

# Pengembangan Perangkat Asesmen Mata Kuliah Psikologi Pendidikan Menggunakan Software Anates

## Assessment Tool Development Educational Psychology Course Using Anates Software

Dessy Rif'a Anzani<sup>1\*</sup>, Achmad Kholif Rosyidi<sup>2</sup>, Rahmat Aziz<sup>3</sup>, Iqbal Ali Wafa<sup>4</sup>, Ilham A.E. Zaeni<sup>5</sup>  
<sup>1,2,3,4</sup> Psikologi, UIN Maulana Malik Ibrahim, Malang, Indonesia  
<sup>5</sup> Teknik Elektro, Universitas Negeri Malang, Malang, Indonesia  
[dr.anzani90@gmail.com](mailto:dr.anzani90@gmail.com)<sup>1</sup>, [19831003@student.uin-malang.ac.id](mailto:19831003@student.uin-malang.ac.id)<sup>2</sup>, [azira@uin-malang.ac.id](mailto:azira@uin-malang.ac.id)<sup>3</sup>,  
[iqbalaliwafa2@gmail.com](mailto:iqbalaliwafa2@gmail.com)<sup>4</sup>, [ilham.ari.ft@um.ac.id](mailto:ilham.ari.ft@um.ac.id)<sup>5</sup>

### Abstrak

Evaluasi hasil belajar siswa merupakan usaha yang perlu dilaksanakan oleh seorang pendidik. Dalam rangka mengukur ketercapaian tujuan yang telah ditetapkan. Penelitian ini dilakukan untuk asesmen pada mata kuliah Psikologi Pendidikan. Penelitian ini bertujuan untuk meneliti tentang bagaimanakah daya beda, tingkat kesukaran, dan kualitas butir soal mata kuliah Psikologi Pendidikan di UIN Maulana Malik Ibrahim Malang. Soal dalam asesmen disusun berdasarkan 4 tema dalam materi ini dan dikembangkan menjadi 24 butir soal. Penelitian ini dilaksanakan di Jurusan Psikologi UIN Maulana Malik Ibrahim dengan mengikutsertakan 23 mahasiswa. Penelitian ini dilaksanakan dengan cara memberikan interpretasi deskriptif atas data kualitatif yang ada. Soal yang telah dibuat akan dilakukan analisa yang meliputi reliabilitas soal, daya beda soal, derajat kesukaran, korelasi skor item terhadap skor total, dan kualitas pengecoh. Berdasarkan hasil analisa, didapati bahwa soal yang telah dibuat dapat digunakan lagi. Beberapa perubahan masih diperlukan, seperti penggantian soal dengan daya pembeda rendah. Kualitas pengecoh pada pilihan jawaban harus ditingkatkan agar pertanyaan menjadi lebih berkualitas. Soal-soal tersebut memiliki tingkat kesulitan yang lumayan. Soal dibagi secara merata dalam kompleksitas, mulai dari yang sangat sulit hingga yang sangat mudah.

**Kata kunci:** Asesmen, Psikologi Pendidikan, Anates

### Abstract

*Evaluation of student learning outcomes is an effort that needs to be carried out by an educator. In order to measure the achievement of the goals that have been set. This research was conducted for assessment in Educational Psychology course. This study aims to examine how the differentiating power, level of difficulty, and quality of items in the Educational Psychology course at UIN Maulana Malik Ibrahim Malang are. The questions in the assessment are arranged based on 4 themes in this material and developed into 24 items. This research was conducted at the Department of Psychology UIN Maulana Malik Ibrahim by involving 23 students. This research was conducted by providing a descriptive interpretation of the existing qualitative data. The questions that have been made will be analyzed which includes the reliability of the questions, the differentiating power of the questions, the degree of difficulty, the correlation of the item scores to the total score, and the quality of distractors. Based on the results of the analysis, it was found that the questions that have been made can be used again. Some changes are still needed, such as the replacement of questions with low discriminating power. The quality of the distractors in the answer choices must be improved so that the questions are of higher quality. These questions have a decent level of difficulty. Problems are evenly divided in complexity, ranging from very difficult to very easy.*

**Keywords:** Assessment, Educational Psychology, Anates

## Pendahuluan

Melakukan evaluasi hasil belajar siswa merupakan salah satu tanggung jawab seorang pendidik. Penilaian adalah suatu usaha atau tindakan untuk menentukan ketercapaian tujuan pembelajaran yang telah ditetapkan. Penilaian (asesmen) digunakan untuk menilai keberhasilan proses dan hasil belajar siswa. Asesmen ini dapat dilakukan dalam tiga domain yang meliputi ranah kognitif, emosional, dan psikomotorik. Domain kognitif merupakan domain yang paling sering dinilai oleh pengajar dari ketiga ranah lainnya karena memiliki kaitan terhadap kemampuan siswa untuk menguasai isi materi pendidikan.

Penilaian ranah kognitif dapat dilakukan melalui beberapa teknik. Ujian harian, tes pertengahan semester, tes pada akhir semester, dan ujian akhir nasional merupakan beberapa ujian untuk mengukur domain kognitif (Gunawan & Palupi, 2016). Ujian tersebut dapat dilakukan menggunakan soal esai maupun pilihan ganda. Dalam penggunaan soal pilihan ganda, evaluasi seorang guru mungkin diperlukan dalam beberapa keadaan untuk dapat membuat soal ujian pilihan ganda dengan tepat dan akurat, serta untuk mengevaluasi apakah soal ujian tersebut dapat digunakan dalam penilaian atau tidak.

Soal pilihan ganda merupakan soal yang siswa harus menjawab dengan cara memilih dari daftar pilihan ganda (Alwi, 2015). Pilihan ganda merupakan jenis ujian objektif yang semakin populer dan sering dimanfaatkan dalam kegiatan evaluasi pendidikan. Pertanyaan pilihan ganda memiliki dua bagian dalam hal struktur: materi pelajaran (batang), yang terdiri dari masalah yang sedang dievaluasi, dan sejumlah pilihan atau alternatif tanggapan (opsi). Di antara banyak alternatif jawaban tersebut terdapat satu saja jawaban yang benar, yang disebut sebagai kunci jawaban. Pilihan selain kunci jawaban berfungsi sebagai pengecoh.

Terdapat berbagai aturan yang perlu dipertimbangkan saat merancang pertanyaan pilihan ganda, yang meliputi topik, struktur, dan bahasa. Menurut Slamet & Maarif (2014), bentuk ujian pilihan ganda memiliki keunggulan menilai berbagai tingkat kognisi (dari memori hingga evaluasi); penilaiannya itu sederhana, cepat, dan objektif, dan dapat digunakan untuk mencakup berbagai konten dalam ujian untuk berbagai tingkat pendidikan; dan dapat diterapkan untuk ujian dengan jumlah peserta yang banyak atau yang bersifat massal, dan yang hasilnya perlu untuk segera diumumkan, contohnya ujian nasional, ujian sekolah, dan tes akhir semester, adalah kandidat ideal untuk jenis penilaian ini.

Sedangkan kekurangan dari bentuk tes pilihan ganda adalah sebagai berikut: pembuatan soal membutuhkan waktu yang tidak singkat, tidak mudah untuk menyediakan distraktor yang homogen dan fungsional, dan ada kemungkinan siswa untuk menebak jawaban yang tepat. Guru pada umumnya percaya bahwa soal ujian yang telah mereka siapkan sudah baik, dan dengan demikian berharap hasil yang diperoleh siswa akan baik pula. Namun pada praktiknya tidak selalu sesuai dengan keinginan guru, sehingga guru harus mengkaji hasil tes untuk menilai kualitas perangkat tes serta keefektifan setiap butir tes dalam perangkat tes.

Untuk mencapai butir-butir yang sangat baik, analisis butir soal dalam suatu ujian perlu dilakukan untuk menginspeksi/mengevaluasi setiap butir soal. Pertanyaan berkualitas tinggi dapat menghasilkan informasi yang tepat seperti yang dimaksudkan. Ini menyiratkan bahwa analisis item dapat memberikan informasi tentang apakah suatu pertanyaan ujian sangat baik atau tidak serta dapat juga memberikan saran tentang cara meningkatkannya.

Dalam melaksanakan kegiatan analisis butir ditujukan untuk menyelesaikan pertanyaan-pertanyaan berikut, menurut Linn dan Gronlund dalam Sidin & Khaeruddin (2012). Pertanyaan tersebut adalah 1) Apakah tingkat kesulitan sesuai dengan pertanyaan? 2) Apakah ada hal lain dalam pertanyaan yang tidak relevan? 3) Apakah opsi jawaban berfungsi dengan baik? Menurut Arikunto (2010), tujuan dari analisis tes adalah untuk membantu pengajar dalam mendeteksi item yang bermasalah, memperoleh informasi yang dapat digunakan untuk meningkatkan pertanyaan untuk digunakan lagi di masa depan, dan mendapatkan gambaran singkat tentang status pertanyaan yang telah dikumpulkan. Tujuan dari analisis item menurut Sidin & Khaeruddin (2012) adalah untuk mengategorikan pertanyaan (baik, buruk, dan membutuhkan perbaikan), meningkatkan kemandirian tanggapan alternatif terhadap pertanyaan (terutama pengecoh), membantu meningkatkan pertanyaan yang perlu diperbaiki, dan memilih pertanyaan. Dalam persiapan akhir untuk ujian tertentu, pertanyaan yang bagus dapat digunakan.

Evaluasi item pada dasarnya dibagi menjadi beberapa kategori, yaitu evaluasi item secara kualitatif dan kuantitatif (Mansyur & Harun, 2015). Evaluasi kualitatif dikaitkan dengan materi isi dan bentuk pertanyaan, sedangkan evaluasi kuantitatif menekankan evaluasi sifat-sifat internal dari soal

melalui statistik yang dapat diterima secara empiris (Mansyur & Harun, 2015) dan terkait dengan sifat statistik yang digunakan oleh (Sidin & Khaeruddin, 2012). Evaluasi item yang benar adalah evaluasi yang memadukan 2 bentuk evaluasi, evaluasi kualitatif dilakukan dengan memanfaatkan teknik moderator dan panel atau *expert judgement*. Komponen yang dianalisis sangat terkait dengan materi/isi materi, konstruksi, dan bahasa. Sementara itu, evaluasi kuantitatif terutama didasarkan sepenuhnya pada statistik empiris dari item-item tersebut. Metode ini pertanyaan-pertanyaan yang sudah disusun diuji coba terlebih dahulu pada siswa untuk mencapai statistik empiris.

Berdasarkan uraian di atas, Pengembangan perangkat asesmen merupakan topik yang penting untuk di teliti dan perlu mendapatkan perhatian khusus. Salah satu mata kuliah yang perlu dikembangkan perangkat asesmennya adalah Psikologi Pendidikan. Artikel ini bertujuan untuk meneliti tentang bagaimanakah daya beda, tingkat kesukaran, dan kualitas butir soal matakuliah Psikologi Pendidikan di UIN Maulana Malik Ibrahim Malang.

## Metode Penelitian

Metode Kuantitatif digunakan dalam penelitian yang telah dilaksanakan ini. Penelitian ini merupakan penelitian deskriptif kuantitatif dengan memberikan interpretasi deskriptif atas data kualitatif yang ada. Data penelitian kualitatif diolah dengan menggunakan persentase, kemudian dideskripsikan dan dievaluasi secara kualitatif Data kuantitatif diolah dengan memanfaatkan program aplikasi ANATES untuk mendeskripsikan temuan atas hasil tanggapan siswa (Arif, 2014). Analisa yang dilakukan dalam penelitian ini meliputi: Reliabilitas soal, Daya beda, Derajat kesukaran, Korelasi skor item terhadap skor total, dan Kualitas pengecoh

Tempat dilaksanakan penelitian ini adalah di Jurusan Psikologi UIN Maulana Malik Ibrahim. Sedangkan waktu penelitian adalah dari bulan November sampai dengan Desember 2020. Penelitian ini dilakukan untuk asesmen pada mata kuliah Psikologi Pendidikan. Jumlah peserta dalam penelitian ini sebanyak 23 mahasiswa.

Jumlah soal yang disusun dalam asesmen ini adalah 24 butir soal. Soal yang dikembangkan terbagi menjadi 4 tema dengan jumlah pokok bahasan antara 4 sampai dengan 5 pokok bahasan untuk masing-masing tema. Jumlah pokok bahasan berjumlah 21 pokok bahasan. Kisi-kisi soal ditunjukkan pada Tabel 1.

**Tabel 1.**

Kisi-kisi soal

No	Tema	Pokok Bahasan
1	Perkembangan Kognitif dan bahasa	a. Pengertian b. Gambaran umum perkembangan anak c. Perkembangan kognitif dan bahasa d. Aplikasi dalam praktik pendidikan
2	Perbedaan Individu dan Pendidikan Khusus	a. Pengertian b. Anak dengan keterbatasan c. Anak berbakat d. Mengenal sekolah inklusi
3	Perkembangan Emosi dan Sosial	a. Pengertian b. Gambaran umum c. Perkembangan anak d. Perkembangan kognitif dan Bahasa e. Faktor keluarga teman dan sekolah
4	Perbedaan Kelompok	a. Pengertian b. Perbedaan inteligensi c. Perbedaan gaya belajar d. Perbedaan kepribadian

## Hasil Penelitian

### a. *Reliabilitas Soal*

Tujuan penelitian yang telah dilaksanakan ini adalah untuk melihat apakah soal-soal yang telah disusun dapat dipercaya (reliabel). Ketika diberikan kepada kelompok yang sama pada banyak waktu dan pada kesempatan yang berbeda, sebuah tes dianggap dapat dipercaya jika temuannya selalu sama. Kalau sebuah pertanyaan yang disediakan tidak dapat diandalkan, pertanyaan tersebut tidak boleh ditawarkan lagi pada kesempatan evaluasi lain.

Besarnya konsistensi atau stabilitas hasil pengukuran suatu tes disebut sebagai reliabilitasnya (Bhakti, 2015). Alat ukur yang terpercaya merupakan alat yang menghasilkan hasil yang cukup konsisten ketika digunakan untuk menguji objek yang sama berulang kali. Koefisien reliabilitas atau error pengukuran standar merupakan parameter yang diturunkan berdasarkan koefisien reliabilitas, adalah cara umum untuk mewakili keandalan tes. Teknik belah dua, yang menggunakan rumus Spearman-Brown untuk menghitung reliabilitas semua tes (Eisinga et al., 2013), merupakan salah satu cara untuk mencari nilai koefisien reliabilitas. Hasil tes reliabilitas soal ditunjukkan sebagai berikut:

$$\begin{aligned} \text{Rata2} &= 14.39 \\ \text{Simpang Standar} &= 3.49 \\ \text{Korelasi XY} &= 0.54 \\ \text{Reliabilitas Uji} &= 0.70 \end{aligned}$$

Jumlah soal pada asesmen ini adalah 24 item soal. Skor maksimal dari asesmen ini adalah 24. Jumlah peserta dalam asesmen ini adalah 23 mahasiswa. Hasil analisa pada jawaban mahasiswa menunjukkan bahwa rata-rata skor yang diperoleh mahasiswa adalah 14,39 dengan simpangan baku sebesar 3,49. Hasil tes reliabilitas menunjukkan bahwa skor reliabilitas tes adalah 0,70. Apabila nilai  $\alpha$  lebih besar atau sama dengan 0,70 menunjukkan bahwa soal hasil belajar yang sedang diuji dapat disimpulkan kalau mempunyai reliabilitas yang tinggi. Apabila nilai  $\alpha$  kurang dari pada 0,70 menunjukkan bahwa soal hasil belajar yang sedang diuji dapat disimpulkan kalau belum mempunyai reliabilitas yang tinggi.

### b. *Daya beda*

Daya beda soal adalah kemampuan suatu pertanyaan untuk membedakan antara mahasiswa yang pandai (memiliki kemampuan tinggi) dengan mahasiswa yang kurang pintar (memiliki kemampuan rendah) (Solichin, 2017). Daya beda dapat dihitung menggunakan persamaan berikut:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

Keterangan

*D* : Daya beda

*JA* : Banyaknya peserta yang masuk ke dalam kelompok atas

*JB* : Banyaknya peserta yang masuk ke dalam kelompok bawah

*Ba* : Banyaknya peserta dari kelompok atas memilih jawaban benar

*Bb* : Banyaknya peserta dari kelompok bawah memilih jawaban benar

Adapun kriteria daya beda (DB) ditunjukkan pada tabel 2. Sedangkan hasil uji daya pembeda soal ditunjukkan pada Tabel 3.

**Tabel 2.**

Kriteria Daya Beda Soal

DP	Kualifikasi
<0,00	Sangat buruk, harus dibuang
0,00 s/d 0,19	Buruk
0,20 s/d 0,39	Cukup
0,40 s/d 0,69	Baik
0,70 s/d 1,00	Sangat Baik

**Tabel 3.**

## Hasil dari Uji Daya Beda Soal

No Butir	Kelompok Atas	Kelompok Bawah	Beda	Indeks DB (%)	Kualifikasi
1	6	4	2	33.33	Cukup
2	5	5	0	0	Buruk
3	6	1	5	83.33	Baik sekali
4	2	1	1	16.67	Buruk
5	6	3	3	50	Baik
6	3	1	2	33.33	Cukup
7	6	4	2	33.33	Cukup
8	6	5	1	16.67	Buruk
9	2	0	2	33.33	Cukup
10	4	1	3	50	Baik
11	6	2	4	66.67	Baik
12	5	5	0	0	Buruk
13	4	2	2	33.33	Cukup
14	1	1	0	0	Buruk
15	6	3	3	50	Baik
16	1	0	1	16.67	Buruk
17	2	4	-2	-33.33	Sangat Buruk, harus dibuang
18	6	2	4	66.67	Baik
19	6	1	5	83.33	Baik sekali
20	6	4	2	33.33	Cukup
21	6	4	2	33.33	Cukup
22	5	4	1	16.67	Buruk
23	5	0	5	83.33	Baik sekali
24	5	2	3	50	Baik

Berdasarkan hasil uji di atas terdapat 3 soal mempunyai daya beda sangat baik, 6 soal mempunyai daya beda baik, 7 soal mempunyai daya beda yang cukup, 7 soal mempunyai daya beda buruk, dan 1 soal memiliki daya beda sangat buruk (harus dibuang). Dari hasil uji tentang daya beda, dapat dimaknai bahwa soal nomor 17 merupakan soal yang mempunyai daya beda yang sangat buruk sehingga harus dibuang.

### c. Tingkat Kesukaran

Analisis tentang tingkat kesukaran bertujuan untuk menilai apakah sebuah item pertanyaan tersebut Termasuk kriteria mudah atau sukar (Boopathiraj & Chellamani, 2013). Tingkat kesukaran adalah angka yang mengisyaratkan mudah atau sukarnya sebuah item pertanyaan (Arikunto, 2010). Tingkat kesukaran suatu butir soal dapat dihitung menggunakan persamaan:

$$P = \frac{B}{J}$$

Keterangan:

$P$  = indeks kesukaran,

$B$  = jumlah peserta ujian yang memilih jawaban benar, dan

$J$  = jumlah peserta ujian.

Hasil perhitungan derajat kesukaran soal yang telah dibuat ditunjukkan pada Tabel 4.

**Tabel 4.**

#### Hasil Tingkat Kesukaran Soal

No Butir	Jumlah Betul	Tingkat Kesukaran (%)	Keterangan
1	19	82.61	Mudah
2	19	82.61	Mudah
3	13	56.52	Sedang
4	3	13.04	Sangat Sukar
5	19	82.61	Mudah

No Butir	Jumlah Betul	Tingkat Kesukaran (%)	Keterangan
6	6	26.09	Sukar
7	19	82.61	Mudah
8	20	86.96	Sangat Mudah
9	5	21.74	Sukar
10	11	47.83	Sedang
11	17	73.91	Mudah
12	20	86.96	Sangat Mudah
13	12	52.17	Sedang
14	4	17.39	Sukar
15	20	86.96	Sangat Mudah
16	1	4.35	Sangat Sukar
17	12	52.17	Sedang
18	17	73.91	Mudah
19	13	56.52	Sedang
20	21	91.3	Sangat Mudah
21	19	82.61	Mudah
22	20	86.96	Sangat Mudah
23	8	34.78	Sedang
24	13	56.52	Sedang

Dari hasil uji tingkat kesukaran soal, maka dapat diketahui bahwa terdapat 2 soal mempunyai kategori kesukaran sangat sukar, 3 soal mempunyai kategori kesukaran sukar, 7 soal mempunyai kategori kesukaran sedang, 7 soal mempunyai kategori kesukaran mudah, dan 5 soal mempunyai kategori kesukaran sangat mudah. Hal ini dapat disimpulkan bahwa tingkat kesukaran soal relatif merata mulai dari kategori sangat sukar hingga sangat mudah.

#### **d. Korelasi Skor Item dengan Skor Total**

Suatu item tes dikatakan valid menurut Alpusari (2014) jika memberikan dukungan yang kuat terhadap skor keseluruhan. Skor keseluruhan mungkin tinggi atau rendah tergantung pada hasil setiap ujian. Dengan kata lain validitas suatu item tes dikatakan tinggi jika skor item tersebut berkorelasi dengan skor total. Keselarasan ini dapat dievaluasi dengan menggunakan korelasi, maka rumus korelasi digunakan untuk menentukan validitas item tes. Untuk pertanyaan bentuk objektif, skor 1 (untuk jawaban yang benar) atau 0 (untuk jawaban yang salah) umumnya diberikan, dengan skor keseluruhan dihitung dengan menjumlahkan skor untuk setiap item dalam perangkat tes. Rumus korelasi product moment dapat digunakan untuk menghitung besarnya koefisien korelasi, yaitu:

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

Dimana:

$r_{xy}$  = koefisien korelasi product moment

$X$  = skor item

$Y$  = skor keseluruhan

$N$  = banyaknya siswa yang mengikuti ujian

**Tabel 5.**  
Hasil Korelasi Skor Butir dengan Skor Total

No Butir	Korelasi	Signifikansi
1	0.49	Signifikan
2	0.053	-
3	0.615	Sangat Signifikan
4	0.069	-
5	0.49	Signifikan
6	0.251	-

No Butir	Korelasi	Signifikansi
7	0.322	-
8	0.272	-
9	0.125	-
10	0.401	Signifikan
11	0.475	Signifikan
12	-0.031	-
13	0.391	Signifikan
14	0.048	-
15	0.537	Sangat Signifikan
16	0.413	Signifikan
17	-0.12	-
18	0.62	Sangat Signifikan
19	0.718	Sangat Signifikan
20	0.352	-
21	0.389	Signifikan
22	0.12	-
23	0.639	Sangat Signifikan
24	0.564	Sangat Signifikan

Dari hasil analisis korelasi skor butir dengan skor total diperoleh 6 soal memiliki korelasi sangat signifikan, dan 7 soal memiliki korelasi signifikan.

#### e. *Kualitas Pengecoh*

Perhitungan tentang banyaknya siswa yang memilih setiap pilihan pertanyaan dapat digunakan untuk menentukan kemanjuran setiap pilihan pertanyaan. Selain itu, pengecoh mana yang berhasil, mana yang kurang atau tidak efektif, dan mana yang menipu dapat diperhatikan. Jika mayoritas siswa yang memilih jawaban pengecoh tertentu sementara hanya sedikit yang memilih jawaban kunci, mungkin saja pengajar yang salah dalam menentukan kunci jawaban dan pengecoh itu benar-benar kunci jawaban. Namun, ada kemungkinan bahwa kunci jawabannya benar, tetapi opsi pengecoh sangat menarik untuk dilewatkan.

**Tabel 6.**  
Hasil Uji Kualitas Pengecoh

No Butir	a	b	c	d	*
1	19**	0--	2+	2+	0
2	2+	1+	1+	19**	0
3	4++	13**	4++	1-	0
4	3**	12--	8++	0--	0
5	0--	19**	4--	0--	0
6	4+	3+	6**	10--	0
7	1+	3--	0-	19**	0
8	1++	2--	20**	0--	0
9	1--	4+	5**	13--	0
10	6+	1--	5++	11**	0
11	17**	1-	5--	0-	0
12	1++	20**	0-	2--	0
13	5+	12**	5+	1-	0
14	4+	4**	15--	0-	0
15	20**	2--	0-	1++	0
16	18--	4+	1**	0--	0
17	0--	12**	1-	10--	0
18	3+	2++	17**	1-	0

No Butir	a	b	c	d	*
19	6--	2+	13**	2+	0
20	0--	21**	0--	2---	0
21	0--	0--	19**	4---	0
22	1++	1++	20**	1++	0
23	8**	8-	1--	6++	0
24	4++	4++	2+	13**	0

*Keterangan:*

*\*\* : Jawaban yang benar*

*++ : Sangat Bagus*

*+ : Bagus*

*- : Kurang Bagus*

*-- : Buruk*

*--- : Sangat Buruk*

Berdasarkan data di atas, beberapa pengecoh masih termasuk kategori buruk dan sangat buruk. Pilihan jawaban pengecoh ini memerlukan perbaikan sehingga dapat meningkatkan kualitas soal yang telah dibuat.

## Pembahasan

Arikunto (2014) menekankan perlunya menganalisis pertanyaan yang telah diajukan untuk menentukan pertanyaan mana yang sangat baik, agak buruk, dan buruk sehingga pertanyaan yang dianggap agak buruk atau buruk dapat diperbaiki.

Dilihat dari analisis tingkat kesukarannya, butir-butir yang termasuk dalam kategori bagus (dalam arti derajat kesukaran butir soal cukup atau sedang) harus segera disimpan dalam buku soal, dan kemudian soal-soal tersebut dapat dimanfaatkan kembali dalam asesmen hasil belajar di lain waktu. Penguji kemudian harus memeriksa kembali, dan mengevaluasi soal dalam kategori mudah sehingga dapat mengetahui unsur-unsur yang membuat soal tersebut mampu dijawab oleh hampir semua peserta tes (Asri & Burhan, 2014).

Jika item pertanyaan merupakan kategori mudah, dapat dilakukan prediksi item tersebut sebagai berikut: pertanyaan pengecoh tidak berfungsi atau mayoritas siswa menjawab dengan benar butir soal yang berarti bahwa mayoritas siswa sudah memiliki pemahaman terhadap materi yang ditanyakan. Untuk soal yang merupakan kategori sulit, penguji harus memeriksa kembali, menelusuri, dan mengevaluasi hal-hal yang membuat soal sulit dikerjakan oleh siswa (Kadir, 2015).

Jika suatu item pertanyaan merupakan kategori sulit, kemungkinan besar terjadi hal-hal berikut: item tersebut mungkin memiliki kunci jawaban yang salah; item memiliki dua atau lebih jawaban benar; item pertanyaan yang bersangkutan belum disampaikan atau pembelajaran belum tuntas, akibatnya kompetensi minimal yang seharusnya dimiliki siswa belum dapat dicapai; kemampuan yang diukur tidak sesuai untuk diuji dengan memanfaatkan bentuk pertanyaan yang diberikan; atau kalimat pertanyaan yang diajukan terlalu panjang dan kompleks (Kadir, 2015).

Dari hasil analisis reliabilitas tes menggunakan teknik belah dua, yang menggunakan rumus Spearman-Brown (Eisinga et al., 2013) dapat diketahui bahwa skor reliabilitas tes yang telah dibuat adalah 0,70. Nilai ini lebih dari atau sama dengan 0,70 maka dapat diketahui bahwa yang telah dibuat diujikan kalau memiliki reliabilitas yang tinggi. Skor reliabilitas yang tinggi menunjukkan bahwa pengukuran menggunakan soal yang dibuat memiliki konsistensi atau stabilitas hasil (Bhakti, 2015). Soal yang dibuat dapat dipercaya akan menghasilkan hasil yang cukup konsisten ketika digunakan untuk menguji materi yang sama berulang kali. Koefisien reliabilitas atau eror pengukuran standar merupakan parameter yang diturunkan berdasarkan koefisien reliabilitas, adalah cara umum untuk mewakili keandalan tes. Dari hasil analisis korelasi skor item terhadap skor total diperoleh 6 soal memiliki hubungan yang sangat signifikan, dan 7 soal memiliki korelasi signifikan.

Berdasarkan hasil uji pada soal yang telah dibuat, terdapat 3 soal yang mempunyai daya beda sangat baik, 6 soal yang mempunyai daya beda baik, 7 soal yang mempunyai daya beda cukup, dan 7 soal memiliki daya beda buruk. Dari soal yang dibuat terdapat 1 soal memiliki daya beda sangat buruk sehingga harus dibuang. Dari hasil uji daya beda, dapat diketahui bahwa soal nomor 17 merupakan soal yang mempunyai daya beda negatif yang tidak baik sehingga harus dibuang.



Dari hasil uji tingkat kesukaran soal, maka dapat ditunjukkan bahwa terdapat 2 soal mempunyai kategori kesukaran sangat sukar, 3 soal mempunyai kategori kesukaran sukar, 7 soal mempunyai kategori kesukaran sedang, 7 soal mempunyai kategori kesukaran mudah, dan 5 soal mempunyai kategori kesukaran sangat mudah. Sebaran tingkat kesukaran soal yang telah dibuat dapat ditunjukkan pada Gambar 1.



Gambar 1. Sebaran Tingkat Kesukaran Soal yang Dibuat

Gambar 1 di atas menunjukkan bahwa sebaran tingkat kesukaran soal yang dibuat menyerupai kurva normal. Mayoritas soal memiliki derajat kesukaran mudah dan sedang. Jumlah soal yang mempunyai tingkat kesukaran sangat sukar, sukar, dan sangat mudah lebih sedikit dibandingkan soal yang memiliki tingkat kesukaran sedang dan mudah. Maka hal ini dapat disimpulkan bahwa tingkat kesukaran soal tersebar secara proporsional. Tingkat kesukaran soal relatif merata mulai dari kategori sangat sukar hingga sangat mudah.

Berdasarkan data di atas, beberapa pengecoh masih termasuk kategori buruk dan sangat buruk. Pilihan jawaban pengecoh ini memerlukan perbaikan sehingga mampu meningkatkan kualitas soal yang telah dibuat. Jawaban pengecoh yang sudah mampu melaksanakan fungsi yang diharapkan dengan baik dapat digunakan lagi pada asesmen yang akan dilakukan di masa datang, sedangkan jawaban pengecoh yang belum dapat melaksanakannya dengan baik dapat hendaknya direvisi atau diganti dengan jawaban pengecoh lain.

## Kesimpulan

Berdasarkan uraian yang telah dijabarkan pada tahap analisa hasil, dapat ditarik kesimpulan bahwa soal yang telah dibuat dapat digunakan lagi. Beberapa perbaikan masih perlu dilakukan seperti mengganti soal yang memiliki daya pembeda yang buruk. Kualitas pengecoh yang ada pada pilihan jawaban perlu diperbaiki sehingga kualitas soal akan menjadi lebih baik. Adapun tingkat kesukaran soal sudah baik. Tingkat kesukaran soal sudah merata mulai dari sangat sukar hingga sangat mudah. Soal-soal yang memiliki kualitas yang baik dari segi validitas, reliabilitas, tingkat kesukaran, daya beda, maupun pengecohnya selanjutnya dapat dihimpun ke dalam sebuah bank soal untuk dapat dipergunakan pada soal tes berikutnya.

Pengembangan perangkat asesmen dan Analisa perangkat ini perlu dilakukan oleh pengajar karena pengajar perlu untuk melakukan pengukuran pemahaman siswa terhadap materi. Pengukuran akan memberikan hasil yang tepat apabila perangkat asesmen yang digunakan memiliki kualitas yang baik. Dengan menggunakan perangkat pengukuran yang tepat, hasil asesmen akan memberikan gambaran yang tepat mengenai pemahaman siswa sehingga pengajar dapat mengetahui ketercapaian pembelajaran yang telah dilakukan.

## Referensi

- Alpusari, M. (2014). Analisis butir soal konsep dasar IPA 1 melalui penggunaan program komputer anates versi 4.0 for Windows. *Primary: Jurnal Pendidikan Guru Sekolah Dasar*, 3(2), 106–115.
- Alwi, I. (2015). Pengaruh Jumlah Alternatif Jawaban Tes Obyektif Bentuk Pilihan Ganda Terhadap Reliabilitas, Tingkat Kesukaran Dan Daya Pembeda. *Faktor Exacta*, 3(2), 184–193.

- Arif, M. (2014). Penerapan aplikasi anates bentuk soal pilihan ganda. *Edutic-Scientific Journal of Informatics Education*, 1(1).
- Arikunto, S. (2010). Metode peneltian. *Jakarta: Rineka Cipta*.
- Arikunto, S. (2014). *Dasar-Dasar Evaluasi pendidikan: Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: PT Bumi Aksara.
- Asri, A. F., & Burhan, A. (2014). Analisis Tingkat Kesukaran, Daya Pembeda dan Fungsi Distraktor Soal Ujian Semester Ganjil Mata Pelajaran Produktif di SMK Negeri 1 Indralaya Utara Tahun Pelajaran 2012/2013. *Jurnal Pendidikan Teknik Mesin*, 1(2).
- Bhakti, Y. B. (2015). Pengaruh jumlah alternatif jawaban dan teknik penskoran terhadap reliabilitas tes. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, 5(1).
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*, 2(2), 189–193.
- Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58(4), 637–642.
- Gunawan, I., & Palupi, A. R. (2016). Taksonomi Bloom–revisi ranah kognitif: kerangka landasan untuk pembelajaran, pengajaran, dan penilaian. *Premiere Educandum: Jurnal Pendidikan Dasar Dan Pembelajaran*, 2(02).
- Kadir, A. (2015). Menyusun dan menganalisis tes hasil belajar. *Al-TA'DIB: Jurnal Kajian Ilmu Kependidikan*, 8(2), 70–81.
- Mansyur, S., & Harun, R. (2015). Asesmen pembelajaran di sekolah: Panduan bagi guru dan calon guru. *Yogyakarta: Pustaka Pelajar*.
- Sidin, A., & Khaeruddin, K. (2012). *Evaluasi Pembelajaran*. Universitas Negeri Makassar.
- Slamet, S., & Maarif, S. (2014). Pengaruh bentuk tes formatif asosiasi pilihan ganda dengan reward dan punishment score pada pembelajaran matematika siswa SMA. *Infinity Journal*, 3(1), 59–80.
- Solichin, M. (2017). Analisis daya beda soal, taraf kesukaran, validitas butir tes, interpretasi hasil tes dan validitas ramalan dalam evaluasi pendidikan. *Dirasat: Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 192–213.

This page is intentionally left blank