

Analisis Kualitas Butir Soal *HOTS* pada Mata Pelajaran Biologi yang Diberikan pada Siswa Kelas 12 dan Dievaluasi Menggunakan Aplikasi Anates

Iing Dwi Lestari¹, Ryke Zain Yumna K², Siti Alda Lusia Dewi³, Early Sri M⁴, Putri Nurhalizah⁵, Siti Rosalina⁶

¹²³⁴⁵⁶Program Studi Pendidikan Biologi, Fakultas Keguruan Dan Ilmu Pendidikan, Universitas Sultan Ageng Tirtayasa

iingdwiles@gmail.com¹, 2224210025@untirta.ac.id², 2224210035@untirta.ac.id³, 2224210076@untirta.ac.id⁴, 2224210038@untirta.ac.id⁵, 2224210007@untirta.ac.id⁶

ABSTRACT

The aim of this research is to develop and describe an effective assessment instrument. This type of research is Research and Development, while the data collection technique uses test techniques. Data analysis techniques use qualitative and quantitative analysis. The results of this study are as follows: there are 13 significant assessment items and 12 items that are not yet significant. Therefore, there are still some items that need to be improved to be used as a suitable assessment instrument, both theoretically and empirically. Suitable theoretically means that it has fulfilled the aspects of construction, language, and material. While suitable empirically means through validity testing, reliability, difficulty level, discriminant power, and distractor quality analysis. An assessment instrument can be considered suitable if it meets the characteristics of Higher Order Thinking Skills; and the assessment instrument can be used as a tool to evaluate learning outcomes because it can differentiate between students with high, medium, and low abilities.

Keywords: *assessment, higher order thinking skills*

ABSTRAK

Tujuan penelitian ini adalah untuk mengembangkan dan mendeskripsikan instrumen penilaian yang efektif. Jenis penelitian ini adalah *Research and Development*, sedangkan teknik pengumpulan data menggunakan teknik tes. Teknik analisis data menggunakan analisis kualitatif dan kuantitatif. Hasil penelitian ini, yaitu: instrumen penilaian sebanyak 13 soal sudah signifikan dan 12 soal belum signifikan. Sehingga masih ada beberapa soal yang harus diperbaiki agar bisa dijadikan instrumen penilaian yang layak digunakan, baik secara teoritis maupun empiris. Adapun yang dimaksud layak secara teoritis yaitu telah memenuhi aspek konstruksi, bahasa, dan materi. Sedangkan layak secara empiris yaitu melalui uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan analisis kualitas pengecoh. Instrumen penilaian dapat dikatakan layak jika memenuhi karakteristik *Higher Order Thinking Skills*; dan instrumen penilaian dapat digunakan sebagai alat evaluasi hasil belajar karena instrumen penilaian dapat membedakan siswa yang berkemampuan tinggi, sedang, dan rendah.

Kata kunci: *penilaian, higher order thinking skills*

PENDAHULUAN

Di era globalisasi saat ini, diperlukan sumber daya manusia yang berdaya saing untuk memenuhi tuntutan zaman yang semakin maju. Kualitas sumber daya

manusia suatu negara ditentukan oleh tingkat pendidikan negara tersebut. Pendidikan memegang peranan penting, karena pendidikan merupakan sarana untuk meningkatkan dan mengembangkan kualitas sumber daya manusia. Dalam dunia pendidikan khususnya pendidikan sekolah, Biologi merupakan mata pelajaran yang sangat penting, karena Biologi merupakan ilmu yang dapat melatih berpikir kritis, sistematis, logis, dan kreatif. Biologi juga memiliki struktur yang kuat dan jelas serta hubungan antar konsep yang memungkinkan pemikiran rasional. Itulah mengapa penting tidak hanya mempelajari Biologi, tetapi juga mencoba memahaminya dan dapat menerapkannya pada masalah lain.

Berdasarkan hasil tes dan survei yang dilakukan PISA tahun 2015, nilai siswa Biologi Indonesia masih tergolong rendah yaitu di 63 dari 69 negara yang disurvei. Siswa Indonesia memiliki sedikit pengetahuan tentang materi dan sulit menjawab pertanyaan yang membutuhkan pembenaran. Kemampuan memahami materi Biologi dan menggunakannya untuk memecahkan masalah sehari-hari disebut literasi Biologi. Hasil tersebut menunjukkan bahwa pendidikan Biologi dasar siswa di Indonesia masih belum memuaskan berdasarkan studi internasional. Hal ini dikarenakan siswa belajar menghafal rumus tanpa memahami konsepnya. Sehingga ketika diberikan pertanyaan yang berhubungan dengan konsep Biologi yang sama sekalipun, siswa menjadi bingung dan menganggap soal tersebut sulit. Menurut OECD (2014), literasi Biologi adalah kemampuan seseorang untuk merumuskan, menerapkan, dan menginterpretasikan Biologi dalam berbagai konteks, termasuk kemampuan bernalar secara matematis dan menggunakan konsep, prosedur, dan fakta untuk menggambarkan fenomena/peristiwa atau memprediksi. Siswa diharapkan dapat memanfaatkan kemampuannya dengan sebaik-baiknya sehingga lebih cerdas dan kritis dalam menyerap dan mengolah informasi. Berdasarkan standar isi kurikulum (2013) dibuat pengembangan keterampilan membaca untuk pembelajaran Biologi. Dalam kurikulum 2013, jurusan Biologi diharapkan tidak hanya mampu menggunakan perhitungan atau rumus dalam soal-soal ujian, tetapi juga mengetahui kemampuan berpikir logis dan analitis untuk memecahkan masalah sehari-hari. Pemecahan masalah ini bukan hanya masalah berupa masalah rutin, melainkan masalah yang dihadapi setiap hari. Sebagian besar soal Biologi pada silabus 2013 berjenis *Higher Order Thinking Skill (HOTS)*. Soal-soal bergaya *HOTS* adalah soal-soal yang membutuhkan kemampuan berpikir tingkat tinggi dan melibatkan proses berpikir untuk meningkatkan kemampuan berpikir kritis, logis, reflektif, metakognitif, dan kreatif. Soal-soal ala *HOTS* melatih siswa berpikir pada tingkat analisis, evaluasi, dan kreasi. Tujuan dari makalah ini adalah untuk menyelidiki apakah soal-soal bertipe *Higher Order Thinking Skill (HOTS)* yang membutuhkan berpikir tingkat tinggi dapat menunjang pendidikan Biologi siswa.

METODE PENELITIAN

Jenis penelitian ini adalah penelitian dan pengembangan atau *Research and Development (R&D)*. Produk yang dikembangkan merupakan alat untuk menilai kemampuan berpikir tingkat tinggi. Untuk mendapatkan prototipe pengembangan, model Dick & Carey diadaptasi dalam penelitian ini. Model Dick & Carey adalah model

pengembangan yang dikembangkan mengikuti pendekatan sistem. Model sistem pembelajaran yang dikembangkan oleh Dick, Carey & Carey terdiri dari beberapa komponen yang harus dilaksanakan ketika merencanakan kegiatan pembelajaran yang lebih besar. Penerapan model desain sistem pembelajaran ini membutuhkan proses sistematis yang menyeluruh. Hal ini diperlukan untuk menciptakan suatu sistem pembelajaran yang dapat digunakan secara optimal untuk mengatasi permasalahan pembelajaran. Tahapan model Dick & Carey (Qorih et al, 2017) yaitu (1) analisis kebutuhan untuk mengidentifikasi tujuan, (2) analisis pembelajaran, (3) analisis peserta didik dan konteks, (4) perumusan tujuan pencapaian, (5) pengembangan alat penilaian, (6) pengembangan strategi pembelajaran, (7) pengembangan dan pemilihan materi pembelajaran, (8) melakukan penilaian formatif, (9) merevisi, (10) merencanakan dan melakukan penilaian sumatif.

HASIL DAN PEMBAHASAN

1. Analisis Daya Pembeda

Menurut Elviana (2020) menyatakan bahwa Daya Pembeda merupakan suatu indeks yang menunjukkan kemampuan butir soal untuk membedakan kelompok atas (kelompok unggul, siswa yang nilai tesnya tinggi) dan kelompok bawah (kelompok asor, siswa yang nilai tesnya rendah). aspek Daya Pembeda memiliki tujuan untuk dapat menentukan butir soal termasuk ke dalam kategori skor tinggi atau skor rendah pada keseluruhan tes. Daya Pembeda dibagi menjadi 5 tingkatan yaitu Sangat Buruk, Buruk, Agak Baik atau cukup, Baik, dan Sangat Baik, lebih lanjutnya klasifikasi kategori ini berdasarkan tabel di bawah ini:

Indeks Daya Pembeda	Kriteria Daya Pembeda
Negatif sampai 9%	Sangat Buruk (Harus tidak dipakai)
10% sampai 19%	Buruk (sebaiknya tidak dipakai)
20% sampai 29%	Agak Baik atau Cukup
30% sampai 49%	Baik
50% ke atas	Sangat Baik

(Elviana, 2020)

Hasil analisis butir soal *HOTS* Biologi Kelas 12 dengan menggunakan *software* Anates pada aspek Daya Pembeda dibagi menjadi 5 (lima) tingkatan, yaitu sangat Buruk, Buruk, Agak Baik atau Cukup, Baik, dan Sangat Baik. Berikut merupakan tabel hasil persebaran soal *HOTS* Biologi Kelas 12 pada aspek Daya Pembeda soal Pilihan Ganda dan soal Uraian:

Tabel. Hasil Daya Pembeda

Kategori Soal	Jumlah	Nomor Soal
Sangat Buruk	2	Butir soal pilihan ganda nomor 5 dan uraian nomor 2
Buruk	6	Butir soal pilihan ganda nomor 1, 4, 5, 11, 14 dan uraian nomor 7 dan 10
Cukup atau Agak Baik	4	Butir soal pilihan ganda nomor 2, 3, 13, 15
Baik	7	Butir soal pilihan ganda nomor 8, 9, 12 dan uraian nomor 1, 4, 5, 9
Sangat Baik	6	Butir soal pilihan ganda nomor 6, 7, 10 dan uraian nomor 3, 6, 8
Jumlah	25	

Berdasarkan tabel hasil analisis di atas butir soal pilihan ganda nomor 5 dan uraian nomor 2 yang ada di kategori Sangat Buruk yang harus tidak digunakan, artinya soal ini harus diganti karena memiliki Daya Pembeda yang sangat buruk yaitu hanya di rentang negatif sampai 9% saja, selanjutnya soal pilihan ganda nomor Butir soal pilihan ganda nomor 1, 4, 5, 11, 14 dan uraian nomor 7 dan 10 termasuk ke dalam kategori Buruk yang seharusnya tidak digunakan, artinya soal ini juga perlu diperbaiki kembali karena rentang Daya Pembedanya yang hanya berkisar di angka 10% sampai 19% saja, selanjutnya soal pilihan ganda nomor 2, 3, 13, 15 termasuk ke dalam kategori Cukup yang rentang Daya Pembedanya berkisar di angka 20% sampai 29%, kemudian soal pilihan ganda nomor 8, 9, 12 dan uraian nomor 1, 4, 5, 9 termasuk ke dalam kategori Baik yang rentang daya pembedanya berkisar di angka 30% sampai 49%, sedangkan untuk butir soal pilihan ganda nomor 6, 7, 10 dan uraian nomor 3, 6, 8 termasuk ke dalam kategori Sangat Baik karena rentang Daya Pembedanya berkisar di angka lebih dari 50%.

2. Analisis Tingkat Kesukaran

Tingkat kesukaran merupakan suatu perbandingan peserta tes menjawab benar dengan butir soal. Suatu tes dikatakan baik apabila butir soal pada tes tersebut tidak terlalu sukar dan juga tidak terlalu mudah. 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 kemampuannya. (Elviana. 2020)

Adapun klasifikasi indeks tingkat kesukaran yang digunakan sebagai acuan secara garis besar meliputi tiga kategori yaitu Sukar, Sedang, dan Mudah. Lebih lanjutnya klasifikasi kategori ini berdasarkan tabel di bawah ini:

Indeks Tingkat Kesukaran	Kriteria Tingkat Kesukaran
Pada rentang 0,00 – 0,29	Sukar
Pada rentang 0,30-0,70	Sedang
Pada rentang 0,70-1,00	Mudah

(Warju. et al. 2020)

Hasil analisis butir soal *HOTS* Biologi Kelas 12 dengan menggunakan *software* Anates pada aspek Tingkat Kesukaran dibagi menjadi 5 (lima) tingkatan yang berbeda yaitu Sangat Mudah, Mudah, Sedang, Sukar, Sangat Sukar. Berikut merupakan tabel hasil persebaran soal *HOTS* Biologi Kelas 12 pada aspek Tingkat Kesukaran soal Pilihan Ganda dan soal Uraian.

Tabel. Hasil Tingkat Kesukaran

Kategori Soal	Jumlah	Nomor Soal
Sangat Mudah	4	Butir soal pilihan ganda nomor 1, 8, 11, 12
Mudah	4	Butir soal pilihan ganda nomor 7, 9, 13, 15
Sedang	7	Butir soal pilihan ganda nomor 2, 6, 10, 14 dan uraian nomor 1, 6, 8
Sukar	5	Butir soal pilihan ganda nomor 5 dan uraian nomor 3, 4, 5, 9
Sangat Sukar	5	Butir soal pilihan ganda nomor 3, 4 dan uraian nomor 2, 7, 10
Jumlah	25	

Berdasarkan tabel hasil analisis di atas terdapat sebanyak 8 butir soal pada kategori mudah, untuk lebih rincinya butir soal pilihan ganda nomor 1, 8, 11, 12 termasuk ke dalam soal berkategori sangat mudah karena siswa banyak menjawab benar pada butir soal ini, begitu pula pada butir soal pilihan ganda nomor 7, 9, 13, 15, hal ini artinya 8 butir soal ini indeks tingkat kesukarannya berada di rentang 0,70-1,00. Selanjutnya, sebanyak 7 butir soal pada kategori sedang, yaitu butir soal pilihan ganda nomor 2, 6, 10, 14 dan uraian nomor 1, 6, 8 dengan indeks tingkat kesukarannya berada di rentan 0,30-0,70. Selanjutnya terdapat 10 butir soal pada kategori Sukar, untuk lebih rincinya butir soal pilihan ganda nomor 5 dan uraian nomor 3, 4, 5, 9 termasuk ke dalam soal berkategori sukar begitu pula pada Butir soal pilihan ganda nomor 3, 4 dan uraian nomor 2, 7, 10 yang termasuk ke dalam kategori sangat sukar, hal ini mengartikan bahwa 10 butir soal ini memiliki indeks tingkat kesukaran pada rentang 0,00–0,29, karena hanya sedikit siswa yang menjawab benar pada butir soal kategori sukar yang telah disebutkan di atas.

3. Analisis Validitas

Validitas atau (kesahihan) tes dapat diartikan sebagai ketetapan dan kecermatan tes dalam menjalankan fungsi pengukurannya. Semakin tinggi koefisien maka semakin cermat suatu tes. Sebuah tes dikatakan valid apabila minimal koefisiennya sebesar 0,5. Pada suatu tes jika Semakin tinggi koefisiennya maka semakin cermat suatu tes. (Elviana. 2020)

Di *software* Anates ini untuk nilai *r* tabel dapat langsung diketahui yaitu terletak pada bagian korelasi dengan skor total yang ada di bawah tabel. Sesuai dengan jumlah soal PG yaitu sebanyak 15 butir soal dan nilai *r* tabel yang tertera pada taraf signifikansi 5%, *n*-2 dari 15 butir soal didapatkan nilai sebesar 0,482. Lalu, Sesuai

dengan jumlah soal uraian yaitu sebanyak 10 butir soal dan nilai r_{tabel} yang tertera pada taraf signifikansi 5%, $n-2$ dari 10 butir soal didapatkan nilai sebesar 0,576.

Sehingga hasil analisis untuk aspek Validitas soal pilihan ganda berpatok pada: jika nilai r_{pbi} lebih besar sama dengan 0,482 maka soal tersebut valid, sedangkan jika r_{pbi} lebih kecil dari 0,482 maka soal tidak valid.

Begitu pula pada hasil analisis untuk aspek Validitas soal uraian berpatok pada: jika nilai r_{pbi} lebih besar sama dengan 0,576 maka soal tersebut valid, sedangkan jika r_{pbi} lebih kecil dari 0,576 maka soal tidak valid.

Hasil analisis butir soal *HOTS* Biologi Kelas 12 dengan menggunakan *software* Anates pada aspek Validitas, Menurut Fiska *et al.* (2021) Pada *software* Anates, validitas dapat ditentukan pada korelasi skor butir dengan skor total yang terbagi dalam dua kategori yaitu signifikan (yang berarti soal valid) dan tanda (-) yang berarti tidak signifikan (soal tidak valid).

Tabel. Hasil Validitas

Kategori Soal	Jumlah	Nomor Soal
Valid (Soal Signifikan)	13	Butir soal pilihan ganda nomor 1, 6, 7, 8, 11, 12, 15 dan uraian nomor 3, 5, 6, 8, 9, 10
Tidak Valid (Soal Tidak Signifikan)	12	Butir soal pilihan ganda nomor 2, 3, 4, 5, 9, 10, 13, 14 dan uraian nomor 1, 2, 4, 7
Jumlah	25	

Berdasarkan tabel hasil analisis di atas, terdapat 13 butir soal dengan rincian sebanyak 7 butir soal pilihan ganda dan 6 butir soal uraian yang termasuk ke dalam kategori valid atau signifikan, yaitu pada butir soal pilihan ganda nomor 1, 6, 7, 8, 11, 12, 15 artinya ke-7 butir soal pilihan ganda ini memiliki nilai r_{pbi} lebih besar sama dengan 0,482 sehingga dikatakan signifikan atau valid, dan uraian nomor 3, 5, 6, 8, 9, 10 artinya memiliki nilai r_{pbi} lebih besar sama dengan 0,576 sehingga dikatakan signifikan atau valid. Sedangkan untuk butir soal yang tidak valid atau signifikan sebanyak 12 butir soal dengan rincian sebanyak 8 butir soal pilihan ganda dan 4 butir soal uraian, yaitu pada Butir soal pilihan ganda nomor 2, 3, 4, 5, 9, 10, 13, 14 artinya butir soal tersebut memiliki nilai r_{pbi} lebih kecil dari 0,482 sehingga tidak valid atau signifikan, dan uraian nomor 1, 2, 4, 7 memiliki nilai r_{pbi} lebih kecil dari 0,576 sehingga termasuk ke dalam soal tidak valid atau signifikan.

4. Analisis Reliabilitas

Gulliksen menyatakan bahwa reability adalah perbandingan antara varians skor sebenarnya (T^2) dan varians skor yang diamati (X^2) dalam Zimmerman (2009). Selain itu, pengurangan rumus akan dilakukan dengan menggunakan persamaan yang disediakan oleh Hopkin dan Antes dan rumus Guliksen, menghasilkan rumus reliabilitas yang dipengaruhi oleh tebakan probabilitas. Akibatnya, perbandingan antara definisi varians skor riil dan skor amatan mengarah pada pengembangan rumus untuk koefisien reliabilitas KR-20 dan koefisien reliabilitas berbasis probabilitas.

Menurut teori klasik, reabilitas adalah sejauh mana temuan pengukuran dapat dipercaya. Jika pengukuran dilakukan pada kelompok orang yang sama beberapa kali dan temuannya pada dasarnya sama, implikasi pengukuran dapat dianggap serius selama karakteristik subjek yang diukur tetap konstan. Konotasi yang secara substansial identik menyiratkan bahwa ada toleransi untuk sedikit variasi dalam skor beberapa pengukuran. Gulliksen mendefinisikan dependabilitas dalam Zimmerman sebagai perbedaan antara varians skor yang diamati (σ_X^2) dan varians skor aktual (σ_T^2).

Hasil Nilai Uji Reliabilitas Pilihan Ganda sebesar 0,79, pada kategori yang digunakan, nilai Reliabilitas sebesar 0,79 untuk pilihan ganda termasuk dalam kategori *High*, sesuai dengan hasil analisis butir soal *HOTS* Biologi Kelas 12 menggunakan perangkat lunak anates pada aspek Reliabilitas pada soal dan deskripsi *Multiple Choice*. Distribusi soal dari bagian Keandalan Pilihan Ganda ujian *HOTS* Biologi Kelas 12 ditunjukkan pada tabel berikut.

Tabel. Hasil Reliabilitas soal Pilihan Ganda

Rata rata	8,86
Simpang Baku	2,13
KorelasiXY	0,66
Reliabilitas Tes	0,79

Sedangkan untuk nilai Reliabilitas Tes Uraian bernilai sebesar 0,61 dengan kategori yang digunakan maka nilai Reliabilitas sebesar 0,61 untuk Uraian termasuk ke kategori Tinggi. Berikut merupakan tabel hasil persebaran soal *HOTS* Biologi Kelas 12 pada aspek Reliabilitas Tes soal Uraian.

Tabel. Hasil Reliabilitas soal Uraian

Rata rata	11,52
Simpang Baku	7,52
KorelasiXY	0,43
Reliabilitas Tes	0,61

5. Analisis Kualitas Pengecoh

Hasil pemeriksaan butir-butir *HOTS* Biologi Kelas 12 dengan menggunakan perangkat lunak Anates dipecah menjadi 5 (lima) tingkatan terpisah, yaitu *Very Bad*, *Bad*, *Less*, *Good*, dan *Very Good*. Jumlah siswa yang memilih setiap opsi jawaban terlihat di aplikasi Anates untuk tabel Kualitas Pengecoh, tetapi nama dan asal kelompok siswa disembunyikan. Pembagian soal *HOTS* Biologi Kelas 12 pada aspek Soal Pilihan Ganda Kualitas 15 soal dengan 5 kemungkinan respon pada setiap butir soal ditunjukkan pada tabel di bawah ini.

Tabel. Hasil Kualitas Pengecoh

Kategori Soal	Jumlah	Nomor Soal
Sangat Buruk	5	Butir soal pilihan ganda nomor 4, 5, 7, 11, 14
Buruk	4	Butir soal pilihan ganda nomor 1, 3, 12, 13
Kurang	3	Butir soal pilihan ganda nomor 8, 9, 10
Baik	2	Butir soal pilihan ganda nomor 2, 15
Sangat Baik	1	Butir soal pilihan ganda nomor 6
Jumlah		15

Sebuah pengecoh dikatakan telah dapat menjalankan fungsinya dengan baik apabila pengecoh tersebut mempunyai daya tarik sedemikian rupa sehingga siswa yang kurang menguasai konsep merasa bimbang dan pada akhirnya mereka terkecoh untuk memilih pengecoh sebagai jawaban benar. Efektivitas Pengecoh dapat dilihat dengan menghitung indeks pengecoh.

Untuk menentukan Efektivitas Pengecoh setiap butir soal digunakan skala dengan rentang antara sangat baik sampai sangat tidak baik.

- 1) Efektivitas Pengecoh butir soal dikatakan sangat baik apabila keempat pengecoh berfungsi.
- 2) Efektivitas Pengecoh butir soal dikatakan baik apabila tiga pengecoh berfungsi.
- 3) Efektivitas Pengecoh butir soal dikatakan kurang baik apabila dua pengecoh berfungsi.
- 4) Efektivitas Pengecoh butir soal dikatakan tidak baik apabila hanya satu pengecoh berfungsi.
- 5) Efektivitas Pengecoh butir soal dikatakan sangat tidak baik apabila semua pengecoh tidak berfungsi.

Kriteria berikut digunakan untuk menilai validitas, tingkat kesulitan, kekuatan pembeda, dan kemanjuran yang menyesatkan dari item pertanyaan:

- a. Suatu item pertanyaan dikatakan berkualitas tinggi jika memenuhi kriteria berikut:
 - 1) Dalam hal validitas, rpbj harus lebih besar dari atau sama dengan rtabel pada tingkat signifikansi 5% berdasarkan ukuran sampel.
 - 2) Item pertanyaan yang sangat baik, dalam hal tingkat kesulitan, adalah yang memiliki tingkat kesulitan sedang dengan indeks kesulitan 0,31 hingga 0,70.
 - 3) Item pertanyaan dianggap baik dalam hal daya pembeda jika termasuk dalam kategori daya pembeda sedang dengan indeks daya pembeda 0,20 hingga 0,40; kategori daya pembeda baik dengan indeks daya pembeda 0,40 hingga 0,70; dan kategori daya pembeda sangat baik dengan indeks daya pembeda 0,71 hingga 1,00.
 - 4) Suatu butir pertanyaan dianggap baik dalam hal kemanjuran pengecoh jika termasuk dalam kategori baik atau memiliki 3 deseptor yang berfungsi, dan dianggap sangat baik jika memiliki 4 deseptor yang berfungsi.

- b. Butir soal dikatakan memiliki kualitas yang kurang baik apabila hanya memenuhi tiga dari keempat kriteria butir soal yang baik.
- c. Butir soal dikatakan memiliki kualitas yang tidak baik apabila hanya memenuhi dua atau kurang dari dua kriteria dari keempat kriteria butir soal yang baik.

KESIMPULAN DAN SARAN

Menurut Elviana (2020) menyatakan bahwa Daya Pembeda merupakan suatu indeks yang menunjukkan kemampuan butir soal untuk membedakan kelompok atas (kelompok unggul, siswa yang nilai tesnya tinggi) dan kelompok bawah (kelompok asor, siswa yang nilai tesnya rendah). Aspek Daya Pembeda memiliki tujuan untuk dapat menentukan butir soal termasuk ke dalam kategori skor tinggi atau rendah pada keseluruhan tes. Daya Pembeda dibagi menjadi 5 tingkatan yaitu, sangat buruk, buruk, agak baik atau cukup, baik dan sangat baik. Hasil analisis butir soal *HOTS* Biologi di kelas 12 dengan menggunakan *software* Anates, pada aspek daya pembeda soal pilihan ganda dan soal uraian menunjukkan bahwa pilihan ganda nomor 5 dan uraian nomor 2 termasuk kategori soal yang buruk. Artinya, soal ini harus diganti karena memiliki Daya Pembeda yang sangat buruk yaitu hanya di rentang negatif sampai 9% saja. Untuk soal pilihan ganda nomor 1, 2, 5, 11, 14 dan uraian nomor 7 dan 10 termasuk ke dalam kategori buruk, artinya soal ini juga perlu diperbaiki kembali karena tentang Daya Pembedanya hanya berkisar 10-19% saja. Selanjutnya pilihan ganda nomor 2, 3, 13, 15 termasuk ke dalam katgeori cukup dengan rentang Daya Pembedanya sebesar 20-29%. Kemudian pilihan ganda no 8, 9, 12 dan uraian nomor 1, 4, 5, 9 termasuk kategori baik karena daya pembeda berkisar 30-49% dan untuk soal pilihan ganda nomor 6, 7, 10 dan uraian nomor 3, 6, 8 termasuk kategori sangat baik karena Daya Pembedanya di atas angka 50%.

Pada aspek tingkat kesukaran, Kesukaran soal Pilihan Ganda dan soal Uraian. Butir soal pilihan ganda nomor 1, 8, 11, 12 termasuk ke dalam soal berkategori sangat mudah. Pada butir soal pilihan ganda nomor 7, 9, 13, 15 termasuk soal mudah, lalu kategori butir soal pada tingkat sedang yaitu butir soal pilihan ganda nomor 2, 6, 10, 14 dan uraian nomor 1, 6, 8. Selanjutnya butir soal pada kategori Sukar, terdapat di soal pilihan ganda nomor 5 dan uraian nomor 3, 4, 5, 9 dan begitu pula pada Butir soal pilihan ganda nomor 3, 4 dan uraian nomor 2, 7, 10 yang termasuk ke dalam kategori sangat sukar. Pada aspek validitas, yang termasuk kategori valid atau signifikan, yaitu pada butir soal pilihan ganda nomor 1, 6, 7, 8, 11, 12, 15 dan uraian nomor 3, 5, 6, 8, 9, 10, sedangkan untuk butir soal yang tidak valid atau signifikan pada butir soal pilihan ganda nomor 2, 3, 4, 5, 9, 10, 13, 14 dan uraian nomor 1, 2, 4, 7. Pada aspek pengecoh soal, hasil pemeriksaan butir-butir *HOTS* Biologi Kelas 12 dibagi menjadi 5 (lima) tingkatan terpisah, yaitu *Very Bad*, *Bad*, *Less*, *Good*, dan *Very Good*. Jumlah siswa yang memilih setiap opsi jawaban terlihat di aplikasi Anates untuk tabel Kualitas Pengecoh, tetapi nama dan asal kelompok siswa disembunyikan.

Dari pembahasan materi di atas, diharapkan pembaca dapat mengetahui kualitas dari soal soal yang dibuat, agar ketika membuat soal baru menjadi lebih baik.

DAFTAR PUSTAKA

- Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. *Jurnal MUDARRISUNA*, 10 (2).
- Fiska, Joelan M., Yunin H., Nur Qomaria., dan Wiwin Puspita H. (2021). Analisis Butir Soal Ulangan Harian IPA Menggunakan Software Anates pada Pendekatan Teori Tes. *Jurnal Natural Science Educational Research*. 4(1).
- N, Aisyah., Sabdaningtyas L., Ambarita A. (2017). Pengembangan Instrumen Penilaian Tematik An-Nahl Berbasis Higher Order Thingking Skills untuk Peserta Didik Kelas III. *Parole; Jurnal Pendidikan Bahasa dan sastra Indonesia*, 2(4).
- N, Hidayah., Tureni D, Astija., & Shamdas G. (2022). Analisis Soal Ujian Berbasis HOTS (Higher Order Thingking Skills). *Jurnal Natural Science Educational Research*. 4(1).
- Nusantari, D. O. (2016). Meningkatkan Estimasi Reliabilitas Instrumen Pilihan Ganda dengan Memperhatikan Probabilitas Guessing Pada Soal Ujian Sekolah Kelas XII Mata Pelajaran Bahasa Indonesia Di SMK. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, 6(2).
- Qoriaha, Y., Sumarno., & Nurul, U. (2017). The Development Prehistoric of Jember Tourism Module Using Dick and Carey Model. *Jurnal Historica*. 1(1).
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Praujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1).
- Suryapusparini, BK., Wardono, W., & Kartono, K. (2018, Februari). Analisis Soal-Soal Matematika Tipe Higher Order Thinking Skill (HOTS) pada Kurikulum 2013 untuk Mendukung Kemampuan Literasi Siswa. *PRISMA, Prosiding Seminar Nasional Matematika*, 1, 876-884.
- T, Wiradani., Kasih C., & Latifah. (2019). Analisis Butir Soal HOTS (High Order Thingking Skills). *Jurnal Pendidikan Bahasa dan sastra Indonesia*, 2(4).
- Warju., Sudirman R., Ariyanto., Soeryanto., dan Rio Adi T. (2020). Analisis Kualitas Butir Soal Tipe HOTS pada Kompetensi Sistem Rem Siswa di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi dan Kejuruan*. 17(1).