



Analisis Kualitas Soal Sumatif Biologi SMA

Yusnaeni^{1*}, Fransiska S. Herin Samon¹, Mario J. Santrum¹, Sudirman²

¹ Program Studi S-1 Pendidikan Biologi, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Nusa Cendana Kupang, Indonesia

² Program Studi S-1 Pendidikan Kimia, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Nusa Cendana Kupang, Indonesia

*Email: yusnaeni_75@yahoo.co.id

Info Artikel	Abstrak
<p>Diterima: 9 Maret 2023 Direvisi: 2 Mei 2023 Diterbitkan: 29 Mei 2023</p> <hr/> <p>Keywords: Analisis soal, asesmen sumatif, HOTS, kualitatif, kuantitatif.</p>	<p>Analisis kualitas soal dimaksudkan untuk melihat sejauh mana kualitas soal yang disusun guru dalam kaitannya dengan tuntutan assessmen yang diharapkan dalam pembelajaran biologi di SMA. Sampel diambil pada soal UAS semester ganjil TA 2019/2020 dan 2020/2021 SMA Negeri 1 Adonara Timur NTT. Instrumen pengumpul data berupa lembar wawancara dan lembar validasi instrument tes. Data analisis kualitas soal mencakup analisis kualitas dan kuantitas, selanjutnya data tersebut dianalisis secara deskriptif. Hasil penelitian menunjukkan bahwa secara kualitas, soal TA 2019/2020 dan 2020/2021 dikategorikan dalam kriteria baik, sedangkan secara kuantitas menunjukkan hasil yang beragam baik dari segi validitas, reliabilitas, tingkat kesukaran, daya pembeda, efektivitas pengecoh dan persentase soal HOTS. Secara umum disimpulkan bahwa distribusi hasil analisis kuantitatif belum mengacu pada kelayakan instrument yang dianjurkan dalam assessment penilaian peserta didik dan masih perlu dilakukan revisi. Sedangkan persentase soal HOTS belum dikategorikan baik karena soal yang berada pada level kognitif C4 sampai dengan C6 kurang dari 35%.</p>

© 2023 Yusnaeni. This is an open-access article under the CC BY-SA license (<https://creativecommons.org/licenses/by/4.0>)

PENDAHULUAN

Pendidikan adalah sebuah usaha untuk mewujudkan potensi seseorang dimana pendidikan bermutu dapat menjadi indikator kemajuan suatu bangsa. Mutu pendidikan tidak lepas dari mutu pembelajaran dan mutu penilaian di kelas. Oleh karena itu diperlukan upaya yang maksimal dari seorang guru untuk menciptakan pembelajaran yang berkualitas dan penilaian yang berkualitas. Dalam Kemendikbud (2005) secara tegas dijelaskan bahwa kemampuan menyusun alat evaluasi menjadi salah satu kompetensi profesionalisme seorang guru. Evaluasi yang berkualitas akan mendorong guru untuk menyusun alat penilaian yang baik. Ketika guru melakukan penilaian yang



baik, tentunya akan memberi dampak dalam proses dan hasil pembelajaran (Popham, 2009 *dalam* Setiadi, 2016). Dan ketika penilaian memberi dampak, maka dapat dijadikan sebagai dasar dalam penentuan kebijakan berikutnya.

Ada beberapa syarat yang dijadikan acuan untuk menentukan baik tidaknya kualitas alat evaluasi yang telah disusun yakni: 1) validitas, 2) reliabilitas, 3) objektivitas, 4) praktikabilitas, dan 5. ekonomis (Arikunto, 2018). Selain itu, secara kualitatif juga perlu didasarkan atas pertimbangan substansi, konstruksi dan bahasa. Alat evaluasi juga harus mampu mengidentifikasi kemampuan dari setiap peserta didik dalam pembelajaran yang diperoleh. Tes merupakan contoh alat evaluasi yang biasa digunakan dalam menentukan kemampuan pencapaian materi pelajaran yang telah diberikan guru. Analisis soal mutlak dilakukan agar bebas dari bias dan terkontrol, serta dapat menginterpretasikan pencapaian belajar peserta didik. Menurut Lissitz & Samuels (2007), analisis soal dapat menjadi salah satu upaya validasi berbasis data empirik. Gierl & Lai, (2013) juga menambahkan bahwa umumnya kualitas butir soal ditentukan melalui uji empiric. Oleh karena itu, dalam penyusunan soal idealnya mengacu pada kaidah penyusunan soal agar mampu menjangkau kompetensi yang ditetapkan dan dapat mengemban fungsinya secara tepat.

Penulisan soal sebagai salah satu alat evaluasi sejatinya mengacu pada merupakan pedoman atau kaidah dalam penulisan soal sehingga soal tersebut mampu mengukur informasi yang seharusnya diukur. Tes ujian akhir termasuk salah satu contoh tes sumatif. Menurut Sudijono (2012), tes sumatif merupakan tes yang dilaksanakan setelah penyampaian materi pembelajaran dalam satu topik atau program. Lebih lanjut oleh Kemendikbud (2007), diungkap bahwa UAS juga adalah penilaian berbentuk tes yang dilakukan oleh pihak sekolah yang bertujuan untuk mengetahui tingkat keberhasilan peserta didik selama satu semester.

Penyusunan soal UAS biasanya dibuat oleh kelompok guru yang mengacu pada kisi-kisi standar yang dikeluarkan oleh badan standar nasional. Berdasarkan hasil interview pada sampel guru biologi di SMA Negeri 1 Adonara Timur diperoleh informasi bahwa soal-soal yang diujikan dalam UAS adalah soal yang dikonstruksi sendiri oleh gurunya, dimana proses penyusunannya sudah mengikuti kaidah yang ditetapkan dan sudah disesuaikan dengan isi kurikulum yang berlaku, demikian pula dengan kontennya. Hanya guru belum pernah menganalisis kualitas soal yang diujikan baik dari aspek kuantitatif maupun kualitatif. Oleh karena itu soal UAS tersebut perlu dianalisis apakah soal tersebut memenuhi standar sebagai alat ukur pencapaian belajar peserta didik. Selain itu juga dilakukan survey dalam proses pembelajaran dan arsip document yang ada terlihat bahwa 1) dalam kegiatan pembelajaran, peserta didik tidak dibiasakan untuk mengerjakan soal berbasis High Order Thinking Skill (HOTS) sehingga kemampuan berpikir tingkat tinggi peserta didik, 2) soal UAS yang digunakan di sekolah tersebut adalah soal pilihan ganda, yang sepiantas terlihat baik secara teknis penyusunan, namun secara kasat mata terlihat masih belum HOTS. Dimana sekitar 60%-70% item soal masih di level kognitif C1 sampai C3 sementara menurut Setiawati *dkk* (2018) standar evaluasi sekarang menghendaki soal dengan level kognitif C4 sampai C6 dan berbasis konseptual. Level kognitif kategori HOTS menurut Anderson dan Karthwohl (2001) terdistribusi pada *level kemampuan* analisis (C4), evaluasi (C5) dan mencipta (C6). Level C4 sampai C6 adalah tuntutan soal yang perlu dikembangkan oleh seorang guru ketika ingin menumbuhkan keterampilan berpikir tingkat tinggi peserta didik (Aydin & Yilmaz, 2010).

Atas hal tersebut dipandang perlu untuk dilakukan pengkajian soal UAS tersebut, terlebih lagi beberapa hasil penelitian sebelumnya tentang analisis soal juga telah dilaporkan bahwa soal-soal UAS pada beberapa sekolah masih banyak yang perlu diperbaiki karena belum layak untuk digunakan. Diantara hasil penelitian tersebut dilaporkan oleh Subari *dkk* (2021), dimana analisis secara kualitatif dan kuantitatif soal UAS semester ganjil Biologi di MAN 2 Jambi secara keseluruhan tidak ada soal yang dapat langsung dipakai, namun harus dilakukan perbaikan pada beberapa soal, dan sisanya tidak dapat dipakai. Terdapat 11 soal yang perlu dilakukan perbaikan, dan sisanya 14 soal yang tidak dapat dipakai. Hal serupa juga dilaporkan oleh Rahmani *dkk* (2013) bahwa ada beberapa soal buatan guru biologi di SMA Negeri 1 Tanah Pino yang masih perlu diperbaiki dari segi konstruksi. Selanjutnya oleh Arifin (2015), juga melaporkan bahwa soal

ulangan akhir semester genap mata pelajaran biologi siswa kelas X MAN Filial Pontianak tahun ajaran 2014/2015 termasuk kriteria yang kurang baik kualitasnya.

METODE

Penelitian ini tergolong penelitian deskriptif, dilakukan di SMA Negeri 1 Adonara Timur. Sampel tes yang dianalisis adalah tes UAS biologi pada kelas X, XI, dan XII pada TA 2019/2020 dan 2020/2021. Total sampel siswa untuk keperluan Analisis kuantitatif tes UAS sebanyak 486 siswa, yang terdistribusi untuk soal UAS TA 2019/2020 kelas X 80 siswa, kelas XI 78 siswa, kelas XII 68 siswa. Untuk soal UAS TA 2020/2021 kelas X, XI dan XII jumlah sampel berturut-turut adalah 86; 82 dan 74 sampel. Penentuan sampel dilakukan secara acak. Sementara untuk keperluan analisis kualitas instrument dilakukan oleh 3 orang validator ahli instrumen. Prosedur pengumpulan data dilakukan teknik tes, dan dokumentasi. Dalam teknik ini, dokumen berupa hasil tes yang telah dikerjakan oleh peserta didik seperti sampel yang diungkap sebelumnya dianalisis kualitasnya. Analisis tersebut mencakup analisis kuantitatif dan kualitatif. Analisis kuantitatif instrument tes meliputi tingkat kesukaran, daya pembeda, validitas, reliabilitas, dan efektifitas pengecoh. Sedangkan analisis kualitatif dari segi materi, konstruksi, dan bahasa yang dilakukan oleh validator ahli instrumen. Hasil validasi ditentukan dengan kriteria yang mengacu pada Arikunto (2018) yakni: 1,00 - 1,50 (kurang), 1,56 - 2,50 (cukup), 2,56 - 3,50 (baik), 3,56 - 4,00 (sangat baik). Untuk teknik analisis data hasil analisis kuantitatif terkait tingkat kesukaran mengacu pada Sudaryono (2012) dengan kriteria: 0,00 - 0,30 (Sukar), 0,31 - 0,70 (Sedang), dan 0,71 - 1,00 (mudah), daya pembeda mengacu pada Arikunto (2018) dengan kriteria: 0,00 - 0,20 (jelek), 0,21 - 0,40 (cukup), 0,41 - 0,70 (baik), dan 0,71 - 1,00 (baik sekali), validitas dengan harga kritik untuk validitas butir instrumen adalah 0,3. Artinya apabila r_{xy} lebih besar atau sama dengan 0,3 ($r_{xy} \geq 0,3$) maka nomor butir tersebut dapat dikatakan valid (Widoyoko, 2012), reliabilitas Spearman-Brown (Widoyoko, 2012) dimana bila r hitung lebih besar atau sama dengan r tabel ($r_h \geq r_t$) diartikan ada korelasi yang signifikan, instrumen dianggap reliabel. Sementara efektifitas pengecoh diukur dengan menggunakan mengacu pada Arifin (2013) dengan kriteria pengecoh pada butir soal dikatakan baik jika pengecoh tersebut telah dipilih sekurang-kurangnya 5% dari jumlah peserta tes. Untuk persentase soal HOTS ditentukan dari hasil analisis validator saat memvalidasi instrument tes UAS berdasarkan rubrik penelaahan butir soal HOTS bentuk pilihan ganda (Depdiknas, 2017).

HASIL DAN PEMBAHASAN

1. Analisis Kualitatif

Hasil analisis kualitatif yang mencakup materi, konstruksi dan bahasa pada instrument tes ujian akhir buatan guru dari 3 orang validator yang dapat dilihat pada Tabel 1.

Tabel 1.

Data Hasil Validasi 3 Validator Ahli Pada Tes Buatan Guru Untuk 3 aspek pada Soal TA 2019/2020 dan TA 2020/2021.

No	Aspek	Tahun Ajaran					
		2019/2020			2020/2021		
		X	XI	XII	X	XI	XII
1	Materi	2,50	2,66	2,38	2,50	2,17	2,50
2	Konstruksi	2,63	2,82	2,74	2,70	2,67	2,89
3	Bahasa	3,25	3,08	3,34	3,25	3,17	3,42
Rata-rata skor		2,80	2,85	2,82	2,82	2,67	2,94
Kriteria		Baik	Baik	Baik	Baik	Baik	Baik

Tabel 1 menunjukkan bahwa secara kualitas instrument tes ujian akhir buatan guru pada TA 2019/2020 dan TA 2020/2021 untuk semua kelas sudah dikategorikan baik, karena nilai rata-rata skornya berada pada rentang 2,56 - 3,50. Acuan ini berdasarkan pada Arikunto (2018). Soal tersebut dikategorikan baik dikarenakan syarat-syarat yang diperlukan dalam penyusunan soal

sudah terpenuhi misalnya dari segi materi sudah sesuai dengan indikator yang ditetapkan, menggunakan stimulus dan tidak mengandung unsur SARAPP. Sementara dari segi konstruksi sudah memenuhi konstruksi misalnya pokok soal sudah dirumuskan dengan tegas dan jelas serta tidak mengandung pernyataan bersifat negative ganda. Dari segi bahasa juga sudah komunikatif dan menggunakan kaidah bahasa yang benar. Namun demikian tetap juga masih banyak kekurangan yang harus dilakukan perbaikan terutama dari segi materi dan konstruksi soal. Rata-rata soal yang dibuat belum menggunakan stimulus yang menarik dan kontekstual, termasuk belum mengakomodir ranah di atas level C3. Sehingga saat dilakukan analisis lebih jauh tentang berapa persentase soal HOTS, maka diperoleh hasil seperti pada Tabel 2.

Tabel 2.

Persentase soal HOTS Soal UAS Biologi TA 2019/2020 dan TA 2020/2021.

Tahun Ajaran	Kelas	Jmlah Soal	Tingkat Kognitif						% HOTS
			C1	C2	C3	C4	C5	C6	
2019/2020	X	40	(15%)	(32%)	(42,5%)	(10%)	(0%)	(0%)	10,00%
	XI	30	(40%)	(26,66%)	(16,67%)	(16,67%)	(0%)	(0%)	16,67%
	XII	40	(10%)	(30%)	(27,5%)	(30%)	(2,5%)	(0%)	32,5%
2020/2021	X	30	(10%)	(33,33%)	(46,67%)	(10%)	(0%)	(0%)	3%
	XI	25	(76%)	(12%)	(12%)	(0%)	(0%)	(0%)	0%
	XII	35	(25,71%)	(17,14%)	(34,29%)	(22,86%)	(0%)	(0%)	22,86%

Tabel 2 menunjukkan bahwa persentase soal HOTS untuk soal UAS buatan guru masih rendah (masih dibawah 50%), bahkan ada soal di TA 2020/2021 yang sama sekali belum memiliki soal HOTS (persentasi 0%) yakni pada kelas XI. Hasil ini kategori rendah karena soal yang disusun tersebut lebih banyak mengandalkan ingatan, belum melibatkan kemampuan berpikir peserta didik saat menentukan jawaban. Di sisi lain, kehadiran stimulus yang diberikan belum berfungsi sepenuhnya untuk memandu kemampuan berpikir. Sejatinya soal HOTS itu bertujuan melatih siswa memahami konsep, menerapkan konsep, menganalisis dan memecahkan masalah dan bukan menghafal saja. Soal HOTS adalah soal-soal yang mengukur kemampuan berpikir tingkat tinggi dalam Taksonomi Bloom yang telah disempurnakan oleh Anderson & Krathwohl (2001). HOTS dapat meningkatkan kemampuan peserta didik dalam memunculkan ide atau gagasan, mengargumentasikan ide dengan terstruktur dan jelas (Dinni, 2018). Soal HOTS umumnya mengukur dimensi metakognitif tidak hanya dari segi faktual, konseptual, prosedural. HOTS dapat melatih berpikir logis, kritis, kreatif dan problem solving (Setiawati et al., 2019). Fanani (2018) menambahkan bahwa keuntungan dari penilaian HOTS adalah dapat meningkatkan motivasi dan pencapaian hasil belajar.

Hasil penelitian ini tidak jauh berbeda dari beberapa hasil penelitian terdahulu yang dilakukan oleh Acesta (2020) yakni 5% dari 20 item soal buatan guru kategori HOTS. Muzayyanah et al. (2020) juga melaporkan bahwa dari 40 item hanya terdapat 27% soal HOTS. Hal yang sama juga diperoleh Zohratunaqiyah dkk (2023) yakni hanya terdapat 7,5 % soal HOTS dari 40 item. Khan dan Inamullah (2011) juga mengungkapkan bahwa dalam pembelajaran sebagian besar guru ketika mengajukan pertanyaan berada kategori tingkat rendah.

Soal HOTS memiliki beberapa karakteristik yaitu mengukur kemampuan berpikir tingkat tinggi, berbasis permasalahan kontekstual, tidak rutin dan menggunakan bentuk soal beragam. Karakteristik pertama mengukur kemampuan berpikir tingkat tinggi. Kemampuan berfikir tingkat tinggi mencakup kemampuan pemecahan masalah, berfikir kritis, berfikir kreatif, kemampuan membuat keputusan, serta kemampuan memberikan pendapat. Kemampuan berpikir tingkat tinggi merupakan salah satu kompetensi penting dalam abad 21, sehingga wajib dimiliki oleh setiap peserta didik (Kunanti, 2020).

2. Analisis Kuantitatif

Hasil analisis kuantitatif dari jawaban siswa atas tes UAS yang diberikan berupa tingkat kesukaran, daya pembeda, validitas, reliabilitas, dan efektifitas pengecoh selengkapnya dapat dilihat pada Tabel 3.

Tabel 3.

Indikator Kuantitatif Soal UAS Ganjil TA 2019/2020 dan TA 2020/2021.

Tahun Ajaran	Kelas	Item	Indikator Kuantitatif				
			Tingkat Kesukaran	Daya Pembeda	Validitas	Reliabilitas	Efektifitas pengecoh
2019/2020	X	40	4 sedang, 36 mudah	1 tidak baik, 5 jelek, 21 cukup, 13 baik	34 valid, 6 tidak valid	0,89 (Reliabel)	Semua item jelek
	XI	30	2 sedang, 28 mudah	8 jelek, 11 cukup, 11 baik	23 valid, 7 tidak valid	0,84 (Reliabel)	Semua item jelek
	XII	40	5 sedang, 35 mudah	3 tidak baik, 8 jelek, 15 cukup, 14 baik	26 valid, 14 tidak valid	0,83 (Reliabel)	Semua item jelek
2020/2021	X	30	6 sukar, 19 sedang, 5 mudah	2 tidak baik, 2 jelek, 8 cukup, 8 baik, 10 baik sekali	26 valid, 4 tidak valid	0,82 (Reliabel)	20 jelek, 10 kurang baik
	XI	25	14 sukar, 10 sedang, 1 mudah	3 jelek, 9 cukup, 12 baik, 1 baik sekali	21 valid, 4 tidak valid	0,73 (Reliabel)	13 jelek, 10 kurang baik, 2 baik
	XII	35	25 sukar, 10 sedang	2 tidak baik, 17 jelek, 8 cukup, 7 baik, 1 baik sekali	12 valid, 23 tidak valid	0,10 (tdk reliabel)	12 jelek, 20 kurang baik, 3 baik

Tabel 3 menunjukkan bahwa analisis kuantitas soal UAS buatan guru TA 2019/2020 dan TA 2020/2021 untuk kelas X sampai kelas XII sangat variatif. Ditinjau dari tingkat kesukaran yang belum proporsional dari segi perbandingan antara soal sukar, mudah dan sedang. Tidak proporsionalnya hasil tersebut karena pada tingkatan kelas ada soal yang hanya terdistribusi pada kategori mudah dan sedang, ataupun sedang sukar. Padahal idealnya soal yang diberikan ada soal yang sukar, ada yang mudah dan ada yang sedang sehingga bisa membedakan kemampuan peserta didik. Bagi peserta didik dengan kemampuan akademik yang tinggi tentu bisa menjawab semua soal yang diberikan, akan tetapi bagi peserta didik yang berkemampuan rendah tentunya hanya mampu menyelesaikan soal kategori yang mudah. Pada tabel juga jelas terlihat bahwa tingkat kesukaran soal TA 2019/2020 kelas X (90% mudah, 10% sedang), kelas XI (93,33% mudah, 6,77% sedang), kelas XII (87,5% mudah, 12,5% sedang), soal TA 2020/2021 kelas X 16,67% mudah, 73,33% sedang, 20,0% sukar, kelas XI 4,0% mudah, 40,0% sedang, 56,0% sukar, kelas XII 28,57% sedang, 71,43% sukar). Menurut Arikunto (2018) dalam pemberian tes pada peserta didik, idealnya mengikuti proporsi soal 30% mudah, 40% sedang dan 30% sukar. Pernyataan tersebut mengindikasikan bahwa soal dengan perbandingan yang dimaksud mampu membedakan antara siswa berkemampuan akademik rendah, sedang dan tinggi. Siswa yang berkemampuan akademik tinggi tentu mampu menjawab semua soal tersebut dengan baik. Sebaliknya, siswa yang berkemampuan akademik rendah hanya mampu menjawab sekitar 30% dari total soal yang ada.

Indikator daya pembeda juga menunjukkan bahwa soal UAS ganjil biologi tahun ajaran 2019/2020 dan 2020/2021 pada kelas X sampai kelas XII, bekerja tidak maksimal sehingga belum bisa membedakan peserta didik kelompok atas (sudah memahami materi) dan kelompok bawah (belum memahami materi), padahal menurut Sofiah dkk (2009), soal dikategorikan bermutu jika soal yang digunakan dalam penilaian dan evaluasi hasil belajar akan memberikan informasi yang tepat sesuai dengan tujuannya misal, mampu membedakan peserta didik yang sudah atau belum kompeten. Daya pembeda butir soal bergantung pada besar kecilnya nilai indeks diskriminasi, jika memiliki angka indeks diskriminasi dengan tanda positif ($D > 0$), maka dapat dikatakan bahwa butir soal tersebut telah memiliki daya pembeda. Hal tersebut dapat diartikan bahwa peserta tes yang termasuk kelompok Atas lebih banyak yang dapat menjawab dengan benar terhadap butir soal yang bersangkutan, sedangkan peserta tes yang termasuk kelompok Bawah lebih banyak yang menjawab salah. Hal ini diperkuat oleh Elfrida (2017) bahwa daya beda tes

berhubungan dengan kemampuan peserta tes dalam kelompok tertentu menjawab tes dihitung dari hasil tes kelompok peserta ujian tertentu.

Sementara dari segi validitas soal, terlihat bahwa UAS ganjil biologi tahun ajaran 2019/2020 pada kelas X sampai kelas XII, dan tahun ajaran 2020/2021 kelas X dan XI sudah baik karena hanya sedikit saja item soal yang tidak valid (soal pada kelas XII). Ini berarti bahwa soal tersebut sudah dapat mengukur apa yang seharusnya diukur pada peserta didik. Menurut Sani dkk (2020), sebuah tes disebut valid apabila tes itu dapat dengan tepat mengukur apa yang hendak diukur. Validitas tes mencakup isi, konstruk, ramalan, dan kriteria. Validitas isi merupakan kemampuan alat ukur yang tepat untuk mengukur yang seharusnya diukur tentunya pada materi yang diajarkan. Ketika tes itu tepat mengukur konsep maka sudah valid secara. Sedangkan manakala terdapat hubungan skor pada kelompok yang dibandingkan maka telah memenuhi validitas bandingan. Demikian pula ketika alat ukur itu dapat memperdiksikemampuan ke depan dan relevan maka memenuhi komponen validitas prediksi dan kriteria. Lebih lanjut Djemari (2012) *dalam* Hairun (2020) mengungkapkan bahwa ketika penggunaan tes mendukung mendukung penafsiran skor tes yang merujuk pada derajat bukti dan teori maka tes tersebut dikatakan valid.

Selanjutnya untuk indikator reliabilitas, terlihat bahwa hampir semua tes UAS tersebut reliable, kecuali pada tes UAS kelas XII tahun ajaran 2020/2021. Tes tersebut reliable memberikan data pengukuran hasil yang tetap walaupun diberikan pada waktu yang berbeda kepada responden yang sama. Kestabilan data skor yang diperoleh orang yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda atau dari satu pengukuran ke pengukuran lainnya. Sebaliknya untuk tes TA 2020/2021 tidak reliable karena terdapat beberapa data yang tidak konsisten sehingga saat pengukuran diperoleh hasil yang tidak reliable. Hal ini dapat terjadi karena adanya variabilitas kelompok yang diuji, level kelompok yang diuji dan panjang pendeknya tes. Hasil ini mengindikasikan bahwa tes UAS tersebut dianggap sudah konsisten dan dapat digunakan untuk pengukuran di sekolah ini pada tahun selanjutnya maupun digunakan oleh sekolah lain. Menurut Arikunto (2018), reliabilitas adalah ketetapan suatu tes apabila diteskan pada subjek yang sama. Lebih lanjut, Jannah (2019) menambahkan bahwa reliabilitas merupakan suatu pengukuran yang dilakukan secara berulang dalam melakukan evaluasi agar mendapatkan hasil pengukuran di setiap evaluasinya dengan skor nilai yang sama. Indikator yang terakhir terkait kuantitas tes adalah efektivitas pengecoh. Dari Tabel 3 terlihat bahwa soal UAS ganjil biologi TA 2019/2020 dan TA 2020/2021 untuk semua jenjang kelas hampir semuanya tidak berfungsi, meskipun ada sedikit pengecoh bisa bekerja karena ada beberapa alternatif pengecoh yang sudah berfungsi namun belum maksimal sehingga belum mampu membedakan kemampuan peserta didik. Tidak berfungsinya pengecoh karena distribusi jawaban peserta didik pada option tidak merata antara option yang merupakan jawaban dengan option yang bukan jawaban. Hal ini disebabkan oleh kurang menariknya alternatif pengecoh sehingga membuat peserta didik tidak tertarik untuk memilih jawaban tersebut. Menurut Daryanto (2012), efektifitas pengecoh dapat dikatakan berfungsi dengan baik apabila pengecoh yang dibuat mampu menarik peserta tes untuk memilihnya.

KESIMPULAN

Berdasarkan hasil analisis soal UAS ganjil mata pelajaran biologi SMA Negeri 1 Adonara Timur Kabupaten Flores Timur, dapat disimpulkan bahwa: (1) Kualitas soal UAS mata pelajaran biologi SMA Negeri 1 Adonara Timur secara kualitatif dari tiga aspek yang diteliti yaitu materi, konstruksi dan bahasa dikategorikan dalam kriteria baik. Sedangkan secara kuantitatif dari segi tingkat kesukaran baik tahun ajaran 2019/2020 maupun 2020/2021 dikategorikan kurang baik. Untuk daya pembeda perlu juga masih dalam kriteria kurang baik dan perlu banyak perbaikan. Validitas dan reliabilitas pada tahun ajaran 2019/2020 termasuk kategori baik sedangkan pada tahun ajaran 2020/2021 pada kelas XII perlu diperbaiki. Kemudian untuk efektifitas pengecoh semua perlu diteliti kembali dan diperbaiki sehingga mampu menjadi alat penilaian yang lebih berkualitas. (2) Presentase soal HOTS pada UAS ganjil mapel biologi SMA N 1 Adonara Timur belum dikategorikan baik karena soal yang berada pada level kognitif C4 - C6 kurang dari 35%.

DAFTAR PUSTAKA

- Acesta, A. (2020). Analisis Kemampuan Higher Order Thingking Skills (Hots) Siswa Materi IPA di Sekolah Dasar. *Quagga: Jurnal Pendidikan dan Biologi*, 12(2), 170-175. p-ISSN 1907-3089, e-ISSN2651-5869
- Anderson, L. W., dan Krathwohl, D.R. (2001). *Revisi Taksonomi Bloom*. Jakarta: Rineka Cipta
- Arifin, M. (2015). Analisis Butir Soal Ulangan Akhir Semester Genap Mata Pelajaran Biologi Kelas X MAN Filial Pontianak Tahun Ajaran 2014/2015. *Skripsi*. Program Studi Pendidikan Biologi FKIP Universitas Muhammadiyah Pontianak
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- Aydin, N., Yilmaz, A. (2010). The effect of constructivist approach in chemistry education on students' higher order cognitive skills. *Journal of Education*, (39), 57-68
- Daryanto. (2012). *Evaluasi pendidikan*. Jakarta: PT Rineka Cipta
- Depdiknas. (2017). *Modul Penyusunan Soal HOTS*. Jakarta: Direktorat Jendral Pembinaan Sekolah Dasar dan Menengah
- Dinni, H. N. (2018). HOTS (High Order Thinking Skills) dan kaitannya dengan kemampuan literasi matematika. *PRISMA, Prosiding Seminar Nasional Matematika*, 1, 170–176.
- Elfrida, A. (2017). *Analisis Kualitas Butir Soal Ujian Akhir Semester Genap Mata Pelajaran Ekonomi Kelas XI SMA Negeri 1 Purworejo Tahun Pelajaran 2016/2017*. *Skripsi*. FKIP Universitas Muhammadiyah Purworejo
- Fanani, M. Z. (2018). Strategi Pengembangan Soal Higher Order Thinking Skill (HOTS) dalam Kurikulum 2013. *Journal of Islamic Religious Education*. 2(1): 57-76.
- Gierl, M. J., & Lai, H. (2013). Instructional Topics in Educational Measurement (ITEMS) Module: Using Automated Processes to Generate Test Items. *Educational Measurement: Issues and Practice*, 32(3), 36–50. <https://doi.org/10.1111/emip.12018>
- Hairun, Y. (2020). *Evaluasi dan Penilaian dalam Pembelajaran*. Ternate: Deepublish
- Jannah, M. (2019). Analisis Butir Soal Ujian Akhir Semester (UAS) Mata Pelajaran Fisika pada Ujian Semester Genap Kelas XI Tahun Ajaran 2017/2018 di SMAN 16 Banda Aceh. *Skripsi*. Program Studi Pendidikan Fisika Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Ar-Raniry Darussalam, Banda Aceh
- Khan, W.B. & Inamullah, H.M. (2011). A Study of Lower-order and Higher-order Questions at Secondary Level. *Canadian Center of Science and Education Asian Social Science*, (7), 149-157. DOI:10.5539/ass.v7n9p149
- Kemendikbud. (2005). *Undang-Undang Republik Indonesia Nomor 14 Tahun 2005 Tentang Guru dan Dosen*. Jakarta: Kemendikbud
- Kemendikbud. (2007). *Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 Tentang Standar Penilaian Pendidikan*. Jakarta: Kemendikbud
- Kunanti, S. E. (2020). Penyusunan Pengembangan Penilaian Berbasis HOTS. Inovasi Pembelajaran Bahasa dan Sastra Indonesia Guna Mendukung Merdeka Belajar pada Era Revolusi Industry 4.0 dan Society. *Prosiding Seminar Nasional PBSI-III*. 19.26.
- Lissitz, R. W., & Samuels, K. (2007). Further Clarification Regarding Validity and Education. *Educational Researcher*, 36(8), 482–484. <https://doi.org/10.3102/0013189X07311612>
- Muzayyanah, A., Wijayanti, A., & Ardiyanto, A. (2020). Pengembangan lembar kerja peserta didik (LKPD) tematik berbasis HOTS (Higher Order Thinking Skill) kelas IV Sekolah Dasar. *Jurnal Pijar MIPA*, 15(5), 452–457. DOI: 10.29303/jpm.v15i5.1961
- Rahmani, M., Ningsih, K., dan Nurdini, A. (2013). Analisis Kualitas Butir Soal Buatan Guru Biologi Kelas X SMA Negeri 1 Tanah Pinoh. *Skripsi*. Program Studi Pendidikan Biologi Jurusan Pendidikan MIPA FKIP UNTAN
- Sani, A. R., Arafah, K., Aziz, I., Suswanto, H., dan Tanjung, R. (2020). *Evaluasi Proses dan Penilaian Hasil Belajar*. Bandung: PT Remaja Rosdakarya
- Setiadi, H. (2016). Pelaksanaan Penilaian pada Kurikulum 2013. *Jurnal Penelitian dan Evaluasi Pendidikan*. 20 (2), 167-178. DOI: <http://dx.doi.org/10.21831/pep.v20i2.7173>

- Setiawati, W., Ariyana, Y., Asmira, O., Bestary, R., dan Pudjiastuti, A. (2018). *Buku Penilaian Berorientasi Higher Order Thinking Skill (HOTS)*. Jakarta: Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan
- Sopiah, A., Sidaurak, S., dan Asi, B. N. (2019). Kualitas Soal Penilaian Akhir Semester (PAS) Buatan Guru Mata Pelajaran Kimia Kelas X IPA SMA Negeri di Kabupaten Seruyan pada Semester Ganjil Tahun Ajaran 2018/2019. *Jurnal Ilmiah Kanderang Tingang*. 10 (2). 110-126
- Subari, A., Lufri., dan Syamsurizal. (2021). Analisis Butir Soal Ujian Akhir Semester Ganjil Biologi Kelas XI MAN 2 Kota Jambi. *Bioscientist: Jurnal Ilmiah Biologi*. 9 (1). 45-53
- Sudaryono, A. (2012). *Dasar-Dasar Evaluasi Pembelajaran*. Yogyakarta: Graha Ilmu
- Sudijono, A. (2013). *Pengantar Evaluasi Pendidikan*. Jakarta: PT. Raja Grafindo Persada
- Widoyoko, E. (2012). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Belajar
- Zahrotunnaqiyah, E., L. Nulhakim., & A. N. Taufik. (2023). Analisis Higher Order Thinking Skills (HOTS) Pada Soal Penilaian Akhir Semester Mata Pelajaran IPA Kelas VII Di Kabupaten Serang. *Jurnal Pendidikan MIPA*. 13 (1), 67 - 72. doi.org/10.37630/jpm.v13i1.810