

ANALISIS BUTIR SOAL PELAJARAN BAHASA ARAB DI MA ROUDLOTUL ULUM PAGAK MALANG

Siti Fathimah Al Fathiyah^{1*}

¹Dosen Sekolah Tinggi Ilmu Tarbiyah Ibnu Sina Malang

Abstract: The focus of this research is to analyze the results of the final exam of the Arabic subjects of the class XI-IPS students in MA Roudlotul Ulum Pagak Malang at odd semester Year lesson 2018/2019. The goal is to allow teachers of Arabic language to make good measuring instruments in measuring students' level of understanding. The type of research used is quantitative with the data collection model obtained from both, students' answer and question sheets. The Data obtained is analyzed using anathes and described in a descriptive basis based on the level of validity, reliability, difficulty, power difference and tricky in each multiple choice. The results of the study showed: (1) based on the results of the validity test, it was found that from item question number 1-15, which belongs to the category of valid question only 3 (2.85%) And that belongs to the category of invalid 12 questions (97.15%). (2) Reliability test result indicated that the level of the reliability of mulltiple choice questions 0.54, indicating medium category. (3) The result of difficulty analysis has been known between Question number 1 to 15, there are 8 questions that belong to the category of "Easy", 3 items that belong to the category of "Medium", and 4 items that belong to the category of "difficult" question. (4) Distinguishing power test results show there are 3 questions that include the category of good problem, 3 items including category of medium problem, and the rest is include category of question that can not distinguish between students from Upper group and lower group. Even, the difference in power level is 0 (zero) so there is no discrimination. consequently, the question is revised or unused. (5) The quality of the tricky question has been found there are many problems of tricky questions that need to be revised.

Keywords: Item Question, Validity & Reliability Test, Difficulty Test, Tricky Question Test

*Korespondensi Penulis: faelfath@gmail.com

Abstrak: Fokus penelitian ini adalah menganalisis hasil ulangan semester ganjil mata pelajaran Bahasa Arab siswa kelas XI-IPS di MA Roudlotul Ulum Pagak Malang Tahun Pelajaran 2018/2019. Tujuannya, agar pengajar Bahasa Arab mampu membuat alat ukur yang baik dalam mengukur tingkat pemahaman siswa. Penelitian menggunakan pendekatan kuantitatif dengan model pengumpulan data yang didapatkan dari dokumen lembar jawaban beserta soal siswa. Data yang didapatkan dianalisis dengan menggunakan anathes dan di jabarkan secara deskriptif berdasarkan tingkat validitas, reliabilitas, kesukaran, daya beda dan pengecoh pada tiap-tiap pilihan ganda. Hasil penelitian menunjukkan: (1) Hasil uji validitas menemukan bahwa dari nomor soal 1-15, yang termasuk dalam kategori soal valid hanya 3 butir soal (2.85%) dan yang termasuk dalam kategori soal tidak valid 12 butir soal (97.15%). (2) Hasil uji reliabilitas didapatkan bahwa tingkat reliabilitas tes soal pilihan ganda sebesar 0,54, menunjukkan kategori sedang. (3) Hasil analisis tingkat kesukaran diketahui antara soal no 1 sampai 15 terdapat 8 butir soal yang termasuk dalam kategori soal “Mudah”, 3 butir soal termasuk dalam kategori soal “Sedang”, dan 4 butir soal termasuk dalam kategori soal “Sukar”. (4) Hasil uji daya pembeda menunjukkan terdapat 3 butir soal yang termasuk kategori soal baik, 3 butir soal termasuk kategori soal sedang, dan sisanya termasuk kategori soal yang tidak bisa membedakan antara siswa dari kelompok atas dan kelompok bawah. Bahkan tingkat daya bedanya adalah 0 (nol) sehingga tidak ada diskriminasi. Implikasinya, butir soal tersebut direvisi atau tidak digunakan. (5) Kualitas dari soal pengecoh telah ditemukan masih banyak soal pengecoh yang perlu direvisi.

Kata Kunci: Butir Soal, Uji Validitas & Reliabilitas, Uji Tingkat Kesukaran, Uji Soal Pengecoh

A. Pendahuluan

Dalam dunia pendidikan, penilaian merupakan bagian yang tak terpisahkan dari proses belajar mengajar. Sistem penilaian yang baik akan mendorong guru menggunakan strategi mengajar yang lebih baik dan memotivasi anak untuk belajar lebih giat. Penilaian biasanya dimulai dengan kegiatan pengukuran. Pengukuran (*measurement*) merupakan cabang ilmu statistika terapan yang bertujuan untuk membangun dasar-dasar pengembangan tes yang lebih baik sehingga menghasilkan tes yang berfungsi secara optimal, valid, dan reliabel.

Proses belajar mengajar dilaksanakan tidak hanya untuk kesenangan atau bersifat mekanis saja tetapi mempunyai misi atau tujuan bersama. Dalam usaha untuk mencapai misi dan tujuan, itu perlu diketahui apakah usaha yang dilakukan sudah sesuai dengan tujuan? Untuk mengetahui apakah tujuan pendidikan sudah tercapai perlu diadakan tes.

Analisis butir soal dilakukan untuk menstandarisasi butir-butir soal yang telah dibuat agar diperoleh butir-butir soal yang bermutu baik.¹ Sebuah tes yang dapat baik sebagai alat pengukur harus dianalisis terlebih dahulu. Dalam menganalisis butir soal dalam tes harus memperhatikan daya serap, tingkat kesukaran, daya beda, fungsi pengecoh, validitas dan reabilitas. Hal tersebut dilakukan agar tes yang diberikan kepada siswa sesuai dengan daya serap siswa, tingkat kesukarannya, dan soal yang diberikan pun harus valid. Sehingga, tujuan dari pembelajaran dapat tercapai.

Tes ini diterapkan pada Ulangan Umum Semester Ganjil TP. 2018/2019 bidang studi Bahasa Arab siswa kelas XI-IPS MA Roudlotul Ulum Pagak Malang, jenis soal ini adalah tes obyektif dengan 50 butir soal bentuk pilihan ganda yang mempunyai lima alternatif pilihan jawaban yaitu a, b, c, d, dan e.

Jumlah siswa kelas XI-IPS adalah 13 orang, peneliti hanya akan menganalisis butir soal nomor 1-15 dengan mengukur tingkat validitas, reliabilitas, tingkat kesukaran setiap butir soalnya, daya pembeda dan tingkat daya pengecoh antar pilihan alternatif jawaban.

B. Landasan Teori

1. Analisis Butir Soal

Kita telah ketahui bersama bahwa analisis butir soal merupakan salah satu bagian pembahasan dari evaluasi pendidikan. Oleh karena itu sebelum kita membahasnya lebih jauh, alangkah bijaknya bagi kita untuk sekedar

¹ Tobari. *Evaluasi Soal-soal Penerimaan Pegawai Baru Dilengkapi Dengan Hasil Penelitiannya*. (Yogyakarta: Penerbit Deepublish, 2015). Hlm. 70

mengingat kembali tentang makna evaluasi. Kata evaluasi menurut kamus besar bahasa Indonesia yaitu penilaian.²

Sedangkan evaluasi itu sendiri menurut Eveline Siregar dan Hartini Nara adalah suatu proses menentukan nilai prestasi belajar pembelajaran dengan menggunakan patokan-patokan tertentu agar mencapai tujuan pengajaran yang telah ditentukan sebelumnya.³

Analisis butir soal diartikan sebagai menyelidiki atau penelitian terhadap suatu bagian dari keseluruhan sesuatu yang harus dijawab oleh peserta didik. Analisis soal digunakan untuk menilai tes yang telah dibuat baik oleh guru maupun tes standar yang dibuat oleh tim.⁴

Dari beberapa definisi di atas dapat saya disimpulkan, bahwa analisis butir soal yaitu suatu proses yang dilakukan untuk menyelidiki, meneliti dan mengkaji pertanyaan-pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai.

2. Syarat-Syarat Instrumen Penilaian yang Baik

a. Valid

Suatu instrumen dikatakan valid atau mempunyai validitas yang tinggi apabila alat itu betul-betul mampu mengukur dan menilai apa yang ingin diukur dan/atau dinilai.⁵

Oleh karena itu, validitas suatu instrumen merujuk kepada ketepatan suatu instrumen menilai apa yang dinilai. Ketepatan dan kebermaknaan itu disimpulkan berdasarkan bukti-bukti skor instrumen tiap individu peserta ujian atau subjek yang dinilai.

² Departemen Pendidikan dan Kebudayaan, *Kamus Besar Bahasa Indonesia*. (Jakarta: Balai Pustaka, 1996). Hlm. 272

³ Eveline Siregar Dan Hartini Nara. *Teori Belajar dan Pembelajaran*. (Bogor: Ghalia Indonesia, 2010). Hlm. 142

⁴ Darwyan Syah dkk. *Pengembangan Evaluasi Sistem Pendidikan Agama Islam*, (Jakarta: Diadit Media, 2009). Hlm. 147

⁵ Syaifuddin Azwar, *Dasar-dasar Psikometri*. (Yogyakarta: Pustaka Pelajar, 2012). Hlm. 51

Dalam hal ini, untuk mengetahui apakah butir soal tersebut bisa dikatakan valid atau tidaknya maka peneliti menggunakan rumus analisis validitas berdasarkan daya pembeda dengan indeks sebagai berikut:⁶

DB \geq 0.30 = Valid

DB \leq 0.30 = Gugur (tidak valid)

b. Reliabel

Suatu instrumen yang baik harus valid dan reliabel. Namun perlu dicermati dengan baik pernyataan berikut: Sesuatu yang valid/sahih adalah reliabel, tetapi suatu instrumen yang reliabel belum tentu valid. Reliabilitas suatu instrumen menunjuk kepada ketetapan, konsistensi, atau stabilitas instrumen/suatu pengukuran yang dilakukan.⁷

Reliabilitas adalah tingkat atau derajat konsisten dari suatu instrumen. Reliabilitas tes berkenaan dengan pertanyaan, apakah suatu tes teliti dan dapat dipercaya sesuai kriteria yang telah ditetapkan. Suatu tes dapat dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda⁸.

Disini peneliti menggunakan teknik reliabilitas yang dikemukakan oleh **Kuder** dan **Richardson** dengan rumus K-R.20.⁹

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

Keterangan:

p = proporsi peserta didik yang menjawab betul dari suatu butir soal

q = proporsi peserta didik yang menjawab item dengan salah (q = 1-p)

$$S^2 = \frac{n \sum x^2 - (\sum x)^2}{n(n-1)}$$

$\sum pq$ = jumlah hasil perkalian antara p dan q

⁶Allen, M. dan W.M. Yen. 1979. *Introduction to Measurement Theory*. (California:Brooks/Cole Publishing Company). Hlm. 122

⁷ Muri Yusuf, *Assesmen dan Evaluasi Pendidikan*. (Jakarta: Kencana, 2017)Hlm, 59

⁸ Zainal Arifin. *Evaluasi Pembelajaran ;Prinsip, Teknik, Prosedur*. (Bandung: PT Remaja Rosdakarya, 2014). Hlm. 258

⁹ Suharsimi Arikunto. *Dasar-Dasar Evaluasi Pendidikan*. (Jakarta: Bumi Askara, 2009). Hlm. 100

c. Objektif

Objektif suatu instrumen atau alat ukur menunjukkan kepada kesamaan skor atau diagnosa yang diperoleh dari data yang sama apabila dilakukan oleh penskor/penilai dengan kualitas yang sama. Dengan kata lain penskor hendaklah menilau/menskor apa adanya, tanpa dipengaruhi oleh subjektivitas penskor atau faktor-faktor lainnya di luar data yang tersedia.¹⁰

d. Praktis dan Mudah Dilaksanakan

Suatu alat ukur atau instrumen dikatakan praktis apabila biaya alat ukur itu mudah dan murah. Mudah dapat diartikan dalam konteks administrasi, penskoran dan penginterpretasian. Alat yang disusun mudah diadministrasikan, mudah di skor, dan mudah diinterpretasikan. Mudah diadministrasikan berarti para pelaksana instrumen dapat melaksanakan instrumen itu dengan baik dan para peserta ujian dengan mudah dapat memahaminya, tidak rumit bentuknya, dan sederhana bahasanya. Adapun murah merujuk pada biaya atau beban pelaksana dan peserta ujian tersebut. Ini berarti, beban biaya peserta yang mengikuti ujian tidak terlalu tinggi dan dapat dilaksanakan dalam periode-periode tertentu.¹¹

C. Metode Penelitian

Jenis dari penelitian ini adalah Kuantitatif, metode penelitian kuantitatif dapat diartikan sebagai metode penelitian yang berdasarkan pada filsafat positivisme, penggunaan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistik, dengan tujuan untuk menguji hipotesis yang telah ditetapkan.¹²

Adapun pengumpulan data didapatkan dari dokumen lembar jawaban siswa beserta soal siswa kelas XI-IPS MA Roudlotul Ulum Pagak. Kemudian data yang

¹⁰ Muri Yusuf, ..., Hlm. 59

¹¹ Muri Yusuf, ..., Hlm. 60

¹² Sugiyono. *Metode Penelitian Kuantitatif Kualitatif dan R&D*. (Bandung; Alfabeta, 2011). Hlm. 8

didapatkan dianalisis dengan menggunakan anathes dan di jabarkan berdasarkan tingkat kevalidan, kerealibilitas, tingkat kesukaran, daya beda dan pengecoh pada tiap-tiap pilihan ganda.

D. Hasil Analisis dan Pembahasan

1. Validitas

Analisis yang digunakan adalah berdasarkan daya beda jawaban peserta ujian 10 orang dengan jumlah soal 10 buah.

Tabel 1. Analisis Validitas Butir Soal Berdasarkan Daya Beda

| No Butir Soal | Daya Beda | Kriteria Validasi Butir Soal |
|---------------|-----------|------------------------------|
| 1 | 0 | Tidak valid |
| 2 | 0 | Tidak valid |
| 3 | 0 | Tidak valid |
| 4 | 0 | Tidak valid |
| 5 | 0.25 | Tidak valid |
| 6 | 0.25 | Tidak valid |
| 7 | 0 | Tidak valid |
| 8 | 0 | Tidak valid |
| 9 | 0.75 | Valid |
| 10 | 0.25 | Tidak valid |
| 11 | 0.50 | Valid |
| 12 | -0.25 | Tidak valid |
| 13 | 0 | Tidak valid |
| 14 | 1 | Valid |
| 15 | 0 | Tidak valid |

Berdasarkan tabel 1, hasil perhitungan validitas sebagaimana yang mengacu pada pendapat Allen dan Yen, maka dapat dikatakan bahwa tingkat validitas tes soal pilihan ganda Ulangan Umum Semester Ganjil TP. 2018/2019 bidang studi Bahasa Arab siswa kelas XI-IPS MA Roudlotul Ulum Pagak Malang dari nomor soal 1-15 yang termasuk dalam kategori soal yang **valid** hanya 3 butir soal (2.85%) dan yang termasuk dalam kategori soal **tidak valid** sebanyak 12 butir soal (97.15%).

2. Reliabilitas

Disini peneliti menggunakan teknik reliabilitas yang dikemukakan oleh **Kuder** dan **Richardson** dengan rumus K-R.20.¹³

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

Keterangan:

p = proporsi peserta didik yang menjawab betul dari suatu butir soal

q = proporsi peserta didik yang menjawab item dengan salah (q = 1-p)

$$S^2 = \frac{n \sum X^2 - (\sum x)^2}{n(n-1)}$$

$\sum pq$ = jumlah hasil perkalian antara p dan q

Adapun langkah-langkah perhitungan reliabilitasnya adalah sebagaimana yang terlihat pada tabel 2. Dari tabel, selanjutnya dihitung nilai S^2 sebagai berikut:

Menghitung S^2 :

$$\begin{aligned} \frac{n \sum X^2 - (\sum x)^2}{n(n-1)} &= \frac{13(1142) - (120)^2}{13(13-1)} = \frac{14846 - 14400}{156} = \frac{446}{156} \\ &= 2.858 \end{aligned}$$

Dimasukkan dalam rumus K-R. 20:

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

$$r_{11} = \left(\frac{15}{15-1} \right) \left(\frac{2.858 - 1.408}{2.858} \right)$$

$$r_{11} = 1.071 \times 0.507$$

$$r_{11} = 0.54$$

¹³ Suharsimi Arikunto. Dasar-Dasar Evaluasi Pendidikan. (Jakarta: Bumi Askara, 2009). Hlm. 100

Tabel 2. Perhitungan Reliabilitas

| | A | B | C | D | E | F | G | H | I | J | K | L | M | Σ | p | Q | Σpq |
|----------------|----|----|----|----|-----|-----|-----|-----|-----|----|----|-----|-----|------|-----------|-----------|-------------|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 | 1 | 0 | 0 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 | 1 | 0 | 0 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 | 1 | 0 | 0 |
| 4 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0.1538462 | 0.8461538 | 0.130177515 |
| 5 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 12 | 0.9230769 | 0.0769231 | 0.071005917 |
| 6 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 12 | 0.9230769 | 0.0769231 | 0.071005917 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 12 | 0.9230769 | 0.0769231 | 0.071005917 |
| 9 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 6 | 0.4615385 | 0.5384615 | 0.24852071 |
| 10 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 7 | 0.5384615 | 0.4615385 | 0.24852071 |
| 11 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 0.7692308 | 0.2307692 | 0.177514793 |
| 12 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 7 | 0.5384615 | 0.4615385 | 0.24852071 |
| 13 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.0769231 | 0.9230769 | 0.071005917 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 15 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 12 | 0.9230769 | 0.0769231 | 0.071005917 |
| X | 7 | 8 | 7 | 8 | 11 | 12 | 11 | 10 | 10 | 9 | 7 | 10 | 10 | 120 | | | |
| X ² | 49 | 64 | 49 | 64 | 121 | 144 | 121 | 100 | 100 | 81 | 49 | 100 | 100 | 1142 | | | |

Hasil perhitungan tingkat reliabilitas, selanjutnya dianalisis menggunakan kategori tingkat reliabilitas menurut Djiwandono, sebagaimana tabel 3.

Tabel 3. Kategori Tingkat Reliabilitas Menurut Djiwandono:

| Tingkat Realibilitas | = | 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 |

Berdasarkan hasil perhitungan reliabilitas dan Mengacu pada pendapat Djiwandono, maka dapat dikatakan bahwa tingkat realibilitas tes soal pilihan ganda Ulangan Umum Semester Ganjil TP. 2018/2019 bidang studi Bahasa Arab siswa kelas XI-IPS MA Roudlotul Ulum Pagak Malang ini adalah (0,54) menunjukkan kategori **Sedang**.

3. Tingkat Kesukaran

Tingkat perhitungan soal adalah pengukuran seberapa besar derajat kesukaran suatu soal. Jika suatu soal memiliki tingkat kesukaran seimbang (proposional), maka dapat dikatakan bahwa soal tersebut baik.¹⁴ Dalam menganalisis tingkat kesukaran bisa menggunakan tiga cara yaitu: *Pertama*, Menggunakan rumus Tingkat Kesukaran (TK). *Kedua*, Menggunakan table batas tingkat kesukaran. *Ketiga*, menggunakan Proporsi menjawab benar (*proportion correct*). Dan dalam hal ini peneliti menggunakan cara yang ketiga dengan rumus sebagai berikut:

$$\text{Rumus:}^{15} p = \frac{\sum B}{N}$$

Keterangan:

P = Tingkat Kesukaran

$\sum B$ = jumlah peserta didik yang menjawab benar

N = jumlah peserta didik

Untuk menafsirkan tingkat kesukaran dapat digunakan criteria sebagai berikut:

| | |
|-------------------------|----------|
| $P > 0,70$ | = mudah |
| $0,30 \leq p \leq 0,70$ | = sedang |
| $P < 0,30$ | = sukar |

Tabel 4 merupakan hasil tabulasi hasil jawaban siswa yang benar dan yang salah.

¹⁴ Zainal Arifin, ..., Hlm. 266

¹⁵ Zainal Arifin, ..., Hlm. 272

Tabel 4. Jawaban Benar (+) Salah (-)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|----|---|---|---|---|---|---|---|---|---|----|----|----|----|
| 1 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| 2 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| 3 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| 4 | - | - | + | - | - | + | - | - | - | - | - | - | - |
| 5 | - | + | + | + | + | + | + | + | + | + | + | + | + |
| 6 | + | + | - | + | + | + | + | + | + | + | + | + | + |
| 7 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 8 | + | + | + | + | + | + | + | + | + | + | - | + | + |
| 9 | - | + | - | - | + | + | + | + | - | + | - | - | - |
| 10 | - | - | - | - | + | + | + | + | + | - | - | + | + |
| 11 | - | + | - | - | + | + | + | + | + | + | + | + | + |
| 12 | + | - | - | - | + | + | + | - | + | - | - | + | + |
| 13 | - | - | - | + | - | - | - | - | - | - | - | - | - |
| 14 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 15 | + | - | + | + | + | + | + | + | + | + | + | + | + |

Berdasarkan hasil tabulasi jawaban benar dan salah sebagaimana tabel 4, maka tingkat kesukaran tiap soal dapat dihitung sebagai berikut:

- 1) Untuk soal nomer 1 : $P = \frac{13}{13} = 1$
- 2) Untuk soal nomer 2 : $P = \frac{13}{13} = 1$
- 3) Untuk soal nomer 3 : $P = \frac{13}{13} = 1$
- 4) Untuk soal nomer 4 : $P = \frac{2}{13} = 0.152$
- 5) Untuk soal nomer 5 : $P = \frac{12}{15} = 0.923$
- 6) Untuk soal nomer 6 : $P = \frac{12}{13} = 0.923$

- 7) Untuk soal nomer 7 : $P = \frac{0}{13} = 0$
- 8) Untuk soal nomer 8 : $P = \frac{12}{13} = 0.923$
- 9) Untuk soal nomer 9 : $P = \frac{8}{13} = 0.615$
- 10) Untuk soal nomer 10 : $P = \frac{9}{13} = 0.692$
- 11) Untuk soal nomer 11 : $P = \frac{10}{12} = 0.769$
- 12) Untuk soal nomer 12 : $P = \frac{9}{13} = 0.629$
- 13) Untuk soal nomer 13 : $P = \frac{1}{13} = 0.076$
- 14) Untuk soal nomer 14 : $P = \frac{0}{13} = 0$
- 15) Untuk soal nomer 15 : $P = \frac{12}{13} = 0.923$

Hasil penghitungan tingkat kesukaran tiap soal, selanjutnya dianalisis menggunakan tafsiran hasil tingkat kesukaran sebagaimana tabel 5.

Tabel 5. Penafsiran Hasil Tingkat Kesukaran

| No | Tingkat Kesukaran | Penafsiran |
|----|-------------------|------------|
| 1 | 1 | Mudah |
| 2 | 1 | Mudah |
| 3 | 1 | Mudah |
| 4 | 0.152 | Sukar |
| 5 | 0.923 | Mudah |
| 6 | 0.923 | Mudah |
| 7 | 0 | Sukar |
| 8 | 0.923 | Mudah |
| 9 | 0.615 | Sedang |
| 10 | 0.692 | Sedang |
| 11 | 0.769 | Mudah |

| | | |
|----|--------|--------|
| 12 | 0.629 | Sedang |
| 13 | 0.076 | Sukar |
| 14 | 0 | Sukar |
| 15 | 0.923 | Mudah |
| | Mudah | |
| | Sedang | |
| | Sukar | |

Berdasarkan tafsiran pengkategorian kesukaran sebagaimana tabel 5. Hasil analisis tingkat kesulitan soal pilihan ganda Ulangan Umum Semester Ganjil TP. 2018/2019 bidang studi Bahasa Arab siswa kelas XI-IPS MA Roudlotul Ulum Pagak Malang, dapat dilihat pada tabel 6.

Tabel 6. Hasil Analisis Tingkat Kesulitan Soal Pilihan Ganda

| Analisis Tingkat Kesulitan | Kategori | Jumlah | Keterangan/nomor soal |
|----------------------------|----------|--------|--------------------------|
| | Mudah | 8 | 1, 2, 3, 5, 6, 8, 11, 15 |
| | Sedang | 3 | 9, 10, 12 |
| | Sukar | 4 | 4, 7, 13, 14 |

Tabel 6 menunjukkan bahwa diantara soal no 1 sampai 15 terdapat 8 butir soal yang termasuk dalam kategori soal "*Mudah*", 3 butir soal yang termasuk dalam kategori soal "*Sedang*", dan 4 butir soal yang termasuk dalam kategori soal "*Sukar*".

4. Daya Pembeda

Daya pembeda soal adalah kemampuan sesuatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh (berkemampuan rendah).¹⁶

¹⁶ Suharsimi Arikunto, ..., Hlm. 211

Suatu kelas terdiri dari 13 orang teste, maka di ambil 27% sebagai kelompok atas dan 27% kelompok bawah. 4 orang dari kelompok atas dan 4 orang dari kelompok bawah untuk menentukan daya beda (D) suatu butir tes dapat digunakan dengan rumus berikut:¹⁷

$$D = \frac{\sum \text{Benar upper} - \sum \text{Benar lower}}{\sum \text{kelompok}} \times 100\%$$

Dalam menentukan rentangan indeks Daya Beda (D) dari suatu tes, dapat digunakan rumus yang dikemukakan oleh Djiwandono (1996: 144) sebagai berikut:

- $D > 0.50$: Baik
 $0.20 \leq D \leq 0.50$: Sedang
 > 0.20 : Kurang
 0 : tidak ada deskriminasi
 $-$ (negative) : negatif

Tabel 7 dan 8 merupakan hasil tabulasi jawaban siswa yang dikategorikan dalam kelompok atas dan kelompok bawah.

Tabel 7. Jawaban dari Kelompok Atas

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 2 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 3 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| 4 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 |

¹⁷ Imam Asrori, dkk. *Evaluasi Pembelajaran Bahasa Arab*. (Malang: Misykat, 2012). Hlm. 74

Tabel 8. Jawaban dari kelompok bawah

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 3 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| 4 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |

Berdasarkan pengelompokan atas dan bawah sebagaimana tabel 7 dan 8, dapat dianalisis:

- 1) Untuk soal nomor 1 pada kelompok bawah tidak ada yang salah, dan pada kelompok atas tidak ada yang salah juga.
- 2) Untuk soal nomor 2 pada kelompok bawah tidak ada yang salah, dan pada kelompok atas tidak ada yang salah juga.
- 3) Untuk soal nomor 3 pada kelompok bawah tidak ada yang salah, dan pada kelompok atas tidak ada yang salah juga.
- 4) Untuk soal nomor 4 pada kelompok bawah yang salah 3 orang, dan pada kelompok atas yang salah 3 orang.
- 5) Untuk soal nomor 5 pada kelompok bawah yang salah 1 orang, dan pada kelompok atas tidak ada yang salah.
- 6) Untuk soal nomor 6 pada kelompok bawah yang salah 1 orang, dan pada kelompok atas tidak ada yang salah.
- 7) Untuk soal nomor 7 pada kelompok bawah semuanya salah, dan pada kelompok atas semuanya juga salah.
- 8) Untuk soal nomor 8 pada kelompok bawah semua menjawab benar, dan pada kelompok atas semua menjawab benar.
- 9) Untuk soal nomor 9 pada kelompok bawah semuanya salah, dan pada kelompok atas yang salah 1 orang.
- 10) Untuk soal nomor 10 pada kelompok bawah yang salah 2 orang, dan pada kelompok atas yang salah 1 orang.

- 11) Untuk soal nomor 11 pada kelompok bawah yang salah 2 orang, dan pada kelompok atas jawabannya semua benar.
- 12) Untuk soal nomor 12 pada kelompok bawah yang salah 3 orang, dan pada kelompok atas salah semua.
- 13) Untuk soal nomor 13 pada kelompok bawah salah semua, dan pada kelompok atas salah semua.
- 14) Untuk soal nomor 14 pada kelompok bawah salah semua, dan pada kelompok atas benar semua.
- 15) Untuk soal nomor 15 pada kelompok bawah benar semua, dan pada kelompok atas benar semua.

Selanjutnya, dilakukan penghitungan tingkat daya beda sebagai berikut:

- | | |
|--|--|
| 1) $D = \frac{4-4}{4} \times 100 = 0$ | 9) $D = \frac{3-0}{4} \times 100 = 0.75$ |
| 2) $D = \frac{4-4}{4} \times 100 = 0$ | 10) $D = \frac{3-2}{4} \times 100 = 0.25$ |
| 3) $D = \frac{4-4}{4} \times 100 = 0$ | 11) $D = \frac{4-2}{4} \times 100 = 0.50$ |
| 4) $D = \frac{1-1}{4} \times 100 = 0$ | 12) $D = \frac{0-1}{4} \times 100 = -0.25$ |
| 5) $D = \frac{4-3}{4} \times 100 = 0.25$ | 13) $D = \frac{0-0}{4} \times 100 = 0$ |
| 6) $D = \frac{4-3}{4} \times 100 = 0.25$ | 14) $D = \frac{4-0}{4} \times 100 = 1$ |
| 7) $D = \frac{0-0}{4} \times 100 = 0$ | 15) $D = \frac{4-4}{4} \times 100 = 0$ |
| 8) $D = \frac{4-4}{4} \times 100 = 0$ | |

Berdasarkan penghitungan tingkat daya beda tersebut, dilakukan penafsiran hasil perhitungan sebagaimana tabel 9.

Tabel 9. Penafsiran Hasil Perhitungan Tingkat Daya Beda

| No | Tingkat Daya Beda | Penafsiran |
|----|-------------------|------------------------|
| 1 | 0 | Tidak ada deskriminasi |
| 2 | 0 | Tidak ada deskriminasi |
| 3 | 0 | Tidak ada deskriminasi |
| 4 | 0 | Tidak ada deskriminasi |
| 5 | 0.25 | Sedang |
| 6 | 0.25 | Sedang |
| 7 | 0 | Tidak ada deskriminasi |
| 8 | 0 | Tidak ada deskriminasi |
| 9 | 0.75 | Baik |
| 10 | 0.25 | Sedang |
| 11 | 0.50 | Baik |
| 12 | -0.25 | Negative |
| 13 | 0 | Tidak ada deskriminasi |
| 14 | 1 | Baik |
| 15 | 0 | Tidak ada deskriminasi |

Merujuk penafsiran pada tabel 9, maka dapat diketahui bahwa hasil analisis daya pembeda soal pilihan ganda Ulangan Umum Semester Ganjil TP. 2018/2019 bidang studi Bahasa Arab siswa kelas XI-IPS MA Roudlotul Ulum Pagak Malang, dapat dikategorikan sebagaimana tabel 10.

Tabel 10. Hasil Analisis Daya Pembeda Soal

| Analisis Daya Pembeda | Kategori | Jumlah | Keterangan/nomor soal |
|-----------------------|------------------------|--------|-----------------------------|
| | Baik | 2 | 9 dan 14 |
| | Sedang | 3 | 5, 6 dan 10 |
| | Tidak ada deskriminasi | 8 | 1, 2, 3, 4, 7, 8, 13 dan 15 |
| | Negatif | 1 | 12 |

Berdasarkan tabel 10. Didapati hasil bahwa dari butir soal no 1-15 yang termasuk dalam kategori soal yang baik adalah soal no 9, 11 dan 14

dengan indeks 0.75, 0.50, dan 1. Dan yang termasuk dalam kategori soal yang sedang adalah butir soal no 5, 6, dan 10 dengan indeks semuanya 0.25.

Terdapat 3 butir soal yang termasuk dalam kategori soal yang baik, dan terdapat 3 butir soal juga yang termasuk dalam kategori soal sedang, sisanya adalah termasuk kategori soal yang tidak bisa membedakan antara siswa dari kelompok atas dan kelompok bawah. Bahkan tingkat daya bedanya adalah 0 (nol) sehingga tidak ada deskriminasi, Implikasinya, butir soal tersebut direvisi atau tidak digunakan.

5. Analisis Pengecoh

Pengecoh adalah alternatif jawaban (opsi) yang ada pada bentuk soal pilihan ganda. Butir soal yang baik, pengecohnya akan dipilih secara merata oleh peserta didik yang menjawab salah. Sebaliknya, butir soal yang kurang baik, pengecohnya akan dipilih secara tidak merata.¹⁸ Indeks Pengecoh dapat dihitung dengan rumus:¹⁹

$$IP = \frac{P}{(N - B) : (n - 1)} \times 100\%$$

Keterangan:

IP = indeks pengecoh

P = jumlah peserta didik yang memilih pengecoh

N = jumlah peserta didik yang ikut tes

B = jumlah peserta didik yang menjawab benar pada setiap soal

n = jumlah alternatif jawaban (opsi)

1 = bilangan tetap

Untuk kualitas pengecoh berdasarkan indeks pengecoh adalah:

| | | |
|--------------|----------------------|------------------------|
| Sangat baik | IP = 76% | - 125% |
| Baik | IP = 51% | - 75% atau 126% - 150% |
| Kurang baik | IP = 26% | - 50% atau 151% - 175% |
| Jelek | IP = 0% | - 25% atau 176% - 200% |
| Sangat jelek | IP = lebih dari 200% | |

¹⁸ Zainal Arifin, ..., Hlm. 279

¹⁹ Zainal Arifin, ..., Hlm. 279

Langkah pertama adalah mengetahui perhitungan alternatif jawaban, sebagaimana tabel 11.

Tabel 11. Perhitungan Alternatif Jawaban 1

| | a | b | c | d | e |
|----|----|----|----|---|---|
| 1 | 13 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 13 | 0 | 0 |
| 3 | 13 | 0 | 0 | 0 | 0 |
| 4 | 7 | 0 | 2 | 1 | 2 |
| 5 | 1 | 12 | 0 | 0 | 0 |
| 6 | 0 | 0 | 12 | 0 | 1 |
| 7 | 0 | 0 | 12 | 1 | 0 |
| 8 | 0 | 0 | 12 | 1 | 0 |
| 9 | 1 | 1 | 5 | 5 | 1 |
| 10 | 7 | 6 | 0 | 0 | 0 |
| 11 | 1 | 2 | 10 | 0 | 0 |
| 12 | 0 | 1 | 0 | 7 | 5 |
| 13 | 1 | 10 | 2 | 0 | 0 |
| 14 | 13 | 0 | 0 | 0 | 0 |
| 15 | 12 | 0 | 0 | 1 | 0 |

Setelah dihitung dengan menggunakan rumus sebagaimana disebutkan sebelumnya, maka di peroleh hasil dengan analisis sebagai berikut:

Keterangan kualitas pengecoh:

- ** : Kunci jawaban
- ++ : sangat baik
- + : baik
- : kurang baik
- : jelek
- : Sangat jelek

tabel 12 merupakan analisis hasil hitung indeks pengecoh untuk soal nomor 1 sampai 3.

Tabel 12. Analisis Indeks Pengecoh pada Soal Nomer 1-3

| No Soal | Alternatif Jawaban | a | b | c | d | e |
|---------|--------------------|----|----|----|----|----|
| 1 | Distribusi Jwb | 13 | 0 | 0 | 0 | 0 |
| | Indeks pengecoh | ** | 0% | 0% | 0% | 0% |
| | Kualitas pengecoh | ** | -- | -- | -- | -- |
| 2 | Distribusi Jwb | 0 | 0 | 13 | 0 | 0 |
| | Indeks pengecoh | 0% | 0% | ** | 0% | 0% |
| | Kualitas pengecoh | -- | -- | ** | -- | -- |
| 3 | Distribusi Jwb | 13 | 0 | 0 | 0 | 0 |
| | Indeks pengecoh | ** | 0% | 0% | 0% | 0% |
| | Kualitas pengecoh | ** | -- | -- | -- | -- |

Berdasarkan tabel 12, didapatkan hasil bahwa pada soal no 1, IP butir b, c, d, dan e adalah 0%. Semuanya perlu diganti karena termasuk pengecoh yang jelek. Pada soal no 2, IP butir a, b, d, dan e adalah 0%. Semuanya perlu diganti karena termasuk pengecoh yang jelek. Begitupun pada soal no 3, IP butir b, c, d, dan e adalah 0%. Semuanya perlu diganti karena termasuk pengecoh yang jelek.

Selanjutnya dianalisis indeks pengecoh soal nomor 4 sampai dengan nomor 6 sebagaimana tabel 13. Merujuk hasil tabulasi perhitungan indeks pengecoh sebagaimana tabel 13, didapatkan hasil bahwa pada soal no 4, IP pengecoh (a) menyesatkan karena persentasenya di atas 200%, pengecoh (b) tidak berfungsi sama sekali karena mempunyai 0%, pengecoh (c) masih tergolong pengecoh yang kurang baik karena masih mempunyai 36%, dan pengecoh (e) termasuk dalam kategori pengecoh yang baik, maka pengecoh a, b, dan c harus direvisi. Selanjutnya, pada soal no 5, IP butir c, d, dan e adalah 0%. Pengecoh (a) adalah 400%. Semuanya perlu diganti karena termasuk pengecoh yang jelek. Dan, pada soal no 6, IP butir a, b, dan d adalah 0%.

Pengecoh (e) adalah 400%. Semuanya perlu diganti karena termasuk pengecoh yang jelek.

Tabel 13. Analisis Indeks Pengecoh pada Soal Nomer 4-6

| No Soal | Alternatif Jawaban | a | b | c | d | e |
|---------|--------------------|------|----|----|-----|------|
| 4 | Distribusi Jwb | 7 | 0 | 2 | 1 | 2 |
| | Indeks pengecoh | 255% | 0% | ** | 36% | 73% |
| | Kualitas pengecoh | --- | -- | ** | - | + |
| 5 | Distribusi Jwb | 1 | 12 | 0 | 0 | 0 |
| | Indeks pengecoh | 400% | ** | 0% | 0% | 0% |
| | Kualitas pengecoh | --- | ** | -- | -- | -- |
| 6 | Distribusi Jwb | 0 | 0 | 12 | 0 | 1 |
| | Indeks pengecoh | 0% | 0% | ** | 0% | 400% |
| | Kualitas pengecoh | -- | -- | ** | -- | --- |

Berikutnya, analisis hasil penghitungan indeks pengecoh pada soal nomor 7 sampai dengan nomor 9 sebagaimana tabel 14. Tabel 14 menunjukkan bahwa, pada soal no 7, IP butir a, dan b adalah 0%. pengecoh (c) adalah 369%, dan pengecoh (d) adalah 31%. Semuanya perlu diganti karena termasuk pengecoh yang jelek. Analisis: selanjutnya, pada soal no 8, IP butir a, b, dan e adalah 0%. Pengecoh (c) adalah 400%. Semuanya perlu diganti karena termasuk pengecoh yang jelek. Dan, pada soal no 9, IP butir a, b, dan e adalah 50%. Pengecoh (a) adalah 250%. Semuanya perlu diganti karena termasuk pengecoh yang jelek.

Tabel 14. Analisis Indeks Pengecoh pada Soal Nomer 7-9

| No Soal | Alternatif Jawaban | a | b | c | d | e |
|---------|--------------------|-----|-----|------|------|-----|
| 7 | Distribusi Jwb | 0 | 0 | 12 | 1 | 0 |
| | Indeks pengecoh | 0% | 0% | 369% | 31% | ** |
| | Kualitas pengecoh | -- | -- | --- | - | ** |
| 8 | Distribusi Jwb | 0 | 0 | 12 | 1 | 0 |
| | Indeks pengecoh | 0% | 0% | ** | 400% | 0% |
| | Kualitas pengecoh | -- | -- | ** | --- | -- |
| 9 | Distribusi Jwb | 1 | 1 | 5 | 5 | 1 |
| | Indeks pengecoh | 50% | 50% | ** | 250% | 50% |
| | Kualitas pengecoh | - | - | ** | --- | -- |

Analisis selanjutnya adalah analisis indeks pengecoh soal nomor 10 sampai dengan nomor 12, sebagaimana tabel 15.

Tabel 15. Analisis Indeks Pengecoh pada Soal Nomer 10-12

| No Soal | Alternatif Jawaban | a | b | c | d | e |
|---------|--------------------|------|------|----|----|------|
| 10 | Distribusi Jwb | 7 | 6 | 0 | 0 | 0 |
| | Indeks pengecoh | ** | 400% | 0% | 0% | 0% |
| | Kualitas pengecoh | ** | --- | -- | -- | -- |
| 11 | Distribusi Jwb | 1 | 2 | 10 | 0 | 0 |
| | Indeks pengecoh | 133% | 267% | ** | 0% | 0% |
| | Kualitas pengecoh | + | --- | ** | -- | -- |
| 12 | Distribusi Jwb | 0 | 1 | 0 | 7 | 5 |
| | Indeks pengecoh | 0% | 67% | 0% | ** | 333% |
| | Kualitas pengecoh | -- | + | -- | ** | --- |

Berdasarkan tabulasi penghitungan indeks pengecoh sebagaimana tabel 15, pada soal no 10, IP butir c, d, dan e adalah 0%. Pengecoh (b) adalah 400%. Semuanya perlu diganti karena termasuk pengecoh yang jelek. Sedangkan pada soal no 11, IP butir d dan e adalah 0%. pengecoh (b) adalah 267% Semuanya perlu diganti karena termasuk pengecoh yang jelek. Dan pengecoh (a) adalah 133% yang termasuk kualitas pengecoh baik. Dan, pada soal no 12, IP butir a dan c adalah 0%. Pengecoh (e) adalah 333%. Semuanya perlu diganti karena termasuk pengecoh yang jelek. Dan pengecoh (b) adalah 67% yang termasuk kualitas pengecoh baik.

Terakhir, dilakukan analisis indeks pengecoh pada soal nomor 13-15 sebagaimana tabel 16. Berdasarkan tabel 16 didapatkan hasil bahwa pada soal no 13, IP butir d dan e adalah 0%. Pengecoh (b) adalah 333%. Semuanya perlu diganti karena termasuk pengecoh yang jelek. Dan pengecoh (c) adalah 67% yang termasuk kualitas pengecoh baik. Selanjutnya, Pada soal no 14, IP butir b, d, dan e adalah 0%. Pengecoh (a) adalah 400%. Semuanya perlu diganti karena termasuk pengecoh yang jelek. Terakhir, Pada soal no 15, IP butir b, c, dan e adalah 0%. Pengecoh (d) adalah 400% Semuanya perlu diganti karena termasuk pengecoh yang jelek.

Tabel 16. Analisis indeks pengecoh pada soal nomer 14

| No Soal | Alternatif Jawaban | a | b | c | d | e |
|---------|--------------------|------|------|-----|------|----|
| 13 | Distribusi Jwb | 1 | 10 | 2 | 0 | 0 |
| | Indeks pengecoh | ** | 333% | 67% | 0% | 0% |
| | Kualitas pengecoh | ** | --- | + | -- | -- |
| 14 | Distribusi Jwb | 13 | 0 | 0 | 0 | 0 |
| | Indeks pengecoh | 400% | 0% | ** | 0% | 0% |
| | Kualitas pengecoh | --- | -- | ** | -- | -- |
| 15 | Distribusi Jwb | 12 | 0 | 0 | 1 | 0 |
| | Indeks pengecoh | ** | 0% | 0% | 400% | 0% |
| | Kualitas pengecoh | ** | -- | -- | --- | -- |

E. Kesimpulan

Hasil penelitian dalam menganalisis butir soal pilihan ganda Ulangan Umum Semester Ganjil TP. 2014/2015 bidang studi Bahasa Arab siswa kelas XI-IPS MA Roudlotul Ulum Pagak Malang, dapat disimpulkan beberapa poin. *Pertama*, tingkat validitas soal no 1-15 masih dikategorikan dalam soal yang kurang valid, sehingga untuk semester berikutnya agar guru mengganti soal tersebut. *Kedua*, dalam tingkat reliabilitas soal ini masuk dalam kategori Sedang dengan nilai 0.54. untuk tingkat kesukarannya antara butir soal nomor 1-15 ada 8 butir soal yang *Mudah* dan 3 butir soal termasuk soal yang *Sedang* dan 4 butir soal yang termasuk dalam soal yang *Sukar*. *Ketiga*, untuk tingkatan daya pembeda soal ini masih banyak yang tidak bisa membedakan antara siswa dari kelompok atas dengan siswa kelompok bawah. Dan yang terakhir *keempat* adalah karena ini soal pilihan ganda maka harus menghitung kualitas dari pengecoh, dan telah ditemukan masih banyak pengecoh yang perlu direvisi. Dengan demikian, hasil analisis tersebut memberikan implikasi bagi guru, khususnya kepada guru pembuat soal, agar memperbaiki soal yang masih belum valid dan membuat pengecoh yang benar-benar bisa mengecoh siswa. *WaAllahu A'lam bisshowab*

Daftar Rujukan

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks.
- Arifin, Z. (2014). *Evaluasi Pembelajaran: Prinsip, Teknik, dan Prosedur*, Bandung: Remaja Rosda Karya.
- Arikunto, S. (2009). *Dasar-dasar Evaluasi Pendidikan (edisi revisi)*. Jakarta: Bumi Aksara.
- Asrori, I. dkk. (2012). *Evaluasi Dalam Pembelajaran Bahasa Arab. Malang: Misykat Indonesia*.
- Azwar, S. (2012). *Dasar-dasar psikometri*. Yogyakarta: Pustaka Pelajar.

Siti Fathimah Al Fathiyah

Darwyan Syah dkk. (2009). *Pengembangan Evaluasi Sistem Pendidikan Agama Islam*. Jakarta: Diadit Media.

Departemen Pendidikan dan Kebudayaan. (1996). *Kamus Besar Bahasa Indonesia*. Jakarta: Balai Pustaka.

Eveline Siregar, Hartini Nara. (2010). *Teori Belajar dan Pembelajaran*. Bogor: Ghalia Indonesia.

Sugiyono. (2011). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung; Alfabeta.

Tobari, H. (2015). *Evaluasi Soal-soal Penerimaan Pegawai Baru Dilengkapi dengan Hasil Penelitiannya*. Deepublish.

Yusuf, A. M. (2017). *Asesmen dan Evaluasi Pendidikan*. Prenada Media.