dalam mengukur dimensi yang seharusnya diukur, sehingga hasil pengukuran dapat dianggap dapat dipercaya. Suatu butir soal dianggap memiliki reliabilitas tinggi jika hasil pengukuran pada percobaan pertama dan kedua menunjukkan kesamaan yang signifikan, sedangkan reliabilitas dianggap rendah jika hasil antara percobaan pertama dan kedua memiliki perbedaan yang besar.[23] Metode untuk menghitung reliabilitas butir soal ada tiga cara, yaitu: Pertama, metode bentuk paralel (equivalent). Kedua, metode tes ulang (Test- retest method). Ketiga, metode belah dua (Split-half method). Pada aplikasi Anates menggunakan metode yang ketiga yaitu metode belah dua (Split-half method), Metode belah dua dibuat untuk mengatasi kekurangan yang timbul pada metode paralel dan metode uji ulang karena metode ini memungkinkan estimasi reliabilitas tanpa perlu mengadakan pengujian dua kali.[22] Metode belah dua terbagi menjadi dua yaitu pembelahan ganjil-genap dan pembelahan awal dan akhir, setelah metode pembelahan dilakukan hanya diketahui reliabilitas separuh tes. dan untuk mengetahui reliabilitas seluruh tes maka menggunakan rumus Spearman-Brown.[21] Contoh pembelahan ganjil genap:

langkah pertama:

o	no	Nama	Item ganjil (X)	Item genap (Y)
1	1	Yuni	5	5
2	2	Yuri	4	3
3	3	Yura	4	4
4	4	Yuta	3	2
5	5	Yuzi	3	3
6	6	Yuli	4	0
7	7	Yuki	4	3
8	8	Yumi	3	5

langkah kedua

o	no	Nama	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
1	1	Yuni	5	5	25	25	25
2	2	Yuri	4	3	16	9	12
3	3	Yura	4	4	25	16	20
4	4	Yuta	3	2	9	4	6
5	5	Yuzi	3	3	9	9	9
6	6	Yuli	4	0	16	0	0
7	7	Yuki	4	3	16	9	12
8	8	Yumi	3	5	9	25	15
			$\Sigma X=31$	$\Sigma Y=25$	$\Sigma X^2=125$	$\Sigma Y^2=97$	$\Sigma XY=99$

Rumus belah dua:

$$\begin{aligned}
 r_{xy} &= \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{(N \sum X^2 - (\sum X)^2)\} \{N \sum Y^2 - (\sum Y)^2\}}} \\
 &= \frac{8 \times 99 - (31)(25)}{\sqrt{\{(8 \times 125 - (31)^2)\} \{(8 \times 97 - (25)^2)\}}} \\
 &= \frac{792 - 775}{\sqrt{(1000 - 961)(776 - 625)}} \\
 &= \frac{17}{\sqrt{(39)(151)}} \\
 &= \frac{17}{76,74} \\
 r_{xy} &= 0,22
 \end{aligned}$$

Setelah koefisien korelasi diperoleh maka selanjutnya dihitung dengan formula Spearman-Brown sebagai berikut:

$$\begin{aligned}
 r_{11} &= \frac{2r \frac{1}{2} \frac{1}{2}}{1 + r \frac{1}{2} \frac{1}{2}} \\
 r_{11} &= \frac{2 \times 0,22}{1 + 0,22} \\
 &= 0,36
 \end{aligned}$$

Untuk menentukan kategori reabilitas maka perlu adanya indeks untuk mengklasifikasikannya, berikut adalah indeks reliabilitas penilaian:

**Tabel 2.** Indeks Reliabilitas dan kategorinya

Indeks	Kategori
0,90 – 1,00	Amat tinggi
0,70 – 0,89	Tinggi
0,50 – 0,69	Sedang
0,30 – 0,49	Rendah
< 0,30	Amat rendah

Berikut hasil uji reliabilitas pada butir soal:



Rata2= 13.07  
 Simpang Baku= 4.78  
 KorelasiXY= 0.78  
 Reliabilitas Tes= 0.87  
 Nama berkas: E:\UMSIDA\TUGAS AKHIR\REFERENSI\REFERENSI BIMBINGAN PERTAMA\DATA HASIL UJIAN SAMAMDA.ANA

No.Urut	No. Subyek	Kode>Nama Subyek	Skor Ganjil	Skor Genap	Skor Total
1	1	Faizah atha n...	6	8	14
2	2	muhammad agas...	0	4	4
3	3	gege miftah r...	2	1	3
4	4	radix ahmad l...	2	1	3
5	5	rega nararya ...	1	3	4
6	6	m djibril A B	8	9	17
7	7	mevlana brahm...	6	7	13
8	8	utiy hasna	6	8	14
9	9	faizal maulan...	6	9	15
10	10	arya giriputra	7	8	15
11	11	M. bayu priyo...	6	5	11
12	12	alfatir prama...	4	7	11
13	13	ryan anantama...	5	9	14
14	14	gusti adrian ...	4	4	8
15	15	nafeeza reswara	9	8	17
16	16	Thirza Anindy...	10	9	19
17	17	Alyn Aurelia ...	10	9	19
18	18	Aldofan Mahen...	9	9	18
19	19	Nadhif Sidqi ...	9	9	18
20	20	Achmad Trista...	9	8	17
21	21	Chandra Khali...	4	5	9
22	22	Firdaus Nuzul...	5	5	10
23	23	Zaskia Zahrot...	6	7	13
24	24	Aghnia Zharfa...	8	7	15
25	25	Adelina Aji p...	8	6	14
26	26	Annisa farhat...	8	8	16
27	27	Fahrizal fidz...	4	7	11
28	28	Farah hanun a...	10	8	18
29	29	Ikhwan dwi ha...	8	8	16
30	30	Iqbal firdaus	9	7	16

**Gambar 2.** Hasil uji reliabilitas penilaian dengan Anates

Gambar 2 menunjukkan hasil yang didapatkan dari aplikasi Anates, diketahui bahwa reliabilitas butir soal sebesar 0,87. Apabila disesuaikan dengan pengkategorian reliabilitas pada tabel 2 diatas maka hasil uji reliabilitas tes tersebut dikategorikan tinggi karena termasuk dalam rentang 0,70 – 0,89. Dari hasil tersebut dapat dikatakan bahwa soal bahasa Arab mempunyai koefisiensi yang baik dan handal dalam pengukuran kemampuan siswanya sebab reliabilitasnya ada pada kategori **Tinggi**, hal ini menunjukkan bahwa soal tersebut layak untuk digunakan.[24]

### C. Tingkat Kesukaran

Tingkat kesukaran suatu butir soal dapat diukur melalui perbandingan antara jumlah peserta yang memberikan jawaban benar dengan total peserta. Artinya, semakin banyak peserta tes yang mampu menjawab butir soal dengan benar maka indeksnya akan semakin tinggi, menunjukkan bahwa butir soal tersebut lebih mudah. Sebaliknya, jika sedikit peserta tes yang dapat menjawab butir soal dengan benar, maka indeksnya akan rendah dan menunjukkan bahwa tingkat kesukaran butir soal tersebut akan semakin tinggi. Hasil butir soal yang terlalu mudah akan membuat siswa tidak terangsang kemampuannya dalam menjawab soal tersebut, begitupun sebaliknya apabila butir soal terlalu sukar maka akan membuat siswa mudah berputus asa dalam menjawabnya karena di luar kemampuannya. Berikut adalah hasil analisa tingkat kesukaran pada penilaian Sumatif tengah semester:

TINGKAT KESUKARAN					
=====					
Jumlah Subyek= 30					
Butir Soal= 20					
Nama berkas: E:\UMSIDA\TUGAS AKHIR\REFERENSI\REFERENSI BIMBINGAN PERTAMA\DATA HASIL UJIAN SAMAMDA.ANA					
No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran(%)	Tafsiran	
1	1	22	73.33	Mudah	
2	2	20	66.67	Sedang	
3	3	19	63.33	Sedang	
4	4	18	60.00	Sedang	
5	5	23	76.67	Mudah	
6	6	26	86.67	Sangat Mudah	
7	7	21	70.00	Sedang	
8	8	2	6.67	Sangat Sukar	
9	9	25	83.33	Mudah	
10	10	26	86.67	Sangat Mudah	
11	11	14	46.67	Sedang	
12	12	22	73.33	Mudah	
13	13	14	46.67	Sedang	
14	14	15	50.00	Sedang	
15	15	18	60.00	Sedang	
16	16	24	80.00	Mudah	
17	17	19	63.33	Sedang	
18	18	25	83.33	Mudah	
19	19	14	46.67	Sedang	
20	20	25	83.33	Mudah	

**Gambar 3.** Hasil uji tingkat kesukaran dengan Anates

Untuk hasil uji tingkat kesukaran berupa presentase sehingga perlu diubah kedalam bentuk desimal, agar mempermudah pembacaan data maka hasil tersebut dikelompokkan berdasarkan indeksnya:

Indeks	Penilaian	jumlah
P 0,00 – 0,30 (Sukar)	8	1
P 0,31 – 0,70 (Sedang)	2,3,4,7,11,13,14,15 ,17,19	10
P 0,71 – 1,00 (Mudah)	1,5,6,9,10,12,16,18,20	9

Dari tabel 3 pada bagian kiri terdapat indeks tingkat kesukaran yang dibagi menjadi 3 yaitu sukar, sedang, dan mudah, dari hasil uji tingkat kesukaran tersebut diketahui butir soal sukar sebanyak 1 soal, kemudian soal yang sedang sebanyak 10 butir soal dan butir soal yang mudah sebanyak 9 soal. Dari hasil tersebut dapat disimpulkan bahwa tingkat kesukaran yang digunakan dalam butir soal dinyatakan masih belum proporsional, sebab dalam menyusun butir soal pada ujian sebaiknya butir soal yang digunakan mempunyai tingkat kesukaran yang berimbang atau proporsional. Terdapat beberapa perbandingan yang dapat digunakan sebagai ukuran dalam mengukur proporsionalnya suatu soal. Ada yang menggunakan 1:3:1, apabila jumlah soal sebanyak 5 nomor maka perbandingannya satu persen soal mudah, tiga persen soal sedang dan satu persen soal sukar. Ada juga yang memakai ukuran 3:4:3, jika soal sebanyak 10 maka tiga persen soal mudah, empat persen sedang dan tiga persen sukar. Ada juga yang memakai ukuran 3:5:2, tiga persen soal mudah, lima persen soal sedang dan dua persen soal sukar.[25]

Dari pernyataan tersebut diketahui bahwa tingkat kesukaran soal tersebut tidak menunjukkan jumlah yang proporsional. yaitu 10 soal berkategori sedang, 1 soal berkategori sukar dan 9 soal berkategori mudah. Apabila kita merujuk pada pernyataan paragraf sebelumnya bahwa soal dikatakan proporsional apabila sekiranya 1:3:1, 3:5:2

maupun 3:4:3. Berbeda dengan soal yang tingkat kesukarannya sedang apabila kita menggunakan skala 3:5:2 maka sudah termasuk kedalam kategori soal yang proporsional sehingga tidak perlu ada perbaikan, akan tetapi secara garis besar soal masih menunjukkan proporsi tingkat kesukaran yang masih kurang sehingga perlu diperbaiki lagi. Banyaknya soal mudah dikarenakan banyak soal yang dengan mudah dijawab oleh sebagian besar siswa tanpa adanya kesulitan, berbeda dengan jumlah butir soal sukar yang hanya menunjukkan presentase yang sangat kecil, padahal jumlah butir soal sukar yang proporsional mampu meningkatkan kemampuan berpikir siswa.

#### D. Daya Pembeda

Daya beda butir soal adalah indikator yang menggambarkan sejauh mana butir soal mampu membedakan antara siswa yang memiliki penguasaan materi yang tinggi (klompok atas) dan rendah (kelompok bawah).[26] Hal ini tercermin dalam fakta bahwa siswa yang berpemahaman tinggi cenderung memberikan jawaban yang benar lebih sering pada butir soal tersebut, sedangkan siswa yang berpemahaman rendah cenderung memberikan jawaban yang kurang tepat. Tujuan menilai daya pembeda adalah untuk mengukur kinerja butir soal dan menentukan apakah penilaian tersebut termasuk dalam kategori skor tinggi atau skor rendah dalam keseluruhan tes. Selain itu, daya pembeda butir juga mencakup kemampuan butir soal untuk membedakan kelompok berdasarkan aspek yang diukur, sesuai dengan perbedaan antara kedua kelompok tersebut. Berikut hasil analisis daya beda:

DAYA PEMBEDA						
=====						
Jumlah Subyek= 30						
Klp atas/bawah(n)= 8						
Butir Soal= 20						
Nama berkas: E:\UMSIDA\TUGAS AKHIR\REFERENSI\REFERENSI BIMBINGAN PERTAMA\DATA HASIL UJIAN SAMAMDA.ANA						
No Butir Baru	No Butir Asli	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)	
1	1	8	2	6	75.00	
2	2	8	1	7	87.50	
3	3	8	1	7	87.50	
4	4	7	2	5	62.50	
5	5	8	3	5	62.50	
6	6	8	4	4	50.00	
7	7	8	2	6	75.00	
8	8	1	0	1	12.50	
9	9	8	3	5	62.50	
10	10	8	4	4	50.00	
11	11	8	1	7	87.50	
12	12	8	3	5	62.50	
13	13	7	1	6	75.00	
14	14	6	2	4	50.00	
15	15	7	4	3	37.50	
16	16	8	4	4	50.00	
17	17	8	3	5	62.50	
18	18	8	4	4	50.00	
19	19	4	2	2	25.00	
20	20	7	6	1	12.50	

**Gambar 4.** Hasil analisis daya pembeda

Indeks daya pembeda masih berupa presentase, sehingga perlu diubah terlebih dahulu dalam bentuk desimal. Kemudian data perlu dikelompokkan untuk mempermudah pembacaan data:

**Tabel 4.** Pengelompokan indeks daya pembeda

DP	Penilaian	jumlah
Negative (jelek)	-	-
<0,20 (lemah)	8,20	2

0,20 – 0,39 (cukup)	15,19	2
0,40 – 0,69 (baik)	4,5,6,9,10,12,14,16,17,18	10
0,70 – 1,00 (Sangat baik)	1,2,3,7,11,13	6

Berdasarkan tabel 4 hasil daya pembeda penilaian Sumatif Tengah Semester dengan aplikasi Anates diketahui:

1. 2 butir soal atau sebesar 10% menunjukkan daya beda lemah yang bermakna bahwa 2 butir soal ini jelek karena lebih banyak dijawab benar oleh kelompok bawah dibandingkan dengan jawaban kelompok atas. Ini berarti bahwa untuk menjawab soal dengan benar dapat dilakukan dengan menebak.
2. 2 butir soal atau 10% menunjukkan daya beda yang cukup yang bermakna bahwa 2 butir soal ini tergolong mampu untuk memilah kemampuan murid walaupun masih belum lebih baik dari soal berkategori baik.
3. 10 butir soal atau 50% daya pembeda baik yang bermakna bahwa 10 butir soal ini tingkat akurasi baik dalam membedakan kemampuan murid karena mayoritas murid kelompok atas menjawab soal tersebut dengan benar, sehingga dapat membedakan murid yang asal menjawab dan yang memiliki pengetahuan untuk menjawab soal.
4. 6 soal atau 30% daya pembedanya sangat baik yang bermakna 6 butir soal ini sangat bisa diandalkan dalam memilah kemampuan siswa dengan tingkat akurasi sempurna. dengan demikian hanya murid yang berkemampuan atas yang mampu menjawab dengan benar pada butir soal tersebut.

## VII. SIMPULAN

Dari hasil penelitian yang dilakukan mengenai analisis pada butir soal pilihan ganda penilaian Sumatif Tengah Semester SMA Muhammadiyah 2 Sidoarjo menunjukkan bahwa terdapat beberapa poin yang dapat disimpulkan. 1) pada butir soal pilihan ganda terdapat 16 butir soal yang dinyatakan valid dan 4 butir soal dinyatakan tidak valid. 2) untuk hasil perhitungan tingkat reliabilitas soal tersebut menunjukkan bahwa reliabilitasnya tergolong Tinggi dengan perolehan 0,87. 3) untuk tingkat kesukaran penilaian tersebut menunjukkan nilai yang tidak proporsional dengan 1 butir soal sukar, 10 butir soal sedang dan 9 butir soal mudah. 4) hasil daya pembeda menunjukkan bahwa 2 butir soal memiliki daya beda lemah sehingga perlu diganti, 2 butir soal memiliki daya beda yang cukup, 10 butir soal memiliki daya beda baik dan 6 butir soal memiliki daya beda yang sangat baik.

## UCAPAN TERIMA KASIH

Syukur Alhamdulillah penulis ucapkan atas berkat rahmat, ridho dan karunia Allah SWT yang telah memberikan kemudahan dalam menyelesaikan penelitian ini dengan baik dan tepat waktu. Tak lupa penulis juga mengucapkan terimakasih kepada orang tua yang memotivasi sepenuhnya, dan juga kepada pihak-pihak lain yang membantu dalam menyelesaikan penelitian baik dari sekolah SMA Muhammadiyah 2 Sidoarjo maupun teman-teman “Jazākumullahu khairan”.

## REFERENSI

- [1] usamah as-sayyid ali Zaki, *Al-ikhtibar Al-lughawiah*, 1st ed. Riyadh: Wojooh, 2016.
- [2] K. Abdul Aziz, *Mu'jam ikhtibar*, Pertama. Riyadh: Madaralwatan, 2011.
- [3] N. Sa'diyah, “Taqwimut ta'lim lughah arabiyah (dirosatu ikhtibar lughawiyah),” *Islam. Rev. J. Ris. dan Kaji. Keislam.*, vol. 8, no. 2, pp. 259–279, 2019.
- [4] M. A. al khuli, *Asalib tadrīs al-lughah*, 2nd ed. Riyadh, 1986.
- [5] U. Ridho, U. Islam, N. Syarif, and K. Kunci, “Evaluasi dalam pembelajaran bahasa arab,” vol. 20, no. 01.
- [6] S. Arikunto, *Dasar-dasar evaluasi pendidikan*, 3rd ed. Jakarta: Sinar Grafika Offset, 2018.
- [7] A. Kurniawan, A. N. Febrianti, R. Risan, D. Merris, and M. Sari, *Evaluasi pembelajaran*.

- [8] M. A. al khuli, *Al-ikhtibarat Al-lughawiyah*, 1st ed. Al-urdun: An-nashir, 2000.
- [9] I. R. N. Hula, "EVALUASI DAN TES BAHASA ARAB : Tinjauan Teori Abstrak," no. December, pp. 0–11, 2021.
- [10] S. Fathimah and A. Fathiyah, "ANALISIS BUTIR SOAL PELAJARAN BAHASA ARAB," vol. 4, no. 1, 2019.
- [11] A. Anates, V. Dalam, A. Interdiana, C. Sari, and M. Herawati, "MENGANALISIS BUTIR SOAL," vol. I, no. 2, pp. 203–214, 2014.
- [12] M. Arif, U. T. Madura, and Y. Wibisono, "Penerapan aplikasi anates bentuk soal pilihan ganda," vol. 1, no. 1, pp. 1–9, 2014.
- [13] R. Husna, "tahlilu bunudi asilati lughatul arabiatu fi madrasah Aliyah," UIN Ar-raniry, 2022.
- [14] I. Negeri and A.-R. Banda, "ANALISIS BUTIR SOAL EVALUASI PEMBELAJARAN PENDIDIKAN AGAMA ISLAM MENGGUNAKAN PROGRAM PENDAHULUAN Evaluasi merupakan proses mengukur dan menilai , dua hal ini sangat erat kaitannya dengan sistem evaluasi . Pengukuran merupakan sebuah proses perbandingan an," vol. 10, no. 2, pp. 58–74, 2020.
- [15] D. Maulana and A. Sanusi, "Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Bersama Daerah (Uambd) Madrasah Ibtidaiyah Tahun 2017-2018," *Ta'lim al-'Arabiyyah J. Pendidik. Bhs. Arab Kebahasaaraban*, vol. 4, no. 1, pp. 12–24, 2020, doi: 10.15575/jpba.v4i1.8054.
- [16] C. N. Aulia, Arief Rahman; Eva, *EVALUASI PEMBELAJARAN*, 1st ed. Ponorogo: Uwais Inspirasi Indonesia, 2019.
- [17] F. Hikmawati, *METODOLOGI PENELITIAN*, 4th ed. Depok: RajaGrafindo persada, 2020.
- [18] Hardani, *Buku Metode Penelitian Kualitatif dan Kualitatif*, no. April. 2020.
- [19] Rukminingsih, G. Adnan, and M. A. Latief, *Metode Penelitian Pendidikan. Penelitian Kuantitatif, Penelitian Kualitatif, Penelitian Tindakan Kelas*, vol. 53, no. 9. 2020.
- [20] A. Munip, U. Islam, N. Sunan, and K. Yogyakarta, *Penilaian pembelajaran bahasa arab*, no. May. 2019.
- [21] Asrul, R. Ananda, and Rosinta, *Evaluasi Pembajalaran*. 2014.
- [22] satria wiguna, *APLIKASI ANATES DALAM EVALUASI PEMBELAJARAN*, 1st ed. Banyumas: CV. Pena persada, 2021.
- [23] B. Hanan, "As-sidqu wa tsabat fil buhusi ijtima"i," *J. Univ. Jijl*, vol. 3, no. 2, p. 125, 2020.
- [24] M. Yusuf, *ASMEN DAN EVALUASI PENDIDIKAN*, Edisi pert. K E N C A N A, 2017.
- [25] Asdar, "Evaluasi Pembelajaran." pp. 1–164, 2020.
- [26] W. Buska and R. Widayanti, "Taqniyyah Al Ikhtibārah Fi Al-Ta'Līmi Al-Lughah Al-'Arabiyyah Biljāmi'Ah Al Islāmiyyah," *Nazharat J. Kebud.*, vol. 26, no. 02, pp. 457–480, 2020, doi: 10.30631/nazharat.v26i02.38.

**Conflict of Interest Statement:**

*The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.*