

**ANALISIS KUALITAS BUTIR SOAL TOAFL
UNHASY TEBUIRENG**

Lailatul Qomariyah¹, Isniyatun Niswah, Mz²

¹Pendidikan Bahasa Arab, Fakultas Agama Islam, UNHASY Tebuireng

² Pendidikan Bahasa Arab, Fakultas Agama Islam, UNHASY Tebuireng

E-mail : lailatulqomariyah.unhasy@gmail.com¹

Abstrak

Test of Arabic Foreign Language merupakan tes terstandar yang diujikan untuk mengukur sejauhma kompetensi peserta didik dalam berbahasa Arab. Agar tes bisa berfungsi sebagaimana mestinya, maka tes harus dianalisis kualitasnya dengan harapan tes tersebut mampu menjadi tolak ukur guru dalam peningkatan kegiatan pembelajaran. Artikel ini menyajikan analisis kualitas butir soaln tes dari sisi validitas dan reliabilitas dengan metode kuantitatif. Teknik pengumpulan data dilakukan melalui dokumen lembar jawaban peserta tes dan dianalisis dengan rumus validitas berdasarkan indeks daya beda dan analisis reliabilitas tes dengan melalui rumus KR. 20. Hasil dari penelitian ini yaitu, 1) Berdasarkan tabel diatas, maka diketahui bahwa tingkat validitas butir soal TOAFL UNHASY pada tes ketrampilan membaca dari nomor 1-50 bagian ada 17 butir soal dan 33 butir soal berkategori valid.; 2) Tingkat reliabilitas butir soal pilihan ganda TOAFL diketahui 0,91. Hal ini menunjukkan bahwa tingkat reliabilitasnya berkategori amat tinggi.

Kata kunci: kualitas tes, validitas, reliabilitas

Abstract

The Test of Arabic Foreign Language is a standardized test that is tested to measure the extent of students' competence in Arabic. In order for the test to function properly, the quality of the test must be analyzed in the hope that the test can become a benchmark for teachers in improving learning activities. This article presents an analysis of the quality of the test items in terms of validity and reliability using quantitative methods. The data collection technique was carried out through the test taker's answer sheet document and analyzed using the validity formula based on the discriminating power index and the test reliability analysis using the KR formula. 20. The results of this study are, 1) Based on the table above, it is known that the level of validity of the TOAFL UNHASY items on the reading skills test from number 1-50 sections there are 17 items and 33 items categorized as valid; 2) The level of reliability of TOAFL multiple choice items is known to be 0.91. This shows that the level of reliability is categorized as very high.

Key words: quality of the test, validity, reliability

1. PENDAHULUAN

Dalam pembelajaran bahasa Arab ada tiga komponen yang saling berkaitan satu sama lain yakni dosen, mahasiswa dan tujuan pembelajaran. Dosen memiliki peran sentral dalam kegiatan pembelajaran. Hal ini sesuai dengan pendapat Copper (1979) dalam (Ainin, 2016) bahwa dosen merupakan pendidik profesional yang memiliki peran dalam penentuan keberhasilan kegiatan pembelajaran. Dia berperan dalam pengambil keputusan pada kegiatan pembelajaran mulai merencanakan, mengimplementasi dan melakukan penilaian. Kegiatan penilaian ini sebagai dasar dosen untuk menentukan apa yang harus diperbaiki dan dipertahankan dalam proses pembelajaran sehingga diketahui sejauhmana ketercapaian keberhasilan pembelajarannya. Arifin (Arifin, 2009), (Ma'arif, n.d.), Eko (ساميياجي, n.d.) dan Solichin (Solichin, 2017) memberikan pendapat yang sama bahwa penilaian merupakan kegiatan yang bersifat sistematis dan berkesinambungan dalam pemerolehan informasi terkait proses dan hasil belajar sehingga dosen mendapat informasi tentang perkembangan mahasiswa, kurikulum pembelajarannya dan serta mengetahui berhasil tidaknya program pembelajaran yang telah dilaksanakan sehingga akan memberikan stimulus bagi dosen untuk meningkatkan kualitas aktivitas pembelajarannya. Mutholib (Mutholib, 2013) menambahkan bahwa kegiatan penilaian ini adalah komponen yang terpadu dengan seluruh aktyivitas proses pembelajaran yang dengan kegiatan tersebut akan diperoleh informasi tentang gambaran penguasaan mahasiswa atas apa yang telah dipelajari sehingga guru menjadi termotivasi untuk merancang

program apa yang seharusnya diupayakan untuk memperbaikinya. Fitriawanati (Fitriawanati, 2015) dan Kadek (Primayana, Kadek Hengki, 2020) memberikan pendapat yang sama bahwa fungsi dari kegiatan penilaian untuk mendorong mahasiswa untuk lebih bersungguh-sungguh dalam belajar dan juga menjadikan guru termotivasi dalam peningkatan kualitas pembelajaran dan pengembangan aktivitas belajar mengajar selanjutnya.

Keberhasilan kegiatan penilaian ini tergantung pada ketepatan instrumen yang digunakan. Apakah instrumen tersebut dapat memberikan informasi yang tepat sesuai tujuan pembelajaran yang telah ditetapkan. Beberapa instrumen yang sering dipakai oleh guru dalam pembelajaran yakni tes. Hal ini sesuai dengan pendapat Zainal (Arifin, 2012) dan Arikunto (Arikunto, 2013) bahwa tes merupakan alat atau prosedur yang digunakan sebagai alat pengukuran. Salah satu bentuk tes yang sudah dinilai kelayakannya dari segi validitas dan reliabilitasnya dan sangat praktis penggunaannya yakni tes standar (الاختبار المقنن). Tes standar ini memiliki ciri khusus sebagai instrumen penilaian yang memberikan perlakuan yang sama pada setiap peserta tes dan materi yang diujikan bersifat komprehensif yang menilai pemahaman dan penguasaan mufrodad Bahasa Arab yang cukup banyak. Tes ini juga memiliki karakteristik yang baik karena dikembangkan oleh tim ahli dan telah diuji kelayakannya baik validitas maupun reliabilitasnya. Hal ini sesuai dengan pendapat Grounlond and Linn, 1995 dalam (Latifah, 2020) dan Patomuan (Harahap, 2018) bahwa tes standar itu harus memenuhi kelayakannya baik validitas (المصدق), reliabilitas (الثبات), dan Usabilitas.

TOAFL merupakan salah satu tes standar yang berfungsi menilai kompetensi seseorang dalam bidang Bahasa Arab. Tes ini memuat butir soal yang menilai kemampuan mendengar siswa dalam Bahasa Arab, kemampuan membacanya, kemampuan berbicaranya dan kemampuan menulisnya. Tes TOAFL ini pada awalnya dikembangkan oleh UIN Syarif Hidayatulloh Jakarta pada tahun 1999 untuk diujikan pada mahasiswa S1, S2 dan S3. Kemudian tes ini dikembangkan selanjutnya oleh setiap perguruan tinggi keagamaan di Indonesia Universitas Hasyim Asy'ari merupakan salah satu universitas di Indonesia yang juga mengembangkan TOAFL. Tes ini digunakan untuk mengukur kemampuan mahasiswa dalam berbahasa Arab dan memberikan standarisasi pada kemampuan Bahasa Arab pada semua prodi dibawah naungan universitas tersebut. Tes ini terdiri dari 164 butir soal yang memuat 160 soal pilihan ganda dan 4 soal dengan bentuk essay yang telah dikembangkan tim Dosen Bahasa Arab dibawah naungan Unit pengembangan Bahasa.

Pada awalnya tes ini dilaksanakan untuk menguji ketercapaian pembelajaran intensif bahasa Asing yang telah dilaksanakan selama satu tahun untuk mahasiswa Baru. Dalam perkembangannya tes ini juga digunakan oleh dosen sebagai prasyarat mengikuti peningkatan kompetensinya dan dalam penyelesaian studi doktoralnya. Oleh sebab itu perlulah dilakukan telaah kembali pada butir soal TOAFL. Hal ini dilakukan untuk mengetahui kualitas butir soal yang tersusun. Kegiatan ini perlu dilakukan dengan tujuan untuk meningkatkan mutu soal dan membantu meningkatkan kualitas tes baik dengan cara revisi atau membuang soal yang tidak efektif, serta untuk mengetahui informasi tingkat kemampuan apakah mereka sudah/belum memahami materi yang telah diajarkan.

Kegiatan analisis kualitas butir soal ini difokuskan pada telaah validitas dan reliabilitasnya yang mana diawali dengan cara mengumpulkan data dari hasil jawaban mahasiswa, kemudian diringkas dan dianalisis sehingga menjadi pedoman dalam menentukan keputusan mana butir soal yang masih layak untuk dipertahankan, atau direvisi bahkan dibuang.

Penelitian sejenis tentang analisis Butir Soal telah banyak ditemukan antara lain penelitian Fathiyah (Al Fathiyah, 2019), dan Rizka (Anugerah, 2017).

Fathiyah (2019) menulis artikel dengan judul "Analisis Butir Soal Pelajaran Bahasa Arab di MA Roudlotul Ulum Pagak Malang". Hasil penelitian menjelaskan bahwa, 1) Butir soal yang diujikan pada ulangan semester ganjil mata pelajaran Bahasa Arab siswa kelas Sebelas IPS ada 3 soal yang dikatakan valid dengan prosentase 2.85% dan yang dikategorikan tidak valid ada 12 soal dengan prosentase 97.15%; 2) Tingkat reliabilitas tes soal pilihan ganda pada ulangan Semester tersebut sebesar 0,54 dengan berkategori sedang, 3) Dari 15 butir soal pilihan ganda yang diujikan ada tiga kategori yang ditemukan dari sisi tingkat kesukaran yakni 8 soal berkategori Mudah, 3 soal berkategori sedang dan 4 soal berkategori sukar, 3) Ditemukan 3 soal yang berkategori baik, 3 butir berkategori sedang, dan sisanya termasuk kategori soal tingkat daya bedanya adalah 0 (nol) dan negatif. sehingga tidak ditemukan diskriminasi. 4) Butir soal yang telah disusun perlu untuk direvisi dari segi daya pengecohnya.

Rizka Anugerah Menulis artikel dengan judul "Taqwiim Bunud al-As'ilah fi ikhtibari al lughoh al Arobiyyah al-Syamil al-Wathoniy ala mustawa Madrasah ath-thanawiyah al Islamiyah li sanah dirosiyyah

2014-2015. Hasil dari penelitian ini yaitu 1) Dari segi materi, Butir soal yang dikembangkan berkategori validitasnya baik karena butir soal sudah sesuai dengan tujuan pembelajaran, 2) Butir soal UAMBN yang disusun setelah dianalisis reliabilitasnya pada tingkat tinggi, 3) Dari segi tingkat kesukaran, terdapat 50% soal yang berkategori sedang, 40 % soal berkategori mudah dan 10% soal berkategori sulit, 4) Daya beda pada butir soal UAMBN terdapat lebih dari 50% berdaya beda baik.

Berdasarkan uraian latar belakang penelitian diatas, artikel ini akan mengkaji, 1) Bagaimana kualitas butir soal TOAFL Universitas Hasyim Asy'ari jika dianalisis dari segi validitas butir soalnya?, dan 2) Bagaimana kualitas butir soal TOAFL Universitas Hasyim Asy'ari jika dianalisis dari segi reliabilitasnya?

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif. Peneliti menggunakan rumus uji validitas dengan menilai daya bedanya per butir soal dan menggunakan uji reliabilitas dengan menggunakan rumus reliabilitas Kuder dan Richardson. Penelitian ini menggunakan data yang diperoleh dari teknik dokumentasi yakni berupa dokumen naskah hasil TOAFL, dan lembar soal TOAFL Bahasa Arab; dan data kurikulum program Intensif Bahasa Arab. Penelitian menggunakan teknik butir soal yang dilakukan secara kuantitatif. Analisis kuantitatif dilakukan untuk mengetahui dan menganalisis tingkat validitas dan reliabilitas.

3. HASIL PENELITIAN DAN PEMBAHASAN

Penelitian ini dilaksanakan untuk menganalisa kualitas butir soal tes TOAFL Bahasa Arab yang telah disusun oleh Tim pengembang Bahasa Arab di UPB Universitas Hasyim Asy'ari Tebuireng Jombang yang berfokus pada 50 butir soal yang untuk mengukur kemampuan membaca. Peneliti menganalisis validitas setiap butir soalnya dan tingkat reliabilitasnya.

3.1. Validitas

Validitas butir soal sangat diperlukan untuk menentukan ketepatan suatu alat ukur bisa menilai apa yang dinilai. Dalam penelitian ini butir soal kemahiran membaca harus mampu mengukur kemampuan mahasiswa dalam ketrampilan membaca Bahasa Arab yakni dalam memahami bacaan yang telah disediakan. Peneliti menggunakan indeks tingkat daya beda soal untuk mengetahui kevalidan butir soal tersebut dengan menggunakan indeks berikut
Sebelum mengukur kevalidan alat tes, peneliti menghitung terlebih dahulu daya beda pada tiap butir soal dengan rumus dan sebagai berikut.

$$D (\text{معامل التمييز}) = \frac{\sum \text{Benar Upper} - \sum \text{Benar lower}}{\sum \text{Kelompok}}$$

Kemudian setelah diketahui daya beda tiap soal kemudian dianalisis tingkat validitasnya dengan rumus kategori validitas berikut.

$DB \geq 0,30$ jika daya beda lebih dari atau sama dengan 30, maka dikatakan valid

$DB \leq 0,30$ jika daya beda kurang dari atau sama dengan 30, maka butir soal dikatakan gugur (tidak valid)

Berikut akan disajikan tabel hasil daya beda tiap butir soal yang kemudian dianalisis kevalidan alat tes berdasar pada indeks daya bedanya.

No	Tingkat Daya Beda	Kriteria Validasi Butir Soal
1	0,5	Valid
2	0,5	Valid
3	0,5	Valid
4	0,25	Tidak Valid
5	1	Valid
6	0	Tidak Valid
7	0,25	Tidak Valid
8	0,25	Tidak Valid
9	0	Tidak Valid
10	0,25	Tidak Valid

11	0,5	Valid
12	0,5	Valid
13	0,5	Valid
14	0,5	Valid
15	0,5	Valid
16	0,5	Valid
17	0	Tidak Valid
18	0,5	Valid
19	0,5	Valid
20	0,5	Valid
21	0,75	Valid
22	0,25	Tidak Valid
23	0,5	Valid
24	0,25	Tidak Valid
25	0,75	Valid
26	0,75	Valid
27	0,5	Valid
28	0,5	Valid
29	0,75	Valid
30	0,25	Tidak Valid
31	0,5	Valid
32	0,75	Valid
33	0,5	Valid
34	0	Tidak Valid
35	0,25	Tidak Valid
36	0,75	Valid
37	0,75	Valid
38	0,25	Tidak Valid
39	0,25	Tidak Valid
40	0,5	Valid
41	0,75	Valid
42	-0,25	Tidak Valid
43	0	Tidak Valid
44	1	Valid
45	0,25	Tidak Valid
46	0,75	Valid
47	0,5	Valid
48	0,5	Valid
49	0,5	Valid

50	0,5	Valid
----	-----	-------

Berdasarkan tabel diatas, maka diketahui bahwa tingkat validitas butir soal TOAFL UNHAS pada tes ketrampilan membaca dari nomor 1-50 ada 17 butir soal tidak valid karena indeks daya bedanya kurang dari atau sama dengan 0,30 dan 33 butir soal berkategori valid karena indeks daya bedanya lebih dari 0,3. Butir soal yang telah berkategori valid maka tidak perlu untuk merevisinya. Sedangkan untuk 17 butir soal yang tidak valid harus direvisi atau diganti dengan butir soal yang lain.

Hal ini sesuai dengan Allen dalam (Al Fathiyah, 2019) bahwa Untuk mengetahui kevalidan butir soal tes bisa dianalisis dengan berdasarkan pada indeks daya pembeda.

3.2. Reliabilitas

Reliabilitas merupakan konsistensi sebuah alat instrumen dalam memebrikan data. Jadi butir soal itu dianggap reliabel jika butir soal tersebut menghasilkan data yang konsisten meskipun diujikan berulang kali dalam waktu yang berbeda. Untuk menghitung reliabilitas, peneliti menggunakan rumus dari Kuder dan Richardson. Berikut, tabel tingkat reliabilitas butir soal pilihan ganda TOAFL Bahasa Arab.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	n	n-1	Σ	p	q	pq	Σpq
1	1	0	0	1	1	1	1	0	1	0	0	0	0	1	1	50	49	8	0.5	0.5	0.2	9.1
2	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1			13	0.9	0.1	0.1	
3	0	0	1	0	0	1	1	1	1	1	1	0	1	1	1			10	0.7	0.3	0.2	
4	0	1	1	1	1	1	0	1	1	1	1	1	0	0	1			11	0.7	0.3	0.2	
5	0	0	1	1	1	1	1	0	1	1	1	0	0	1	1			10	0.7	0.3	0.2	
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			15	1.0	0.0	0.0	
7	1	0	0	0	1	1	1	1	1	0	1	0	1	1	1			10	0.7	0.3	0.2	
8	1	0	1	0	0	1	1	0	0	1	1	0	1	0	1			8	0.5	0.5	0.2	
9	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0			3	0.2	0.8	0.2	
10	0	0	1	1	1	1	0	1	1	1	1	0	1	1	1			11	0.7	0.3	0.2	
11	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1			13	0.9	0.1	0.1	
12	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1			14	0.9	0.1	0.1	
13	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1			4	0.3	0.7	0.2	
14	0	0	0	1	1	1	1	1	1	0	1	1	1	1	1			11	0.7	0.3	0.2	
15	0	0	1	1	0	1	1	1	1	0	1	1	1	1	1			11	0.7	0.3	0.2	
16	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1			13	0.9	0.1	0.1	
17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			15	1.0	0.0	0.0	
18	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1			12	0.8	0.2	0.2	
19	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1			13	0.9	0.1	0.1	
20	1	0	0	0	1	1	0	0	1	0	1	0	0	0	1			6	0.4	0.6	0.2	
21	0	0	1	1	0	1	1	1	1	0	1	1	0	1	1			10	0.7	0.3	0.2	
22	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1			13	0.9	0.1	0.1	
23	0	1	0	1	1	1	1	1	1	0	1	1	0	1	1			11	0.7	0.3	0.2	
24	1	0	1	1	1	1	1	1	1	1	0		0	1	1			11	0.7	0.3	0.2	
25	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1			12	0.8	0.2	0.2	
26	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1			12	0.8	0.2	0.2	

27	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	13	0.9	0.1	0.1
28	0	1	1	1	1	1	1	1	1	1	1	1		1	1	13	0.9	0.1	0.1
29	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	12	0.8	0.2	0.2
30	0	1	1	0	1	0	0	0	1	1	1	1	0	0	1	8	0.5	0.5	0.2
31	0	1	1	1	0	1	1	1	1	1	1	1	0	1	1	12	0.8	0.2	0.2
32	0	1	0	1	1	1	1	0	1	0	1	1	0	1	0	9	0.6	0.4	0.2
33	0	0	1	0	1	1	0	0	1	0	0	0	0	1	0	5	0.3	0.7	0.2
34	1	0	0	1	0	0	1	0	0	0	1	0	1	1	1	7	0.5	0.5	0.2
35	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	3	0.2	0.8	0.2
36	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	11	0.7	0.3	0.2
37	0	0	1	1	1	0	1	0	1	0	1	0	0	1	0	7	0.5	0.5	0.2
38	0	1	0	1	0	1	0	0	1	0	0	1	0	1	0	6	0.4	0.6	0.2
39	1	0	1	1	0	0	1	1	1	0	1	1	0	1	1	10	0.7	0.3	0.2
40	1	0	1	0	1	1	1	1	1	1	1	1	0	0	0	10	0.7	0.3	0.2
41	1	0	1	0	1	1	1	0	1	0	1	1	0	0	1	9	0.6	0.4	0.2
42	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.1	0.9	0.1
43	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	5	0.3	0.7	0.2
44	0	0	1	1	1	1	1	0	1	1	1	1	0	1	1	11	0.7	0.3	0.2
45	0	0	0	0	1	1	1	1	1	0	0	1	1	0	1	8	0.5	0.5	0.2
46	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	5	0.3	0.7	0.2
47	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	3	0.2	0.8	0.2
48	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	4	0.3	0.7	0.2
49	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	10	0.7	0.3	0.2
50	0	1	0	1	1	1	1	1	1	0	1	0	0	1	0	9	0.6	0.4	0.2
X	17	13	32	34	34	40	40	26	44	27	39	32	19	37	37	471			
S2	84.97																		
r11	0.91																		

Untuk mengetahui derajat konsistensi butir soal sebagai alat instrument untuk menilai kemampuan Bahasa Arab mahasiswa, peneliti menggunakan teknik reliabilitas Kuder dan Richardson

$$\text{yakni, } r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S^2 - \sum pq}{S^2} \right)$$

n: total butir soal

p: total skor butir soal – jumlah siswa

q : 1-p

Berdasarkan tabel di atas diketahui tingkat reliabilitasnya yakni 0,91. Kemudian data tersebut dianalisis menggunakan tingkat kategori reliabilitas menurut Djiwandjono yang mana menunjukkan bahwa bahwa tingkat reliabilitasnya berkategori amat tinggi. Hal ini sesuai dengan pendapat Arikunto ((Arikunto, 2013) bahwa reliabilitas tes bisa dianalisis menggunakan rumus K-R 20 yang dikemukakan oleh Kuder dan Richardson.

4. Kesimpulan

Berdasarkan analisa butir soal TOAFL UNHAS Y Tebuireng pada materi Ketrampilan Membaca yang telah dilakukan peneliti, maka dapat ditarik kesimpulan berikut, 1) Berdasarkan tabel diatas, maka diketahui bahwa tingkat validitas butir soal TOAFL UNHAS Y pada tes ketrampilan membaca dari nomor

1-50 bagian ada 17 butir soal dan 33 butir soal berkategori valid.; 2) Tingkat reliabilitas butir soal pilihan ganda TOAFL diketahui 0,91. Hal ini menunjukkan bahwa tingkat reliabilitasnya berkategori amat tinggi. Hasil analisis ini memberikan masukan bagi Tim pengembang soal TOAFL UNHAS Y untuk merevisi dan mengembangkan kembali butir soal yang telah terstandar dan sudah melalui uji kelayakan. sehingga alat tes TOAFL tersebut mampu mengukur ketrampilan berbahasa Arab dengan benar.

5. DAFTAR RUJUKAN

- Ainin, M. (2016). Kesahihan Dalam Penyusunan Tes Bahasa Arab Di Madrasah/Sekolah. *Prosiding Konferensi Nasional Bahasa Arab*, 1(2). <http://prosiding.arab-um.com/index.php/konasbara/article/view/75>
- Al Fathiyah, S. F. (2019). Analisis Butir Soal Pelajaran Bahasa Arab di MA Roudlotul Ulum Pagak Malang. *Tarbiyatuna*, 4(1), 78.
- Anugerah, R. (2017). Taqwim Bunud al As'ilah fi ikhtibar al-Lughoh al-Arabiyyah al Syamil al Wathoniy ala mustawa al Madrah at Tsanawiyah al Islamiyah li as sanah al dirosiyyah 2014-2015. *Al-Ta'rib: Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, 5(2), 150–175.
- Arifin, Z. (2009). *Evaluasi Pembelajaran*. PT. Remaja Rosdakarya.
- Arifin, Z. (2012). *Evaluasi Pembelajaran* (2nd ed.). Direktorat Jenderal Pendidikan Islam, Kementerian Agama RI.
- Arikunto, S. (2013). *Dasar-dasar Evaluasi Pendidikan*. Bumi Aksara.
- Fitrianawati, M. (2015). Peran analisis butir soal guna meningkatkan kualitas butir soal, kompetensi guru dan hasil belajar peserta didik. *Prosiding Seminar Nasional Pendidikan PGSD UMS & HDPGSDI Wilayah Jawa*, 5(3), 282-295. ISBN 978-602-70471-2-9 PERAN. <http://hdl.handle.net/11617/9117>
- Harahap, P. (2018). Analisis Soal TOAFL IAIN CURUP. *Ihya' AL-Arabiyyah*, 2, 18–33.
- Latifah, U. (2020). Optimalisasi Pelaksanaan Program Kolektif Untuk Meningkatkan Kompetensi Guru Sebagai Evaluator Dalam Menyusun Soal UAS di SDN Bakalan Kecamatan Sumobito Tahun Pelajaran 2018/2019. *Jurnal Dinamika Manajemen Pendidikan*, 4(2), 93. <https://doi.org/10.26740/jdmp.v4n2.p93-101>
- Ma'arif, A. S. (n.d.). Shiyagoh Ikhtibarot al Arabiyyah fi Dhou al Ikhtibar Al Muttaqin(TOAFL). *Al Bayan*, 9(2), 160–173.
- Mutholib, A. (2013). "Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Berstandar Nasional (UAMBN) Bagi Siswa Madrasah Ibtidaiyah (MI) Tahun Pelajaran 2012-2013." *Jurnal Arabia*, 5(2).
- Primayana, Kadek Hengki, I. P. S. (2020). Peran Desain Evaluasi Pembelajaran. *Widyacarya*, 4(2), 88–100. <http://jurnal.stahnmpukuturan.ac.id/index.php/widyacarya/article/view/796>
- Solichin, M. (2017). Analisis Daya Beda Soal Taraf Kesukaran, Butir Tes, Validitas Butir Tes, Interpretasi Hasil Tes Validitas Ramalan dalam Evaluasi Pendidikan. *Jurnal Manajemen Dan Pendidikan Islam* 2, 2(2), 192–213.
- سامياجي, إ. (n.d.). تحليل بنود الاختبار المدرسي النهائي الوطني في مادة اللغة العربية بناء على أساس معايير الاختبار الجيد بمدرسة المتوسطة الإسلامية الحكومية مالانج للسنة الدراسية 2018-2019م. كلية الدراسات العليا جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج.