

**ANALISIS BUTIR TES : TINGKAT KESUKARAN, DAYA PEMBEDA DAN EFEKTIVITAS PENGECOH**

**Abdul Qadir**

UIN Antasari Banjarmasin  
[abdulqadir240195@gmail.com](mailto:abdulqadir240195@gmail.com),

**Nuril Huda**

UIN Antasari Banjarmasin  
[nurilhuda@uin-antasari.ac.id](mailto:nurilhuda@uin-antasari.ac.id)

**Dina Hermina**

UIN Antasari Banjarmasin  
[dinahermina@uin-antasari.ac.id](mailto:dinahermina@uin-antasari.ac.id)

**Abstract**

This research aims to analyze the level of difficulty, distinguishing power, and effectiveness of distracting test items. The analysis was carried out based on various scientific sources relevant to the research topic, including scientific journals, books and seminar proceedings. The analysis results show that the ideal test item difficulty level is between 30% and 70%. Test items with a difficulty level outside this range need to be revised. A good test item's discriminating power is 0.40 or more. Test items with low discriminating power need to be revised in order to differentiate between test takers who are able and unable to answer the questions. The effectiveness of a good distractor is if it is chosen by at least 5% of test participants. Ineffective distractors need to be revised so as not to confuse test takers. This research provides several important implications for test development practice. First, it is important to conduct test item analysis to ensure that the test items have a good level of difficulty, discriminating power, and distractor effectiveness. Second, the results of test item analysis can be used to improve the quality of the test and increase its validity. Third, this research can be a reference for test developers to study various methods and techniques for analyzing test items

Keywords : Test Item Analysis; Difficulty Level; Differentiating Power

**Abstrak**

Penelitian ini bertujuan untuk menganalisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh butir tes. Analisis dilakukan berdasarkan berbagai sumber ilmiah yang relevan dengan topik penelitian, termasuk jurnal ilmiah, buku, dan prosiding seminar. Hasil analisis menunjukkan bahwa tingkat kesukaran butir tes yang ideal adalah antara 30% dan 70%. Butir tes dengan tingkat kesukaran di luar rentang tersebut perlu direvisi. Daya pembeda butir tes yang baik adalah 0,40 atau lebih. Butir tes dengan daya pembeda yang rendah perlu direvisi agar dapat membedakan antara peserta tes yang mampu dan tidak

mampu mengerjakan soal. Efektivitas pengecoh yang baik adalah apabila dipilih oleh minimal 5% peserta tes. Pengecoh yang tidak efektif perlu direvisi agar tidak membingungkan peserta tes. Penelitian ini memberikan beberapa implikasi penting bagi praktik pengembangan tes. Pertama, penting untuk melakukan analisis butir tes untuk memastikan bahwa butir tes memiliki tingkat kesukaran, daya pembeda, dan efektivitas pengecoh yang baik. Kedua, hasil analisis butir tes dapat digunakan untuk memperbaiki kualitas tes dan meningkatkan validitasnya. Ketiga, penelitian ini dapat menjadi referensi bagi pengembang tes untuk mempelajari berbagai metode dan teknik analisis butir tes. Kata kunci : Analisis Butir Tes; Tingkat Kesukaran; Daya Pembeda.

## **PENDAHULUAN**

Tes adalah suatu pernyataan, tugas atau seperangkat tugas yang direncanakan untuk memperoleh informasi tentang trait atau atribut pendidikan dan psikologi. Setiap butir pertanyaan atau tugas tersebut mempunyai jawaban atau ketentuan yang dianggap benar. Tes dapat diklasifikasikan menurut bentuk, tipe dan ragamnya (Asmawi Zainul, dkk :1997).

Pengukuran adalah pemberian angka kepada suatu atribut atau karakteristik tertentu yang dimiliki oleh orang, hal atau obyek tertentu menurut aturan atau formulasi yang jelas. Karakteristik dari pengukuran adalah penggunaan angka atau skala tertentu dan menggunakan aturan atau formula tertentu (Asmawi Zainul, dkk :1997).

Penilaian adalah suatu proses untuk mengambil keputusan dengan menggunakan informasi yang diperoleh melalui pengukuran hasil belajar, baik yang menggunakan instrumen tes atau non tes. Dengan kata lain, penilaian adalah pemberian nilai terhadap kualitas sesuatu. (Asmawi Zainul, dkk :1997).

Input evaluation relates to the question "What should be done?" Input evaluation includes resources available to meet needs and achieve goals (Nuril Huda, 2023)

Evaluasi Masukan menjawab persoalan “apa yang harus dilakukan? (what should be done)”. Evaluasi ini mengidentifikasi masalah, aset, dan peluang untuk membantu para pengambil kebijakan menentukan tujuan, prioritas, memilih rencana, penyusunan proposal pendanaan, alokasi berbagai sumber, penempatan staf, dan menjadwalkan program. Tujuannya adalah untuk menentukan perubahan-perubahan yang dibutuhkan (Nuril Huda D. D., 2019)

kegiatan dan produk terhadap standar dan program yang sudah diterapkan. Keterkaitan antara tes, pengukuran dan penilaian adalah penilaian hasil belajar baru dapat dilakukan dengan baik dan benar bila menggunakan informasi yang diperoleh melalui pengukuran hasil belajar yang menggunakan tes sebagai alat ukurnya. Kegunaan tes, pengukuran dan penilaian dalam pendidikan antara lain adalah untuk seleksi,

penempatan, diagnosa, remedial, umpan balik, memotivasi dan membimbing, perbaikan kurikulum, program pendidikan serta pengembangan ilmu.

Perencanaan dalam pengujian sangat penting karena tes baru akan berarti bila terdiri dari butir-butir soal yang menguji tujuan yang penting dan mewakili ranah pengetahuan, kemampuan dan keterampilan secara representatif. Ada enam hal yang perlu dipertimbangkan dalam perencanaan tes yaitu: pengambilan sampel dan pemilihan butir soal, tipe tes yang akan digunakan, aspek yang akan diuji, format butir soal, jumlah butir soal dan distribusi tingkat kesukaran butir soal (Asmawi Zainul, dkk :1997).

This research was concerned with the whole with test items. This includes test item analysis. And the item analysis is used to determine the level of difficulty, discrimination power, and distractor item. (Shohamy, 1995)

Furthermore, the test should possess a distinct and well defined time constraint and be easily controllable. It is crucial to establish an accurate and effective scoring system. The test result should consistently provide reliable outcomes across various conditions (Fulcher & Davidson, 2007).

Hence, the test result is reliable. Reliability refers to the quality of being dependable, whereas validity refers to the capability of tests to accurately measure what should be measured in accordance with the intended learning goals or abilities. To ensure the test's quality, it is necessary to do item analysis in order to identify its quality. To analyze the semester test items; there are some criteria of a good test according to some experts. A good test should have 1) Validity 2) Reliability 3) Level of difficulty 4) Discrimination Power, and 5) The Quality of Options.

Kelemahan butir soal tidak terletak pada bentuk atau tipe butir soal, tetapi lebih banyak ditentukan oleh butir soal yang dikonstruksi dengan baik atau tidak baik. Butir soal obyektif akan sama baiknya dengan butir soal uraian untuk mengukur keberhasilan belajar yang dikonstruksi secara baik. Bahkan dalam beberapa hal butir soal uraian jauh lebih besar resikonya daripada butir soal obyektif. Hal ini disebabkan mutu butir soal uraian tidak hanya terletak pada kemampuan siswa untuk menjawab soal tersebut, tetapi lebih banyak ditentukan oleh kemampuan dan obyektifitas pembuat soal dalam memberikan skor pada hasil tes tersebut.

Butir soal obyektif dapat dianalisa secara lebih akurat dan bertanggung jawab sehingga dapat diketahui kelemahannya secara tepat. Butir soal tes obyektif dapat digunakan

berulang-ulang, asalkan tidak dalam perangkat tes yang sama. Oleh karena itu ada manfaat atau kegunaan analisis butir soal, kemudian direvisi sehingga butir soal yang kurang baik konstruksinya dapat diperbaiki. Akhirnya akan diperoleh butir soal yang telah teruji dan secara akurat mengukur hasil belajar yang ingin diukur.

Ada beberapa alasan mengapa diperlukan analisis butir soal. Menurut (Asmawi Zainul, dkk :1997) alasan tersebut antara lain :

- a. Untuk mengetahui kekuatan dan kelemahan butir tes, sehingga dapat dilakukan seleksi dan revisi butir soal.
- b. Untuk menyediakan informasi tentang spesifikasi butir soal secara lengkap, sehingga akan lebih memudahkan bagi pembuat soal dalam menyusun perangkat soal yang akan memenuhi kebutuhan ujian dalam bidang dan tingkat tertentu.
- c. Untuk segera dapat mengetahui masalah yang terkandung dalam butir soal, seperti: kemenduaan butir soal, kesalahan meletakkan kunci jawaban, soal yang terlalu sukar dan terlalu mudah, atau soal yang mempunyai daya beda rendah. Masalah ini bila diketahui dengan segera akan memungkinkan bagi pembuat soal untuk mengambil keputusan apakah butir soal yang bermasalah itu akan digugurkan atau direvisi guna menentukan nilai peserta didik.
- d. Untuk dijadikan alat guna menilai butir soal yang akan disimpan dalam kumpulan soal.
- e. Untuk memperoleh informasi tentang butir soal sehingga memungkinkan untuk menyusun beberapa perangkat soal yang paralel. Penyusunan perangkat seperti ini sangat bermanfaat bila akan melakukan ujian ulang atau mengukur kemampuan beberapa kelompok peserta tes dalam waktu yang berbeda.

## **METODE PENELITIAN**

Penelitian ini menggunakan pendekatan studi literatur, dimana peneliti berusaha mengumpulkan sumber-sumber buku, jurnal ataupun artikel ilmiah guna dijadikan tinjauan pustaka dalam analisis sehingga dapat menarik kesimpulan nantinya. Teknik analisis data yang digunakan adalah reduksi data, penyajian data dan penarikan kesimpulan.

## **HASIL DAN PEMBAHASAN**

### **1. Tingkat Kesukaran**

Menurut Asmawi Zainul, dkk (1997) tingkat kesukaran butir soal adalah proporsi

peserta tes menjawab benar terhadap butir soal tersebut. Tingkat kesukaran butir soal biasanya dilambangkan dengan p. Makin besar nilai p yang berarti makin besar proporsi yang menjawab benar terhadap butir soal tersebut, makin rendah tingkat kesukaran butir soal itu. Hal ini mengandung arti bahwa soal itu makin mudah, demikian pula sebaliknya.

Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang mahasiswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan mahasiswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya (Suharsimi Arikunto : 2001).

Tingkat kesukaran (difficulty index) atau kita singkat TK dapat didefinisikan sebagai proporsi siswa peserta tes yang menjawab benar (Cronker dan Algina, 1986). Definisi itu dapat dinyatakan dengan sebuah rumus dimana TK adalah jumlah peserta yang menjawab benar dibagi dgn jumlah peserta.

$$TK = \frac{\sum B}{\sum P}$$

Keterangan:

TK=tingkat kesukaran

$\sum B$ =jumlah siswa yang menjawab benar

$\sum P$ =jumlah siswa peserta tes

To analyze the data, the researchers collect the calculations of item difficulty, item discrimination, and distractors.. The researcher needs the key answers of the test and the answersheets from the students.

First the researcher will analyze the Item Facility (IF). The Item Facility can be obtained by using the following formula: (Brown \*, 2004)

Misalnya:

dari 10 siswa yang mengikuti uji coba THB, pada butir 1 terdapat 7 orang dapat menjawab benar dan pada butir 2 terdapat 2 orang dapat menjawab benar. berapakah tingkat kesukaran kedua butir soal tersebut?

$$TK (1) = \frac{7}{10} = 0,70.$$

2

$$TK(2) = \frac{2}{10} = 0,20.$$

10

Nilai TK butir merentang antara 0 sampai 1 TK sebuah butir sama dengan nol terjadi bila semua peserta tidak ada yang menjawab benar, sebaliknya TK sebuah butir akan sama dengan 1 (satu) apabila semua peserta menjawab benar pada butir tersebut. Semakin tinggi indeks TK maka butir soal semakin mudah. Dalam THB, TK butir-butir soal diusahakan sedang. Kalau butir soal terlalu mudah atau terlalu sukar bagi dua atau lebih peserta maka sekor tidak lagi dapat membedakan kemampuan para peserta sekiranya di antara mereka terdapat perbedaan kemampuan. Butir yang sangat sukar sehingga tidak ada siswa yang dapat menjawab dengan benar menyebabkan butir tersebut kehilangan kemampuannya membedakan siswa yang mempunyai kemampuan tinggi dan rendah. Begitu pula dengan butir yang sangat mudah sehingga semua peserta dapat menjawab benar. Oleh karenanya, butir sebaiknya mempunyai TK yang sedang.

TK butir yang sedang berada dalam suatu rentang nilai TK. Kriteria untuk menentukan rentang untuk TK sedang sangat tergantung jumlah kategori yang diinginkan. Misalnya kategori TK meliputi sukar, sedang dan mudah maka kriteria sedang adalah antara 0,33 sampai 0,66. Berikut pembagian kategori TK ke dalam tiga kelompok:

<b>Rentang TK</b>	<b>Kategori</b>
0,00 – 0,32	Sukar
0,33 – 0,66	Sedang
0,67 – 1,00	Mudah

Namun, bila TK di klasifikasikan ke dalam lima kelompok : sangat sukar ,sukar ,sedang ,mudah dan sangat mudah ,maka butir soal di katakana mempunyai TK sedang bila indeks TK berada antara 0,40-0,59. secara keseluruhan pembagian rentang TK diatur sebagai berikut :

<b>Rentang TK</b>	<b>Kategori</b>
0,00 – 0,19	Sangat sukar
0,20 – 0,39	Sukar
0,40 – 0,59	Sedang

0,60 – 0,79	Mudah
0,80 – 1,00	Sangat mudah

Dalam beberapa situasi, TK butir soal tidak diusahakan sedang. Pada keadaan di mana diinginkan sebanyak mungkin peserta tes dapat dinyatakan lulus maka butir diusahakan sangat mudah, misalnya penerimaan siswa di mana diperkirakan jumlah daya tampung lebih banyak daripada pelamar yang mendaftar. Sebaliknya, pada keadaan diinginkan peserta tes sekecil mungkin dapat dinyatakan lulus, maka butir soal diusahakan sesukar mungkin.

## 2. Daya Pembeda

Daya pembeda (discriminating power) atau kita singkat DB adalah kemampuan butir soal THB membedakan siswa yang mempunyai kemampuan tinggi dan rendah. DB berhubungan dengan derajat kemampuan butir membedakan dengan baik perilaku pengambil tes dalam tes yang dikembangkan (Anstasi dan Urbina, 1997:179).

Daya beda butir soal ialah indeks yang menunjukkan tingkat kemampuan butir soal membedakan kelompok yang berprestasi tinggi (kelompok atas) dari kelompok yang berprestasi rendah (kelompok bawah) di antara para peserta tes (Asmawi Zainul, dkk : 1997). Suryabrata (1999) menyatakan tujuan pokok mencari daya beda adalah untuk menentukan apakah butir soal tersebut memiliki kemampuan membedakan kelompok dalam aspek yang diukur, sesuai dengan perbedaan yang ada pada kelompok itu.

DB harus diusahakan positif dan setinggi mungkin. Butir soal yang mempunyai DB positif dan tinggi berarti butir tersebut dapat membedakan dengan baik siswa kelompok atas dan bawah. Siswa kelompok atas adalah kelompok siswa yang tergolong pandai atau mencapai skor total hasil belajar yang tinggi dan siswa kelompok bawah adalah kelompok siswa yang bodoh atau memperoleh skor total hasil belajar yang rendah. DB itu dapat ditentukan dasarnya dengan rumus sebagai berikut:

$$DB = PT - PR$$

Atau

$$DB = \frac{\sum TB}{\sum T} - \frac{\sum RB}{\sum R}$$

Keterangan:

PT = proporsi siswa yang menjawab benar pada kelompok siswa yang mempunyai kemampuan tinggi .

PR = proporsi siswa yang menjawab benar pada kelompok siswa yang mempunyai kemampuan rendah.

$\Sigma$ TB = jumlah peserta yang menjawab benar pada kelompok siswa yang mempunyai kemampuan tinggi.

$\Sigma$ T = jumlah kelompok siswa yang mempunyai kemampuan tinggi .

$\Sigma$ RB = Jumlah peserta yang menjawab benar pada kelompok siswa yang mempunyai kemampuan rendah .

$\Sigma$ R = jumlah siswa yang mempunyai kemampuan rendah .

Sebagai sebuah penjelasan di berikan contoh sebagai berikut :

Sebanyak 10 orang mengikuti uji coba THB berbentuk objektif dengan hasil sebagai berikut :

Siswa	Butir soal										Jumlah
	1	2	3	4	5	6	7	8	9	10	
<b>A</b>	1	0	1	1	1	1	1	1	1	1	<b>9</b>
<b>B</b>	0	1	0	0	1	0	0	1	0	0	<b>3</b>
<b>C</b>	1	1	1	0	1	1	1	1	1	0	<b>8</b>
<b>D</b>	1	0	1	1	1	1	1	1	1	1	<b>9</b>
<b>E</b>	0	0	0	1	1	1	0	0	0	1	<b>4</b>
<b>F</b>	1	0	1	1	1	1	1	1	1	1	<b>9</b>
<b>G</b>	0	1	1	1	0	1	1	0	1	0	<b>6</b>
<b>H</b>	0	1	0	0	0	1	0	0	0	1	<b>3</b>
<b>I</b>	1	1	0	0	0	0	0	0	1	0	<b>3</b>
<b>J</b>	1	1	1	1	1	1	1	1	1	1	<b>10</b>

Perhitungan DB dapat di lakukan dengan langkah-langkah sebagai berikut:

1. Menentukan suswa kelompok atas dan bawah. Kelompok atas adalah setengah kelompok siswa (5 orang) yang mamperoleh skor terendah. Penentuan kelompok atas dan kelompok bwah dapat di sajikan dalam tabel berikut:

Kelompok atas		Kelompok bawah	
Siswa	Skor	Siswa	Skor
A	9	B	3
C	8	E	4
D	9	G	6
F	9	H	3
J	10	I	3

2. Menghitung perolehan sekor butir pada kelompok atas dan kelompok bawah.

Kelompok atas										
Siswa	Butir soal									
	1	2	3	4	5	6	7	8	9	10
A	1	0	1	1	1	1	1	1	1	1
C	1	1	1	0	1	1	1	1	1	0
D	1	0	1	1	1	1	1	1	1	1
F	1	0	1	1	1	1	1	1	1	1
J	1	1	1	1	1	1	1	1	1	1
<b>Jumlah</b>	<b>5</b>	<b>2</b>	<b>5</b>	<b>4</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>4</b>

Kelompok bawah										
Siswa	Butir soal									
	1	2	3	4	5	6	7	8	9	10
B	0	1	0	0	1	0	0	1	0	0
E	0	0	0	1	1	1	0	0	0	1
G	0	1	1	1	0	1	1	0	1	0
H	0	1	0	0	0	0	0	0	1	1
I	1	1	0	0	0	0	0	0	0	0
<b>Jumlah</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>2</b>

### 3. Menghitung DB

DB dihitung sebagai mana rumusnya sebagai berikut:

a. Butir 1

$$\mathbf{DB (1) = \frac{5}{5} - \frac{1}{5} = \frac{4}{5} = 0,80}$$

b. Butir 2

$$\mathbf{DB (2) = \frac{2}{5} - \frac{4}{5} = \frac{2}{5} = - 0,40}$$

Sebuah butir THB yang baik adalah butir soal yang mempunyai DB positif dan signifikan. DB akan positif apabila jumlah siswa kelompok atas yang dapat menjawab dengan benar lebih banyak dari pada jumlah siswa kelompok bawah. DB yang signifikan dimaksudkan sebagai mempunyai indeks minimal +0,30 yang artinya pada butir yang baik jumlah siswa kelompok atas yang dapat menjawab benar minimal 30% lebih banyak dari pada jumlah siswa kelompok bawah yang dapat menjawab benar. Nilai DB akan merentang antara -1,00 hingga +1,00. Dengan mengambil contoh soal diatas, beberapa kondisi ekstrim dapat di jelaskan sebagai berikut:

a. Bila semua siswa kelompok atas dapat menjawab benar dan semua siswa kelompok bawah menjawab salah, maka DB akan +1,00.

$$\mathbf{DB = \frac{5}{5} - \frac{0}{5} = + 1,00}$$

b. Bila semua siswa kelompok atas dapat menjawab salah dan semua siswa kelompok bawah menjawab benar, maka DB -1,00.

$$\mathbf{DB = \frac{0}{5} - \frac{5}{5} = - 1,00}$$

c. Bila baik siswa kelompok atas maupun kelompok bawah dapat menjawab dengan benar maka DB akan 0,00.

$$\mathbf{DB = \frac{5}{5} - \frac{5}{5} = 0,00}$$

- d. Bila baik siswa kelompok atas maupun kelompok bawah menjawab salah maka DB akan 0,00.

$$DB = \frac{0}{5} - \frac{0}{5} = 0,00$$

Berdasar nilai rentang DB diatas dapat ditarik kesimpulan sebagai berikut:

1. Bila semua siswa baik kelompok atas maupun kelompok bawah sama-sama menjawab benar atau sama-sama menjawab salah maka butir soal tidak mempunyai kemampuan membedakan yang ditunjukkan oleh  $DB=0,00$ .
2. Bila siswa kelompok atas yang dapat menjawab benar lebih banyak dari pada kelompok bawah yang menjawab benar maka DB akan positif.
3. Bila siswa kelompok atas yang dapat menjawab benar lebih sedikit dari pada kelompok bawah yang menjawab benar maka DB akan negative.
4. Butir soal mempunyai DB tinggi apabila siswa kelompok atas yang dapat menjawab benar lebih banyak dibandingkan siswa kelompok bawah yang dapat menjawab benar dengan perbandingan tertentu hingga DB minimal +0,30.

Dalam menghitung DB terdapat beberapa kejadian khusus yang harus diperhatikan:

- a. Bila data di tengah sama maka data yang sama di dikeluarkan dari analisis. Misalnya: data skor hasil belajar enam orang siswa di urutkan dari tinggi ke rendah adalah sebagai berikut: 10, 9, 7, 7, 4 dan 2. Data skor yang sama adalah 7 dan di dikeluarkan dari analisis, sehingga perhitungan DB melibatkan siswa yang memperoleh skor 10 dan 9 sebagai kelompok atas dan siswa yang memperoleh skor 4 dan 2 sebagai kelompok bawah.
- b. Dalam hal jumlah siswa uji coba sangat banyak maka penentuan kelompok atas dan bawah adalah dengan mengambil 27% siswa yang memperoleh skor tertinggi sebagai kelompok atas dan 27% siswa yang memperoleh skor terendah sebagai kelompok bawah. Sebanyak 46% siswa di tengah distribusi di dikeluarkan dan tidak di analisis. Perhitungan daya beda butir di dasarkan pada "aturan 27%". Menurut Kelly, pada kondisi normal, titik optimum di mana dua kondisi seimbang di capai pada 27% kelompok atas dan bawah (Anastasi dan Urbina, 1997: 182).

Perhitungan DB butir juga dapat dilakukan dengan mengkorelasikan skor butir dengan skor total. Korelasi butir dengan total menunjukkan kesejajaran nilai antara butir dengan total. Bila skor butir bervariasi sejalan dengan variasi skor total maka butir tersebut mampu membedakan dengan baik siswa yang mempunyai kemampuan tinggi dan rendah. Butir dikatakan mempunyai DB yang tinggi apabila korelasi butir itu dengan total minimal +0,30. Adapun korelasi antara butir dengan total dapat dilakukan menggunakan rumus product moment, biserial, poin biserial. Phi atau tetrakorik (Crocker dan Algina, 1986: 317-319).

Menurut Dali S Naga (1992) kriteria besarnya koefisien daya beda diklasifikasikan menjadi empat kategori. Secara lebih rinci dijelaskan dalam tabel di bawah ini.

Kategori Daya Beda	Koefisien Korelasi
Baik	0,40 – 1,00
Sedang (tidak perlu revisi)	0,30 – 0,39
Perlu direvisi	0,20 – 0,29
Tidak baik	-1,00 – 0,19

### 3. Efektifitas Pengecoh

Analisis butir juga dilakukan dengan memperhatikan pengecoh. Pengecoh (distractor) yang juga dikenal dengan istilah penyesat atau penggoda adalah pilihan jawaban yang bukan merupakan kunci jawaban. pengecoh bukan sekedar pelengkap pilihan. Pengecoh diadakan untuk menyesatkan siswa agar tidak memilih kunci jawaban pengecoh menggoda siswa yang kurang begitu memahami materi pelajaran untuk memilihnya. Agar dapat melakukan fungsinya untuk mengecoh maka pengecoh harus dibuat semirip mungkin dengan kunci jawaban.

Pengecoh dikatakan berfungsi efektif apabila paling tidak ada siswa yang terkecoh memilih. Pengecoh yang sama sekali tidak dipilih tidak dapat melakukan fungsinya sebagai pengecoh karena terlalu mencolok dan dimengerti oleh semua siswa sebagai pengecoh soal. Pengecoh yang berdasarkan hasil uji coba tidak efektif direkomendasikan untuk diganti dengan pengecoh yang lebih menarik.

Dilihat dari konstruksi butir soal terdiri dari dua bagian, yaitu pokok soal dan

alternatif jawaban. Alternatif jawaban jawaban juga terdiri dari dua bagian, yaitu kunci jawaban dan pengecoh. Pengecoh dikatakan berfungsi apabila semakin rendah tingkat kemampuan peserta tes semakin banyak memilih pengecoh, atau makin tinggi tingkat kemampuan peserta tes akan semakin sedikit memilih pengecoh. (Depdikbud : 1997).

The last step the researcher will analyze the distractor efficiency. No mathematical formula is needed to tell you that this item is bad. The item is called bad because no one choose the item and some of students choose but from the high-ability Ss. A good distractor is one that can be avoided by students who are clever and picked by students who are not smart enough. Then, if selected at least 5% of number of participants (Alfath, 2019).

### Contoh pengujian karakteristik butir

Sehubungan dengan analisis butir secara klasik dapat diberikan contoh sebagai berikut: THB uji coba adlah 10 butir soal tes obyektif pilihan ganda dengan empat pilihan.jawaban 10 orang siswa uji coba dilaporkan hasilnya sebagai berikut:

Siswa	Butir soal									
	1	2	3	4	5	6	7	8	9	10
<b>A</b>	B	D	C	D	B	B	C	A	C	D
<b>B</b>	B	C	C	A	B	A	C	A	B	C
<b>C</b>	C	D	A	D	A	D	D	B	C	B
<b>D</b>	D	B	A	A	C	B	A	A	C	A
<b>E</b>	A	C	B	B	A	C	D	C	D	B
<b>F</b>	A	D	B	A	D	B	D	B	C	B
<b>G</b>	C	D	D	C	A	D	A	A	C	A
<b>H</b>	B	D	B	A	B	C	D	A	C	D
<b>I</b>	D	D	D	B	C	D	A	C	D	B
<b>J</b>	B	D	A	B	C	D	D	A	C	B
<b>Kunci</b>	<b>B</b>	<b>D</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>B</b>	<b>D</b>	<b>A</b>	<b>C</b>	<b>B</b>

Dari sebaran jawaban tersebut,penghitungan skor uji coba dan analisis butir dapat diringkaskan dalam tabel sebagai berikut:

Siswa	Butir soal										Jumlah
	1	2	3	4	5	6	7	8	9	10	
<b>A</b>	1	1	1	0	0	1	0	1	1	0	<b>6</b>
<b>B</b>	1	0	1	1	0	0	0	1	0	0	<b>4</b>
<b>C</b>	0	1	0	0	1	0	1	0	1	1	<b>5</b>
<b>D</b>	0	0	0	1	0	1	0	1	1	0	<b>4</b>
<b>E</b>	0	0	0	0	1	0	1	0	0	1	<b>3</b>
<b>F</b>	0	1	0	1	0	1	1	0	1	1	<b>6</b>
<b>G</b>	0	1	0	0	1	0	0	1	1	0	<b>4</b>
<b>H</b>	1	1	0	1	0	0	1	1	1	0	<b>6</b>
<b>I</b>	0	1	0	0	0	0	0	0	0	1	<b>2</b>
<b>J</b>	1	1	0	0	0	0	1	1	1	1	<b>6</b>
<b>ΣB</b>	<b>4</b>	<b>7</b>	<b>2</b>	<b>4</b>	<b>3</b>	<b>3</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>5</b>	
<b>TK</b>	<b>0,40</b>	<b>0,70</b>	<b>0,20</b>	<b>0,40</b>	<b>0,30</b>	<b>0,30</b>	<b>0,50</b>	<b>0,60</b>	<b>0,70</b>	<b>0,50</b>	
<b>DB</b>	<b>0,40</b>	<b>0,60</b>	<b>0</b>	<b>0</b>	<b>-0,20</b>	<b>0,20</b>	<b>0,60</b>	<b>0</b>	<b>0,60</b>	<b>0,20</b>	
<b>EP</b>	<b>E</b>	<b>TE</b>	<b>E</b>	<b>E</b>	<b>E</b>	<b>E</b>	<b>E</b>	<b>E</b>	<b>E</b>	<b>E</b>	

Keterangan:

SB = Jumlah siswa yang menjawab benar pada butir ke-I

TK = Tingkat kesukaran

DB = Daya beda

EP = Efektivitas pengecoh

E = Efektif

TE = Tidak efektif

Bila ditetapkan kriteria untuk memberikan penelitian butir adalah sebagai berikut:

1. TK butir harus sedang yaitu antara 0,33 sampai 0,66
2. DB harus tinggi yaitu minimal +0,30
3. Pengecoh paling tidak seorang siswa ada yang memilih.

Berdasarkan ringkasan analisis butir pada tabel di atas dan kriteria penilaian butir yang baik maka dapat ditarik kesimpulan:

1. Butir 3, 5 dan 6 terlalu sukar

2. Butir 3, 4, 5, 6, 8, dan 10 tidak mampu membedakan kemampuan siswa kelompok atas dan bawah.
3. Pada butir 2 pengecoh A tidak efektif.

Perhitungan analisis butir itu selengkapnya dilakukan sebagai berikut:

1. Tingkat kesukaran

Misalnya TK butir 1 di hitung sebagai berikut:

$$TK (1) = \frac{4}{10} = 0,40$$

2. Daya beda

Perhitungan DB dilakukan dengan langkah sebagai berikut:

- a. Menentukan siswa kelompok atas dan bawah.

Kelompok atas		Kelompok bawah	
Siswa	Skor	Siswa	Skor
A	6	B	4
C	5	D	4
F	6	E	3
H	6	G	4
J	6	I	2

- b. Menghitung perolehan skor tiap-tiap butir pada siswa kelompok atas dan bawah.

Kelompok atas										
Siswa	Butir soal									
	1	2	3	4	5	6	7	8	9	10
A	1	1	1	0	0	1	0	1	1	0
C	0	1	0	0	1	0	1	0	1	1
F	0	1	0	1	0	1	1	0	1	1
H	1	1	0	1	0	0	1	1	1	0
J	1	1	0	0	0	0	1	1	1	1
<b>Jumlah</b>	<b>3</b>	<b>5</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>4</b>	<b>3</b>	<b>5</b>	<b>3</b>

Kelompok bawah										
Siswa	Butir soal									
	1	2	3	4	5	6	7	8	9	10
<b>B</b>	1	0	1	1	0	0	0	1	0	0
<b>D</b>	0	0	0	1	0	1	0	1	1	0
<b>E</b>	0	0	0	0	1	0	1	0	0	1
<b>G</b>	0	1	0	0	1	0	0	1	1	0
<b>I</b>	0	1	0	0	0	0	0	0	0	1
<b>Jumlah</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>3</b>	<b>2</b>	<b>2</b>

c. Menghitung DB butir.

Misalnya DB untuuk butir 1 dapat dihitung sebagai berikut:

$$DB (1) = \frac{3}{5} - \frac{1}{5} = \frac{2}{5} = 0,40$$

3. Efektivitas pengecoh

Efektivitas pengecoh dapat dianalisis sebagaimana ditabulasikan dan sebagai contoh dianalisis tiga butir soal sebagai berikut:

Butir	Kunci	Pemilih	Pengecoh	Pemilih	Efektifitas pengecoh
1	B	4	A	2	Efektif
			C	2	Efektif
			D	2	Efektif
2	D	7	A	0	Tidak efektif
			B	1	Efektif
			C	2	Efektif
3	C	2	A	3	Efektif
			B	3	Efektif
			D	2	Efektif
Dan seterusnya					

## **KESIMPULAN DAN SARAN**

### **1. Kesimpulan**

Tes merupakan salah satu bentuk instrumen evaluasi untuk mengukur seberapa besar kemampuan siswa dalam memahami dan menguasai pokok-pokok materi yang sudah diajarkan, untuk itu seorang guru harus melakukan tes kepada peserta didik setiap pemberian materi yang telah dianggap selesai untuk mengetahui apakah dia bisa menerima atau memahami materi yang sudah diberikan oleh seorang guru.

Ada dua bentuk tes yang digunakan, antara lain tes standar dan tes buatan guru. Tes terstandar adalah tes yang disusun oleh suatu tim ahli, atau disusun oleh lembaga yang khusus menyelenggarakan secara profesional. Sedangkan tes buatan guru adalah tes yang dibuat seorang guru untuk merumuskan bahan dan tujuan khusus untuk kelasnya sendiri dan masih dalam ruang lingkup sekolah tempat dia mengajar.

Analisis soal antara lain bertujuan untuk mengadakan identifikasi soal-soal yang baik, kurang baik, soal yang jelek. Dengan analisis soal dapat diperoleh informasi tentang kejelekan sebuah soal dan petunjuk untuk mengadakan perbaikan. Dalam analisis soal ada beberapa hal yang penting yaitu mencari taraf kesukaran, daya pembeda, dan pola jawaban soal.

### **2. Saran**

Sebagai manusia biasa setiap orang pasti mempunyai kekurangan dan kelebihan, diantaranya adalah pola pikir tiap individu yang berbeda-beda, ada yang cerdas, pintar dan kurang pintar, untuk itu dalam suatu lembaga pendidikan hal inilah yang sangat diperhatikan. Cara pengukuran pola pikir yang dilakukan lembaga pendidikan (sekolah) yaitu melalui tes. tes-tes yang diberikan oleh pendidik harus mempunyai bobot soal yang dianggap baik, dan soal-soal itu harus benar-benar diperhatikan cara penyusunannya.

## **DAFTAR REFERENSI**

- Alfath, K. (2019). Assessment results processing techniques: Processing techniques using the norm reference approach (PAN) and the benchmark reference approach (PAP). *AL-MANAR: Journal of Islamic Communication and Education*, 8(1), 1–28.
- Asmawi Zainul dan Noehi Nasoetion. 1997. *Penilaian Hasil Belajar*. Pusat Antar Universitas, Direktorat Jenderal Pendidikan Tinggi: Departemen Pendidikan Dan kebudayaan.
- Brown \*, G. T. L. (2004). *Teachers' conceptions of assessment: Implications for policy and*

- professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301–318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G. (2001). *Assessment: A guide for lecturers (Vol. 3)*. LTSN Generic Centre York. <https://blogs.shu.ac.uk/teaching/files/2016/09/No3-Brown-G-Assessment-A-guide-for-lecturers.pdf>
- Crocker, L. dan Algina, J. (1986) *Pengantar Teori Tes Klasik dan Modern*. Harcourt, New York, 527.
- Dali, S Naga. 1992. *Pengantar Teori Sekor Pada Pengukuran Pendidikan*. Gunadarma: Jakarta.
- Departemen Pendidikan Dan Kebudayaan. 1997. *Manual Item And Test Analysis (Iteman)*. Badan Penelitian dan Pengembangan Pendidikan dan Kebudayaan: Pusat Penelitian dan Pengembangan Sistem Pengujian.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. Routledge London and New York.
- Nuril Huda, Difi Dahliana (2019) *Evaluasi Program Pendidikan Akhlak Berbasis Ma'had Bagi Mahasiswa Baru Universitas Islam Negeri Antasari Banjarmasin*, Penerbit Universitas Islam Negeri Antasari Banjarmasin
- Nuril Huda<sup>1</sup>, Difi Dahliana<sup>2</sup>, Akhmad Fauzi Aseri (2023) *Improving The Quality of Prophetic Intelligence-Based Moral Education Programs with the CIPP Model*, *Eurasian Journal of Educational Research* Hal. 1-16
- Purwanto. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar. 2009
- Shohamy, E. (1995). Performance assessment in language testing. *Annual Review of Applied Linguistics*, 15, 188–211.
- Suharsimi Arikunto. 2001. *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara: Jakarta.
- Suryabrata, S. 1999. *Pengembangan Alat Ukur Psikologis*. Direktorat jenderal Pendidikan Tinggi: Departemen Pendidikan dan Kebudayaan.