

ANALISIS BUTIR SOAL ULANGAN HARIAN HASIL PENGEMBANGAN GURU MATERI SISTEM GERAK MANUSIA KELAS XI IPA

Analysis of Daily Exam Test Items Developed by Teacher on the Human Motion System Material for Class XI Natural Science

Wikha Ayu Rismaulhijjah

Program Studi S1 Pendidikan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
E-mail: wikha.17030204075@mhs.unesa.ac.id

Nur Kuswanti

Jurusan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
E-mail: nurkuswanti@unesa.ac.id

Abstrak

Analisis butir soal merupakan proses mengkaji kualitas butir soal untuk menentukan kelayakan penggunaannya. Tujuan penelitian ini untuk menganalisis kualitas butir soal yang dikembangkan guru pada materi Sistem Gerak Manusia kelas XI IPA berdasarkan validitas teoritis dan validitas empiris. Validitas teoritis ditentukan berdasarkan hasil validasi ahli pada aspek tingkat kognitif, materi, konstruksi, dan bahasa. Validitas empiris ditentukan berdasarkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal. Penelitian adalah penelitian deskriptif kualitatif dan kuantitatif. Hasil penelitian menunjukkan bahwa berdasarkan aspek materi dan bahasa, soal dinyatakan valid, sedangkan berdasarkan aspek konstruksi, banyak butir soal tidak valid, serta didominasi tingkat kognitif C1, C2, dan C3. Secara empiris, butir soal tipe pilihan ganda termasuk kategori valid (25%) dan tidak valid (75%), reliabilitas rendah, serta berkategori sangat mudah (20%), mudah (15%), sedang (50%), dan sukar (15%), berdaya beda baik (35%), cukup (35%), dan buruk (30%). Butir soal tipe uraian memiliki kategori valid (60%), tidak valid (40%), reliabilitas rendah, serta termasuk kategori sangat mudah (20%), mudah (40%), dan sedang (40%), berdaya beda sangat baik (20%), baik (20%), dan cukup (40%). Dapat disimpulkan bahwa berdasarkan validitas teoritis, butir soal berkategori valid, namun didominasi tingkat kognitif C1, C2, dan C3, serta tidak valid secara empiris.

Kata Kunci: Analisis butir soal, validitas teoritis, validitas empiris, aplikasi ANATES V4.

Abstract

Test item analysis is a process of assessing the quality of the item to determine its feasibility to be used. The purpose of this study was to analyze the quality of test items developed by a high school biology teacher on the Human Motion System material for Class XI Natural Science based on theoretical and the empirical validities. Their theoretical validity was determined based on the results of expert validation referring to cognitive level, material, construction, and language aspects. Their empirical validity was determined based on their validity, reliability, difficulty levels, and discrimination index. This was a descriptive research with qualitative and quantitative analysis. The results showed that based on the material and language aspects the test items were valid, whereas based on construction aspects, many test items were not valid, and dominated by the cognitive domains at the level of C1, C2, and C3. Empirically, multiple choice test items were valid (25%) and not valid (75%), less reliable, very easy (20%), easy (15%), moderately difficult (50%), and difficult (15%). Based on the discrimination index, the items got good (35%), fair (35%), and bad (30%) categories. The essay test items were valid (60%), not valid (40%), less reliable, very easy (20%), easy (20%), and moderately difficult (40%). Based on the discrimination index, the items got very good (20%), good (20%), and fair (40%) categories. It can be concluded that based on theoretical validity the test items were valid, however, they were dominated by the low level domains (C1, C2, and C3). In addition, they were not valid empirically.

Keywords: Item analysis, theoretical validity, empirical validity, ANATES V4 application.

PENDAHULUAN

Pendidikan adalah hal pokok bagi kemajuan suatu negara. Penerapan pendidikan yang sistematis dan terencana diharapkan mampu mengembangkan bakat dan potensi diri setiap peserta didik, sehingga mampu menghasilkan generasi penerus bangsa yang unggul.

Pendidikan memiliki tiga komponen penting, yaitu kurikulum, proses pembelajaran, dan evaluasi (Amalia dan Widayanti, 2012). Evaluasi ialah kegiatan akhir pembelajaran dalam rangka mengumpulkan dan mengelola informasi, serta bertujuan memperbaiki dan melakukan perencanaan ulang terhadap kegiatan

pembelajaran sesuai dengan Standar Penilaian Pendidikan (Permendikbud Nomor 22 Tahun 2016). Dalam hal ini kegiatan evaluasi dapat digunakan guru untuk mengetahui seberapa tingkat ketercapaian Kompetensi Dasar (KD) dalam proses kegiatan pembelajaran (Ulum, 2017). Kegiatan evaluasi memiliki dua teknik, yaitu non tes dan tes. Non tes digunakan oleh guru untuk mengevaluasi dari ranah sikap dan keterampilan, sedangkan tes dari ranah pengetahuan. Teknik tes salah satunya berupa ulangan harian. Permendikbud Nomor 66 Tahun 2013 menyebutkan ulangan harian adalah kegiatan secara berkala oleh guru untuk mengetahui informasi mengenai pencapaian kompetensi peserta didik setelah menuntaskan satu KD atau lebih dalam bentuk ulangan atau penugasan. Perencanaan dan pelaksanaan ulangan harian disesuaikan dengan silabus yang dijabarkan dalam bentuk RPP.

Terdapat dua macam tes (Arikunto, 2012), yaitu objektif dan subjektif. Tes objektif salah satunya berupa tes pilihan ganda, sedangkan tes subjektif berupa esai atau uraian singkat (Ulum, 2017).

Kegiatan evaluasi membutuhkan teknik, prosedur analisis, dan instrumen evaluasi. Dalam kegiatan evaluasi diperlukan butir soal. Soal yang dikembangkan harus baik untuk mampu mengukur pencapaian indikator pencapaian kompetensi dan tujuan pembelajaran. Baik tidaknya butir soal ditentukan melalui analisis butir soal, yaitu proses pengkajian kualitas butir soal untuk mengidentifikasi soal yang baik dan buruk (Sitorus dkk., 2017). Dengan analisis butir soal ini diharapkan dapat dilakukan perbaikan terhadap soal kategori buruk, sehingga dapat meningkatkan kualitas setiap soal. Selain itu, Karim dkk (2021) menambahkan, kegiatan analisis butir soal ini bertujuan untuk mengetahui efektif tidaknya butir soal tersebut untuk mengevaluasi hasil belajar peserta didik. Pengembangan butir soal yang baik dan berkualitas harus melewati proses validasi, reliabilitas, penentuan tingkat kesukaran soal, dan daya pembeda.

Berkaitan dengan pembelajaran, menurut Muhson dkk (2017), soal digunakan untuk mengetahui sejauh mana tingkat penguasaan materi dan membantu guru dalam mengidentifikasi materi mana yang sulit, yang penyusunannya antara lain didasarkan pada ranah kognitif Taksonomi Bloom Revisi. Penentuan tingkatan berpikir peserta didik ini berdasarkan pada aspek pengetahuan dan keterampilan berpikir. Menurut Aziz dkk (2017), penggunaan tingkatan kognitif Taksonomi Bloom dalam penyusunan soal adalah dalam rangka mengidentifikasi kemampuan peserta didik mana yang mampu menguasai setiap tingkatan, mulai dari tingkatan terendah hingga tingkatan tertinggi. Tingkatan kognitif Taksonomi Bloom meliputi mengingat (C1), memahami (C2), mengaplikasikan (C3), menganalisis (C4), mengevaluasi

(C5), dan mencipta (C6) (Anderson and Krathwohl (2001) dalam Effendi (2017)).

Pada dasarnya, Kurikulum 2013 adalah kurikulum berbasis kompetensi, KD sebagai kompetensi minimal yang harus dikuasai oleh peserta didik, di mana seorang guru menjabarkannya ke dalam beberapa indikator pencapaian kompetensi yang dijadikan acuan dalam melakukan kegiatan evaluasi. KD 3.5 yang tercantum dalam Permendikbud Nomor 37 Tahun 2018 (revisi) bagi kelas XI Biologi merupakan salah satu acuan penyusunan indikator aspek pengetahuan yang diperoleh melalui aktivitas “mengingat, memahami, mengaplikasikan, menganalisis, mengevaluasi, dan mencipta” (Permendikbud Nomor 22 Tahun 2016). Oleh karena itu, KD 3.5 menyiratkan tingkat kognitif (pengetahuan) yang harus dicapai peserta didik sesuai dengan tuntutan kurikulum 2013, yaitu memiliki keterampilan berpikir kritis dan berpikir tingkat tinggi minimal menguasai level menganalisis. KD tersebut berbunyi, “Menganalisis hubungan antara struktur jaringan penyusun organ pada sistem gerak dalam kaitannya dengan bioproses dan gangguan fungsi yang dapat terjadi pada sistem gerak.” Hal ini sebagai acuan penyusunan indikator pencapaian kompetensi yang nantinya dijadikan acuan dalam penyusunan butir soal. Menurut Permendikbud RI Nomor 23 Tahun 2016, salah satu pengukuran ketercapaian indikator pencapaian kompetensi dalam aspek pengetahuan dilakukan oleh guru dalam bentuk Ulangan Harian (UH).

Materi Sistem Gerak Manusia yang tertuang dalam KD 3.5 ialah materi biologi yang diajarkan di kelas XI pada semester ganjil, mencakup topik rangka manusia beserta fungsinya, alat gerak aktif dan pasif, persendian dan macamnya, gerak yang dapat dilakukan oleh manusia, serta gangguan dan penyakit yang mungkin terjadi. Menurut keterangan guru SMA Negeri 1 Tarik, butir soal Ulangan Harian materi Sistem Gerak Manusia belum pernah dianalisis sebelumnya. Selain itu, jika dilihat dari aspek materi masih banyak dijumpai butir soal dalam ranah kognitif level mengingat (C1) dan memahami (C2), sehingga belum sesuai dengan tuntutan KD tersebut, yaitu minimal menguasai level menganalisis. Kondisi ini senada dengan pernyataan Amalia dan Widayanti (2012), bahwa berdasarkan hasil penelitiannya, pembuat soal seringkali tidak melakukan analisis butir soal, antara lain karena keterbatasan dan kekuranghandalannya dalam melakukan analisis butir soal. Jika butir - butir soal materi Sistem Gerak Manusia yang disusun guru tidak dianalisis, sementara digunakan berulang kali dalam jangka waktu lama tanpa perbaikan, maka dikhawatirkan butir soal menjadi kurang berarti, lemah, dan tidak bisa digunakan

untuk mengukur ketercapaian indikator pembelajaran berdasarkan tuntutan kurikulum (Suzana, 2017).

Menurut Permendikbud RI Nomor 22 Tahun 2016, soal berfungsi sebagai acuan untuk perbaikan pembelajaran, sehingga butir soal harus baik dan bermutu. Soal dikatakan baik jika dapat mengukur dan menilai kemampuan peserta didik serta dapat memberikan informasi yang sebenarnya, antara lain mengenai peserta didik yang sudah maupun belum menguasai materi yang telah dipelajari (Suzana, 2017).

Ulum (2017) menjelaskan bahwa analisis butir soal ialah kegiatan wajib bagi guru dalam menjamin kualitas butir soal dengan tujuan mengetahui seberapa tingkat ketercapaian indikator pencapaian kompetensi dalam proses pembelajaran. Penelitian terdahulu terkait analisis butir soal dilakukan oleh Effendi (2017) mengenai kesesuaian tingkatan kognitif setiap butir soal dengan Taksonomi Bloom Revisi. Penelitian tersebut bertujuan untuk mengidentifikasi tingkat kognitif setiap butir soal mulai dari tingkatan terendah hingga tertinggi. Dalam hal ini, untuk dapat mencapai tingkatan tertinggi, peserta didik harus mampu menguasai tingkatan yang berada di bawahnya. Penelitian lain oleh Sitorus dkk (2017) bertujuan untuk mendeskripsikan kualitas setiap butir soal ulangan harian biologi kelas X di SMA Negeri 1 Remboken berdasarkan aspek teoritik dan empirik. Analisis aspek teoritik dilakukan berdasarkan pada kaidah dan ketentuan penulisan soal yang baik dan benar. Analisis secara empirik dilakukan melalui perolehan data dari penerapan atau ujicoba butir soal kepada peserta didik, lalu dianalisis berdasarkan validitas, reliabilitas, penentuan tingkat kesukaran soal, daya pembeda, dan fungsi jawaban pengecoh (distraktor) pada pilihan jawaban.

Witarsa dkk (2017) dalam penelitiannya menjelaskan bahwa analisis validitas teoritis mencakup isi (materi yang diujikan), bentuk alat tes, serta penggunaan kata dalam kalimat menggunakan bahasa Indonesia yang baik dan benar sesuai kaidah penulisan butir soal. Suatu tes memiliki validitas teoritis jika sesuai dengan indikator pencapaian kompetensi, tujuan pembelajaran, serta materi pembelajaran yang telah dibelajarkan. Selain itu, Fitriawanati (2017) dalam penelitiannya menambahkan bahwa untuk meningkatkan kualitas setiap butir soal perlu dianalisis secara empiris. Validitas empiris didasarkan pada data empirik yang didapatkan dari hasil ujicoba kepada peserta didik. Dalam penelitian tersebut, validitas empiris ditentukan berdasarkan penentuan tingkat kesukaran soal, daya pembeda, fungsi jawaban pengecoh (distraktor) pada pilihan jawaban, reliabilitas, dan validitas butir soal.

Berdasarkan uraian penjelasan di atas, tujuan penelitian yaitu menganalisis kualitas butir soal Ulangan Harian hasil pengembangan guru pada mata pelajaran biologi materi Sistem Gerak Manusia kelas XI IPA di SMA Negeri 1 Tarik berdasarkan validitas teoritis dan validitas empiris. Analisis validitas teoritis meliputi aspek tingkat kognitif, materi, konstruksi, dan bahasa yang berdasarkan pada kaidah penulisan butir soal yang ditunjang oleh kisi-kisi soal, kurikulum yang digunakan, buku sumber, dan kamus bahasa Indonesia. Analisis validitas teoritis ini dilakukan dengan menggunakan format penelaahan soal, baik untuk butir soal bentuk pilihan ganda maupun uraian. Analisis validitas empiris ditentukan berdasarkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda butir soal, yang mana data diambil dari hasil uji kepada peserta didik.

METODE

Penelitian dilaksanakan mulai bulan Juli 2020, bertempat di SMA Negeri 1 Tarik dan di Universitas Negeri Surabaya. Sasaran penelitian adalah butir soal Ulangan Harian dalam bentuk pilihan ganda dan uraian hasil pengembangan guru biologi materi Sistem Gerak Manusia kelas XI IPA di SMA Negeri 1 Tarik.

Jenis penelitian adalah penelitian deskriptif kualitatif dan kuantitatif. Penelitian diawali dengan pengumpulan data menggunakan metode dokumentasi, berupa butir-butir soal Ulangan Harian hasil pengembangan guru Biologi SMA Negeri 1 Tarik dalam bentuk pilihan ganda dan uraian, beserta kunci jawaban, kisi-kisi soal, dan hasil ujian peserta didik. Butir soal dianalisis untuk mendapatkan validitas teoritis dan validitas empirisnya.

Validitas Teoritis

Validitas teoritis merupakan hasil validasi setiap butir soal oleh dosen ahli berdasarkan aspek tingkat kognitif, materi, konstruksi, dan bahasa yang tercantum di Lembar Validasi Soal yang berpedoman pada kartu telaah butir soal yang ditulis Pusat Studi Pengembangan Sistem Pengujian (Mulyatiningsih, 2011). Tingkat kognitif berkaitan dengan indikator kognitif setiap butir soal berdasarkan Taksonomi Bloom Revisi ranah kognitif. Aspek materi berkaitan dengan substansi materi Sistem Gerak Manusia. Aspek konstruksi terkait kaidah penulisan butir soal yang baik dan benar. Aspek bahasa terkait penggunaan bahasa atau kalimat dalam setiap butir soal yang menggunakan Pedoman Umum Ejaan Bahasa Indonesia (PUEBI). Validasi teoritis butir soal dilakukan dengan cara mencocokkan rumusan soal dengan aspek penilaian yang terdapat pada Lembar Validasi Soal sesuai jenis soal, yaitu pilihan ganda atau uraian. Menurut

Riduwan (2013), penilaian dalam memvalidasi butir soal ditentukan dengan menggunakan skala Guttman (Tabel 1).

Tabel 1. Skala Pengukuran Guttman

Nilai	Keterangan
1	Butir soal sesuai kriteria penilaian
0	Butir soal tidak sesuai kriteria penilaian

Validitas masing – masing butir soal secara teoritis dihitung dengan rumus berikut.

$$\text{Validitas (\%)} = \frac{\sum \text{skor hasil pengumpulan data}}{\sum \text{skor maksimal}} \times 100\%$$

Menurut Riduwan (2013), hasil perhitungan % validitas masing – masing butir soal kemudian diinterpretasikan seperti tercantum pada Tabel 2.

Tabel 2. Interpretasi Validitas Butir Soal

Validitas (%)	Kategori
≥ 70	Valid
< 70	Tidak Valid

Setelah penentuan validitas teoritis, langkah selanjutnya yaitu identifikasi butir soal berdasarkan validitas empiris.

Validitas Empiris

Validitas empiris diperoleh dari hasil ujicoba setiap butir soal kepada peserta didik.

1. Validitas Butir Soal

Validitas butir soal merupakan angka hasil perhitungan indeks korelasi biserial antara skor setiap butir soal dengan skor total. Menurut Arikunto (1986) dalam Amelia (2016), validitas butir soal dihitung menggunakan rumus sebagai berikut.

$$r_{pbi} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

$$p = \frac{\text{banyaknya peserta didik menjawab benar}}{\text{jumlah seluruh peserta didik}}$$

$$q = 1 - p$$

Keterangan :

- r_{pbi} = koefisien korelasi biserial
- M_p = rata – rata skor dari subjek yang menjawab benar bagi item yang dicari validitasnya
- M_t = rata – rata skor total
- S_t = standar deviasi skor total

- p = banyaknya peserta didik menjawab benar
- q = banyaknya peserta didik menjawab salah

Hasil perhitungan kemudian dibandingkan dengan batas signifikansi koefisien korelasi (Tabel 3) sebagai berikut.

Tabel 3. Batas Signifikansi Koefisien Korelasi

Tabel r untuk df = 1 – 50					
Tingkat signifikansi untuk uji dua arah					
df (N-2)	p=0,05	p=0,01	df (N-2)	p=0,05	p=0,01
10	0,576	0,708	60	0,250	0,325
15	0,482	0,606	70	0,233	0,302
20	0,423	0,549	80	0,217	0,283
25	0,381	0,496	90	0,205	0,267
30	0,349	0,449	100	0,195	0,254
40	0,304	0,393	125	0,174	0,228
50	0,273	0,354	> 150	0,159	0,208

Keterangan :

- r = koefisien korelasi
- df = derajat kebebasan
- N = jumlah peserta didik (responden)
- p = probabilitas
- Bila koefisien = 0,000 berarti tidak dapat dihitung

Tingkat signifikansi untuk uji dua arah dengan $p = 0,05$ artinya tingkat kesalahan dalam penelitian sebesar 5% dan tingkat kepercayaan penelitian sebesar 95%. Tingkat signifikansi untuk uji dua arah dengan $p = 0,01$ artinya tingkat kesalahan dalam penelitian sebesar 1% dan tingkat kepercayaan penelitian sebesar 99%. Butir soal dikatakan valid jika memenuhi ketentuan $r_{hitung} > r_{tabel}$. Pada penelitian ini, ujicoba dilakukan terhadap 96 responden (N), sehingga $df (N - 2) = 96 - 2 = 94$. Tingkat signifikansi untuk uji dua arah menggunakan $p = 0,05$ dengan responden 96, maka $df (N - 2)$ sebesar 0,2006. Jadi, butir soal dinyatakan valid jika koefisien korelasi/rhitung (Tabel 12) $>$ r_{tabel} , yaitu sebesar 0,2006.

2. Reliabilitas Butir Soal

Reliabilitas butir soal merupakan tetap tidaknya atau keajegan hasil tes yang didapat dari ujicoba tes yang disusun oleh guru sebanyak satu kali (Sugiyono, 2013). Koefisien reliabilitas butir soal suatu tes dihitung dengan rumus *Cronbach Alpha*. Jika satu butir soal dijawab benar memperoleh skor 1 dan jika satu butir soal dijawab salah memperoleh skor 0. Menurut Arikunto (2012) rumus menghitung koefisien reliabilitas butir soal adalah sebagai berikut.

$$r_{11} = \left(\frac{n}{n-1}\right) \left(1 - \frac{\epsilon\sigma_i^2}{\sigma_t^2}\right)$$

Keterangan :

r_{11} = koefisien reliabilitas butir soal

n = banyaknya butir soal

$\epsilon\sigma_i^2$ = jumlah variansi butir soal

σ_t^2 = variansi total

Hasil perhitungan koefisien reliabilitas butir soal kemudian diinterpretasikan seperti tercantum pada Tabel 4 untuk menentukan reliabel tidaknya butir soal.

Tabel 4. Interpretasi Reliabilitas Butir Soal

Reliabilitas	Kategori
$\geq 0,70$	Tinggi
$< 0,70$	Rendah

(Sudijono, 2009 dalam Ulum, 2017)

3. Tingkat Kesukaran Butir Soal

Tingkat kesukaran butir soal merupakan tingkat sukar mudahnya butir soal yang disusun oleh guru dinyatakan dengan indeks kesukaran (Mehta and Mokhasi, 2014). Menurut Arikunto (2012) rumus menghitung indeks kesukaran butir soal adalah sebagai berikut.

$$P = \frac{B}{JS} \times 100\%$$

Keterangan :

P = indeks kesukaran

B = banyaknya peserta didik menjawab benar

JS = jumlah peserta didik

Hasil perhitungan indeks kesukaran butir soal kemudian diinterpretasikan seperti tercantum pada Tabel 5 sebagai berikut.

Tabel 5. Interpretasi Indeks Kesukaran Butir Soal

Indeks Kesukaran (%)	Kategori
$0 \leq IK \leq 30$	Sukar
$30 < IK \leq 70$	Sedang
$70 < IK \leq 100$	Mudah

(Arikunto, 2012)

Keterangan :

IK = Indeks Kesukaran

4. Daya Pembeda Butir Soal

Daya pembeda butir soal merupakan kemampuan suatu butir soal dalam membedakan antara peserta didik

yang sudah maupun belum menguasai kompetensi atau materi yang telah diajarkan berdasarkan kriteria tertentu (Ulum, 2017). Menurut Arikunto (2012) rumus perhitungan daya pembeda adalah sebagai berikut.

$$DP = \frac{BA}{JA} - \frac{BB}{JB}$$

Keterangan :

DP = Daya Pembeda

JA = jumlah peserta didik kelompok atas

JB = jumlah peserta didik kelompok bawah

BA = jumlah peserta didik kelompok atas benar

BB = jumlah peserta didik kelompok bawah benar

Hasil perhitungan daya pembeda butir soal kemudian diinterpretasikan seperti tercantum pada Tabel 6 untuk menentukan kategori daya pembeda setiap butir soal.

Tabel 6. Interpretasi Daya Pembeda Butir Soal

Indeks Daya Pembeda (%)	Kategori
$DP \leq 0$	Sangat Buruk
$0 < DP \leq 20$	Buruk
$20 < DP \leq 40$	Cukup
$40 < DP \leq 70$	Baik
$70 < DP \leq 100$	Sangat Baik

(Haryanto, 2020)

Keterangan :

DP = Daya Pembeda

Validitas empiris butir soal berdasarkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda dihitung menggunakan aplikasi ANATES V4.

HASIL DAN PEMBAHASAN

Analisis butir soal ialah proses pengkajian kualitas butir soal untuk mengidentifikasi soal yang baik dan buruk (Sitorus dkk., 2017). Hasil analisis ini bisa dijadikan acuan untuk memperbaiki soal yang buruk serta meningkatkan kualitas setiap soal, sehingga tujuan penilaian dapat tercapai serta dapat berfungsi sesuai peruntukannya (Hartati and Yogi, 2019)

Validitas Teoritis

Analisis dilakukan terhadap butir soal Ulangan Harian materi Sistem Gerak Manusia dalam bentuk pilihan ganda dan uraian beserta kunci jawaban dan kisi-kisi soal yang dikembangkan oleh guru SMA pada mata pelajaran biologi kelas XI IPA di SMA Negeri 1 Tarik. Selanjutnya, butir soal divalidasi oleh dosen ahli dengan menggunakan

skala Guttman (Tabel 1). Hasil validasi butir soal tipe pilihan ganda berdasarkan aspek tingkat kognitif Taksonomi Bloom Revisi disajikan dalam Tabel 7.

Tabel 7. Hasil Validasi Butir Soal Tipe Pilihan Ganda Berdasarkan Aspek Tingkat Kognitif Taksonomi Bloom Revisi

Tingkat Kognitif Taksonomi Bloom Revisi			
Nomor Soal	Tingkat Kognitif	Nomor Soal	Tingkat Kognitif
1	C1	11	C2
2	C1	12	C1
3	C1	13	C2
4	C1	14	C2
5	C2	15	C3
6	C3	16	C2
7	C1	17	C2
8	C1	18	C2
9	C2	19	C1
10	C3	20	C1

Pada butir soal pilihan ganda, tingkat kognitif berdasarkan Taksonomi Bloom Revisi level mengingat (C1) dan memahami (C2) mendominasi lebih dari 50%. Tingkat kognitif level C1 ditemukan pada soal nomor 1, 2, 3, 4, 7, 8, 12, 19, dan 20. Tingkat kognitif level C2 ditemukan pada soal nomor 5, 9, 11, 13, 14, 16, 17, dan 18. Tingkat kognitif level C3 ditemukan pada soal nomor 6, 10, dan 15. Hasil validasi butir soal tipe uraian berdasarkan aspek tingkat kognitif Taksonomi Bloom Revisi disajikan dalam Tabel 8.

Tabel 8. Hasil Validasi Butir Soal Tipe Uraian Berdasarkan Aspek Tingkat Kognitif Taksonomi Bloom Revisi

Tingkat Kognitif Taksonomi Bloom Revisi	
Nomor Soal	Tingkat Kognitif
1	C2
2	C3
3	C2
4	C2
5	C2

Pada butir soal uraian, tingkat kognitif Taksonomi Bloom Revisi lebih didominasi ranah kognitif level memahami (C2), salah satunya pada butir soal nomor 3. Menurut dosen ahli (Tabel 11), penentuan butir soal nomor 3 termasuk ranah kognitif level menganalisis (C4) tidak sesuai, lebih tepat jika termasuk menjelaskan atau ranah kognitif level memahami (C2). Penentuan ini karena peserta didik diminta untuk menjelaskan tentang keuntungan tulang paha berbentuk tulang pipa, sehingga tidak sesuai dengan ranah kognitif level menganalisis (C4). Tingkat kognitif Taksonomi Bloom Revisi butir

soal uraian level C3 hanya ditemukan pada soal nomor 2 saja.

Butir soal pilihan ganda dan uraian berdasarkan Taksonomi Bloom Revisi masih didominasi tingkat kognitif level C1, C2, dan C3. Hasil ini diperkuat oleh pendapat dan saran serta komentar dosen ahli (Tabel 11) yang mengatakan bahwa butir soal didominasi oleh ranah kognitif level C1 sebanyak lebih dari 50%. Kondisi ini bisa disebabkan karena guru belum terbiasa dalam menyusun butir soal berdasarkan tingkat kognitif Taksonomi Bloom Revisi level C1 sampai C6, khususnya pada level C4, C5, dan C6. Guru dalam menetapkan ranah kognitif tidak disesuaikan dengan ranah kognitif Taksonomi Bloom Revisi (Supandi dan Farikhah, 2016). Penelitian serupa oleh Saraswati dan Agustika (2020) juga mengidentifikasi bahwa pada butir soal Ulangan Harian level kognitif masih pada level C1 sampai C3 saja. Dalam penelitian tersebut dijelaskan bahwa guru – guru telah mengikuti acara workshop ataupun seminar tentang penyusunan butir soal, akan tetapi dalam pelaksanaannya belum sepenuhnya berjalan maksimal. Pernyataan tersebut diperkuat oleh Prasetya (2012) bahwa guru seringkali mengalami kesulitan dalam menyusun instrumen hasil belajar peserta didik. Kondisi ini dimungkinkan tidak tersedianya buku panduan untuk menyusun instrumen hasil belajar yang baik dan benar sesuai ketentuan.

Menurut Aziz dkk (2017), penggunaan tingkatan kognitif Taksonomi Bloom dalam penyusunan soal itu penting dengan tujuan untuk mengidentifikasi kemampuan peserta didik, yang mana harus mampu menguasai setiap tingkatan, mulai dari tingkatan terendah hingga tertinggi. Taksonomi Bloom Revisi ranah kognitif adalah tingkatan berpikir berdasarkan aspek pengetahuan dan keterampilan berpikir, meliputi mengingat (C1), memahami (C2), mengaplikasikan (C3), menganalisis (C4), mengevaluasi (C5), dan mencipta (C6) (Anderson and Krathwohl (2001) dalam Effendi (2017)).

Selain tidak sesuai Taksonomi Bloom Revisi ranah kognitif, butir soal juga tidak sesuai dengan permintaan KD 3.5 Permendikbud Nomor 37 Tahun 2018 (revisi) bagi kelas XI Biologi, yang berbunyi, “Menganalisis hubungan antara struktur jaringan penyusun organ pada sistem gerak dalam kaitannya dengan bioproses dan gangguan fungsi yang dapat terjadi pada sistem gerak.” KD tersebut menyiratkan tingkat kognitif (pengetahuan) yang harus dicapai peserta didik sesuai dengan tuntutan kurikulum 2013, yaitu memiliki keterampilan berpikir kritis dan tingkat tinggi minimal menguasai level menganalisis.

Berdasarkan analisis hasil validasi diketahui butir soal tidak sesuai dengan permintaan KD 3.5 Permendikbud

Nomor 37 Tahun 2018 (revisi) kelas XI Biologi yang mengacu pada Taksonomi Bloom Revisi, sehingga perlu diperbaiki. Perbaikan terkait kesesuaian dengan KD yang digunakan, serta penentuan dan persebaran tingkatan kognitif setiap butir soal.

Butir soal Ulangan Harian materi Sistem Gerak Manusia juga divalidasi berdasarkan aspek materi, konstruksi, dan bahasa dengan cara mencocokkan rumusan soal dengan kriteria yang terdapat pada Lembar Validasi Soal sesuai jenis soal, yaitu pilihan ganda atau uraian. Hasil validasi secara teoritis butir soal tipe pilihan ganda disajikan dalam Tabel 9.

Tabel 9. Hasil Validasi secara Teoritis Butir Soal Tipe Pilihan Ganda

No	Aspek Penilaian	Rata – Rata (%)		Validitas (%)	Kategori
		V1	V2		
A. Aspek Materi					
1.	Butir soal telah sesuai dengan indikator pencapaian kompetensi	90	100	95	Valid
2.	Butir soal telah sesuai dengan tujuan pembelajaran	85	100	92,5	Valid
3.	Pengecoh pada pilihan jawaban soal berfungsi	85	100	92,5	Valid
4.	Pilihan jawaban yang tersedia logis dan homogen	80	100	90	Valid
5.	Mempunyai satu jawaban yang benar (tidak mengandung unsur ganda)	80	100	90	Valid
Rata – rata validitas (%) aspek materi				92	Valid
B. Aspek Konstruksi					
6.	Butir soal mengandung pokok soal yang dirumuskan secara singkat, padat, jelas, dan tegas	95	100	97,5	Valid
7.	Rumusan soal yang disusun merupakan pernyataan atau kalimat yang diperlukan saja dengan maksud yang jelas	95	100	97,5	Valid
8.	Rumusan jawaban yang disusun merupakan pernyataan atau kalimat yang diperlukan saja	95	100	97,5	Valid

No	Aspek Penilaian	Rata – Rata (%)		Validitas (%)	Kategori
		V1	V2		
	dengan maksud yang jelas				
9.	Butir soal mengandung pokok soal yang tidak langsung merujuk pada jawaban yang benar	95	100	97,5	Valid
10.	Butir soal tidak mengandung pokok soal negatif ganda	100	100	100	Valid
11.	Panjang kalimat yang digunakan dalam merumuskan jawaban relatif sama	95	100	97,5	Valid
12.	Pilihan jawaban tidak menggunakan pernyataan “semua jawaban benar atau semua jawaban salah”	95	100	97,5	Valid
13.	Penyajian tabel, grafik, gambar, maupun diagram jelas dan dapat berfungsi sesuai dengan butir soal yang disusun	45	20	32,5	Tidak Valid
14.	Pilihan jawaban apabila dalam bentuk angka disusun berdasarkan urutan	0	100	50	Tidak Valid
15.	Pilihan jawaban apabila dalam bentuk kejadian disusun berdasarkan kronologis	10	40	25	Tidak Valid
16.	Butir soal yang disusun tidak tergantung kepada jawaban sebelumnya	100	100	100	Valid
17.	Terdapat pedoman penyekoran dalam mengerjakan butir soal	100	0	50	Tidak Valid
Rata – rata validitas (%) aspek konstruksi				78,5	Valid
C. Aspek Bahasa					
18.	Butir soal menggunakan kalimat yang sesuai dengan penulisan kaidah bahasa Indonesia yang baku	80	100	90	Valid
19.	Butir soal	95	100	97,5	Valid

No	Aspek Penilaian	Rata – Rata (%)		Validitas (%)	Kategori
		V1	V2		
	menggunakan bahasa yang komunikatif				
20.	Butir soal tidak menggunakan bahasa daerah setempat	100	100	100	Valid
21.	Pilihan bahasa yang digunakan dalam pilihan jawaban tidak mengulang kata atau frasa yang bukan merupakan satu kesatuan pengertian	100	100	100	Valid
22.	Butir soal dan pilihan jawaban menggunakan kalimat dan bahasa yang mudah dipahami	95	100	97,5	Valid
Rata – rata validitas (%) aspek bahasa				97	Valid
Rata – rata validitas teoritis (%)				89,17	Valid

Ditinjau dari aspek materi, secara keseluruhan, butir soal pilihan ganda tergolong dalam kategori valid. Hal ini menandakan butir soal telah sesuai dengan indikator pencapaian kompetensi, tujuan pembelajaran, dan materi yang telah diajarkan (Witarsa dkk., 2017). Suzana (2017) menambahkan bahwa butir soal tersebut dapat digunakan untuk mengukur dan menilai kemampuan peserta didik serta dapat memberikan informasi yang sebenarnya, antara lain mengenai peserta didik mana yang sudah maupun belum menguasai materi yang telah dipelajari. Hal ini dapat dijadikan evaluasi bagi guru untuk mengetahui seberapa tingkat ketercapaian indikator pencapaian kompetensi dalam proses pembelajaran (Ulum, 2017). Akan tetapi, berdasarkan pendapat dan saran serta komentar dosen ahli (Tabel 11) terdapat beberapa butir soal yang kurang sesuai dengan indikator dan tujuan pembelajaran, seperti butir soal nomor 11 dan 13, sehingga validitas belum mencapai 100% walaupun termasuk kategori valid. Kedua butir soal tersebut perlu dilakukan perbaikan agar validitas dapat mencapai 100%. Selain itu, butir soal nomor 3, 5, 7, 12 dan 13 tidak sesuai dengan aspek penilaian, sehingga validitas butir soal belum mencapai 100% walaupun termasuk kategori valid.

Soal nomor 3, 7, dan 13 pengecoh pada pilihan jawaban tidak berfungsi, serta nomor 3, 5, 7, dan 13 pilihan jawaban tidak logis dan homogen. Menurut Fitriawanati (2017), pilihan jawaban pengecoh (distraktor) harus sesuai dengan materi yang diujikan, dalam hal ini materi Sistem Gerak Manusia. Pembuatan pilihan

jawaban pengecoh panjang pendeknya harus relatif sama dengan kunci jawaban, sehingga logis dan homogen, serta berfungsi dengan baik. Pengecoh dinyatakan berfungsi jika dipilih oleh kurang lebih 5% total peserta didik, serta lebih banyak dipilih oleh peserta didik kelompok bawah (Fitriawanati, 2017).

Butir soal nomor 3, 5, 7, dan 12 pilihan jawabannya mengandung makna ganda. Baik butir soal maupun pilihan jawaban sebaiknya menghindari kata atau kalimat yang mengandung arti lebih dari satu. Hal ini dimaksudkan mempermudah memahami maksud dari materi yang ditanyakan serta maksud dari setiap pilihan jawaban yang tersaji (Fitriawanati, 2017).

Ditinjau dari aspek konstruksi, butir soal tidak mengandung pokok soal negatif ganda dan tidak tergantung kepada jawaban sebelumnya. Hal ini dibuktikan dengan nilai validitas mencapai 100%. Artinya, penggunaan kata dalam kalimat soal tidak memiliki arti negatif, serta butir soal tidak tergantung kepada jawaban atau soal sebelumnya (berdiri sendiri). Jika butir soal bergantung kepada jawaban sebelumnya, dikhawatirkan apabila menjawab salah butir soal pertama, maka tidak bisa menjawab dengan benar butir soal selanjutnya (Depdiknas, 2008).

Ditinjau dari aspek konstruksi, terdapat butir soal tidak sesuai aspek penilaian, seperti nomor 17, pokok soal tidak dirumuskan secara singkat, padat, jelas, dan tegas. Butir soal harus mengandung satu pokok persoalan atau gagasan yang jelas, sehingga ketika membaca dan memahami butir soal, peserta didik memiliki penafsiran yang sama dengan guru sebagai pembuat soal. Oleh karena itu, materi Sistem Gerak Manusia yang akan diukur/ditanyakan harus singkat, padat, jelas, dan tegas (Depdiknas, 2008). Rumusan soal dan jawaban yang disusun menggunakan kalimat dengan maksud yang tidak jelas. Pernyataan atau kalimat yang tidak jelas atau tidak diperlukan sebaiknya dihilangkan saja. Selain itu, pokok soal merujuk jawaban yang benar. Dalam hal ini, butir soal hendaknya menghindari kalimat yang dapat menjadi petunjuk ke arah jawaban yang benar (Depdiknas, 2008). Panjang kalimat jawaban relatif tidak sama, seperti nomor 17. Hal ini akan membuat peserta didik memiliki kecenderungan memilih jawaban paling panjang, karena peserta didik memiliki pemikiran bahwa jawaban panjang merupakan jawaban yang paling lengkap serta merupakan kunci jawabannya (Depdiknas, 2008). Selanjutnya, pada butir soal nomor 17 juga, terdapat pilihan jawaban merujuk pada “Semua jawaban benar atau semua jawaban salah”. Jika menggunakan salah satu pilihan jawaban dengan pernyataan tersebut, selain rumusan jawaban tidak homogen, pilihan jawaban (pengecoh) tidak sesuai

dengan materi Sistem Gerak Manusia yang hendak diukur (Depdiknas, 2008).

Tabel, grafik, gambar, maupun diagram berfungsi sebagai rangkuman informasi berisi data yang berkaitan dengan butir soal agar lebih mudah dipahami oleh peserta didik, sehingga penyajiannya harus jelas dan dapat berfungsi. Apabila peserta didik dapat menjawab butir soal tanpa melihat tabel, grafik, gambar, maupun diagram, maka dapat dikatakan tabel, grafik, gambar, maupun diagram tersebut tidak berfungsi (Depdiknas, 2008).

Butir soal dengan pilihan jawaban dalam bentuk angka tidak disusun berdasarkan urutan dan dalam bentuk kejadian tidak disusun berdasarkan kronologis. Menurut Depdiknas (2008), pilihan jawaban dalam bentuk angka sebaiknya disusun berurutan mulai dari terkecil hingga terbesar atau sebaliknya, serta jika dalam bentuk kejadian disusun berdasarkan kronologis. Hal ini bermaksud memudahkan peserta didik dalam menjawab butir soal.

Butir soal tidak dilengkapi pedoman penyekoran. Sesuai dengan pendapat dan saran serta komentar dosen ahli (Tabel 11), butir soal sebaiknya dilengkapi dengan pedoman penyekoran dalam mengerjakan butir soal. Manfaatnya untuk mengetahui deskripsi tentang kemampuan setiap peserta didik dalam menjawab soal dan dapat dilakukan analisis kuantitatif terkait kualitas butir soal, sehingga dapat dijadikan acuan perbaikan untuk kegiatan evaluasi selanjutnya (Khaerudin, 2016).

Ditinjau dari aspek bahasa, secara keseluruhan, butir soal pilihan ganda juga tergolong dalam kategori valid, karena telah sesuai dengan aspek penilaian yang terdapat pada Lembar Validasi Soal. Akan tetapi, berdasarkan pendapat dan saran serta komentar dosen ahli (Tabel 11) terdapat beberapa butir soal yang belum sesuai dengan Pedoman Umum Ejaan Bahasa Indonesia (PUEBI), seperti nomor 2, 3, 17, dan 18, sehingga nilai validitas belum mencapai 100%. Hal ini karena butir – butir soal tersebut tidak menggunakan kalimat yang sesuai dengan penulisan kaidah bahasa Indonesia yang baku. Selain itu, nomor 17 dan 18 juga, tidak menggunakan bahasa yang komunikatif dan mudah dipahami. Penyusunan soal harus memperhatikan pemakaian kata, kalimat, dan ejaan. Kalimat yang baik adalah kalimat yang bersusun Subyek – Predikat – Obyek – Keterangan (Pusat Pembinaan dan Pemasyarakatan, 2015). Jika butir soal menggunakan bahasa yang komunikatif, maka butir soal yang ditanyakan tersebut mudah dipahami dan dicerna oleh peserta didik (Depdiknas, 2008).

Analisis hasil validasi menunjukkan soal pilihan ganda ditinjau dari aspek materi, beberapa butir soal perlu dilakukan perbaikan agar validitas dapat mencapai 100%. Perbaikan dilakukan dengan menyesuaikan butir soal dengan indikator pencapaian kompetensi, tujuan

pembelajaran, serta materi yang telah dibelajarkan (Witarsa dkk., 2017). KD 3.5 Permendikbud Nomor 37 Tahun 2018 (revisi) kelas XI Biologi materi Sistem Gerak Manusia dijabarkan dalam bentuk indikator pencapaian kompetensi dan tujuan pembelajaran yang dijadikan acuan dalam menyusun butir soal ulangan harian (Permendikbud Nomor 66 Tahun 2013). Selain itu, penyusunan pilihan jawaban disesuaikan dengan materi Sistem Gerak Manusia, sehingga logis dan homogen, serta pengecoh pada pilihan jawaban soal harus berfungsi dengan baik. Penyusunan butir soal hendaknya menghindari kata atau kalimat yang mengandung arti lebih dari satu (Fitrianawati, 2017).

Ditinjau dari aspek konstruksi menunjukkan banyak terdapat butir soal yang tidak sesuai dengan aspek penilaian, sehingga perlu dilakukan perbaikan secara signifikan. Aspek ini terkait kaidah penulisan butir soal yang baik dan benar (Mulyatiningsih, 2011). Perbaikan dilakukan khususnya pada soal nomor 17 hingga dinyatakan valid ataupun soal tersebut sebaiknya dibuang atau diganti. Selain itu, sesuai dengan pendapat dan saran serta komentar dosen ahli terhadap butir soal tipe pilihan ganda (Tabel 11), butir soal harus dilengkapi pedoman penyekoran.

Ditinjau dari aspek bahasa, butir soal hendaknya disesuaikan dengan PUEBI. Penyusunan butir soal harus memperhatikan pemakaian kata, kalimat, dan ejaan, serta menggunakan bahasa yang komunikatif agar mudah dicerna oleh peserta didik (Depdiknas, 2008). Hasil validasi secara teoritis butir soal tipe uraian disajikan dalam Tabel 10.

Tabel 10. Hasil Validasi secara Teoritis Butir Soal Tipe Uraian

No	Aspek Penilaian	Rata – Rata (%)		Validitas (%)	Kategori
		V1	V2		
A. Aspek Materi					
1.	Butir soal telah sesuai dengan indikator pencapaian kompetensi	80	100	90	Valid
2.	Butir soal telah sesuai dengan tujuan pembelajaran yang hendak diukur	80	100	90	Valid
3.	Pertanyaan dan jawaban yang diharapkan memiliki batasan yang jelas	60	100	80	Valid
4.	Isi materi yang digunakan untuk menyusun	100	100	100	Valid

No	Aspek Penilaian	Rata – Rata (%)		Validitas (%)	Kategori
		V1	V2		
	pertanyaan sesuai dengan jenjang sekolah atau tingkatan kelas				
Rata – rata validitas (%) aspek materi				90	Valid
B. Aspek Konstruksi					
5.	Butir soal mengandung pokok soal yang dirumuskan secara singkat, padat, jelas, dan tegas	100	100	100	Valid
6.	Rumusan soal yang disusun merupakan pernyataan atau kalimat yang diperlukan saja dengan maksud yang jelas	100	100	100	Valid
7.	Butir soal mengandung rumusan kalimat tanya atau perintah yang menuntut jawaban uraian	100	100	100	Valid
8.	Butir soal tidak mengandung pokok soal negatif ganda	80	100	90	Valid
9.	Terdapat petunjuk dalam mengerjakan butir soal	100	0	50	Tidak Valid
10.	Terdapat pedoman penyekoran dalam mengerjakan butir soal	100	0	50	Tidak Valid
11.	Terdapat durasi lama waktu dalam mengerjakan butir soal	0	0	0	Tidak Valid
Rata – rata validitas (%) aspek konstruksi				70	Valid
C. Aspek Bahasa					
12.	Butir soal menggunakan kalimat yang sesuai dengan penulisan kaidah bahasa Indonesia yang baku, baik, dan benar	100	100	100	Valid
13.	Butir soal menggunakan bahasa yang komunikatif	100	100	100	Valid
14.	Butir soal tidak menggunakan bahasa daerah setempat	100	100	100	Valid
15.	Butir soal menggunakan kalimat dan bahasa yang mudah	100	100	100	Valid

No	Aspek Penilaian	Rata – Rata (%)		Validitas (%)	Kategori
		V1	V2		
	dipahami				
16.	Butir soal tidak menggunakan kalimat yang mengandung unsur ganda atau salah arti	100	100	100	Valid
17.	Butir soal tidak menggunakan kalimat yang menyinggung perasaan peserta didik	100	100	100	Valid
Rata – rata validitas (%) aspek bahasa				100	Valid
Rata – rata validitas teoritis (%)				86,67	Valid

Ditinjau dari aspek materi, secara keseluruhan, butir soal uraian tergolong dalam kategori valid. Hal ini menandakan bahwa butir soal telah sesuai dengan indikator pencapaian kompetensi, tujuan pembelajaran, dan materi yang telah dibelajarkan (Witarsa dkk., (2017). Selain itu, isi materi yang digunakan untuk menyusun soal sesuai dengan jenjang sekolah atau tingkatan kelas. Akan tetapi, berdasarkan pendapat dan saran serta komentar dosen ahli (Tabel 11), terdapat butir soal tidak sesuai dengan indikator dan tujuan pembelajaran, seperti soal nomor 3 dan 4, sehingga validitas belum mencapai 100% walaupun termasuk kategori valid. Kedua soal tersebut perlu diperbaiki agar sesuai dengan indikator pencapaian kompetensi, tujuan pembelajaran, dan materi yang telah diajarkan, sehingga validitas dapat mencapai 100%. Selain itu, jawaban yang diharapkan dari menjawab soal tidak memiliki batasan yang jelas. Butir soal uraian hendaknya menuntut jawaban uraian dengan batasan yang jelas. Batasan – batasan tersebut mengandung kata kunci, konsep, ataupun kriteria tertentu yang didasarkan pada materi Sistem Gerak Manusia sebagai acuan peserta didik dalam menjawab butir soal (Depdiknas, 2008).

Selain adanya ketidaksesuaian dengan aspek penilaian pada Lembar Validasi Soal, menurut dosen ahli (Tabel 11), kunci jawaban pada butir soal nomor 4 kurang tepat. Pertanyaan soal nomor 4 tentang mekanisme terjadinya kontraksi dan relaksasi otot.

Ditinjau dari aspek konstruksi, butir soal mengandung satu pokok persoalan atau gagasan yang jelas, sehingga ketika membaca dan memahami butir soal, peserta didik memiliki penafsiran yang sama dengan guru sebagai pembuat soal, serta rumusan soal yang disusun merupakan pernyataan atau kalimat yang diperlukan saja dengan maksud yang jelas (Depdiknas, 2008). Selain itu,

butir soal mengandung rumusan kalimat tanya atau perintah yang menuntut jawaban uraian.

Butir soal tidak mengandung pokok soal negatif ganda, kecuali nomor 5, sehingga validitas belum mencapai 100%. Kalimat soal yang mengandung arti negatif dapat mempersulit peserta didik dalam memahami maksud dari soal yang ditanyakan, sehingga sebaiknya dihindari (Depdiknas, 2008).

Butir soal uraian tidak ada petunjuk pengerjaan, pedoman penyekoran, dan durasi lama waktu pengerjaan butir soal. Sesuai pendapat dan saran serta komentar dosen ahli (Tabel 11), butir soal sebaiknya dilengkapi dengan petunjuk pengerjaan, pedoman penyekoran, dan durasi lama waktu pengerjaan. Manfaat adanya pedoman penyekoran adalah untuk mengetahui deskripsi tentang kemampuan setiap peserta didik dalam menjawab soal dan dapat dilakukan analisis kuantitatif terkait kualitas butir soal, sehingga dapat dijadikan acuan perbaikan untuk kegiatan evaluasi selanjutnya (Khaerudin, 2016). Selain itu, adanya pedoman pengerjaan soal dan durasi lama waktu pengerjaan dapat membantu peserta didik disiplin, mandiri, dan percaya diri dalam mengerjakan butir soal serta dapat menyelesaikan butir soal dengan baik sesuai waktu yang diberikan (Dafitri, 2017).

Ditinjau dari aspek bahasa, secara keseluruhan, butir soal uraian berkategori valid. Hal ini menandakan bahwa telah sesuai dengan aspek penilaian yang terdapat pada Lembar Validasi Soal, yaitu kalimat soal sesuai penulisan kaidah bahasa Indonesia yang baku, tidak mengandung makna ganda atau salah arti, bahasanya komunikatif dan mudah dipahami, serta tidak menggunakan bahasa daerah setempat dan tidak menyinggung perasaan peserta didik.

Ditinjau dari aspek materi, hasil validasi menunjukkan beberapa butir soal perlu diperbaiki agar validitas dapat mencapai 100%. Perbaikan dilakukan dengan menyesuaikan butir soal dengan indikator pencapaian kompetensi, tujuan pembelajaran, serta materi pembelajaran yang telah dibelajarkan (Witarsa dkk., 2017). Dalam hal ini disesuaikan dengan KD 3.5 Permendikbud Nomor 37 Tahun 2018 (revisi) kelas XI Biologi materi Sistem Gerak Manusia yang dijabarkan dalam bentuk indikator pencapaian kompetensi dan tujuan pembelajaran (Permendikbud Nomor 66 Tahun 2013). Selain itu, penyusunan butir soal uraian hendaknya menuntut jawaban uraian dengan batasan yang jelas, serta perbaikan juga dilakukan terhadap kunci jawaban yang salah.

Ditinjau dari aspek konstruksi perlu dilakukan perbaikan agar validitas dapat mencapai 100%. Perbaikan dilakukan dengan menghindari kata atau kalimat yang memiliki arti negatif agar memudahkan peserta didik dalam memahami maksud dari soal yang ditanyakan

(Depdiknas, 2008). Sesuai dengan pendapat dan saran serta komentar dosen ahli (Tabel 11), butir soal sebaiknya dilengkapi dengan petunjuk pengerjaan, pedoman penyekoran, dan durasi lama waktu pengerjaan. Ditinjau dari aspek bahasa, soal – soal uraian yang disusun telah sesuai dengan aspek penilaian pada Lembar Validasi Soal.

Pendapat dan saran atau komentar terhadap butir soal Ulangan Harian yang dikembangkan oleh guru SMA pada mata pelajaran biologi materi Sistem Gerak Manusia kelas XI IPA di SMA Negeri 1 Tarik disajikan pada Tabel 11.

Tabel 11. Pendapat dan Saran serta Komentar Dosen Ahli terhadap Butir Soal Tipe Pilihan Ganda dan Uraian

No.	Pendapat dan Saran serta Komentar Dosen Ahli	
	Butir Soal Tipe Pilihan Ganda	Butir Soal Tipe Uraian
1.	Dalam menyusun kalimat soal sebaiknya berpedoman pada penggunaan Bahasa Indonesia yang baik dan benar, yaitu bersusun SPOK.	Soal nomor 3 tidak sesuai kalau menganalisis, lebih tepat menjelaskan, untuk indikator dan tujuan pembelajaran.
2.	Terdapat beberapa soal kurang sesuai dengan indikator dan tujuan pembelajaran, seperti soal nomor 11 dan 13.	Kunci jawaban soal nomor 4 kurang tepat.
3.	Soal C1 masih mendominasi, lebih dari 50%.	Tidak dilengkapi dengan pedoman skor, pedoman mengerjakan dan waktu pengerjaan.
4.	Tidak dilengkapi dengan pedoman pengerjaan.	

Setelah dilakukan analisis validitas teoritis, butir soal dianalisis validitas empirisnya. Validitas empiris ditentukan berdasarkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal.

Validitas Empiris

Validitas empiris merupakan angka hasil ujicoba setiap butir soal kepada peserta didik. Hasil ujicoba tersebut dianalisis untuk mendapatkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda butir soal.

1. Validitas Butir Soal

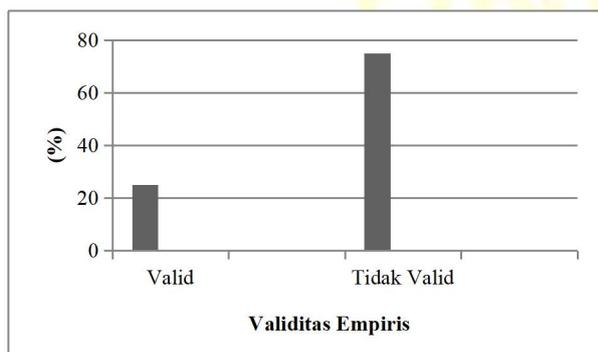
Validitas butir soal merupakan angka hasil perhitungan indeks korelasi biserial antara skor setiap butir soal dengan skor total dengan tujuan untuk mengetahui valid tidaknya suatu item soal (Amalia dan Widayanti, 2012). Pada taraf signifikansi 5% pada rtabel diketahui bahwa batas signifikansi koefisien korelasi butir soal yang diujicoba kepada 96 peserta didik yaitu 0,2006.

Hasil uji validasi dan nilai reliabilitas butir soal pilihan ganda disajikan pada Tabel 12.

Tabel 12. Validitas dan Reliabilitas Butir Soal Pilihan Ganda

No	Korelasi	Signifikansi	Keterangan
1	0,151	-	Tidak Valid
2	0,068	-	Tidak Valid
3	0,486	Signifikan	Valid
4	0,410	-	Tidak Valid
5	0,222	-	Tidak Valid
6	0,435	Signifikan	Valid
7	0,382	-	Tidak Valid
8	0,271	-	Tidak Valid
9	0,221	-	Tidak Valid
10	0,130	-	Tidak Valid
11	0,294	-	Tidak Valid
12	0,331	-	Tidak Valid
13	0,427	Signifikan	Valid
14	0,160	-	Tidak Valid
15	0,375	-	Tidak Valid
16	0,205	-	Tidak Valid
17	0,366	-	Tidak Valid
18	0,266	-	Tidak Valid
19	0,504	Signifikan	Valid
20	0,451	Signifikan	Valid
Reliabilitas Butir Soal			0,57

Tabel 12 memperlihatkan hasil validasi butir soal tipe pilihan ganda. Terdapat 5 soal valid (nomor 3, 6, 13, 19, dan 20) dan 15 soal tidak valid (nomor 1, 2, 4, 5, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, dan 18). Proporsi validitas butir soal tipe pilihan ganda disajikan pada Gambar 1.



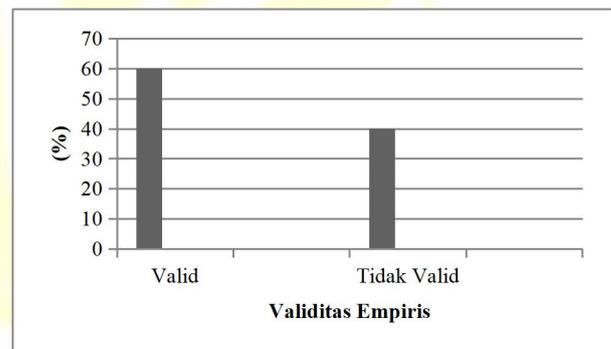
Gambar 1. Validitas empiris butir soal tipe pilihan ganda

Gambar 1., menunjukkan proporsi butir soal pilihan ganda dengan kategori valid sebesar 25% dan tidak valid sebesar 75%. Nilai validitas dan reliabilitas butir soal tipe uraian disajikan pada Tabel 13.

Tabel 13. Validitas dan Reliabilitas Butir Soal Uraian

No	Korelasi	Signifikansi	Keterangan
1	0,273	-	Tidak Valid
2	0,634	Signifikan	Valid
3	0,198	-	Tidak Valid
4	0,768	Sangat Signifikan	Valid
5	0,670	Signifikan	Valid
Reliabilitas Butir Soal			0,48

Tabel 13 memperlihatkan hasil validasi butir soal tipe uraian. Terdapat 3 soal valid (nomor 2, 4, dan 5) dan 2 soal tidak valid (nomor 1 dan 3). Proporsi validitas butir soal tipe uraian disajikan pada Gambar 2.



Gambar 2. Validitas empiris butir soal tipe uraian

Berdasarkan Gambar 2., proporsi butir soal uraian dengan kategori valid sebesar 60% dan tidak valid sebesar 40%. Menurut Sudijono (2011) bahwa butir soal dengan validitas tinggi memiliki ketepatan dan kecermatan yang tinggi, sehingga butir soal tersebut menjalankan fungsi ukurnya sesuai dengan kriteria yang telah ditentukan. Amalia dan Widayanti (2012) menambahkan bahwa suatu tes memiliki validitas jika dapat mengukur apa yang hendak diukur dan memberikan hasil sesuai dengan tujuan pengukuran. Selain itu, butir – butir soal yang dinyatakan valid menandakan bahwa skor yang diperoleh setiap butir soal memiliki kesesuaian arah dengan skor total (Alpusari, 2014).

Butir soal dengan validitas rendah atau tidak valid tidak mampu mengukur kemampuan peserta didik, sehingga perlu diperbaiki sebelum dilakukan ujicoba (Sudijono, 2011). Amalia dan Widayanti (2012)

menambahkan bahwa butir soal yang tidak valid harus dihilangkan atau diperbaiki hingga valid untuk bisa digunakan dan atau dimasukkan ke dalam bank soal. Hal ini menurut Amalia dan Widayanti (2012), suatu tes yang memiliki validitas rendah bahkan tidak valid akan hasilnya tidak relevan dengan tujuan pengukuran. Selain itu, butir – butir soal yang tidak valid menandakan bahwa skor yang diperoleh setiap butir soal memiliki kesesuaian arah yang kecil atau bahkan memiliki korelasi negatif yang signifikan dengan skor total (Alpusari, 2014).

2. Reliabilitas Butir Soal

Reliabilitas butir soal merupakan tetap tidaknya atau keajegan hasil tes yang didapat dari ujicoba tes yang disusun oleh guru sebanyak satu kali (Sugiyono, 2013). Hasil perhitungan diperoleh nilai reliabilitas soal pilihan ganda sebesar 0,57, kemudian dibandingkan dengan Interpretasi Reliabilitas Butir Soal (Tabel 4), diketahui bahwa butir soal tipe pilihan ganda tergolong dalam kategori reliabilitas rendah, karena nilainya 0,57 yaitu lebih kecil dari 0,70 (Sudijono, 2009 dalam Ulum, 2017). Butir soal tipe uraian bahkan memiliki nilai lebih rendah lagi, karena nilai reliabilitasnya 0,48 (Sudijono, 2009 dalam Ulum, 2017).

Setelah dilakukan analisis berdasarkan validitas teoritis, banyak butir soal Ulangan Harian yang dikembangkan guru termasuk kategori tidak valid jika ditinjau dari aspek konstruksi. Selain itu, berdasarkan uji validitas empiris, baik soal pilihan ganda maupun uraian banyak yang berkategori tidak valid. Hal ini dapat berpengaruh terhadap nilai reliabilitas butir soal, yang mana jika butir soal tidak valid maka cenderung tidak reliabel. Menurut Widi (2011) dalam penelitiannya menyatakan bahwa suatu alat ukur harus dinyatakan valid terlebih dahulu sebelum dihitung reliabilitasnya. Jika alat ukur tersebut tidak valid, maka tidak perlu diteruskan untuk uji reliabilitas. Hal ini dikarenakan alat ukur cenderung tidak reliabel. Selain itu, ketidakreliabelan butir soal bisa terkait dengan unsur kejiwaan setiap peserta didik seperti kemampuan, sikap, keahlian, kecakapan, dan lain sebagainya. Suzana (2017) menambahkan bahwa butir soal tidak reliabel dapat dipengaruhi oleh heterogenitas kelompok dan pengalaman serta motivasi peserta didik dalam mengikuti tes. Oleh karena itu, butir soal tersebut sebaiknya diperbaiki untuk meningkatkan kualitasnya sebelum diujicobakan kepada peserta didik.

3. Tingkat Kesukaran Butir Soal

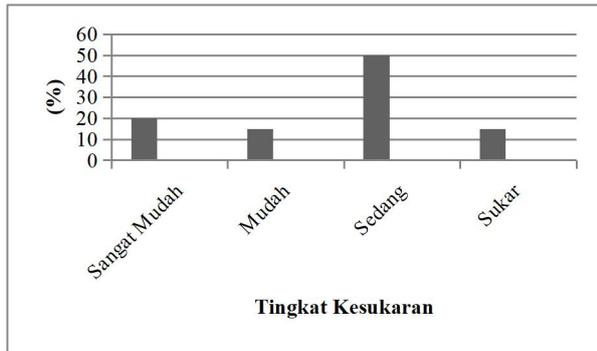
Tingkat kesukaran butir soal merupakan tingkat sukar mudahnya butir soal yang disusun oleh guru yang dinyatakan dengan indeks kesukaran. Hasil perhitungan

indeks kesukaran (%) diinterpretasikan dalam tingkat kesukaran butir soal (Tabel 5). Besarnya indeks kesukaran butir soal berkisar antara 0% sampai dengan 100% (Arikunto, 2012). Hasil uji tingkat kesukaran butir soal tipe pilihan ganda disajikan pada Tabel 14.

Tabel 14. Tingkat Kesukaran Butir Soal Pilihan Ganda

No	Jumlah Benar	Indeks Kesukaran (%)	Interpretasi
1	83	86,46	Sangat Mudah
2	85	88,54	Sangat Mudah
3	67	69,79	Sedang
4	73	76,04	Mudah
5	78	81,25	Mudah
6	65	67,71	Sedang
7	55	57,29	Sedang
8	78	81,25	Mudah
9	20	20,83	Sukar
10	54	56,25	Sedang
11	32	33,33	Sedang
12	44	45,83	Sedang
13	34	35,42	Sedang
14	25	26,04	Sukar
15	15	15,63	Sukar
16	82	85,42	Sangat Mudah
17	31	32,29	Sedang
18	86	89,58	Sangat Mudah
19	47	48,96	Sedang
20	43	44,79	Sedang

Tabel 14 memperlihatkan hasil analisis tingkat kesukaran butir soal pilihan ganda. Diketahui bahwa terdapat 4 soal sangat mudah (soal nomor 1, 2, 16, dan 18), 3 soal mudah (soal nomor 4, 5, dan 8), 10 soal sedang (3, 6, 7, 10, 11, 12, 13, 17, 19, dan 20) serta 3 soal sukar (soal nomor 9, 14, dan 15). Proporsi soal pilihan ganda berdasarkan tingkat kesukaran disajikan pada Gambar 3.



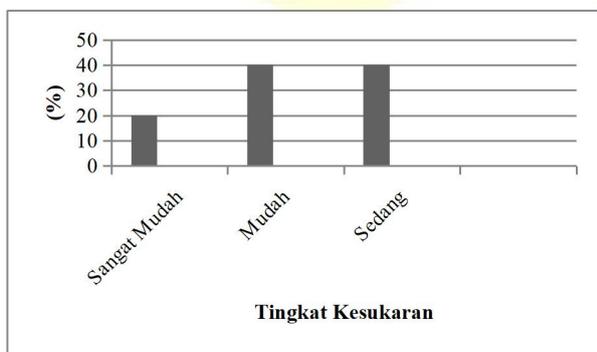
Gambar 3. Proporsi soal tipe pilihan ganda berdasarkan tingkat kesukaran

Berdasarkan Gambar 3., proporsi butir soal pilihan ganda termasuk dalam kategori sangat mudah (20%), mudah (15%), sedang (50%), dan sukar (15%). Hasil uji tingkat kesukaran butir soal uraian disajikan pada Tabel 15.

Tabel 15. Tingkat Kesukaran Butir Soal Uraian

No	Indeks Kesukaran (%)	Interpretasi
1	79,23	Mudah
2	60,38	Sedang
3	95,00	Sangat Mudah
4	58,27	Sedang
5	71,54	Mudah

Tabel 15 memperlihatkan hasil analisis tingkat kesukaran butir soal uraian. Diketahui bahwa terdapat 1 soal sangat mudah (soal nomor 3), 2 soal mudah (soal nomor 1 dan 5), serta 2 soal sedang (soal nomor 2 dan 4). Proporsi butir soal uraian berdasarkan tingkat kesukaran disajikan pada Gambar 4.



Gambar 4. Proporsi soal tipe uraian berdasarkan tingkat kesukaran

Berdasarkan Gambar 4., proporsi butir soal uraian dengan kategori sangat mudah (20%), mudah (40%), dan

sedang (40%). Butir soal tipe pilihan ganda kategori sedang dijumpai pada soal nomor 3, 6, 7, 10, 11, 12, 13, 17, 19, dan 20, sedangkan pada soal uraian dijumpai pada soal nomor 2 dan 4. Arikunto (2012) menyebutkan bahwa derajat kesukaran butir soal kategori sedang dapat diartikan bahwa butir soal tersebut tidak terlalu mudah atau tidak terlalu sukar.

Butir soal tipe pilihan ganda kategori sangat mudah dijumpai pada soal nomor 1, 2, 16, dan 18, sedangkan pada soal uraian dijumpai pada soal nomor 3. Butir – butir soal dengan kategori sangat mudah menandakan bahwa hampir seluruh peserta didik yang mengikuti tes menjawab butir soal tersebut dengan benar (Fatimah dan Alfath, 2019).

Butir soal tipe pilihan ganda kategori mudah dijumpai pada soal nomor 4, 5, dan 8, sedangkan pada soal uraian dijumpai pada soal nomor 1 dan 5. Menurut Supandi dan Farikhah (2016) butir soal kategori mudah menandakan bahwa peserta didik telah menguasai materi, sehingga butir soal dapat diselesaikan dengan baik. Sebaliknya, butir soal kategori sukar menandakan bahwa peserta didik belum menguasai materi, sehingga butir soal tidak dapat diselesaikan dengan baik. Butir soal pilihan ganda dengan kategori sukar dijumpai pada soal nomor 9, 14, dan 15, sedangkan pada soal uraian tidak dijumpai butir soal kategori sukar. Akan tetapi, belum tentu peserta didik belum menguasai materi, hal ini karena berdasarkan hasil analisis menunjukkan bahwa butir soal tersebut tidak valid, bahkan tidak reliabel. Oleh karena itu, sebagian besar peserta didik yang menguasai materi yang ditanyakan beranggapan bahwa terdapat informasi yang salah pada butir soal tersebut, sehingga butir soal kategori sukar bisa disebabkan karena banyak peserta didik menebak jawaban atau menjawab dengan asal – asalan (Solichin, 2017). Hal ini berakibat peserta didik memilih jawaban yang tidak sesuai dengan kunci jawaban.

4. Daya Pembeda Butir Soal

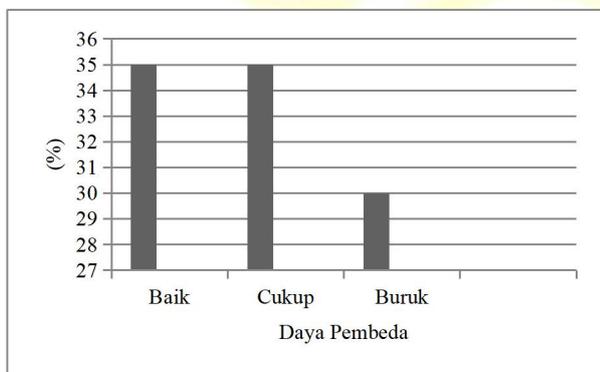
Daya pembeda butir soal digunakan untuk membedakan antara peserta didik yang sudah maupun belum menguasai kompetensi atau materi yang telah diajarkan berdasarkan kriteria tertentu (Ulum, 2017). Hasil perhitungan indeks daya pembeda (%) diinterpretasikan dalam kategori daya beda butir soal (Tabel 6). Hasil uji daya pembeda butir soal tipe pilihan ganda disajikan pada Tabel 16.

Tabel 16. Daya Pembeda Butir Soal Pilihan Ganda

No	Kel. Atas	Kel. Bawah	Indeks Daya Pembeda (%)	Kategori
1	25	21	15,38	Buruk

No	Kel. Atas	Kel. Bawah	Indeks Daya Pembeda (%)	Kategori
2	24	23	3,85	Buruk
3	26	8	69,23	Baik
4	26	13	50,00	Baik
5	24	19	19,23	Buruk
6	25	10	57,69	Baik
7	21	11	38,46	Cukup
8	25	19	23,08	Cukup
9	9	3	23,08	Cukup
10	17	14	11,54	Buruk
11	13	5	30,77	Cukup
12	16	7	34,62	Cukup
13	15	2	50,00	Baik
14	8	3	19,23	Buruk
15	10	1	34,62	Cukup
16	26	22	15,38	Buruk
17	15	4	42,31	Baik
18	26	20	23,08	Cukup
19	19	3	61,54	Baik
20	17	4	50,00	Baik

Tabel 16 memperlihatkan hasil analisis daya pembeda butir soal tipe pilihan ganda. Diketahui bahwa terdapat 7 soal kategori baik (3, 4, 6, 13, 17, 19, dan 20), 7 soal kategori cukup (7, 8, 9, 11, 12, 15, dan 18), serta 6 soal kategori buruk (1, 2, 5, 10, 14, dan 16). Proporsi butir soal pilihan ganda berdasarkan daya pembeda disajikan pada Gambar 5.



Gambar 5. Proporsi butir soal pilihan ganda berdasarkan daya pembedanya

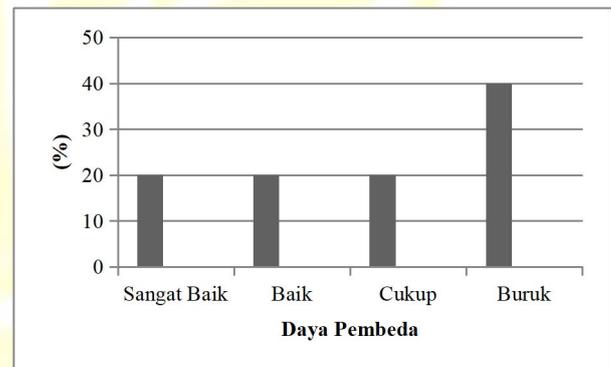
Gambar 5., menunjukkan proporsi butir soal pilihan ganda berdasarkan daya pembeda termasuk kategori baik 35%, cukup 35%, dan buruk 30%. Hasil uji daya

pembeda tiap butir soal tipe uraian disajikan pada Tabel 17.

Tabel 17. Daya Pembeda Butir Soal Uraian

No	Rata - Rata Kel. Atas	Rata - Rata Kel. Bawah	Indeks Daya Pembeda (%)	Kategori
1	4,23	3,69	10,77	Buruk
2	7,15	4,92	22,31	Cukup
3	5,00	4,50	10,00	Buruk
4	9,54	2,12	74,23	Sangat Baik
5	9,62	4,69	49,23	Baik

Tabel 17 memperlihatkan hasil analisis daya pembeda butir soal tipe uraian. Diketahui bahwa terdapat 1 soal kategori sangat baik (soal nomor 4), 1 soal kategori baik (soal nomor 5), 1 soal kategori cukup (soal nomor 2), serta 2 soal kategori buruk (soal nomor 1 dan 3). Proporsi butir soal tipe uraian berdasarkan daya pembeda disajikan pada Gambar 6.



Gambar 6. Proporsi butir soal uraian berdasarkan daya pembedanya

Gambar 6., menunjukkan butir soal uraian memiliki daya pembeda dengan kategori sangat baik sebesar 20%, baik 20%, dan cukup 40%. Butir soal tipe pilihan ganda nomor 3, 6, 13, 19, dan 20, serta tipe uraian nomor 4 termasuk kategori valid dengan tingkat kesukaran kategori sedang dan daya pembeda berkategori baik. Namun sebagai catatan bahwa butir soal tipe uraian nomor 4 kunci jawaban perlu diperbaiki, karena menurut dosen ahli (Tabel 11), kunci jawaban tersebut kurang tepat. Butir soal tipe uraian nomor 5 termasuk kategori valid dengan tingkat kesukaran mudah dan daya pembeda berkategori baik, sedangkan butir soal tipe uraian nomor 2 termasuk kategori valid dengan tingkat kesukaran kategori sedang dan daya pembeda berkategori cukup. Butir soal valid dapat digunakan untuk menguji hasil

belajar peserta didik. Menurut Arikunto (2012), butir soal yang baik memiliki indeks daya pembeda dalam kategori baik, di mana semakin tinggi nilai daya pembeda maka semakin baik butir soal tersebut. Menurut penelitian Sudijono (2011), koefisien daya pembeda semakin tinggi maka butir soal tersebut semakin mampu membedakan antara peserta didik yang sudah maupun belum menguasai kompetensi, di mana peserta didik yang sudah menguasai kompetensi mampu menjawab dengan benar, sedangkan peserta didik yang belum menguasai kompetensi menjawab dengan salah.

Butir soal pilihan ganda nomor 4 dan 17 tidak valid. Butir soal tidak valid tidak mampu mengukur kemampuan peserta didik (Sudijono, 2011). Akan tetapi, kedua soal tersebut memiliki daya pembeda berkategori baik. Menurut dosen ahli (Tabel 7), soal nomor 4 level mengingat (C1) dan nomor 17 level memahami (C2). Effendi (2017) dalam penelitiannya menjelaskan bahwa level mengingat merupakan level terendah dalam Taksonomi Bloom Revisi dan level memahami berada pada tingkatan di atasnya. Pada level mengingat, peserta didik diminta untuk mengingat kembali pengetahuan dari memori jangka panjang, sedangkan pada level memahami, peserta didik diminta untuk mengaitkan informasi baru dengan pengetahuan lama. Soal nomor 4 terkait salah satu ciri – ciri jaringan otot. Untuk dapat menjawabnya, peserta didik diminta untuk mengingat kembali tentang ciri – ciri tersebut termasuk jaringan otot polos, lurik, ataupun jantung. Soal nomor 17 terkait salah satu tanda adanya kelainan pada tulang. Pada soal tersebut peserta didik harus dapat mengaitkannya dengan pengetahuan yang dimiliki terkait kelainan tulang. Oleh karena itu, jika dapat mengingat dan memahami materi, peserta didik dapat menjawab dengan mudah kedua butir soal tersebut, sebaliknya jika tidak dapat mengingat dan memahami materi, peserta didik tidak dapat menjawab dengan benar. Hal ini bisa menjadi salah satu alasan kedua butir soal tersebut memiliki daya pembeda berkategori baik walaupun termasuk soal tidak valid.

Butir soal pilihan ganda nomor 1, 2, 5, 7, 8, 9, 10, 11, 12, 14, 15, 16, dan 18, serta butir soal uraian nomor 1 dan 3 tidak valid, sehingga tidak dapat digunakan untuk menguji hasil belajar peserta didik. Butir soal pilihan ganda nomor 7, 8, 9, 11, 12, 15, dan 18 memiliki daya pembeda berkategori cukup, sedangkan butir soal pilihan ganda nomor 1, 2, 5, 10, 14, dan 16, serta butir soal tipe uraian nomor 1 dan 3 memiliki daya pembeda berkategori buruk. Menurut Arikunto (2012), butir soal dengan daya pembeda kategori cukup bahkan buruk tidak dapat membedakan antara peserta didik yang sudah maupun belum memahami dan menguasai materi yang telah diajarkan berdasarkan kriteria tertentu. Hasil ini menurut

Solichin (2017) bisa disebabkan karena kunci jawaban soal salah, indikator yang diukur tidak jelas, distraktor tidak berfungsi, serta sebagian besar peserta didik yang menguasai materi yang ditanyakan beranggapan bahwa terdapat informasi yang salah pada butir soal tersebut, sehingga banyak peserta didik menebak jawaban atau menjawab dengan asal – asalan.

Berdasarkan hasil penelitian yang telah dilakukan, butir soal pilihan ganda nomor 3, 6, 13, 19, dan 20 serta butir soal uraian nomor 2, 4 dan 5 dapat digunakan untuk menguji hasil belajar peserta didik dengan catatan dilakukan perbaikan. Hal ini dikarenakan butir – butir soal tersebut termasuk kategori valid dan daya pembeda berkategori baik. Butir soal valid dapat mengukur apa yang hendak diukur dan memberikan hasil sesuai dengan tujuan pengukuran (Amalia dan Widayanti, 2012). Butir soal dengan daya pembeda baik mampu membedakan antara peserta didik yang sudah maupun belum menguasai kompetensi (Sudijono, 2011).

Butir soal pilihan ganda nomor 3 dan 13 perlu diperbaiki terkait validitas teoritis, yaitu disesuaikan dengan indikator pencapaian kompetensi, tujuan pembelajaran, dan materi Sistem Gerak Manusia. Perbaikan dilakukan terhadap pengecoh pada pilihan jawaban soal agar dapat berfungsi dengan baik, serta butir soal maupun pilihan jawaban sebaiknya menghindari kata atau kalimat yang mengandung arti lebih dari satu, sedangkan butir soal uraian nomor 4 kunci jawaban perlu diperbaiki sebagai acuan penilaian. Selain itu, perbaikan butir soal perlu dilakukan terhadap tingkat kognitif dan kesesuaian dengan tuntutan Kompetensi Dasar yang digunakan. Tingkatan kognitif butir soal tipe pilihan ganda nomor 3, 19, dan 20 termasuk level C1, nomor 13 level C2, dan nomor 6 level C3, serta pada butir soal tipe uraian nomor 4 dan 5 termasuk level C2. Butir – butir soal tersebut perlu diperbaiki agar sesuai dengan tuntutan KD 3.5 Permendikbud Nomor 37 Tahun 2018 (revisi) bagi kelas XI Biologi yang mengacu pada Taksonomi Bloom Revisi. KD tersebut menyiratkan tingkat kognitif (pengetahuan) yang harus dicapai sesuai dengan tuntutan kurikulum 2013, yaitu memiliki keterampilan berpikir kritis dan berpikir tingkat tinggi minimal menguasai level menganalisis.

Butir soal nomor 1, 4, 8, 9, 10, 14, 15, dan 16 serta butir soal tipe uraian nomor 1 valid secara teoritis. Hal tersebut menandakan bahwa butir soal sesuai dengan kaidah dan ketentuan penulisan soal yang baik dan benar (Mulyatiningsih, 2011). Akan tetapi tidak valid secara empiris. Data empirik diperoleh dari penerapan atau ujicoba butir soal kepada peserta didik, sehingga ketidakvalidan tersebut dapat disebabkan terkait dengan unsur kejiwaan setiap peserta didik seperti kemampuan,

sikap, keahlian, kecakapan, dan lain sebagainya. Butir soal yang tidak valid cenderung tidak reliabel (Widi, 2011). Menurut Suzana (2017), butir soal tidak reliabel dapat dipengaruhi oleh heterogenitas kelompok dan pengalaman serta motivasi peserta didik dalam mengikuti tes. Oleh karena itu, soal - soal tersebut sebaiknya diperbaiki terlebih dahulu untuk meningkatkan kualitasnya sebelum dilakukan ujicoba kepada peserta didik. Perbaikan dilakukan dengan menyesuaikan butir soal dengan kriteria yang telah ditentukan oleh pembuat soal, sehingga butir soal dapat mengukur apa yang hendak diukur dan memberikan hasil sesuai dengan tujuan pengukuran (Amalia dan Widayanti, 2012). Butir soal tipe pilihan ganda nomor 2, 5, 7, 11, 12, 17, dan 18, serta butir soal tipe uraian nomor 3 memiliki kategori tidak valid secara teoritis maupun empiris, sehingga lebih baik dibuang atau diganti dengan butir soal baru.

Perbaikan – perbaikan butir soal didasarkan pada Permendikbud Nomor 22 Tahun 2016, evaluasi bertujuan untuk memperbaiki dan melakukan perencanaan ulang terhadap kegiatan pembelajaran sesuai dengan Standar Penilaian Pendidikan. Dalam hal ini kegiatan evaluasi dapat digunakan guru untuk mengetahui seberapa tingkat ketercapaian Kompetensi Dasar (KD) dalam proses kegiatan pembelajaran (Ulum, 2017). Setelah butir soal melalui proses perbaikan ataupun diganti, diharapkan menghasilkan butir – butir soal yang baik dan bermutu sebagai acuan untuk perbaikan pembelajaran (Permendikbud RI Nomor 22 Tahun 2016).

Guru juga perlu meningkatkan kemampuan untuk memperbaikinya antara lain dengan sering mengikuti acara workshop ataupun seminar tentang penyusunan butir soal, memiliki dan menggunakan buku pedoman, serta mengikuti pelatihan penyusunan butir soal yang baik dan benar sesuai ketentuan, sehingga guru benar – benar siap dalam melakukan kegiatan evaluasi terhadap peserta didik (Saraswati dan Agustika, 2020). Dalam hal ini, guru bisa menyiapkan alat penilaian sebaik mungkin, sehingga hasil yang diperoleh terpercayanya dan akurat serta mampu mengukur kemampuan setiap peserta didik. Oleh karena itu, sebelum digunakan, butir soal dianalisis terlebih dahulu guna mendapatkan validitas teoritis dan validitas empirisnya. Analisis validitas teoritis berdasarkan aspek tingkat kognitif, materi, konstruksi, dan bahasa. Analisis validitas empiris dilakukan berdasarkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Guru dalam melakukan kegiatan analisis butir soal ini perlu didampingi dan dibimbing oleh berbagai ahli, di antaranya ahli materi, ahli analisis butir soal, editor, dan lain sebagainya (Danuwijaya, 2018).

Persebaran jenjang ranah kognitif perlu diperhatikan, minimal sesuai dengan Taksonomi Bloom Revisi ranah

kognitif. Selain itu, menurut Witarsa dkk (2017), penyusunan butir soal hendaknya disesuaikan dengan indikator pencapaian kompetensi, tujuan pembelajaran, dan materi pembelajaran yang beracuan pada KD yang digunakan, serta berpedoman pada kaidah penulisan butir soal dengan memperhatikan penggunaan kalimat Bahasa Indonesia yang baik dan benar. Kalimat yang baik adalah kalimat yang terdiri dari subjek, predikat, objek, dan keterangan (SPOK) (Pusat Pembinaan dan Pemasyarakatan, 2015). Selain itu, butir soal hendaknya dilengkapi dengan petunjuk pengerjaan, pedoman penyekoran, serta waktu yang diperlukan dalam mengerjakan soal.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terimakasih kepada Ibu Dr. Nur Ducha, S.Si., M.Si. dan Ibu Nur Qomariyah, S.Pd., M.Sc. selaku dosen ahli, Bapak Eko Sugiharto, S.Pd. selaku guru SMA Negeri 1 Tarik, Bapak Drs. Digdo Santoso, M.Pd. selaku Kepala Sekolah SMA Negeri 1 Tarik, dan peserta didik kelas XI IPA SMA Negeri 1 Tarik, serta semua pihak yang telah terlibat langsung maupun tidak langsung dalam membantu penelitian ini.

PENUTUP

Simpulan

Berdasarkan hasil penelitian dapat ditarik kesimpulan sebagai berikut.

1. Butir soal Ulangan Harian (UH) dalam bentuk pilihan ganda dan uraian yang dikembangkan oleh guru SMA pada mata pelajaran biologi materi Sistem Gerak Manusia kelas XI IPA di SMA Negeri 1 Tarik berkategori valid berdasarkan aspek materi dan bahasa, sebaliknya banyak butir soal tidak valid (tidak sesuai aspek penilaian) berdasarkan aspek konstruksi. Selain itu, butir soal lebih didominasi tingkat kognitif Taksonomi Bloom Revisi level C1, C2, dan C3.
2. Butir soal Ulangan Harian (UH) dalam bentuk pilihan ganda dan uraian yang dikembangkan oleh guru SMA pada mata pelajaran biologi materi Sistem Gerak Manusia kelas XI IPA di SMA Negeri 1 Tarik banyak di antaranya tidak valid secara empiris.

Saran

Berdasarkan hasil penelitian, terdapat beberapa saran sebagai berikut.

1. Butir soal sebelum diujicobakan kepada peserta didik sebaiknya dianalisis terlebih dahulu berdasarkan validitas teoritis dan validitas empiris.

2. Persebaran jenjang ranah kognitif perlu diperhatikan, sehingga sesuai dengan Taksonomi Bloom Revisi ranah kognitif dan tuntutan Kompetensi Dasar (KD) 3.5 Permendikbud Nomor 37 Tahun 2018 (revisi) kelas XI Biologi.
3. Penyusunan butir soal hendaknya berpedoman pada kaidah penulisan butir soal yang baik dan benar.

DAFTAR PUSTAKA

- Alpusari, Mahmud. (2014). Analisis Butir Soal Konsep Dasar IPA 1 melalui Penggunaan Program Komputer ANATES Versi 4.0 For Windows. *Jurnal Primary Studi Pendidikan Guru Sekolah Dasar Fakultas Keguruan dan Ilmu Pendidikan Universitas Riau*. Vol. 3, No. 2, p. 106 - 115.
- Amalia, Ata Nayla dan Widayati, Ani. (2012). Analisis Butir Soal Tes Kendali Mutu Kelas XII SMA Mata Pelajaran Ekonomi Akuntansi di Kota Yogyakarta Tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*. Vol. 10, No. 1, p. 1 - 26.
- Anderson, L. W., and Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Education Objectives* dalam Effendi, Ramlan. (2017). Konsep Revisi Taksonomi Bloom dan Implementasinya pada Pelajaran Matematika SMP. *Jurnal Ilmiah Pendidikan Matematika*. Vol. 2, No. 2, p. 72 - 78.
- Arikunto, Suharsimi. (1986). *Dasar-Dasar Evaluasi Pendidikan* dalam Amelia, Maria Agustina. (2016). Analisis Soal Tes Hasil Belajar *High Order Thinking Skills* (HOTS) Matematika Materi Pecahan untuk Kelas 5 Sekolah Dasar. *Jurnal Penelitian (Edisi Khusus PDSG)*. Vol. 20, No. 2, p. 123 - 131.
- Arikunto, Suharsimi. (2012). *Dasar-Dasar Evaluasi Pendidikan* dalam Amelia, Maria Agustina. (2016). Analisis Soal Tes Hasil Belajar *High Order Thinking Skills* (HOTS) Matematika Materi Pecahan untuk Kelas 5 Sekolah Dasar. *Jurnal Penelitian (Edisi Khusus PDSG)*. Vol. 20, No. 2, p. 123 - 131.
- Aziz, F., Nurjana, F, dan Sari, D. P. (2017). Aktualisasi TTB (Teori Taksonomi Bloom) melalui Drama Kepahlawanan guna Penanaman Pendidikan Karakter pada Peserta Didik. *Seminar Nasional*. p. 715 - 724.
- Dafitri, Haida. (2017). Pemanfaatan *Wondershare Quiz Creator* dalam Tes Berbasis Komputer. *QUERY : Jurnal Sistem Informasi*. Vol. 1, No. 1, p. 8 - 18.
- Danuwijaya, Ari Arifin. (2018). *Item Analysis of Reading Comprehension Test for Post-Graduate Students*. *Journal of English Education*. Vol. 7, Issue 1, p. 29 - 40.
- Depdiknas. (2008). *Panduan Penulisan Butir Soal*. Jakarta: Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah Direktorat Pembinaan Sekolah Menengah Atas.
- Effendi, Ramlan. (2017). Konsep Revisi Taksonomi Bloom dan Implementasinya pada Pelajaran Matematika SMP. *Jurnal Ilmiah Pendidikan Matematika*. Vol. 2, No. 2, p. 72 - 78.
- Fatimah, Laela Umi dan Alfath, Khairuddin. (2019). Analisis Kesukaran Soal, Daya Pembeda, dan Fungsi Distraktor. *Jurnal Komunikasi dan Pendidikan Islam*. Vol. 8, No. 2, p. 37 - 64.
- Fitrianawati, Meita. (2017). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru dan Hasil Belajar Peserta Didik. *Seminar Nasional Pendidikan PGSD UMS dan HDPGSDI Wilayah Jawa*. p. 282 - 295.
- Hartati, Neti and Yogi, Hendro Pratama Supra. (2019). *Item Analysis for a Better Quality Test. English Language in Focus (ELIF)*. Vol. 2, Issue 1, p. 59 - 70.
- Haryanto. (2020). *Evaluasi Pembelajaran; Konsep dan Manajemen*. Yogyakarta: UNY Press.
- Karim, S. A., Sudiro, S., and Sakinah, S. (2021). *Utilizing test items analysis to examine the level of difficulty and discriminating power in a teacher-made test*. *Journal of English Education, Literature, and Culture*. Vol. 6, No. 2, p. 256 - 269.
- Khaerudin. (2016). Teknik Penskoran Tes Obyektif Model Pilihan Ganda. *Jurnal Madaniyah*. Vol. 2 Edisi XI, p. 185 - 204.
- Mehta, Gyata and Mokhasi, Varsha. (2014). *Item Analysis of Multiple Choice Question- An Assessment of the Assessment Tool*. Vol. 4, Issue 7, p. 197 - 202.
- Muhson, A., Lestari, B., Supriyanto, and Baroroh, K. (2017). *The Development of Practical Item Analysis Program for Indonesian Teachers*. *International Journal of Instruction*. Vol. 10, No. 2, p. 199 - 210.
- Mulyatiningsih, Endang. (2011). *Riset Terapan Bidang Pendidikan dan Teknik*. Yogyakarta: UNY Press.
- Permendikbud. (2013). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 66 Tahun 2013 tentang Standar Penilaian Pendidikan*. Jakarta: Kemendikbud.
- Permendikbud. (2016). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 22 Tahun 2016 tentang Standar Proses Pendidikan Dasar dan Menengah*. Jakarta: Kemendikbud.
- Permendikbud. (2016). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2016 tentang Standar penilaian Pendidikan*. Jakarta: Kemendikbud.

- Permendikbud. (2018). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 37 Tahun 2018 tentang Perubahan Atas Peraturan Menteri Pendidikan dan Kebudayaan Nomor 24 Tahun 2016 tentang Kompetensi Inti dan Kompetensi Dasar Pelajaran pada Kurikulum 2013 pada Pendidikan Dasar dan Pendidikan Menengah*. Jakarta: Kemendikbud.
- Prasetya, Tri Indra. (2012). Meningkatkan Keterampilan Menyusun Instrumen Hasil Belajar Berbasis Modul Interaktif Bagi Guru – Guru IPA SMP N Kota Magelang. *Journal of Education Research and Evaluation*. Vol. 1, No. 2, p. 106 – 112.
- Pusat Pembinaan dan Pemasyarakatan. (2015). *Kalimat*. Jakarta: Kementerian dan Kebudayaan.
- Riduwan. (2013). *Skala Pengukuran Variabel – Variabel Penelitian*. Bandung: Alfabeta.
- Saraswati, Putu Manik Sugiari dan Agustika, Gusti Ngurah Sastra. (2020). Kemampuan Berpikir Tingkat Tinggi dalam Menyelesaikan Soal HOTS Mata Pelajaran Matematika. *Jurnal Ilmiah Sekolah Dasar*. Vol. 4, No. 2, p. 257 – 269.
- Sitorus, Andini P., Wurarah, M., dan Gedoan, S. P. (2017). Analisis Butir Soal Ulangan Harian Buatan Guru Mata Pelajaran Biologi Kelas X SMA N 1 Remboken. *Jurnal Sains, Matematika, dan Edukasi (JSME)*. Vol. 5, No. 1, p. 49 - 54.
- Solichin, Mujiyanto. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan. *Jurnal Manajemen dan Pendidikan Islam*. Vol. 2, No. 2, p. 192 – 213.
- Sudijono, Anas. (2009). *Pengantar Evaluasi Pendidikan dalam Ulum*, Miftahul. (2017). Analisis Butir Soal Ulangan Harian pada Mata Pelajaran Ekonomi KD 3.1 Pendapatan Nasional Kelas XI IPS 1 di SMA Negeri 1 Gresik. *Jurnal Pendidikan Ekonomi*. Vol. 5, No. 3, p. 1 - 5.
- Sudijono, Anas. (2011). *Pengantar Evaluasi Pendidikan*. Jakarta : PT Rajagrafindo Persada.
- Sugiyono. (2013). *Metode Penelitian Kuantitatif Kualitatif dan R dan D*. Bandung: Alfabeta.
- Supandi dan Farikhah, Lailatul. (2016). Analisis Butir Soal Matematika pada Instrumen Uji Coba Materi Segitiga. *Jurnal Ilmiah Pendidikan Matematika*. Vol. 1, No. 1, p. 71 – 78.
- Suzana, Andriani. (2017). Analisis Tingkat Kesukaran dan Daya Beda Butir-Butir Soal Penilaian Akhir Tahun Matematika Kelas X di SMA Negeri 1 Purbalingga. *Jurnal MathGram Matematika*. Vol. 2, No 2, p. 1 - 8.
- Ulum, Miftahul. (2017). Analisis Butir Soal Ulangan Harian pada Mata Pelajaran Ekonomi KD 3.1 Pendapatan Nasional Kelas XI IPS 1 di SMA Negeri 1 Gresik. *Jurnal Pendidikan Ekonomi*. Vol. 5, No. 3, p. 1 - 5.
- Widi, Ristya. (2011). Uji Validitas dan Reliabilitas dalam Penelitian Epidemiologi Kedokteran Gigi. *Stomatognatic (J.K.G. Unej)*. Vol. 8, No. 1, p. 27 – 34.
- Witarsa, B. N., Munawar, W., dan Berman, E., T. (2017). Penyusunan dan Analisis Butir Soal Mata Pelajaran Teknologi Dasar Otomotif di SMK. *Journal of Mechanical Engineering Education*. Vol. 2, No. 2, p. 146 - 155.