

Kualitas Butir Soal dalam Evaluasi Bahasa Arab : Kajian Empiris pada PAS Program Intensif

Roghibah Jadwa Faradisi¹, Mahmudah², M. Baihaqi³, Ajeng Nadlifah⁴, Inayahfidah⁵

UIN Sunan Ampel Surabaya^{1,2,3,4,5}

E-mail : farahjadwa@gmail.com¹, mahmudahsg129@gmail.com², baihaqi@uinsby.ac.id³,
ajengnadlifah@gmail.com⁴, Inayahfidah@gmail.com⁵

Abstract

This study evaluates the quality of multiple-choice test items in the Final Semester Assessment (PAS) of the Arabic Intensive Program at UIN Sunan Ampel Surabaya. Using a descriptive quantitative approach, the research analyzes difficulty index, discrimination index, and distractor efficiency of 25 questions in the writing skill (maharah kitabah) domain. Results reveal that 52% of the questions are categorized as easy, 40% as moderate, and 8% as difficult. Regarding discrimination, only 4% of the items are highly effective, while 48% are categorized as poor, with 8% showing negative indices. Distractor analysis shows that 45% function effectively, while 30% fail to meet the criteria. The findings highlight imbalances in difficulty distribution, limited discrimination capacity, and inconsistent distractor efficiency. Recommendations include revising poorly functioning items to enhance evaluation accuracy and fairness, ultimately supporting effective Arabic language learning and assessment development.

Keywords: *Item analysis, Difficulty index, Discrimination index, Distractor effectiveness.*

Abstrak

Penelitian ini bertujuan untuk mengevaluasi kualitas soal pilihan ganda dalam Penilaian Akhir Semester (PAS) Program Intensif Bahasa Arab di UIN Sunan Ampel Surabaya. Dengan pendekatan kuantitatif deskriptif, penelitian ini menganalisis tingkat kesulitan soal, daya pembeda, dan efektivitas distraktor pada 25 soal keterampilan menulis (maharah kitabah). Hasilnya menunjukkan bahwa 52% soal tergolong mudah, 40% berada pada tingkat sedang, dan 8% termasuk sulit. Dari segi daya pembeda, hanya 4% soal yang dinilai sangat efektif, sementara 48% kurang baik, dan 8% bahkan memiliki indeks negatif. Analisis distraktor mengungkap bahwa 45% distraktor bekerja dengan baik, tetapi 30% lainnya tidak berfungsi sebagaimana mestinya. Temuan ini menunjukkan adanya ketidakseimbangan dalam tingkat kesulitan soal, keterbatasan dalam membedakan kemampuan siswa, serta efektivitas distraktor yang masih bervariasi. Oleh karena itu, diperlukan revisi terhadap soal-soal yang kurang optimal agar evaluasi lebih akurat dan adil, sehingga dapat mendukung proses pembelajaran dan penilaian bahasa Arab yang lebih efektif.

Kata kunci: *Analisis soal, Indeks kesukaran, Indeks diskriminasi, Efektivitas pengecoh, Efisiensi distraktor.*

PENDAHULUAN

Evaluasi dalam pembelajaran bahasa Arab bertujuan untuk memperoleh informasi yang tepat mengenai tingkat pencapaian kompetensi siswa sesuai dengan tujuan instruksional yang telah ditetapkan. Evaluasi ini memiliki beberapa peran, yaitu sebagai instrumen instruksional, administratif, dan bimbingan. Evaluasi instruksional berfungsi untuk menilai keberhasilan siswa dalam mencapai tujuan pembelajaran, sementara evaluasi administratif digunakan dalam pengambilan keputusan terkait kebijakan pendidikan. Adapun fungsi bimbingan bertujuan untuk memberikan arahan kepada siswa berdasarkan hasil evaluasi yang diperoleh¹. Evaluasi juga menjadi barometer untuk mengetahui kekurangan dalam proses pembelajaran, yang kemudian

dapat diperbaiki dan ditingkatkan²³⁴. Salah satu metode evaluasi yang umum digunakan adalah analisis soal sebagai instrumen pembelajaran. Tujuannya adalah memastikan soal memenuhi kriteria kualitas seperti validitas, reliabilitas, dan daya diskriminatif agar dapat mengukur hasil belajar siswa secara efektif dan efisien serta membantu guru dalam menyusun soal yang lebih baik dan memastikan evaluasi berjalan secara akurat dan sistematis.

Instrumen evaluasi dapat berupa tes dan non-tes, di mana tes pilihan ganda menjadi instrumen yang sering digunakan, mulai dari ulangan harian hingga seleksi masuk perguruan tinggi. Tes pilihan ganda terdiri dari satu jawaban benar dan beberapa pengecoh yang berfungsi untuk menilai pemahaman siswa. Di lapangan pendidikan, khususnya dalam konteks ujian dan evaluasi, soal pilihan ganda banyak digunakan untuk mengukur kemampuan dan pemahaman siswa terhadap suatu mata pelajaran tertentu.

Banyak penelitian sebelumnya meneliti efektivitas pertanyaan pilihan ganda untuk tujuan yang beragam, salah satunya adalah untuk memperbaikinya di masa depan. Beberapa penelitian tersebut antara lain adalah penelitian yang dilakukan oleh Yakout dkk. (2024) yang bertujuan

¹ Iis Susiawati, Raswan Raswan, and Dadan Mardani, "The Purpose and Functions of Evaluation of Arabic Learning in Educational Administration Review," *Tarlim : Jurnal Pendidikan Agama Islam* 5, no. 2 (2022): 115–26, <https://doi.org/10.32528/tarlim.v5i2.7613>.

² Mochammad Noor Akhmadi, "Analisis Butir Soal Evaluasi Tema 1 Kelas 4 Sdn Plumbungan Menggunakan Program Anates," *Ed-Humanistics : Jurnal Ilmu Pendidikan* 6, no. 1 (2021): 799–806, <https://doi.org/10.33752/ed-humanistics.v6i1.1464>.

³ Marie Tarrant, James Ware, and Ahmed M. Mohammed, "An Assessment of Functioning and Non-Functioning Distractors in Multiple-Choice Questions: A Descriptive Analysis," *BMC Medical Education* 9, no. 1 (2009), <https://doi.org/10.1186/1472-6920-9-40>.

⁴ Laela Umi Fatimah and Khairuddin Alfath, "Analisis Kesukaran Soal, Daya Pembeda Dan Fungsi Distraktor," *Jurnal Komunikasi Dan Pendidikan Islam* 8, no. 2 (2019): 37–64.

mengevaluasi kualitas soal pilihan ganda dengan menganalisis tingkat kesukaran, daya pembeda, dan efisiensi distraktor di bidang kesehatan. Dari 50 soal yang dianalisis, sebagian besar memiliki tingkat kesukaran yang sesuai, tetapi daya pembeda rata-rata rendah, menunjukkan kurangnya kemampuan soal dalam membedakan siswa berprestasi tinggi dan rendah. Selain itu, 75% pengecoh berfungsi dengan baik, tetapi 25% lainnya tidak efektif dan memerlukan revisi. Selain penelitian tersebut, penelitian yang serupa juga pernah dilakukan oleh Girish R. Chauhan dkk. (2023) yang bertujuan untuk menganalisis kualitas soal pilihan ganda berdasarkan tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Hasil penelitian menunjukkan bahwa soal dengan dua hingga tiga distraktor yang berfungsi memiliki tingkat kesukaran dan daya pembeda yang ideal. Sebaliknya, soal dengan distraktor yang tidak efektif cenderung terlalu mudah dan kurang mampu membedakan kemampuan siswa. Penelitian ini menyimpulkan bahwa distraktor yang berkualitas berperan penting dalam meningkatkan akurasi soal dalam mengevaluasi kemampuan siswa.

Pentingnya evaluasi alat ukur yang digunakan dalam proses pembelajaran, khususnya di lingkungan pendidikan tinggi mendorong peneliti untuk melakukan penelitian ini melalui evaluasi yang dilakukan pada ujian akhir semester yang selanjutnya disebut Penilaian Akhir Semester (PAS) pada Program Intensif Bahasa Arab di Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya, sebuah program khusus yang dirancang untuk meningkatkan kemampuan berbahasa Arab mahasiswa baru. Salah satu indikator keberhasilan program ini adalah kualitas soal ujian akhir semester yang digunakan untuk mengukur pencapaian kompetensi bahasa Arab mengingat soal sebagai alat evaluasi memiliki peran penting dalam menentukan sejauh mana tujuan pembelajaran telah dicapai.

Fenomena yang diangkat dalam penelitian ini mencakup adanya soal-soal yang terlalu sulit atau terlalu mudah, daya pembeda yang rendah, serta distraktor yang tidak efektif. Permasalahan ini tidak hanya menghambat penilaian kemampuan mahasiswa secara objektif tetapi juga mengurangi kualitas pembelajaran secara keseluruhan di mana seharusnya soal yang baik memiliki tingkat kesukaran yang seimbang, dengan distribusi yang tepat antara soal mudah, sedang, dan sulit, mampu membedakan antara siswa yang berprestasi tinggi dan rendah serta memiliki pilihan jawaban yang dapat mengecoh siswa yang kurang memahami materi. Namun, di sisi lain, analisis ini membuka peluang untuk mengidentifikasi area-area yang membutuhkan perbaikan, baik dalam desain soal maupun dalam strategi pengajaran yang mendukung pembelajaran bahasa Arab.

Penelitian ini didasarkan pada teori-teori evaluasi pendidikan dan konsep-konsep ilmiah terkait analisis butir soal. Menurut Suprananto (2012), analisis tingkat kesukaran membantu menentukan proporsi soal yang sesuai dengan kemampuan siswa, sedangkan daya pembeda menjadi indikator kemampuan soal untuk membedakan antara siswa berkemampuan tinggi dan rendah. Selain itu, menurut Daryanto (2007) teori mengenai distraktor yang efektif menyatakan bahwa pengecoh yang baik harus mampu menarik perhatian peserta tes yang kurang memahami konsep yang diuji. Dengan mengacu pada teori-teori ini, penelitian ini menawarkan pendekatan yang sistematis untuk meningkatkan kualitas soal ujian sebagai alat evaluasi pembelajaran.

Dalam penelitian ini, soal yang dianalisis berasal dari materi maharah kitabah dengan jenis soal pilihan ganda. Soal tersebut terdiri dari 25 butir soal dengan empat pilihan jawaban (a, b, c, dan d) yang belum terukur dan tidak dapat dipastikan keefektifitasannya. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kesukaran, daya pembeda, dan fungsi distraktor pada soal-soal tersebut sebagai langkah awal untuk meningkatkan kualitas evaluasi pembelajaran sehingga dapat memberikan gambaran komprehensif mengenai kualitas soal ujian akhir semester berdasarkan tiga aspek utama: tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Penelitian ini diharapkan dapat memberikan rekomendasi yang konkret untuk perbaikan soal, sehingga proses evaluasi pembelajaran menjadi lebih akurat dan adil, memberikan wawasan pedagogis yang mendukung pengajaran yang lebih efektif serta dapat kontribusi nyata bagi pengembangan kualitas evaluasi pembelajaran bahasa Arab, khususnya pada materi maharah kitabah. Hasil penelitian ini juga dapat menjadi dasar untuk memperbaiki sistem evaluasi di masa mendatang, baik dalam konteks program intensif maupun pembelajaran bahasa Arab secara umum.

METODOLOGI PENELITIAN

Penelitian ini menggunakan analisis dokumen dengan pendekatan deskriptif kuantitatif untuk mengevaluasi kualitas soal tes Penilaian Akhir Semester (PAS) pada Program Intensif Bahasa Arab Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya dalam bidang kemahiran menulisa. Data diperoleh dari soal, kunci jawaban, dan hasil tes mahasiswa dari kelas A2 yang terdiri dari 21 mahasiswa pada Desember 2024 melalui metode dokumentasi. Analisis dilakukan menggunakan program Excel untuk mengukur tingkat kesukaran, daya pembeda, serta efektivitas distraktor.

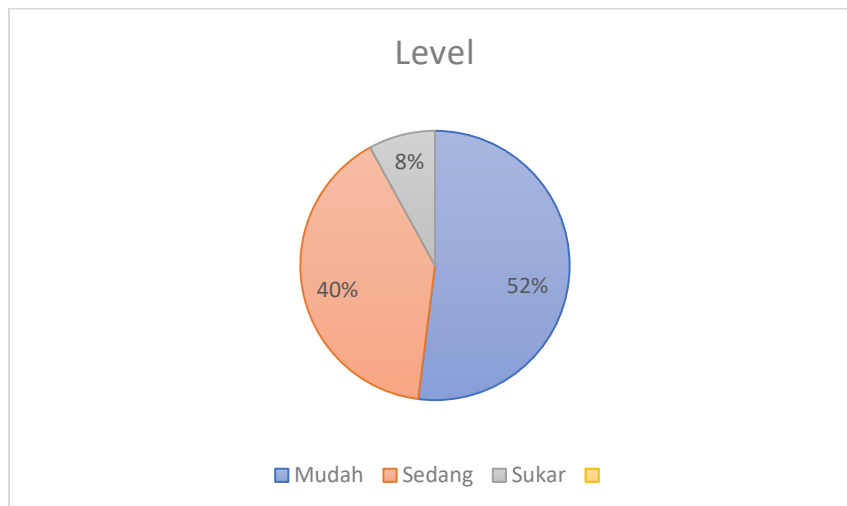
Tingkat kesukaran dikategorikan menjadi sukar ($P < 0,3$), sedang ($0,3 \leq P < 0,7$), dan

mudah ($P \geq 0,7$). Soal dengan tingkat kesukaran sedang ($P 0,3-0,7$) dianggap baik. Daya pembeda digunakan untuk menilai kemampuan soal dalam membedakan siswa berkemampuan tinggi dan rendah, dengan klasifikasi dari jelek ($D < 0,2$) hingga sangat baik ($D \geq 0,7$). Soal ideal memiliki daya pembeda dalam kategori baik ($D 0,4-0,69$).

Efektivitas distraktor diuji dengan melihat sejauh mana pengecoh menarik perhatian peserta tes yang kurang memahami materi. Distraktor dianggap efektif jika dipilih oleh minimal 5% peserta. Hasil penelitian ini diharapkan dapat meningkatkan kualitas soal dengan menyusun butir soal yang memenuhi standar tingkat kesukaran, daya pembeda, dan efektivitas distraktor.

HASIL PENELITIAN DAN PEMBAHASAN

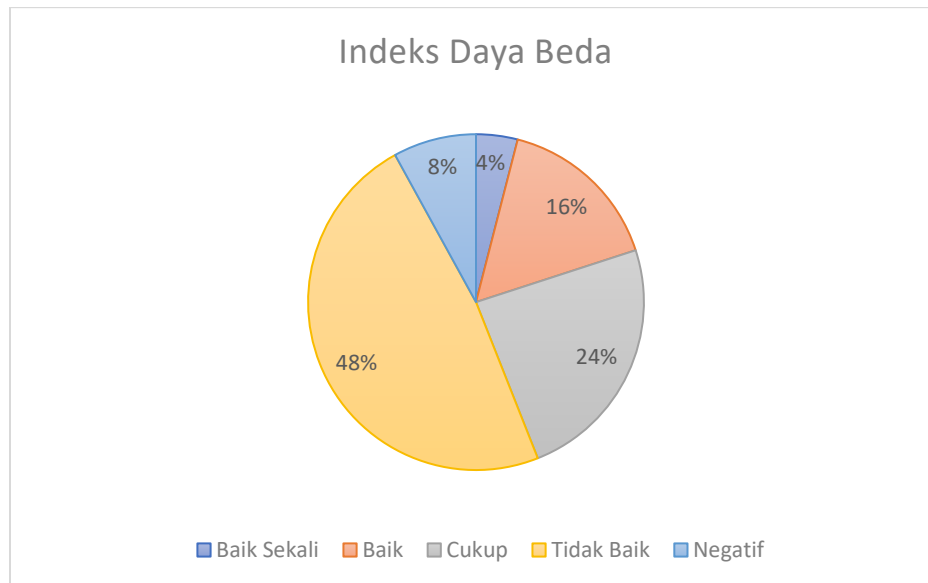
Berdasarkan analisis yang telah dilakukan untuk dapat mengetahui tingkat kesukaran dan daya beda soal serta fungsi distraktor ditemukan bahwa pada analisis tingkat kesukaran, dari 25 soal pilihan ganda terdapat 13 soal yang termasuk kategori mudah (52%) dengan indeks kesukaran antara 0,71-1,00, yaitu soal nomor 3, 5, 7, 11, 12, 13, 14, 16, 18, 19, 20, 22, dan 24. Selanjutnya, terdapat 10 soal dalam kategori sedang (40%) dengan indeks kesukaran antara 0,31-0,70, yaitu soal nomor 1, 2, 4, 6, 9, 10, 15, 17, 21, dan 23. Sementara itu, hanya terdapat 2 soal yang masuk kategori sukar (8%) dengan indeks kesukaran antara 0,00-0,30, yaitu soal nomor 8 dan 25.



Gambar 1. Presentase Tingkat Kesukaran Soal

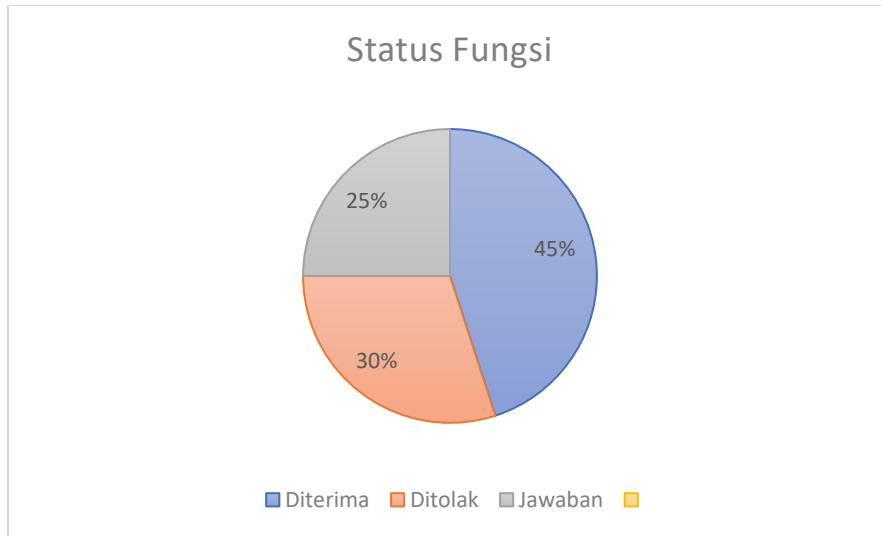
Adapun analisis indeks daya beda memberikan hasil bahwa dari 25 soal pilihan ganda yang ada, hanya 1 soal (4%) yang termasuk kategori "baik sekali" dengan indeks diskriminasi 0,71-1,00, yaitu soal nomor 2. Selanjutnya, terdapat 4 soal (16%) dalam kategori "baik" dengan indeks

0,41-0,70, yaitu soal nomor 4, 6, 7, dan 21. Sebanyak 6 soal (24%) masuk kategori "cukup" dengan indeks 0,21-0,40, yaitu soal nomor 1, 5, 15, 19, 20, dan 23. Sebaliknya, terdapat 12 soal (48%) yang masuk kategori "tidak baik" dengan indeks 0,00-0,20, seperti soal nomor 3, 8, 10, 11, 12, 13, 14, 16, 17, 18, 22, dan 24. Selain itu, 2 soal (8%) memiliki indeks negatif (soal nomor 9 dan 25), sehingga dianggap tidak berfungsi dengan baik.



Gambar 2. Presentase Indeks Daya Bada Soal

Sedangkan dalam analisis keberfungsian distraktor ditemukan bahwa fungsi distraktor pada soal-soal PAS Program Intensif Bahasa Arab kelas A2 Fakultas Sains dan Teknologi menunjukkan bahwa terdapat distraktor yang efektif (45%) maupun tidak efektif (30%). Distraktor dengan tingkat efektivitas baik dan diterima berdasarkan kriteria lebih dari 5% ditemukan pada beberapa nomor soal. Pada pilihan jawaban A, terdapat 15 soal yang diterima, termasuk S1, S2, S3, S4, S5, dan seterusnya hingga S25. Pilihan jawaban B menunjukkan hasil yang serupa, dengan sebagian besar soal diterima. Pilihan jawaban C memiliki beberapa soal yang ditolak, seperti pada nomor S3, S9, S10, S12, dan beberapa lainnya. Adapun pilihan jawaban D menunjukkan tingkat penolakan yang lebih tinggi dibandingkan dengan opsi lainnya, terutama pada nomor-nomor awal seperti S1 hingga S5.



Gambar 3. Presentase Keberfungsian Distraktor

Analisis Tingkat Kesukaran

Salah satu asumsi utama dalam menghasilkan soal berkualitas adalah memastikan keseimbangan tingkat kesulitan soal. Keseimbangan ini tercermin dari adanya distribusi soal dalam kategori mudah, sedang, dan sulit secara proporsional. Tingkat kesulitan suatu soal ditentukan berdasarkan kemampuan siswa dalam menjawabnya, bukan berdasarkan sudut pandang guru sebagai pembuat soal.

Pada soal pilihan ganda, tingkat kesulitan suatu soal diukur menggunakan indeks kesukaran (*difficulty index*). Robert L. Thorndike dan Elizabeth Hagen menginterpretasikan kriteria indeks kesulitan sebagai berikut: soal dengan nilai indeks kurang dari 0,30 tergolong sulit, nilai antara 0,31 hingga 0,70 termasuk kategori sedang, dan nilai lebih dari 0,71 dikategorikan sebagai soal mudah.

Pada data yang dianalisis dalam penelitian ini, sebagaimana telah disebutkan pada bagian hasil, komposisi soal-soal pada PAS Program Intensif Bahasa Arab Fakultas Sains dan Teknologi UINSA menunjukkan ketidakseimbangan dalam tingkat kesukaran. Di mana idealnya, soal dalam tes berkualitas memiliki distribusi yang proporsional, mencakup kategori mudah, sedang, dan sukar. Dalam hal ini, jumlah soal kategori mudah perlu dikurangi, sedangkan soal dalam kategori sedang dan sukar perlu ditambah agar distribusi tingkat kesukaran menjadi lebih seimbang. Proporsi yang lebih merata dapat meningkatkan kualitas evaluasi, sehingga dapat mengukur kemampuan siswa dengan lebih akurat. Adapun gambaran distribusi kategori kesukaran dari data tersebut akan dijelaskan melalui tabel berikut:

Tabel 1. Hasil Analisis Tingkat Kesukaran

| Jenis Soal | Indeks Kesukaran | Jumlah | Presentase | Nomor Soal |
|------------|------------------|--------|------------|---|
| Mudah | 0,71-1,00 | 13 | 52% | 3, 5, 7, 11, 12, 13, 14, 16, 18, 19, 20, 22, 24 |
| Sedang | 0,31-0,70 | 10 | 40% | 1, 2, 4, 6, 9, 10, 15, 17, 21, 23 |
| Sukar | 0,00-0,30 | 2 | 8% | 8, 25 |

Sebagai gambaran dari masing-masing kategori, berikut adalah contoh-contoh dari soal dengan kategori mudah, sedang dan sukar.

٣. سوڤاڤرٲو رٲيسا للذؤلة
أ. أصبڤ
ب. أصبڤت
ج. أصبڤنا
د. أصبڤت

Gambar 4. Contoh Soal Kategori Mudah

Pada contoh soal tersebut, 20 dari total keseluruhan 21 mahasiswa berhasil menjawab dengan benar.

٤. أي الجمل التالية لها التركيب الصحيح من حيث النحو والصرف؟
أ. سافرت سوڤاڤرٲو إلى مكة ليتلقى الدروس الإسلامية
ب. سافرت سوڤاڤرٲو إلى مكة ليتلقى الدروس الإسلامي
ج. سافر سوڤاڤرٲو إلى مكة ليتلقى الدروس الإسلامية
د. سافر سوڤاڤرٲو إلى مكة لتلقى الدروس الإسلامية

Gambar 5. Contoh Soal Kategori Sedang

Soal tersebut berhasil dijawab dengan benar oleh 14 dari total keseluruhan 21 mahasiswa.

٢٥. كُلُّ شَخْصٍ يُكْرِمُ أَنْ..... الْفِيْدِيُوَهَاتِ
أ. تُشْرِكُ
ب. يُشْرِكُ
ج. يُشَارِكُ
د. تُشَارِكُ

Gambar 6. Contoh Soal Kategori Sukar

Adapun pada contoh soal tersebut, hanya ada 2 mahasiswa yang berhasil menjawab dengan benar.

Analisis Daya Beda

Indeks diskriminasi (Discrimination Index/DI), yang juga dikenal sebagai korelasi biserial titik (Point Biserial Correlation/PBS), menggambarkan kemampuan sebuah soal untuk membedakan antara siswa dengan skor tinggi dan skor rendah. Nilai DI berkisar antara -1,00 hingga +1,00. Diharapkan bahwa siswa dengan kinerja tinggi akan lebih sering memilih jawaban yang benar dibandingkan dengan siswa dengan kinerja rendah.

Jika hal ini terjadi, maka tes tersebut dikatakan memiliki DI positif (antara 0,00 hingga +1,00), yang menunjukkan bahwa siswa dengan skor total tinggi lebih sering memilih jawaban yang benar untuk soal tertentu dibandingkan dengan siswa yang memiliki skor rendah. Sebaliknya, Dalam beberapa kasus, indeks ini dapat bernilai negatif, yang menunjukkan adanya anomali pada soal, seperti soal yang terbalik sehingga siswa dengan kemampuan tinggi dinilai rendah, dan sebaliknya. Indeks dikatakan bernilai negatif ketika siswa dengan kinerja rendah lebih sering menjawab benar suatu soal dibandingkan siswa dengan skor tinggi, maka soal tersebut memiliki DI negatif (antara -1,00 hingga 0,00).

Indeks kesulitan dan indeks diskriminasi sering kali memiliki hubungan yang berlawanan. Namun, hal ini tidak selalu berlaku. Soal dengan nilai p tinggi (soal yang mudah) biasanya memiliki kemampuan diskriminasi yang rendah, sedangkan soal dengan nilai p rendah (soal yang sulit) cenderung menjadi alat diskriminasi yang baik dalam membedakan tingkat kemampuan siswa.

Dalam kasus ini, melalui tabel 2 berikut dapat diketahui bahwa sebagian besar soal

memiliki daya pembeda yang rendah atau kurang baik, yang mengindikasikan ketidakmampuan soal untuk secara efektif membedakan siswa berdasarkan tingkat kemampuannya. Untuk meningkatkan kualitas evaluasi, soal-soal dengan daya pembeda yang tidak memadai dan nilai indeks negatif perlu diperbaiki atau diganti. Selain itu, diperlukan peningkatan proporsi soal dengan daya pembeda yang masuk kategori "baik" atau "sangat baik" agar tes menjadi lebih efektif dalam mengukur perbedaan kemampuan siswa.

Tabel 2. Hasil Analisis Daya Beda

| Daya Pembeda | Indeks Deskriminasi | Jumlah | Presentase |
|---------------------|----------------------------|--|-------------------|
| Baik Sekali | 0,71-1,00 | 1 soal (2) | 4% |
| Baik | 0,41-0,70 | 4 soal (4, 6, 7, 21) | 16% |
| Cukup | 0,21-0,40 | 6 soal (1, 5, 15, 19, 20, 23) | 24% |
| Tidak Baik | 0,00-0,20 | 12 soal (3, 8, 10, 11, 12, 13, 14, 16, 17, 18, 22, 24) | 48% |
| Negatif | (-) | 2 soal (9, 25) | 8% |

Analisis Keberfungsian Distraktor

Setiap tes objektif menggunakan alternatif jawaban yang terdiri dari dua elemen, yaitu jawaban yang benar dan jawaban yang salah, yang berfungsi sebagai pengecoh (distraktor). Distraktor dirancang untuk menguji pemahaman peserta tes dengan membingungkan mereka yang kurang mampu atau tidak mengetahui jawaban yang benar, sehingga dapat membedakan peserta dengan tingkat kemampuan yang berbeda. Idealnya, siswa dengan skor rendah, yang belum menguasai materi, akan lebih sering memilih distraktor, sedangkan siswa dengan skor tinggi akan lebih mudah mengabaikan distraktor tersebut dan memilih jawaban yang benar. Dengan menganalisis distraktor, kesalahan dalam penyusunannya dapat lebih mudah diidentifikasi sehingga dapat diperbaiki, diganti, atau dihapus sesuai kebutuhan.

Peran distraktor dalam soal pilihan ganda sangat penting, di antaranya membantu menciptakan konteks yang mengharuskan siswa membedakan jawaban yang benar dan salah. Jika distraktor dirancang dengan baik, yaitu memiliki kemiripan dengan jawaban yang benar, maka mereka dapat meningkatkan pemahaman siswa melalui efek pengujian (testing effect). Namun, jika distraktor tidak relevan, mereka justru dapat menyesatkan peserta tes dan menghambat

pemahaman. Selain itu, distraktor juga berperan dalam membantu pendidik memahami pola kesalahan siswa serta memperbaiki metode pengajaran. Jika banyak siswa memilih jawaban yang salah tertentu, hal ini dapat mengindikasikan adanya kesalahpahaman yang perlu diperbaiki dalam proses pembelajaran.

Pembuatan distraktor memerlukan upaya yang besar. Setiap soal pilihan ganda tidak hanya membutuhkan satu jawaban yang benar, tetapi juga beberapa distraktor yang harus tetap logis meskipun salah. Misalnya, untuk 100 soal dengan lima pilihan jawaban, diperlukan 400 distraktor, yang menjadikan proses penyusunannya sebagai tugas yang kompleks dan memakan waktu bagi pembuat soal.

Distraktor yang sesuai dengan kriteria di atas disebut sebagai distraktor yang efektif, yaitu distraktor yang dapat dihindari oleh siswa dengan kemampuan tinggi, tetapi cenderung dipilih oleh siswa dengan kemampuan rendah. Agar berfungsi secara optimal, distraktor yang efektif sebaiknya dipilih oleh setidaknya 5% dari total peserta tes, dengan kata lain distraktor yang dipilih oleh kurang dari 5% siswa dianggap sebagai *non-functioning distractor* (NF-D) atau distraktor yang tidak berfungsi. Jika sebuah distraktor tidak dipilih oleh peserta sama sekali, hal ini menunjukkan bahwa distraktor tersebut tidak cukup menarik atau tidak cukup membingungkan, sehingga tidak berfungsi sebagaimana mestinya dalam membedakan kemampuan siswa. Sebaliknya, jika distraktor terlalu sulit atau terlalu mirip dengan jawaban yang benar, hal ini dapat menyebabkan kebingungan yang berlebihan dan bahkan membuat siswa dengan kemampuan tinggi salah memilih. Oleh karena itu, keseimbangan dalam merancang distraktor menjadi faktor krusial dalam pembuatan soal pilihan ganda yang berkualitas. Selain itu, efektivitas distraktor juga dapat dievaluasi melalui analisis butir soal, di mana hasilnya dapat membantu pengembang tes dalam memperbaiki atau mengganti distraktor yang kurang efektif. Dengan demikian, distraktor yang baik tidak hanya berkontribusi pada keakuratan penilaian, tetapi juga membantu dalam mengidentifikasi kesenjangan pemahaman siswa dan meningkatkan kualitas pembelajaran.

Pada analisis ini, jumlah pilihan jawaban dari masing-masing soal terdiri dari 4 pilihan yang diwakili dengan abjad A, B, C dan D. Dengan demikian terdapat 1 jawaban benar dan 3 jawaban pengecoh (distraktor).

Tabel 3. Hasil Analisis Fungsi Distraktor 1

| | | Tingkat Distraktor | | | | | | | | |
|------|-----|--------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Item | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 |
| A | 24% | 19% | 95% | 19% | 90% | 19% | 5% | 10% | 62% | 19% |
| B | 52% | 10% | 5% | 14% | 5% | 14% | 71% | 76% | 10% | 24% |
| C | 24% | 10% | 0% | 67% | 5% | 67% | 14% | 14% | 0% | 0% |
| D | 0% | 62% | 0% | 0% | 0% | 0% | 10% | 0% | 29% | 57% |

Tabel 4. Hasil Analisis Fungsi Distraktor 2

| | | Tingkat Distraktor | | | | | | | | |
|------|------|--------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Item | S11 | S12 | S13 | S14 | S15 | S16 | S17 | S18 | S19 | S20 |
| A | 100% | 5% | 86% | 0% | 14% | 71% | 14% | 5% | 14% | 14% |
| B | 0% | 0% | 14% | 95% | 0% | 5% | 19% | 0% | 76% | 76% |
| C | 0% | 95% | 0% | 5% | 67% | 14% | 10% | 0% | 10% | 57% |
| D | 0% | 0% | 0% | 0% | 19% | 10% | 57% | 95% | 0% | 0% |

Tabel 5. Hasil Analisis Fungsi Distraktor 3

| | | Tingkat Distraktor | | | | |
|------|-----|--------------------|-----|-----|-----|--|
| Item | S21 | S22 | S23 | S24 | S25 | |
| A | 33% | 95% | 33% | 95% | 10% | |
| B | 10% | 5% | 29% | 0% | 10% | |
| C | 57% | 0% | 38% | 0% | 52% | |
| D | 0% | 0% | 0% | 5% | 29% | |

Sebagaimana yang tercantum pada tabel 3, 4 dan 5 diketahui bahwa fungsi distraktor pada soal-soal PAS Bahasa Arab tersebut menunjukkan adanya variasi antara distraktor yang efektif dan yang tidak efektif. Distraktor yang dinilai efektif dan memenuhi kriteria (dengan nilai lebih dari 5%) ditemukan pada sejumlah besar soal, terutama pada pilihan jawaban A, yang memiliki distribusi penerimaan cukup merata. Pilihan jawaban B juga menunjukkan hasil yang serupa, dengan mayoritas soal dianggap berfungsi dengan baik. Namun, pada pilihan jawaban C, terdapat beberapa soal yang dinilai kurang efektif, sementara pilihan jawaban D menunjukkan tingkat

ketidakefektifan yang lebih tinggi dibandingkan dengan opsi lainnya, terutama pada soal-soal di awal.

Di sisi lain, beberapa distraktor tidak memenuhi kriteria yang diharapkan dan dinilai tidak efektif. Distraktor pada pilihan jawaban C dan D cenderung menunjukkan kelemahan yang lebih signifikan dibandingkan dengan pilihan A dan B. Hal ini menunjukkan perlunya evaluasi dan perbaikan terhadap kualitas distraktor pada pilihan-pilihan tersebut agar lebih mampu mengukur pemahaman siswa secara akurat. Secara keseluruhan, hasil ini memberikan panduan penting bagi penyusun soal untuk meningkatkan kualitas distraktor yang kurang efektif dan memastikan bahwa soal-soal yang dibuat dapat berfungsi dengan optimal.

Pada butir soal nomor 18 sebagai salah satu soal dengan distraktor yang tidak berfungsi dengan baik di mana 2 distraktor yang dimiliki tidak dapat mengecoh satupun mahasiswa, penulis menyarankan untuk mengubah distraktor yang semula berupa

١٨. إِنَّ الْخُدْعَةَ مُضِرَّةٌ

الكلمة التي تحتها الخطّ بمعنى.....

أ. Jujur

ب. Kabar

ج. Medsos

د. Berbahaya

menjadi:

١٨. إِنَّ الْخُدْعَةَ مُضِرَّةٌ

الكلمة التي تحتها الخطّ بمعنى.....

أ. Bertentangan

ب. Bersembunyi

ج. Beracun

د. Berbahaya

dengan cara mengubah pilihan jawaban lain menjadi dalam bentuk kata sifat dengan konotasi negatif atau lainnya sebagaimana pola kata pada pilihan jawaban yang benar (berbahaya).

Keberfungsian distraktor perlu untuk diperhatikan dengan baik oleh penyusun tes sebab distraktor dapat memengaruhi tingkat kesukaran dan daya pembeda setiap butir soal secara langsung. Apabila tingkat kesukaran dan daya pembeda soal rendah, hal ini biasanya disebabkan oleh efektivitas distraktor yang kurang optimal. Sebaliknya, jika distraktor dirancang dengan baik dan efektif, tingkat kesukaran dan daya pembeda soal cenderung meningkat, sehingga mendukung kualitas tes secara keseluruhan.

KESIMPULAN

Penelitian ini mengungkapkan bahwa kualitas soal pilihan ganda pada Penilaian Akhir Semester (PAS) Program Intensif Bahasa Arab di Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya menunjukkan variasi yang signifikan dalam tingkat kesukaran, daya pembeda, dan keberfungsian distraktor. Meskipun sebagian besar soal memiliki tingkat kesukaran yang sesuai, ditemukan adanya distribusi yang tidak proporsional antara kategori soal mudah, sedang, dan sukar. Analisis daya pembeda menunjukkan bahwa sebagian besar soal masih kurang efektif dalam membedakan siswa berkemampuan tinggi dan rendah. Selain itu, efektivitas distraktor juga masih memerlukan peningkatan, terutama pada pilihan jawaban tertentu yang tidak memenuhi kriteria. Hasil penelitian ini menjadi rekomendasi penting untuk memperbaiki kualitas soal sehingga evaluasi pembelajaran dapat dilakukan secara lebih akurat dan adil.

DAFTAR PUSTAKA

- Amalia, Ata Nayla, and Ani Widayati. "Analisis Butir Soal Tes Kendali Mutu Kelas Xii Sma Mata Pelajaran Ekonomi Akuntansi Di Kota Yogyakarta Tahun 2012." *Jurnal Pendidikan Akuntansi Indonesia* 10, no. 1 (2012): 1–261. <https://doi.org/10.21831/jpai.v10i1.919>.
- Marthese Azzopardi, and Carmel Azzopardi. "Relationship Between Item Difficulty Level and Item Discrimination in Biology Final Examinations." *Education and New Developments* 2019 2, no. 2017 (2019): 3–7. <https://doi.org/10.36315/2019v2end001>.
- Borualogo, Ihsana Sabriani, and Dwi Agustin Nuraini , Kusdiyati, Sulisworo, Susandari , dan Sirodj. "Analisis Item Soal UTS Pedologi Semester Ganjil 2015-2016." *SCHEMA Journal of Psychological Research*, 2016, 46–57.
- Cantabrana, José Luis Lázaro, Mireia Usart Rodríguez, and Mercè Gisbert Cervera. "Assessing Teacher Digital Competence: The Construction of an Instrument for Measuring the Knowledge of Pre-Service Teachers." *Journal of New Approaches in Educational Research* 8, no. 1 (2019): 73–78. <https://doi.org/10.7821/naer.2019.1.370>.
- Chauhan, Girish R, Bhoomika R Chauhan, Jayesh V Vaza, and Pradip R Chauhan. "Relations of the Number of Functioning Distractors With the Item Difficulty Index and the Item Discrimination Power in the Multiple Choice Questions." *Cureus* 15, no. 7 (2023). <https://doi.org/10.7759/cureus.42492>.
- Fatimah, Laela Umi, and Khairuddin Alfath. "Analisi Kesukaran Soal, Daya Pembeda Dan Fungsi Distraktor." *Jurnal Komunikasi Dan Pendidikan Islam* 8, no. 2 (2019): 37–64.
- Gierl, Mark J., Okan Bulut, Qi Guo, and Xinxin Zhang. *Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review. Review of Educational Research*. Vol. 87, 2017. <https://doi.org/10.3102/0034654317726529>.
- Hingorjo, Mozaffer Rahim, and Farhan Jaleel. "Analysis of One-Best MCQs: The Difficulty Index, Discrimination Index and Distractor Efficiency." *Journal of the Pakistan Medical Association* 62, no. 2 (2012): 142–47.
- Magdalena, Ina, Septy Nurul Fauziah, Siti Nur Fазiah, and Fika Sulaehatun Nupus. "Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas Iii Sdn Karet 1 Sepatan." *BINTANG : Jurnal Pendidikan Dan Sains* 3, no. 2 (2021): 198–214.
- Mochammad Noor Akhmadi. "Analisis Butir Soal Evaluasi Tema 1 Kelas 4 Sdn Plumbungan Menggunakan Program Anates." *Ed-Humanistics : Jurnal Ilmu Pendidikan* 6, no. 1 (2021): 799–806. <https://doi.org/10.33752/ed-humanistics.v6i1.1464>.
- Novianti, Noorrachma Chandra. "TEST ITEM ANALYSIS OF THE FINAL EXAMINATION ON ECONOMICS SUBJECT IN GRADE XII IPS SMA NEGERI 1 WONOSARI ACADEMIC YEAR 2014/2015." YOGYAKARTA STATE UNIVERSITY, 2015.
- Nuraisyah, Andini. "Meningkatkan Mutu Evaluasi Pembelajaran : Analisis Soal Pilihan Ganda Bahasa Arab." *Al Mi ' Yar : Jurnal Ilmiah Pembelajaran Bahasa Arab Dan Kebahasaaraban* 7, no. 2 (2024): 904–13.

- Purba, Yoel Octobe, Fadhilaturrehmi, Jesica Triani Purba, and Kevin Wiliam Andri Siahaan. *Teknik Uji Instrumen Penelitian Pendidikan*. Edited by Aas Masruroh. *Widini Bhakti Persada Bandung*. 1st ed. Vol. 01. Bandung: Widina Bhakti Persada, 2021.
- Saputra, Hendra Dani, Wawan Purwanto, Dedi Setiawan, Donny Fernandez, and Rido Putra. "Hasil Belajar Mahasiswa: Analisis Butir Soal Tes." *Edukasi: Jurnal Pendidikan* 20, no. 1 (2022): 15–27. <https://doi.org/10.31571/edukasi.v20i1.3432>.
- Susiawati, Iis, Raswan Raswan, and Dadan Mardani. "The Purpose and Functions of Evaluation of Arabic Learning in Educational Administration Review." *Tarlim : Jurnal Pendidikan Agama Islam* 5, no. 2 (2022): 115–26. <https://doi.org/10.32528/tarlim.v5i2.7613>.
- Tarrant, Marie, James Ware, and Ahmed M. Mohammed. "An Assessment of Functioning and Non-Functioning Distractors in Multiple-Choice Questions: A Descriptive Analysis." *BMC Medical Education* 9, no. 1 (2009). <https://doi.org/10.1186/1472-6920-9-40>.
- Wind, Stefanie A., Meltem Alemdar, Jeremy A. Lingle, Roxanne Moore, and Abdullah Asilkalkan. "Exploring Student Understanding of the Engineering Design Process Using Distractor Analysis." *International Journal of STEM Education* 6, no. 1 (2019). <https://doi.org/10.1186/s40594-018-0156-x>.
- Yakout, Sahar M, Khulud Hadi, and Modhi Alolayan. "Analysis Of Difficulty Index, Discrimination Index And Distractor Efficiency In Multiple Choices Questions." *Journal of Advanced Zoology* 45, no. 3 (2024): 252–60.
- Yani, Muhammad. "Efektivitas Distractor Pada Tes Pilihan Ganda Untuk Mendeteksi Kesalahan Siswa Dalam Menyelesaikan Soal Matematika." *Al Khawarizmi: Jurnal Pendidikan Dan Pembelajaran Matematika* 2, no. 2 (2019): 125. <https://doi.org/10.22373/jppm.v2i2.4502>.