

Pengembangan Instrumen Penilaian Berbasis *Liveworksheet* Pada Materi Sistem Pernapasan Siswa Kelas XI MIA 3 MA Madani Alauddin Pao-Pao

Fahri Amalia Jufri¹, St. Syamsudduha^{2*}, Ahmad Ali³

^{1, 2, 3} Universitas Islam Negeri Alauddin Makassar

*Correspondence email: st.syamsudduha@uin-alauddin.ac.id

ABSTRAK

Penelitian ini dilaksanakan dengan tujuan merancang serta mengembangkan instrumen evaluasi berbasis liveworksheet dan menilai kualitasnya pada materi Sistem Pernapasan siswa kelas XI MIA 3 MA Madani Alauddin Pao-pao. Kualitas instrumen dianalisis berdasarkan aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran, serta efektivitas pengecoh pada setiap butir soal. Penelitian ini mengadaptasi model pengembangan yang dikemukakan oleh Tessmer, termasuk dalam kategori Formative Research. Prosedur pengembangannya meliputi tiga tahapan utama, yakni tahap awal, tahap evaluasi mandiri yang mencakup kegiatan analisis dan perancangan, serta tahap prototyping yang terdiri atas peninjauan oleh ahli, uji coba perorangan, uji kelompok kecil, dan uji lapangan. Subjek penelitian ini adalah peserta didik kelas XI MIA 3 MA Madani Alauddin Pao-pao. Hasil penelitian menunjukkan bahwa koefisien validitas isi mencapai angka 1, yang mengindikasikan tingkat validitas sangat tinggi. Dari total 25 butir soal, sebanyak 44% dinyatakan valid secara empiris, sedangkan 56% tergolong tidak valid. Uji reliabilitas dengan menggunakan rumus KR-20 menghasilkan nilai sebesar 0,99, yang menandakan reliabilitas sangat tinggi. Analisis daya pembeda memperlihatkan bahwa 36% butir soal berada pada kategori sangat buruk, 36% buruk, 24% cukup, dan hanya 4% tergolong baik. Berdasarkan tingkat kesukaran, 84% butir soal dikategorikan mudah dan 16% sedang. Adapun hasil analisis efektivitas pengecoh menunjukkan bahwa sebagian besar pengecoh belum berfungsi secara optimal, dengan rincian 24% sangat buruk, 28% buruk, 8% kurang baik, 20% baik, dan 20% sangat baik.

ABSTRACT: *This study was conducted with the aim of designing and developing a liveworksheet-based evaluation instrument and assessing its quality in the Respiratory System material for grade XI MIA 3 MA Madani Alauddin Pao-pao students. The quality of the instrument was analysed based on the aspects of validity, reliability, discriminating power, level of difficulty, and effectiveness of distractors in each item. This study adapted the development model proposed by Tessmer, which falls under the category of Formative Research. The development procedure consisted of three main stages, namely the initial stage, the self-evaluation stage which included analysis and design activities, and the prototyping stage which consisted of expert review, individual testing, small group testing, and field testing. The subjects of this study were grade XI MIA 3 MA Madani Alauddin Pao-pao students. The results showed that the content validity coefficient reached 1, indicating a very high level of validity. Of the total 25 items, 44% were declared empirically valid, while 56% were classified as invalid. The reliability test using the KR-20 formula produced a value of 0.99, indicating very high reliability. The discrimination power analysis showed that 36% of the items were in the very poor category, 36% were poor, 24% were adequate, and only 4% were good. Based on the level of difficulty, 84% of the items were categorised as easy and 16% as moderate. The results of the distractor effectiveness analysis showed that most distractors were not functioning optimally, with 24% very poor, 28% poor, 8% not very good, 20% good, and 20% very good.*

Keywords: Assessment instruments, *Liveworksheet*, Question item analysis.

PENDAHULUAN

Pada masa sekarang kita telah memasuki era digital, dimana pada era ini kita banyak menjumpai masyarakat telah menggunakan alat-alat digital untuk membantu masyarakat dalam menyampaikan informasi dan mejadikan sebagai alat mencari sumber-sumber yang dibutuhkan terutama dalam proses belajar mengajar. Perkembangan teknologi sangat membantu guru dalam beradaptasi dengan kurikulum yang sesuai dengan perkembangan zaman ini sehingga pembelajaran yang terjadi dalam dunia pendidikan akan menjadi lebih menarik dalam mencapai tujuan pembelajaran (Hidayah et al., 2020). Kondisi tersebut menjadi dasar rasional bagi implementasi pemanfaatan teknologi dalam konteks pendidikan.

Pemanfaatan teknologi dalam ranah pendidikan pada masa kini masih belum mencapai tingkat optimal. Hal tersebut tercermin dari masih terbatasnya integrasi teknologi dalam pelaksanaan proses pembelajaran di berbagai institusi pendidikan. Kondisi tersebut tergambar dari kecenderungan peserta didik yang masih memanfaatkan media berbasis kertas sebagai sarana utama dalam penyelesaian tugas atau soal. Berdasarkan hasil wawancara dengan salah satu pendidik di MA Madani Alauddin Pao-pao, diketahui bahwa instrumen penilaian yang diterapkan masih bersifat konvensional, yakni menggunakan lembar kertas, sehingga dinilai kurang efisien karena pendidik harus membagikan dan menilai jawaban secara manual. Selain itu, metode tersebut juga memiliki potensi terjadinya kehilangan atau tercecernya lembar jawaban.

Penilaian dilakukan melalui instrumen yang dirancang oleh guru guna menilai dan mengevaluasi hasil dari pembelajaran siswa terkait dengan apa yang diajarkan dalam pembelajaran. Penilaian guru terhadap siswa menggunakan berbagai instrumen seperti instrumen tes yang berisikan soal-soal, instrumen observasi dalam bentuk lembar observasi, instrumen penugasan yang terdiri atas lembar kerja sesuai dengan topik pembelajaran, serta jenis instrumen lainnya yang mendukung proses evaluasi hasil belajar. Instrumen-instrumen yang disajikan disertai dengan pedoman penskoran (Febriana, 2019). Instrumen evaluasi pembelajaran dapat dikatakan optimal apabila mampu mengakomodasi secara komprehensif tiga ranah utama dalam proses pendidikan, yaitu ranah afektif, kognitif, dan psikomotorik (Sylvia et al., 2019). Ketiga dimensi tersebut berperan sebagai elemen esensial yang membentuk sistem penilaian komprehensif dalam konteks pendidikan.

Liveworksheet merupakan salah satu aplikasi digital berbasis web yang tersedia secara daring melalui laman resminya <https://www.liveworksheets.com>. Penggunaan

aplikasi ini dapat memudahkan guru dalam membuat LKPD ataupun Instrumen penilaian berbasis konvensional menjadi berbasis online. Penggunaan aplikasi ini sangat praktis karena dapat memudahkan guru dalam mengoreksi jawaban siswa sehingga guru dapat menghemat waktu dan aplikasi serta dapat merangsang motivasi belajar siswa secara lebih aktif. (Ariskasari & Sulisworo, 2021). Selain itu, lembar kerja interaktif ini memanfaatkan teknologi terbaru yang diterapkan dalam bidang pendidikan, seperti suara, video, latihan seret dan lepas, penggabungan dengan panah, pilihan ganda, dan sebagainya, bahkan termasuk latihan berbicara yang dilakukan siswa dengan mikrofon (Andriyani et al., 2020).

METODE PENELITIAN

Penelitian ini dikategorikan sebagai studi *Research and Development* (R&D) dengan penerapan pendekatan Formative Research sebagaimana dikemukakan oleh Tessmer (1993). Proses pengembangan dijalankan melalui tiga tahap utama, yaitu tahap *Preliminary*, *tahap Self-Evaluation* yang mencakup kegiatan analisis dan perancangan, serta tahap Prototyping yang meliputi serangkaian prosedur, antara lain *expert review*, *one-to-one*, *small group*, dan *field test*. Partisipan penelitian terdiri dari 29 siswa kelas XI MIA 3 MA Madani Alauddin Pao-Pao. Instrumen yang digunakan mencakup lembar validasi dan tes evaluasi. Seluruh data yang diperoleh dianalisis melalui pengujian validitas dan reliabilitas, disertai dengan analisis daya pembeda, tingkat kesulitan butir soal, serta efektivitas distractor pada masing-masing item evaluasi.

Analisis data validitas merujuk pada validitas isi dan validitas soal. Analisis validitas isi/dilakukan dengan mempertimbangkan relevansi kedua validator. Hasil relevansi kedua validator kemudian dapat dihitung menggunakan rumus Gregory (Gregory, 2000 dalam Setemen, 2018).

$$\text{Validitas isi} = \frac{D}{(A + B + C + D)}$$

Tabel 1. Relevansi dua ahli (validator)

		Validator 1	
		Tidak relevan Skor (1-2)	Relevan Skor (3-4)
Validator 2	Tidak relevan Skor (1-2)	A	B
	Relevan Skor (3-4)	C	D

Koefisien korelasi yang diperoleh dari hasil penilaian dua validator mengindikasikan tingkat relevansi yang terbagi ke dalam tiga klasifikasi. Koefisien dengan nilai $\geq 0,81$

merepresentasikan tingkat validitas tinggi, sementara nilai yang berada pada rentang $> 0,40$ hingga $< 0,81$ dikategorikan sebagai validitas sedang. Adapun nilai $\leq 0,40$ menunjukkan tingkat validitas yang rendah (Retnawati, 2016).

Uji validitas empiris terhadap setiap butir instrumen dilaksanakan melalui pendekatan statistik yang relevan, yakni dengan memanfaatkan rumus *korelasi Product Moment Pearson* guna mengidentifikasi tingkat Keterkaitan antara skor pada setiap butir pernyataan dengan skor total keseluruhan instrumen secara komprehensif (Yusuf, 2015).

$$r_{xy} = \frac{1N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\}\{N \sum Y^2 - (\sum Y)^2\}}}$$

Keterangan:

r_{xy} = Koefisien korelasi antara variabel X dengan variabel Y

N = Jumlah responden

$\sum X$ = Jumlah variabel X

$\sum Y$ = Jumlah variabel Y

$\sum XY$ = Jumlah perkalian X dan Y

$\sum X^2$ = Jumlah kuadrat dari sebaran X

$\sum Y^2$ = Jumlah kuadrat dari sebaran Y

Koefisien korelasi validitas empiris suatu butir soal dikategorikan valid apabila nilai $r_{xy} \geq r_{tabel}$, sedangkan apabila nilai $r_{xy} \leq r_{tabel}$, Dengan demikian, butir soal tersebut dikategorikan sebagai tidak valid menurut parameter korelasi Pearson (Yusuf, 2015).

Analisis reliabilitas instrumen dapat dilakukan melalui perhitungan menggunakan formula *Kuder-Richardson 20* (KR-20) (Yusuf, 2015).

$$r_{11} = \left(\frac{n}{n-1} \right) \frac{s^2_t - \sum pq}{s^2_t}$$

keterangan:

r_{11} = Reliabilitas keseluruhan instrumen

n = Total butir soal dalam satu instrumen

p = Persentase siswa yang menjawab benar pada setiap item soal

q = Persentase siswa yang menjawab salah pada setiap item soal

S^2_t = Variasi skor dari instrumen (Kuadrat St)

St = Deviasi standar dari total skor

Nilai butir soal dengan $r_{11} \geq 0,70$ dikategorikan memiliki tingkat reliabilitas tinggi, sedangkan nilai $r_{11} \leq 0,70$ termasuk dalam kategori reliabilitas rendah atau tidak reliabel (Yusuf, 2015).

Analisis terhadap kemampuan daya pembeda suatu instrumen dapat dilaksanakan melalui prosedur perhitungan yang mengacu pada penggunaan rumus tertentu (Syamsudduha, 2014 dalam Alwiyah et al., 2020):

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

Keterangan:

ID = Indeks diskriminasi (Daya pembeda).

BA = Jumlah kelompok atas yang menjawab benar.

BB = Jumlah kelompok bawah yang menjawab benar.

JA = Banyaknya peserta kelompok atas.

JB = Banyaknya peserta kelompok bawah.

Indeks diskriminasi suatu instrumen memperoleh nilai $ID \leq 0,00$ menunjukkan kemampuan pembeda yang sangat rendah. Apabila berada pada rentang 0,00 hingga 0,20, instrumen tersebut termasuk dalam kategori daya pembeda rendah. Sementara itu, nilai yang berkisar antara 0,20 hingga 0,40 diklasifikasikan sebagai memiliki daya pembeda yang cukup. Selanjutnya, indeks diskriminasi dengan kisaran 0,40 hingga 0,70 menunjukkan kategori baik, sementara nilai yang berada pada interval 0,70 hingga 1,00 merepresentasikan kualitas daya pembeda yang sangat baik (Syamsudduha, 2014 dalam Alwiyah et al., 2020).

Analisis terhadap derajat kesukaran suatu instrumen dapat dilaksanakan dengan menerapkan perhitungan berdasarkan rumus atau formula tertentu (fitriani, 2021):

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

Keterangan:

P = Indeks kesukaran.

B = Banyaknya subjek yang menjawab benar pada tiap butir soal.

JS = Jumlah subjek yang ikut tes.

Butir instrumen yang menunjukkan indeks kesukaran 0,00 hingga 0,30 dikategorikan sebagai butir dengan tingkat kesukaran tinggi atau tergolong sukar. Sebaliknya, butir yang memiliki indeks kesukaran antara 0,31 sampai 0,70 diklasifikasikan dalam kategori sedang. Butir soal yang memiliki indeks kesukaran dalam rentang 0,71 hingga 1,00 tergolong dalam kategori rendah, mengindikasikan bahwa butir tersebut cenderung mudah untuk diselesaikan oleh peserta tes (Fitriani, 2021).

Analisis keefektifan daya pengecoh instrumen dapat dihitung dengan menggunakan rumus (Arifin, 2013 dalam Alwiyah et al.,2020):

$$IP = \frac{P}{(n - B)/(n - 1)} \times 100\%$$

Keterangan:

IP = Indeks daya pengecoh.

P = Jumlah siswa menjawab salah pada setiap butir soal (terkecoh).

N = Jumlah seluruh siswa yang ikut tes

B = Jumlah seluruh siswa menjawab benar pada setiap butir soal.

n = Jumlah pilihan jawaban (opsi).

I = Bilangan tetap.

Indeks daya pengecoh pada instrumen yang memiliki persentase melebihi 200% dikategorikan dalam kriteria sangat buruk. Adapun rentang nilai antara 0%–25% maupun 176%–200% termasuk dalam kategori buruk, sedangkan rentang 26%–50% atau 151%–175% digolongkan kurang baik. Selanjutnya, nilai indeks antara 51%–75% atau 126%–150% berada pada kategori baik, sementara kisaran 76%–125% diklasifikasikan sebagai kategori sangat baik (Arifin, 2013 dalam Alwiyah et al.,2020).

HASIL DAN PEMBAHASAN

Pengembangan instrumen penilaian mengadaptasi model pengembangan Tessmer, yang berfokus pada tiga komponen dalam *Formative Evaluation*. Prosedur pengembangan dimulai dari tahap pendahuluan (*preliminary*) sebagai landasan awal, kemudian dilanjutkan pada tahap evaluasi diri (*self-evaluation*) yang meliputi proses analisis kebutuhan serta perancangan instrumen penilaian. Tahap berikutnya adalah pengembangan prototipe (*prototyping*), di mana instrumen yang telah dirancang divalidasi melalui beberapa tahapan

sistematis, yakni penelaahan oleh ahli (*expert review*), uji coba individual (*one-to-one*), pengujian dalam kelompok kecil (*small group*), hingga uji lapangan (*field test*).

Penelitian ini dilaksanakan di suatu lembaga pendidikan Madrasah Aliyah Madani Alauddin Pao-Pao. Tahap *one-to-one*, kegiatan berlangsung di ruang guru dengan melibatkan satu peserta non-subjek yang dipilih secara purposif oleh guru Biologi. Tahap berikutnya, yaitu *small group*, mencakup enam peserta, sedangkan tahap *field test* melibatkan 32 siswa, keduanya dilaksanakan di ruang kelas. Evaluasi pada tahap pertama dan *small group* dilaksanakan dengan memanfaatkan media cetak berupa lembar penilaian, sedangkan *field test*, peserta didik mengerjakan instrumen evaluasi melalui perangkat *smartphone* menggunakan aplikasi *Liveworksheet*. Seluruh rangkaian kegiatan penelitian berlangsung di bawah supervisi langsung peneliti bersama guru mata pelajaran yang bersangkutan, sehingga memastikan keterkendalian proses serta keabsahan pelaksanaannya.

Hasil penelitian yang diperoleh dari setiap tahap sebagai berikut:

1) Tahap *Preliminary*

Dalam tahap ini, peneliti mengumpulkan beragam literatur ilmiah yang memiliki keterkaitan signifikan dengan fokus dan lingkup penelitian yang sedang dilaksanakan, khususnya yang membahas pengembangan instrumen evaluasi dan penerapan aplikasi *Liveworksheet*. Hasil penelaahan literatur menunjukkan bahwa perancangan instrumen penilaian berbasis soal menuntut analisis komprehensif terhadap beragam indikator kualitas butir soal.

2) Tahap *Self Evaluation*

Tahapan *Self Evaluation* mencakup dua fase utama yang saling berkaitan dalam proses pelaksanaannya:

a. Analisis

Tahap ini mencakup tiga komponen utama, yakni analisis terhadap kurikulum, identifikasi karakteristik peserta didik, serta telaah komprehensif terhadap substansi materi pembelajaran.

b. Desain

Tahap perancangan merupakan fase fundamental dalam proses pengembangan instrumen penilaian yang berorientasi pada keterpaduan antara materi pembelajaran dan indikator soal dengan kompetensi dasar yang ditetapkan. Pada tahap ini, disusun pula

instrumen validitas isi yang meliputi penyusunan kisi-kisi serta perumusan butir-butir soal yang berfungsi sebagai elemen utama dalam rancangan awal instrumen penilaian.

3) Tahap *prototyping*

Tahap pengembangan prototipe Proses ini melibatkan rangkaian tahapan sistematis yang mencakup peninjauan oleh para ahli (*expert review*), evaluasi individual (*one-to-one evaluation*), penilaian dalam kelompok kecil (*small group evaluation*), serta uji lapangan (*field test*):

a. *Expert Review* (Pakar)

Rancangan yang disusun diuji melalui proses evaluasi dengan melibatkan dua orang ahli yang memiliki kompetensi spesifik di bidang biologi dan pendidikan. Proses validasi tersebut bertujuan untuk menilai tingkat kesesuaian serta relevansi rancangan yang dikembangkan, dengan hasil penilaian para validator dapat diuraikan sebagai berikut.

Tabel 2. Hasil relevansi dua validator

Nomor soal	Penilaian validator		Relevan skor
	Validator 1	Validator 2	
1	3	3	D
2	4	3	D
3	3	3	D
4	3	4	D
5	4	4	D
6	4	3	D
7	3	4	D
8	4	4	D
9	4	3	D
10	4	3	D
11	3	3	D
12	3	3	D
13	3	3	D
14	4	4	D
15	3	4	D

Nomor soal	Penilaian validator		Relevan skor
	Validator 1	Validator 2	
16	4	4	D
17	3	3	D
18	4	3	D
19	3	3	D
20	3	3	D
21	4	3	D
22	4	3	D
23	3	4	D
24	3	3	D
25	3	4	D
26	3	3	D
27	4	3	D
28	3	3	D
29	3	3	D
30	3	3	D

b. *One to one*

Proses pelaksanaan tahap ini dilaksanakan pada 21 Agustus 2023 bertempat di ruang guru MA Madani Alauddin Pao-Pao. Siswa mengerjakan soal sebanyak 30 nomor dengan waktu 45 menit dalam bentuk media cetak berupa kertas dan pengerjaan soal tersebut dalam pengawasan guru dan peneliti.

c. Small Group

Pada fase uji coba kelompok kecil, Prototype II yang telah disusun sebelumnya diaplikasikan kepada enam peserta didik yang mewakili tiga tingkatan kemampuan akademik, yaitu tinggi, menengah, dan rendah. Proses pelaksanaan uji coba tersebut dilaksanakan pada tanggal 11 September 2023 di ruang kelas XI MIA 2, MA Madani Alauddin Pao-Pao. Selama proses berlangsung, peserta didik mengerjakan sebanyak 30 butir soal dalam kurun waktu 45 menit menggunakan media cetak berbentuk lembar kertas, dengan pengawasan langsung dari guru serta peneliti.

d. Field Test

Pada fase ini, Prototype III yang telah melalui rangkaian penyempurnaan secara sistematis diperkenalkan untuk evaluasi lebih lanjut diimplementasikan dalam uji coba berskala kecil dengan melibatkan 29 peserta didik kelas XI MIA 3 sebagai subjek penelitian. Instrumen penelitian berupa butir soal yang termuat dalam *Prototype* III digunakan sebagai bahan pengujian. Setiap butir soal merupakan hasil revisi yang telah melalui serangkaian analisis secara kuantitatif dan kualitatif, meliputi uji validitas, reliabilitas, daya pembeda, efektivitas pengecoh, serta tingkat kesulitan, guna memastikan mutu, konsistensi, dan kelayakan instrumen yang diterapkan dalam penelitian ini.

Pelaksanaan tahap ini dilaksanakan pada tanggal 1 Februari 2024 di kelas XI MIA 3. Alokasi waktu selama 45 menit untuk menyelesaikan 25 butir soal yang disajikan melalui platform *Liveworksheet*, yang diakses menggunakan tautan yang telah dibagikan oleh pendidik. Setelah waktu pengerjaan berakhir, peserta didik mengirimkan hasil pekerjaan mereka dengan menekan tombol submit pada aplikasi tersebut.

Berdasarkan temuan penelitian, dilakukan analisis mendalam yang mencakup berbagai dimensi, termasuk validitas instrumen, reliabilitas pengukuran, kemampuan membedakan item, tingkat kesulitan soal, serta efektivitas distraktor pada butir instrumen. Uraian berikut menyajikan hasil temuan dari analisis terhadap instrumen penilaian yang telah dikembangkan:

1) Validitas

Berdasarkan analisis terhadap 26 peserta didik diperoleh temuan dari 25 soal yang diujikan, sebanyak 11 butir (44%) memenuhi standar validitas, sementara 14 butir (56%) tidak mencapai kriteria tersebut. Hasil ini mengindikasikan bahwa proporsi butir soal yang valid tergolong lebih rendah dibandingkan dengan butir yang dinyatakan tidak valid. Hasil tersebut juga memperlihatkan bahwa tingkat validitas empiris instrumen penilaian yang dikembangkan tidak sejalan dengan hasil penelitian Oktanin dan Sukirno (2015), yang melaporkan bahwa dari 50 butir soal yang disusun, jumlah butir valid lebih dominan dibandingkan yang tidak valid. Berdasarkan situasi tersebut, dilakukan analisis mendalam yang meliputi validitas, reliabilitas, kemampuan membedakan, tingkat kesulitan, serta efektivitas distractor, guna memperoleh pemahaman yang lebih menyeluruh mengenai mutu instrumen penilaian yang dikembangkan.

2) *Reliabilitas*

Analisis reliabilitas mengindikasikan bahwa instrumen penelitian ini memiliki tingkat keandalan yang luar biasa, ditunjukkan oleh koefisien reliabilitas sebesar 0,99, yang secara signifikan melampaui ambang batas standar KR-20. Instrumen tersebut terdiri atas 25 butir soal, dapat ditafsirkan bahwa instrumen evaluasi yang dikembangkan memiliki konsistensi internal yang sangat tinggi dan menunjukkan kesesuaian yang signifikan dengan tujuan serta konteks penelitian. Mengacu pada panduan Arif (2017), Instrumen dianggap memiliki tingkat reliabilitas yang tinggi apabila koefisien reliabilitasnya (r) mencapai atau melebihi 0,70.

3) Daya pembeda

Hasil analisis mengenai daya pembeda tiap butir soal menunjukkan bahwa sebanyak 9 butir (36%) termasuk dalam kategori “sangat rendah”, sedangkan 9 butir lainnya (36%) tergolong dalam kategori “rendah”, 6 butir (24%) termasuk kategori “sedang”, dan hanya 1 butir (4%) yang menunjukkan kategori “tinggi”. Hasil analisis menunjukkan bahwa tidak terdapat satupun item soal yang mencapai kriteria pada kategori “sangat tinggi”, sehingga secara umum instrumen penilaian yang dikembangkan belum menunjukkan kinerja optimal

dalam membedakan kemampuan peserta didik. Jika dibandingkan dengan hasil penelitian Alwiyah (2020), instrumen ini masih berada di bawah standar.

4) Tingkat kesukaran

Analisis yang dilakukan terhadap tingkat kesulitan butir instrumen mengungkapkan bahwa tidak ada satupun butir yang tergolong dalam kategori “sukar” (0%). Sebanyak 4 butir, atau sekitar 16%, dikategorikan sebagai “sedang”, sedangkan mayoritas, yakni 21 butir atau 84% kategori “mudah”. Hasil menunjukkan kaitan dengan temuan Arif (2017), yang menyatakan bahwa instrumen penilaian berbentuk pilihan ganda umumnya memperlihatkan variasi tingkat kesulitan, dengan proporsi 28% soal mudah, 52% soal sedang, dan 20% soal sukar.

5) Daya pengecoh

Hasil analisis terhadap daya pengecoh setiap butir soal yang dikembangkan tidak terdapat soal dengan kategori “sangat buruk” (0%). Sebanyak 9 soal (36%) tergolong “buruk”, 1 soal (4%) diklasifikasikan sebagai “kurang memadai”, 12 soal (48%) masuk dalam kategori “baik”, dan 3 soal (12%) termasuk dalam kategori “sangat baik”. Temuan ini sejalan dengan hasil penelitian Alwiyah (2020) yang mengungkapkan bahwa instrumen evaluasi yang dirancang menunjukkan tingkat efektivitas pengecoh yang signifikan, dengan persentase total mencapai 80%, di mana 33% masuk dalam kategori “cukup” dan 47% tergolong “sedang”.

KESIMPULAN

Berdasarkan temuan penelitian, dapat diinterpretasikan bahwa pengembangan instrumen ini dilaksanakan dengan merujuk pada kerangka model pengembangan yang telah diusulkan oleh Tessmer melalui pendekatan *Formative Research*, yang meliputi tiga tahapan utama. Tahap pertama, *preliminary*, difokuskan pada pengumpulan literatur dan referensi yang relevan guna memperkuat landasan teoretis penelitian. Pada tahap kedua, yaitu *self-evaluation*, difokuskan pada pemeriksaan komprehensif terhadap kurikulum, profil dan karakteristik peserta didik, substansi materi pembelajaran, serta desain awal instrumen evaluasi yang akan dikembangkan. Tahap *prototyping* selanjutnya meliputi rangkaian prosedur sistematis, yang mencakup penilaian oleh para pakar (*expert review*), pengujian individual (*one-to-one*), diskusi kelompok terbatas (*small group*), serta uji coba lapangan

(*field test*). Pada fase one-to-one dan small group, instrumen awal terdiri dari 30 butir soal. Namun, melalui proses evaluasi kritis, lima butir dinyatakan tidak memenuhi kriteria dan dihapus, sehingga tersisa 25 butir yang selanjutnya diterapkan dalam uji coba lapangan. Hasil analisis terhadap instrumen penilaian berbasis *LiveWorksheet* pada materi sistem pernapasan menunjukkan bahwa 44% butir dinyatakan valid dan 56% tidak valid, dengan koefisien reliabilitas sebesar 0,99 yang tergolong sangat tinggi. Berdasarkan kriteria *item analysis*, daya pembeda butir menunjukkan 36% sangat rendah, 36% rendah, 6% cukup, dan 4% baik; tingkat kesukaran soal terdiri atas 84% kategori mudah dan 16% kategori sedang; sedangkan daya pengecoh menunjukkan 36% rendah, 4% kurang baik, 48% baik, dan 12% sangat baik. Secara keseluruhan, dari hasil evaluasi tersebut, 11 butir soal dinyatakan layak digunakan, sementara 14 butir lainnya dinilai tidak layak.

DAFTAR PUSTAKA

- Andriyani, N., Hanafi, Y., Safitri, I. Y. B., & Hartini, S. (2020). Penerapan model problem based learning berbantuan LKPD Liveworksheet untuk meningkatkan keaktifan mental siswa pada pembelajaran tematik kelas V A SD Negeri Nogopuro. In *Prosiding Pendidikan Profesi Guru, Universitas Indonesia, Depok*. 122-130 <http://eprints.uad.ac.id/id/eprint/21216>
- Arif, W. P. (2017) "Pengembangan Tes Kognitif Pada Materi Sistem Pernapasan", *Jurnal Biotek*, 5(2), 108-122.
- Ariskasari, V., & Sulisworo, D. (2021), Developing the Interactive Worksheet Supported by Simulation and Liveworksheet on Physics Learning, *Proceedings of WRS International Conference*. 69-73.
- Febriana, R. (2019), *Evaluasi Pembelajaran*, Jakarta: Bumi Aksara.
- Fitriani, N. (2021). Analisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh soal pelatihan kewaspadaan kegawatdaruratan maternal dan neonatal, *Paedagoria: Jurnal Kajian, Penelitian dan Pengembangan Kependidikan*, 12(2), 199-205. <http://journal.ummat.ac.id/index.php/paedagoria>
- Hidayah, A. N., Winingsih, P. H., & Amalia, A. F. (2020). "Development Of Physics E-LKPD (Electronic Worksheets) Using 3D Pageflip Based on Problem Based Learning on Balancing And Rotation Dynamics," *Jurnal Ilmiah Pendidikan Fisika-COMPTON*, 7 (2), 36-43.
- Retnawati, H. (2016). Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, dan Psikometrian), (Cet I; Yogyakarta: Parama Publishing)
- Setemen, K. (2018). Pengembangan dan pengujian validitas butir instrumen kecerdasan logis-matematis, *Jurnal Pendidikan Teknologi dan Kejuruan*, 15(2). <https://ejournal.undiksha.ac.id/index.php/JPTK/issue/view/851>
- Sylvia, I., Anwar, S., & Khairani, K. (2019). Pengembangan Instrumen Penilaian Autentik Berbasis Pendekatan Authentic Inquiry Learning Pada Mata Pelajaran Sosiologi Di Sekolah Menengah Atas, *Jurnal Socius: Journal of Sociology Research and Education*, 6(2), 103-120. <https://doi.org/10.24036/scs.v6i2.162>

- Umni, A., Syamsudduha, S., & Ali, A.(2020). Pengembangan Instrumen Penilaian Kognitif Berbasis Google Form Pada Materi Pelajaran Biologi, *Jurnal Al-ahya*,2 (3), 129-143.
- Yusuf, M. (2015). Asesmen dan Evaluasi Pendidikan (Pilar Penyedia Informasi dan Kegiatan Pengendalian Mutu Pendidikan), Edisi Pertama. Cet. I; Jakarta: Prenadamedia Group.