

Analisis Butir Soal Bahasa Arab di Madrasah Ibtida’iyah Bengkulu Tengah menggunakan Metode Distinguishing, Difficulty, dan Dispersion

Syintia Dwi Dinanti¹, Prayoga Saputra², Risda Aprilia³ & Nur Qomari⁴

^{1,2,3,4}Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

Corresponding E-mail: 220104210002@student.uin-malang.ac.id

Abstrak : Penelitian ini bertujuan untuk menganalisis butir soal dengan fokus pada validitas, reliabilitas, tingkat kesulitan, daya pembeda, dan analisis pengecoh butir soal. Metode yang digunakan adalah pendekatan kuantitatif dengan pengumpulan data dari lembar jawaban siswa kelas IV Semester 1 di MIN 1 Bengkulu Tengah. Hasil evaluasi menunjukkan bahwa dari 25 butir soal yang diuji, 21 butir soal dinyatakan valid berdasarkan analisis *Product Moment*, sementara 4 butir soal tidak valid. Lebih lanjut, dari 25 butir soal, 21 butir soal dinilai reliabel, sementara 4 butir soal tidak reliabel berdasarkan koefisien *Alpha Cronbach*. Tingkat kesulitan butir soal menunjukkan bahwa sebagian besar butir soal berada pada tingkat kesulitan sedang (3 butir soal) dan sukar (22 butir soal), sementara tidak ada butir soal yang tergolong mudah. Analisis daya pembeda menunjukkan bahwa sebagian besar butir soal memiliki daya pembeda yang baik. Analisis pengecoh butir soal menunjukkan bahwa sebagian besar butir soal memiliki pengecoh yang diterima dengan baik, namun ada beberapa butir soal yang perlu direvisi. Dari hasil evaluasi ini, disimpulkan bahwa evaluasi butir soal Bahasa Arab di Madrasah Ibtida’iyah perlu ditingkatkan. Evaluasi ini diharapkan dapat memberikan wawasan yang mendalam tentang kelemahan dan potensi perbaikan yang diperlukan untuk meningkatkan kualitas pendidikan Bahasa Arab di tingkat MI.

Kata Kunci: Distinguishing; Difficulty; Dispersion; Evaluasi; Bahasa Arab

المستخلص

يهدف هذا البحث إلى تحليل الأسئلة مع التركيز على الصدق والثبات ومستوى الصعوبة والقوة التمييزية والتحليل التشتيتي للأسئلة. الطريقة المستخدمة هي منهج كمي من خلال جمع البيانات من أوراق إجابات طلاب الفصل الرابع من الفصل الأول في مدرسة الإبتدئية باغكولو. تظهر نتائج التقييم أنه من بين الأسئلة الـ 25 التي تم اختبارها، تم الإعلان عن صحة 21 سؤالاً بناءً على تحليل لحظة المنتج، بينما كانت 4 أسئلة غير صالحة. علاوة على ذلك، من أصل 25 سؤالاً، تم تصنيف 21 سؤالاً على أنها موثوقة، في حين كانت 4 أسئلة غير موثوقة على أساس معامل ألفا كرونباخ. ويظهر مستوى صعوبة الأسئلة أن أغلبية الأسئلة متوسطة الصعوبة (3 أسئلة) وصعبة (22 سؤالاً)، بينما لا توجد أسئلة تصنف على أنها سهلة. يوضح تحليل القوة التمييزية أن معظم العناصر تتمتع بقوة تمييزية جيدة. يوضح تحليل تشتت عناصر السؤال أن معظم عناصر السؤال تحتوي على تشتتات مقبولة جيداً، ولكن هناك بعض عناصر الأسئلة التي تحتاج إلى مراجعة. ومن نتائج هذا التقييم يستنتج أن تقييم أسئلة اللغة العربية في المدرسة الإبتدائية بحاجة إلى تحسين. ومن المأمول أن يوفر هذا التقييم نظرة متعمقة حول نقاط الضعف والتحسينات المحتملة اللازمة لتحسين جودة تعليم اللغة العربية على مستوى مدرسة الإبتدئية

كلمات مفتاحية: التمييز، الصعوبة، التشتت، التويم، اللغة العربية

PENDAHULUAN

Madrasah Ibtida'iyah (MI) merupakan lingkungan pendidikan yang memegang peranan penting dalam pembentukan kualitas pendidikan anak-anak dalam bidang keagamaan dan akademik.¹ Salah satu mata pelajaran yang menjadi fokus utama di MI adalah Bahasa Arab, yang menjadi landasan penting dalam memahami ajaran agama Islam serta memperluas pengetahuan siswa dalam aspek bahasa dan kesusasteraan.² Evaluasi butir soal Bahasa Arab menjadi suatu hal yang esensial dalam menilai efektivitas dan keberhasilan proses pendidikan di MI.³ Butir soal merupakan komponen utama dalam asesmen pembelajaran, yang tidak hanya mencerminkan pemahaman siswa terhadap materi pelajaran, tetapi juga menjadi tolok ukur kualitas pengajaran yang diberikan oleh pendidik.

Evaluasi merupakan suatu proses penilaian atau peninjauan terhadap suatu hal. Dalam arti yang lebih sederhana, evaluasi adalah cara kita untuk menilai atau mengetahui sejauh mana sesuatu berjalan dengan melihat dan mempertimbangkan segala hal yang terlibat di dalamnya.⁴ Sedangkan analisis butir tes adalah proses untuk mengevaluasi kualitas dan karakteristik butir-butir tes. Salah satu metode yang umum digunakan adalah analisis *distinguishing index*, *difficulty index*, dan *disperity index*. *Distinguishing index* (indeks *distinguishability*): Indeks ini digunakan untuk mengukur sejauh mana suatu butir tes dapat membedakan antara siswa yang mampu dan tidak mampu. Indeks ini mengacu pada proporsi siswa yang menjawab dengan benar antara kelompok atas dan kelompok bawah. Semakin tinggi nilai indeks, semakin baik butir tes tersebut dalam membedakan antara siswa yang mampu dan tidak mampu.

Difficulty index (indeks kesulitan): Indeks ini digunakan untuk mengukur tingkat kesulitan suatu butir tes. Indeks ini dihitung dengan membandingkan proporsi siswa yang menjawab dengan benar terhadap total siswa yang mengambil tes. Semakin tinggi nilai indeks, semakin sulit butir tes tersebut. *Dispersion index* (indeks dispersi): Indeks ini digunakan untuk mengukur sejauh mana jawaban siswa terhadap suatu butir tes tersebar secara merata. Indeks ini mengacu pada seberapa jauh jawaban siswa dari posisi tengah (misalnya, mean, median). Semakin tinggi nilai indeks, semakin besar penyebaran jawaban siswa.⁵

Dalam konteks ini, "Evaluasi tuntas butir soal Bahasa Arab di Madrasah Ibtida'iyah" merujuk pada upaya komprehensif untuk menelaah dan mengevaluasi kualitas, kecocokan dengan kurikulum, kebermaknaan, serta keberlakuan butir soal yang digunakan dalam pembelajaran Bahasa Arab di MIN 1 Bengkulu Tengah. Evaluasi ini dilakukan dengan tujuan untuk memastikan bahwa butir soal yang digunakan adalah representatif, relevan, dan mampu mengukur pemahaman serta kemampuan siswa sesuai dengan standar yang diharapkan. Melalui evaluasi tuntas butir soal Bahasa Arab di MI, diharapkan dapat teridentifikasi kelemahan atau kekurangan dalam perancangan butir soal, sehingga

¹ Abu Dharin, "Pendidikan Karakter Berbasis Komunikasi Edukatif Religius (KER) Di Madrasah Ibtida'iyah" (2021).

² Lady Farah Aziza and Ariadi Muliensyah, "Keterampilan Berbahasa Arab Dengan Pendekatan Komprehensif," *El-Tsaqafah: Jurnal Jurusan PBA* 19, no. 1 (2020): 56–71.

³ Siti Shalihah, "Teori Evaluasi Pengajaran Bahasa Arab Di MI," *Primary: Jurnal Keilmuan dan Kependidikan Dasar* 4, no. 2 (2012): 143–158.

⁴ Tim Redaksi Kamus Besar Bahasa Indonesia, "Kamus Besar Bahasa Indonesia" (2018). Halaman 272

⁵ Bagiyono, "The Analysis of Difficulty Level and Discrimination Power of Test Items of Radiography Level 1 Examination" Vol. 16 No (2017).

memungkinkan untuk perbaikan yang signifikan guna meningkatkan efektivitas proses pembelajaran. Evaluasi ini juga bertujuan untuk memastikan bahwa butir soal yang digunakan sesuai dengan kebutuhan pembelajaran siswa serta memfasilitasi pencapaian tujuan pendidikan yang diinginkan.

Penelitian sebelumnya menginvestigasi hasil ulangan semester mata pelajaran Bahasa Arab siswa kelas XI-IPS di MA Roudlotul Ulum Pagak Malang pada Tahun Pelajaran 2018/2019. Tujuannya adalah untuk membantu pengajar Bahasa Arab dalam menciptakan alat ukur yang efektif untuk mengukur pemahaman siswa. Metode penelitian yang digunakan adalah pendekatan kuantitatif dengan menggunakan dokumen lembar jawaban dan soal siswa sebagai sumber data. Implikasi dari penelitian ini adalah perlunya revisi pada butir soal yang dinilai tidak valid, serta butir soal yang memiliki daya pembeda rendah atau tidak ada.

Butir soal pengecoh juga perlu diperhatikan untuk meningkatkan kualitasnya. Keseluruhan temuan menyarankan perlunya perbaikan pada butir soal tersebut, bahkan ada kemungkinan untuk meninjau kembali pemakaian butir soal yang tidak memenuhi kriteria evaluasi yang diharapkan.⁶ Dengan demikian, pendekatan evaluasi tuntas butir soal Bahasa Arab di Madrasah Ibtida'iyah diharapkan dapat memberikan wawasan mendalam tentang kekuatan, kelemahan, dan potensi perbaikan yang diperlukan untuk meningkatkan mutu pendidikan Bahasa Arab di tingkat MI.

METODE

Untuk menjawab rumusan masalah yang ada maka penelitian ini memilih menggunakan pendekatan Kuantitatif. Dengan melalui beberapa tahapan diantaranya (1) Perumusan masalah (2) Menganalisa (3) Menguji. (Abdussamad, 2021) Penelitian ini menggunakan metode kuantitatif, yang merujuk pada sebuah metode riset yang berakar pada filsafat positivisme. Metode teknik pengumpulan data melibatkan penggunaan instrumen penelitian serta menerapkan analisis data yang bersifat kuantitatif dan statistik, dengan tujuan untuk menguji hipotesis yang sudah ditetapkan.⁷ Data yang digunakan dalam penelitian diperoleh dari lembar jawaban serta soal siswa kelas IV Semester 1 di MIN 1 Bengkulu Tengah yang berjumlah 21 siswa. Data tersebut kemudian dianalisis menggunakan teknik analisis dan dijabarkan secara rinci berdasarkan tingkat validitas, reliabilitas, tingkat kesulitan, daya pembeda, serta analisis pengecoh pilihan ganda pada setiap butir soal menggunakan aplikasi SPSS dan menggunakan Excel.

HASIL DAN PEMBAHASAN

1. Validitas Butir Soal

Suatu instrumen bisa dikatakan valid untuk menentukan sebuah tes adalah memiliki dampak yang betul-betul bisa untuk mengukur dan menilai apa yang ingin dinilai dengan baik dan benar.⁸ Dalam konteks ini, untuk menentukan apakah butir soal dapat dianggap valid atau tidak, peneliti menggunakan rumus analisis validitas yang berdasarkan pada daya pembeda. Rumus ini menggunakan indeks sebagai ukuran untuk menilai sejauh mana butir

⁶ Siti Fathimah Al Fathiyah, "Analisis Butir Soal Pelajaran Bahasa Arab Di MA Roudlotul Ulum Pagak Malang," *Tarbiyatuna: Jurnal Pendidikan Ilmiah* 4, no. 1 (2019): 77.

⁷ Dr Sugiyono, "Metode Penelitian Kuantitatif Dan R&D," *Bandung: Alfabeta* (2010): hlm. 8.

⁸ Saifuddin Azwar, "Dasar-Dasar Psikometri (Edisi Kedua)," *Yogyakarta: Pustaka Belajar* (2012).

soal tersebut dapat membedakan antara siswa yang memiliki pengetahuan yang berbeda dalam materi ujian.⁹ Dalam hal ini pengujian validitas menggunakan Product Moment yang mana bisa menguji validitas dengan aplikasi SPSS.

Uji validitas product moment pearson correlation menggunakan prinsip mengkorelasikan atau menghubungkan antara masing-masing skor item atau soal dengan skor total yang diperoleh dari jawaban responden atas kuesioner. Dasar Pengambilan Keputusan Uji Validitas Product Moment yaitu:

- a. Membandingkan nilai r hitung dengan nilai r table
 1. Jika nilai r hitung \bar{r} table, maka item soal angket tersebut dinyatakan valid.
 2. Jika nilai r hitung \bar{r} table, maka item soal angket tersebut dinyatakan tidak valid.
- b. Membandingkan Nilai Sig. (2-tailed) dengan Probabilitas 0,05
 1. Jika nilai Sig. (2-tailed) dan Pearson Correlation bernilai positif, maka item soal angket tersebut valid.
 2. Jika nilai sig. (2-tailed) $\bar{0},05$ dan Pearson Correlation bernilai negative, maka item soal angket tersebut tidak valid.
 3. Jika nilai sig. (2-tailed) $\bar{0},05$, maka item soal angket tersebut valid.

Adapun rumus manual uji Validitas menggunakan Product Moment yaitu:

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Keterangan
 r_{xy} : Koefisien validitas
 N : Banyaknya subjek
 X : Nilai pembanding
 Y : Nilai dari instrument yang akan dicari validitasnya.

Dalam penelitian ini peneliti menggunakan SPSS untuk menguji validitas butir soal yang ada di Madrasah Ibtida'iyah Negeri 1 Bengkulu Tengah Kelas 4 yaitu:

⁹ Mary J Allen and Wendy M Yen, "Introduction to Measurement Theory. Monterey, CA: Brooks" (Cole Publishing Company, 1979).

Tabel. 1 Analisis Validitas Butir Soal

No Butir Soal	Hasil r Hitung	Hasil r Tabel (0.05)	Kriteria Validasi Butir Soal
1	1.75	0.41	Valid
2	4.08	0.41	Valid
3	0.79	0.41	Valid
4	0.91	0.41	Valid
5	2.83	0.41	Valid
6	3.61	0.41	Valid
7	2.04	0.41	Valid
8	3.04	0.41	Valid
9	0.28	0.41	Tidak Valid
10	0.28	0.41	Tidak Valid
11	0.91	0.41	Valid
12	0.00	0.41	Tidak Valid
13	0.55	0.41	Valid
14	2.04	0.41	Valid
15	1.13	0.41	Valid
16	2.04	0.41	Valid
17	4.17	0.41	Valid
18	3.61	0.41	Valid
19	3.35	0.41	Valid
20	1.67	0.41	Valid
21	4.40	0.41	Valid
22	1.38	0.41	Valid
23	0.28	0.41	Tidak Valid
24	1.13	0.41	Valid
25	1	0.41	Valid

Menurut hasil pengujian validitas yang diuji menggunakan uji Validitas Product Moment pada butir soal Bahasa Arab di Madrasah Ibtida'iyah Negeri Bengkulu Tengah kelas 4 adalah jumlah validitas soal ada 21 soal dan 4 soal tidak valid berdasarkan hasil r hitung. Dinyatakan valid jika signifikan butir soal melebihi 0,41 dengan tumpuan 0,05, dan dinyatakan tidak valid jika hasil r hitung kurang dari 0,41 dengan tumpuan 0,05. Dalam hal ini butir soal layak digunakan guru dikelas karena dinyatakan valid, lalu untuk butir soal yang tidak valid maka akan dilakukan evaluasi perombakan soal yang tidak valid menjadi valid.

Menurut Sugiyono,¹⁰ validasi data penelitian dapat dilihat sebagai suatu proses sistematis yang akurat mengenai sejauh mana variabel penelitian mengintegrasikan proses penelitian dan objek penelitian dengan menggunakan data berbeda yang dikatakan peneliti. Pada saat yang sama, Utama berpendapat bahwa dukungan data dalam konteks desain penelitian adalah efektif, yang tidak lain adalah kualitas surat kepada para ilmuwan untuk menjelaskan fenomena dan realitas dunia. Keyakinan mengacu pada fakta atau pendapat yang dihasilkan oleh penelitian. Maka dari itu peneliti menguji validitas butir soal yang ada di kelas 4 untuk menguji keabsahan soal.

2. Reabilitas Butir Soal

Uji reliabilitas digunakan untuk mengetahui apakah alat pengumpulan data digunakan Hal ini menunjukkan tingkat ketelitian, tingkat ketelitian, kestabilan atau persistensi dan paparan gejala tertentu pada kelompok individu meskipun diterapkan pada waktu yang berbeda. Menurut Suharsimi Arikunto kepercayaan merujuk pada pemahaman dan alat yang cukup handal untuk digunakan sebagai alat pengumpul data karena alatnya bagus. Aplikasi Data yang dapat diandalkan akan menghasilkan data yang dapat diandalkan. Kepercayaan diri kita bisa percaya. Tujuan dari kepercayaan adalah untuk memahami hal itu Alat ini sangat handal untuk digunakan sebagai alat pengumpul data karena alatnya bagus.

Koefisien alfa Cronbach adalah statistik yang umum digunakan menguji reliabilitas instrumen survei. Alat penelitian diindikasikan memiliki tingkat kepercayaan yang cukup jika koefisien Alpha Cronbach lebih besar atau sama dengan 0,60. Sebuah metode untuk mengukur kereabelitan Butir Soal. Berikut adalah rumus dari Alpha Cornbach secara manual:

$$r_{11} = \left[\frac{k}{(k-1)} \right] \left[1 - \frac{\Sigma \sigma_b^2}{\sigma_t^2} \right]$$

Keterangan :

- r_{11} = koefisien reliabilitas instrument (total tes)
- k = jumlah butir pertanyaan yang sah
- $\Sigma \sigma_b^2$ = jumlah varian butir
- σ_t^2 = varian skor total

Dalam penelitian ini peneliti menggunakan aplikasi SPSS dalam mengukur kereabelitan suatu butir soal dengan menguji menggunakan Alpha Cornbach. Keputusan uji reliabilitas ditentukan dengan ketentuan sebagai berikut:

1. Jika r hitung lebih dari r tabel berarti item pertanyaan dikatakan reliabel
2. Jika r hitung kurang dari r tabel berarti item pertanyaan dikatakan tidak reliabel

Maka Hasil yang diperoleh peneliti dalam menguji butir soal adalah:

¹⁰ Sugiyono, *Memahami Penelitian Kualitatif*. Bandung, 2012.

Tabel 2 Reabilitas Butir Soal

No.	No. Item	Statistics Item		
		Prop. Correct	Biser	Point Biser
1	1	0,810	0,793	0,551
2	2	0,667	0,787	0,489
3	3	0,857	0,681	0,518
4	4	0,714	0,842	0,533
5	5	0,619	1,291	0,797
6	6	0,571	1,374	0,853
7	7	0,667	1,202	0,747
8	8	0,286	0,564	0,442
9	9	0,571	1,003	0,623
10	10	0,571	0,731	0,454
11	11	0,286	0,778	0,610
12	12	0,333	0,849	0,623
13	13	0,476	0,538	0,347
14	14	0,667	0,709	0,441
15	15	0,381	0,990	0,689
16	16	0,333	0,585	0,430
17	17	0,429	0,747	0,498
18	18	0,571	1,350	0,838
19	19	0,714	0,763	0,483
20	20	0,571	1,053	0,654
21	21	0,524	1,115	0,703
22	22	0,524	0,560	0,353
23	23	0,429	0,332	0,222
24	24	0,619	-0,001	-0,001
25	25	0,429	0,539	0,360

Dari hasil uji Reabilitas menggunakan SPSS dengan Alpha Cornbach yang digunakan peneliti dalam menguji reabilitas butir soal ialah 21 soal dinyatakan Reabilitas dan 4 soal dinyatakan tidak Reabilitas berdasarkan data yang sudah diujikan diatas. Menurut Suharsimi Arikunto, reliabilitas mengacu pada pemahaman bahwa alat tersebut cukup dapat diandalkan untuk digunakan sebagai alat pengumpul data karena alat tersebut baik. Alat yang andal akan menghasilkan data yang andal. Kepercayaan berarti kepercayaan. Tujuan dari reliabilitas adalah untuk memahami bahwa alat tersebut cukup dapat diandalkan untuk digunakan sebagai alat pengumpul data karena alat tersebut baik.¹¹

3. Tingkat Kesukaran

Pengukuran tingkat kesulitan suatu soal adalah penilaian sejauh mana tingkat kesulitan soal tersebut. Jika suatu soal memiliki tingkat kesulitan yang seimbang secara proporsional, maka dapat disimpulkan bahwa soal tersebut dinilai baik.¹² Dalam proses analisis tingkat kesulitan, ada tiga metode yang dapat diterapkan. Metode pertama adalah menggunakan rumus Tingkat Kesulitan (TK). Metode kedua adalah dengan menggunakan tabel batas tingkat kesulitan. Metode ketiga yang dipilih peneliti dalam analisis kali ini

¹¹ Arikunto Suharsimi, *Dasar-Dasar Evaluasi Pendidikan*, 2018.

¹² Zainal Arifin, *Evaluasi Pembelajaran*, vol. 8 (Bandung: Remaja Rosdakarya, 2009). Hlm. 266

adalah dengan memanfaatkan Proporsi Jawaban Benar (Proportion Correct) sebagai rumus yang digunakan pada aplikasi Excel adalah:

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

P : Tingkat kesukaran

$\sum B$: Jumlah Peserta didik yang menjawab benar

N : Jumlah peserta didik

Untuk menafsirkan tingkat kesukaran dapat digunakan criteria sebagai berikut:

$P \geq 0,70$ = mudah

$0,30 \leq p < 0,70$ = sedang

$P < 0,30$ = sukar

Maka dari itu hasil penelitian yang sudah di uji adalah:

Tabel. 3 Hasil Tingkat Kesukaran

No Soal	Hasil	Penafsiran
1	0,33	Sedang
2	0,80	Sedang
3	0,80	Sedang
4	1,19	Sedang
5	0,19	Sukar
6	0,80	Sedang
7	0,76	Sedang
8	0,71	Sedang
9	1,14	Sedang
10	0,42	Sedang
11	0,52	Sedang
12	0,80	Sedang
13	0,28	Sukar
14	0,85	Sedang
15	0,42	Sedang
16	0,76	Sedang
17	0,38	Sedang
18	0,28	Sukar
19	0,33	Sedang
20	1,19	Sedang
21	0,57	Sedang
22	0,33	Sedang
23	0,80	Sedang
24	0,80	Sedang

25	1,19	Sedang
Sukar		
Sedang		

Tabel. 4 Hasil Analisis Tingkat Kesulitan Soal

	Kategori	Jumlah	Keterangan
Analisis Tingkat Kesukaran	Sukar	22	1,2,3,4,6,7,8,9,10,11,12,14,16,17,19,20,21,22,23,24,25
	Sedang	3	5,13,19
	Mudah	0	0

Berdasarkan hasil uji tingkat kesukaran butir soal maka dinyatakan bahwa tingkat kesukaran butir soal dengan kategori sedang 3 butir soal, dengan kategori sukar 22 soal, dan dengan kategori mudah 0 soal. Menurut Anas Sudijono,¹³ observasi itu penting. Dibatasi setelah dilakukan analisa tingkat kesulitan sebagai berikut: a) Tingkat kesulitannya tergantung pada hasil tes pada bagian yang benar. (kesulitan sedang atau sedang), lebih disukai itemnya disimpan di bank soal. Pertanyaan-pertanyaan ini dapat digunakan lagi pada tes berikutnya. b) Definisi tingkat kesukaran soal berdasarkan analisis jenis kuat. Maka dari itu dari hasil uji daya kesukaran butir soal yang ada di Madrasah Ibtidaiyah Bengkulu Tengah Kelas 4 harus dievaluasi dikarenakan tingkat kesukaran soal lebih banyak dan mendominasi butir soal dalam 25 soal yang telah diteliti.

4. Daya Pembeda

Daya pembeda suatu soal adalah kemampuan suatu soal dalam membedakan antara siswa cerdas (berdaya tinggi) dan siswa cerdas (berdaya rendah). rumus berikut dapat digunakan:

$$DP = \frac{JB_A - JB_B}{JS_A} \text{ atau } DP = \frac{|JB_A - JB_B|}{JS_B}$$

Dengan:

JB_A = jumlah siswa kelompok atas yang menjawab soal dengan benar

JB_B = jumlah siswa kelompok bawah yang menjawab soal dengan benar

JS_A = jumlah siswa kelompok atas

JS_B = jumlah siswa kelompok bawah

¹³ Sudjiono Anas, *Pengantar Evaluasi Pendidikan*, 2011.

Untuk menentukan indeks kekuatan beda (D) pengujian, Metode yang dikemukakan Djiwandono dapat digunakan sebagai berikut:

$D \geq 0,50$: Baik

$0,20 \leq D < 0,50$: Sedang

$D < 0,20$: lebih sedikit

0: tidak ada diskriminasi - (negatif): negative

Akan tetapi peneliti memilih menggunakan Excel dalam pengolahan data uji data butir soal daya pembeda. Berikut hasil data butir soal daya pembeda:

Tabel 6 Analisis Tingkat Daya Beda

No Soal	Hasil Daya Beda	Keterangan
1	0,36	Cukup
2	0,44	Baik
3	0,27	Cukup
4	0,35	Cukup
5	0,72	Baik Sekali
6	0,81	Baik Sekali
7	0,63	Baik
8	0,21	Cukup
9	0,62	Baik
10	0,43	Baik
11	0,40	Baik
12	0,50	Baik
13	0,23	Cukup
14	0,44	Baik
15	0,41	Baik
16	0,31	Cukup
17	0,32	Cukup
18	0,81	Baik Sekali
19	0,54	Baik
20	0,62	Baik
21	0,71	Baik Sekali
22	0,14	Jelek
23	0,13	Jelek
24	-0,22	Jelek Sekali
25	0,32	Cukup

Tabel 7 Analisis daya beda

	Kategori	Jumlah	Keterangan
Analisis Daya Pembeda	Baik Sekali	4	5,6,18,21
	Baik	10	2,7,9,10,11,12,14,15,19,20
	Cukup	8	1,3,4,8,13,16,17,25
	Jelek	2	22,23
	Jelek Sekali	1	24

Berdasarkan hasil uji Butir soal dengan daya beda menggunakan excel dapat disimpulkan bahwa butir soal daya beda dibagi menjadi 5 kriteria yaitu Baik sekali, Baik, cukup, jelek, dan jelek sekali. Butir soal dengan daya pembeda ini menurut hasil ada 4 butir soal dengan keterangan nomor soal 5,6,18,21 dinyatakan kategori baik sekali, 10 butir soal dengan keterangan nomor 2,7,9,10,11,12,14,15,19,20 dinyatakan kategori baik, 8 butir soal dengan keterangan nomor 1,3,4,8,13,16,17,25 dinyatakan kategori cukup, 2 butir soal dengan keterangan nomor 22,23 dinyatakan kategori jelek, sedangkan ada 1 jumlah butir soal dengan keterangan nomor 24 dinyatakan kategori jelek sekali. Dalam hal ini peneliti menyarankan untuk mengevaluasi beberapa butir soal yang dinyatakan cukup, jelek bahkan jelek sekali.

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. Pengecoh soal ini diberikan guru kepada murid agar murid lebih teliti dengan butir soal yang telah diberikan. Rumus pengecoh butir soal ini dapat dilihat dari:

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

Keterangan:

IP : Banyaknya peserta tes yang memilih pengecoh tertentu

N : Banyaknya seluruh peserta tes

B : Banyaknya peserta tes yang menjawab benar butir soal yang bersangkutan

N : Banyaknya alternatif jawaban

Akan tetapi peneliti memilih menggunakan Excel dalam pengolahan data uji data butir soal pengecoh. Berikut hasil data butir soal pengecoh:

Tabel 8 Butir Soal Daya Pengecoh

No Item	kunci	Jawaban				Status Soal
		A	B	C	D	
1	D	0	4	0	17	Dapat diterima
2	B	1	14	2	4	Dapat diterima
3	A	18	3	0	0	Dapat diterima

4	B	3	15	3	0	Dapat diterima
5	C	2	5	13	1	Dapat diterima
6	A	12	5	3	1	Dapat diterima
7	B	4	14	3	0	Dapat diterima
8	B	9	6	2	4	Soal Sebaiknya Direvisi
9	B	7	12	1	1	Dapat diterima
10	D	4	4	1	12	Dapat diterima
11	A	6	2	9	4	Soal Sebaiknya Direvisi
12	C	9	3	7	2	Soal Sebaiknya Direvisi
13	C	9	0	10	2	Dapat diterima
14	A	14	1	5	1	Dapat diterima
15	C	5	5	8	3	Dapat diterima
16	C	9	2	7	2	Soal Sebaiknya Direvisi
17	C	7	5	9	0	Dapat diterima
18	A	12	4	5	0	Dapat diterima
19	C	4	1	15	1	Dapat diterima
20	B	2	12	3	4	Dapat diterima
21	D	3	5	2	11	Dapat diterima
22	A	11	6	1	3	Dapat diterima
23	B	7	9	1	4	Dapat diterima
24	D	5	2	1	13	Ditolak / Jangan digunakan
25	C	4	2	9	6	Dapat diterima

Berdasarkan hasil uji butir soal daya pengecoh dapat dinyatakan bahwa dari 21 siswa dengan jumlah butir soal bahasa arab 25 maka dapat disimpulkan bahwa soal nomor 1,2,3,4,5,6,7,9,10,13,14,15,17,18,19,20,21,22,23,25 dapat diterima dan siswa mampu mengerjakan dengan baik. Sedangkan butir soal nomor 8,11,12,16,24 dapat direvisi oleh tim pendidik dikarenakan dari uji validitas, reabilitas, tingkat kesukaran, daya pembeda, serta daya pengecoh menyatakan bahwa butir soal tersebut harus direvisi.

KESIMPULAN

Hasil pengujian validitas yang diuji menggunakan uji Validitas Product Moment pada butir soal Bahasa Arab di Madrasah Ibtida'iyah Negeri Bengkulu Tengah kelas 4 adalah jumlah validitas soal ada 21 soal dan 4 soal tidak valid berdasarkan hasil r hitung. Dinyatakan valid jika signifikan butir soal melebihi 0,41 dengan tumpuan 0,05, dan dinyatakan tidak valid jika hasil r hitung kurang dari 0,41 dengan tumpuan 0,05. Dalam hal ini butir soal layak digunakan guru dikelas karena dinyatakan valid, lalu untuk butir soal yang tidak valid maka akan dilakukan evaluasi perombakan soal yang tidak valid menjadi valid. Hasil uji Reabilitas menggunakan SPSS dengan Alpha Cornbach yang digunakan peneliti dalam menguji reabilitas butir soal ialah 21 soal dinyatakan Reabilitas dan 4 soal dinyatakan tidak Reabilitas berdasarkan data yang sudah diujikan. Hasil uji tingkat kesukaran butir soal maka dinyatakan bahwa tingkat kesukaran butir soal dengan kategori sedang 3 butir soal, dengan kategori sukar 22 soal, dan dengan kategori mudah 0 soal. hasil uji Butir soal dengan daya beda menggunakan excel dapat disimpulkan bahwa butir soal daya beda dibagi menjadi 5 kriteria yaitu Baik sekali, Baik, cukup, jelek, dan jelek sekali. Hasil uji butir soal daya pengecoh dapat dinyatakan bahwa dari 21 siswa dengan jumlah butir soal bahasa arab 25 maka dapat disimpulkan bahwa soal nomor 1,2,3,4,5,6,7,9,10,13,14,15,17,18,19,20,21,22,23,25 dapat diterima dan siswa mampu mengerjakan dengan baik, selain dari itu butir soal tidak dapat diterima.

DAFTAR PUSTAKA

- Allen, Mary J, and Wendy M Yen. "Introduction to Measurement Theory. Monterey, CA: Brooks." Cole Publishing Company, 1979.
- Arifin, Zainal. *Evaluasi Pembelajaran*. Vol. 8. Bandung: Remaja Rosdakarya, 2009.
- Arikunto Suharsimi. *Dasar-Dasar Evaluasi Pendidikan*, 2018.
- Bagiyono. "The Analysis of Difficulty Level and Discrimination Power of Test Items of Radiography Level 1 Examination" Vol. 16 No (2017).
- Aziza, Lady Farah, and Ariadi Muliansyah. "Keterampilan Berbahasa Arab Dengan Pendekatan Komprehensif." *El-Tsaqafah: Jurnal Jurusan PBA* 19, no. 1 (2020): 56–71.
- Azwar, Saifuddin. "Dasar-Dasar Psikometri (Edisi Kedua)." *Yogyakarta: Pustaka Belajar* (2012).
- Dharin, Abu. "Pendidikan Karakter Berbasis Komunikasi Edukatif Religius (KER) Di Madrasah Ibtida'iyah" (2021).
- Fathiyah, Siti Fathimah Al. "Analisis Butir Soal Pelajaran Bahasa Arab Di MA Roudlotul Ulum Pagak Malang." *Tarbiyatuna: Jurnal Pendidikan Ilmiah* 4, no. 1 (2019): 77.
- Indonesia, Tim Redaksi Kamus Besar Bahasa. "Kamus Besar Bahasa Indonesia" (2018).
- Shalihah, Siti. "Teori Evaluasi Pengajaran Bahasa Arab Di MI." *Primary: Jurnal Keilmuan dan Kependidikan Dasar* 4, no. 2 (2012): 143–158.
- Sugiyono, Dr. "Metode Penelitian Kuantitatif Dan R&D." *Bandung: Alfabeta* (2010): 26–33.
- Sudjiono Anas. *Pengantar Evaluasi Pendidikan*, 2011.
- Sugiyono. *Memahami Penelitian Kualitatif*. Bandung, 2012.