

Pengembangan instrumen evaluasi mata pelajaran biologi SMA kelas X di SMA Panjura Kota Malang

Tania Puspa Chandra, Triastono Imam Prasetyo*, Agung Witjoro

Universitas Negeri Malang, Jl. Semarang No. 5 Malang, Jawa Timur, Indonesia

*Penulis korespondensi, Surel: triastono.imam.fmipa@um.ac.id

Paper received: 02-01-2021; revised: 15-01-2021; accepted: 30-01-2021

Abstract

2013 curriculum is a competency-based curriculum. An evaluation instrument is used to measure the students' competence in the learning. The evaluation instruments developed in this research were 34 multiple choice questions with a single answer from 5 answer choices, questions with description answer, and 3 performance assessments. The multiple choice question with a single answer refers to the question's indicator that developed from the competency's indicator. Basic competency is a reference for the formulation of competency's indicators. Performance assessment refers to basic competencies. The objectives of the research are: 1) develop the multiple choice questions with a single answer of basic competencies 3.4, 3.5, and 3.6, 2) develop the performance assessment of basic competencies 4.4, 4.5, and 4.6. The research used ADDIE research model which consists of 5 stages (1) analysis, (2) planning, (3) development, (4) implementation, (5) evaluation. Data collection is obtained from logical validation by evaluation experts, material experts, and field experts, and empirical validation which includes difficulty level, differentiation, reliability and distractor analysis.

The result of the research is the multiple choice questions with a single answer is known that the validity of the questions' construct is compatible with the question's indicator. As for the validity of the content, known that the test material is compatible with the question's indicator, Empirical validation Of results on the small group trials results on multiple choice questions with a single answer to the difficulty level obtained the ratio of easy problem-moderate-difficult to achieve 4:4:2, the difference is obtained the largest percentage of 52,95% good category. For the analysis of distractor, the choice of answers outwit percentage is 41.1796. In addition, revisions were made for small group trials, Results of large group test product of difficulty level obtained the ratio of easy problem:medium:difficult to reach 4:4:2. Different power obtained the largest percentage of 41.17 percent good category. Reliability is very high that is 0.855, For the analysis of distractor, the choice of answers outwit percentage is 82.36 percent. Revisions are also made based on empirical analysis results of large groups.

Keywords: assessment; High school biology; basic competencies.

Abstrak

Kurikulum 2013 merupakan kurikulum berbasis kompetensi. Untuk mengukur kompetensi siswa dalam pembelajaran digunakan instrumen evaluasi. Instrumen evaluasi yang dikembangkan dalam penelitian ini adalah 34 soal pilihan ganda jawaban tunggal dengan 5 pilihan jawaban, soal uraian dan 3 asesmen kinerja. Soal pilihan ganda jawaban tunggal mengacu pada indikator soal yang dikembangkan dari indikator kompetensi. Kompetensi dasar merupakan acuan perumusan indikator kompetensi. Asesmen kinerja mengacu pada kompetensi dasar. Tujuan penelitian yaitu: 1) merumuskan soal pilihan ganda jawaban tunggal dari kompetensi dasar 3.4, 3.5, dan 3.6, (2) merumuskan asesmen kinerja dari kompetensi dasar 4.4, 4.5 dan 4.6. Model penelitian menggunakan ADDIE yang terdiri dari 5 tahap yaitu: (1) analisis, (2) perencanaan, (3) pengembangan, (4) implementasi, (5) evaluasi. Pengumpulan data diperoleh dari hasil validitas logis oleh ahli evaluasi, ahli materi, dan ahli lapangan, dan validitas empiris meliputi tingkat kesukaran, daya beda, reliabilitas dan analisis pengecoh.

Hasil penelitian yaitu soal pilihan ganda jawaban tunggal diketahui validitas konstruk soal sudah sesuai dengan indikator soal, validitas isi diketahui materi tes sesuai dengan indikator soal. Hasil

validitas empiris pada uji coba kelompok kecil soal pilihan ganda jawaban tunggal untuk tingkat kesukaran diperoleh rasio soal mudah-sedang:sukar mencapai 4:4:2, daya beda diperoleh persentase terbesar yaitu 52,956 kategori baik. Analisis pengecoh berfungsi sebanyak 41,1796. Berdasarkan uji coba kelompok kecil dilakukan revisi. Setelah dilakukan validitas logis dan empiris dilakukan revisi untuk digunakan uji coba kelompok besar. Hasil uji coba kelompok besar tingkat kesukaran soal diperoleh rasio soal mudahssedangssukar mencapai 4:4:2. Daya beda diperoleh persentase terbesar 41,1796 kategori baik. Reliabilitas tergolong sangat tinggi yaitu 0,855, Analisis pengecoh berfungsi sebanyak 82,366. Revisi dilakukan berdasarkan hasil analisis validasi empiris kelompok besar.

Kata kunci: evaluasi; biologi SMA; kompetensi dasar

1. Pendahuluan

Karakteristik Kurikulum 2013 yaitu isi atau konten kurikulum adalah kompetensi yang dinyatakan dalam bentuk Kompetensi Inti (KI) mata pelajaran dan dirinci lebih lanjut ke dalam Kompetensi Dasar (KD). Kompetensi inti pada kurikulum 2013 merupakan tingkat kemampuan untuk mencapai standar kompetensi lulusan yang harus dimiliki siswa pada setiap tingkat kelas. KD merupakan kemampuan dan materi pembelajaran minimal yang harus dicapai siswa untuk suatu mata pelajaran pada masing-masing satuan pendidikan yang mengacu pada kompetensi inti (Permendikbud No 24, 2016). Indikator merupakan penjabaran dari KD agar kompetensi lebih jelas dan lebih rinci. Indikator kompetensi adalah perilaku yang dapat diukur dan diobservasi untuk menunjukkan ketercapaian kompetensi dasar tertentu yang menjadi acuan penilaian mata pelajaran. Indikator kompetensi selanjutnya akan dikembangkan menjadi indikator penilaian agar dapat digunakan untuk mengukur kompetensi siswa.

Salah satu penekanan di dalam kurikulum 2013 adalah penilaian autentik. Proses penilaian dalam pembelajaran harus dapat mencakup penilaian seluruh kompetensi (Permendikbud Nomor 23 Tahun 2016). Sistem penilaian pada kurikulum 2013 menuntut guru untuk melakukan penilaian baik saat proses maupun di akhir proses pembelajaran (Noverina *dkk*, 2014). Penilaian pada saat proses pembelajaran dilakukan dengan menggunakan asesmen kinerja, sedangkan pada akhir proses pembelajaran dilakukan dengan tes. Asesmen kinerja dilakukan guru berdasarkan model pembelajaran yang digunakan, hal tersebut tercantum pada RPP (Rencana Pelaksanaan Pembelajaran) (Sundari, 2014).

Pengembangan instrumen evaluasi untuk mengukur kompetensi siswa diperlukan, karena tercapainya tujuan pembelajaran dapat dilihat dari kompetensi yang dikuasai oleh siswa. Alat yang digunakan sebagai sarana untuk mengukur kompetensi siswa adalah tes dan non tes. Tes sebagai alat ukur dalam pendidikan mempunyai peranan yang sangat penting dalam mengukur kompetensi siswa. Tes yang baik perlu memperhatikan validitas logis dan validitas empiris (Widyantoro *dkk*, 2009).

Berdasarkan hasil wawancara terhadap guru biologi di SMA PANJURA Kota Malang tanggal 23 November 2016. Berkaitan dengan instrumen evaluasi, di SMA PANJURA Kota Malang dalam pembuatan soal Ulangan Harian belum dilakukan validasi logis dan analisis butir soal dan belum dilengkapi dengan asesmen kinerja dengan rubrik yang jelas.

Wawancara juga dilakukan terhadap siswa SMA PANJURA Kota Malang yang berjumlah enam orang siswa, menyatakan bahwa saat ulangan harian terdapat soal yang tidak sesuai dengan materi yang disampaikan pada pembelajaran. Hal ini dimungkinkan terjadi jika guru

tidak melakukan validitas logis. Berdasarkan latar belakang dan hasil wawancara tersebut maka dilakukan pengembangan instrumen evaluasi mata pelajaran biologi SMA kelas X KD 3.4, 3.5, 3.6, 4.4, 4.5, dan 4.6. Pengembangan instrumen evaluasi ini bertujuan untuk Merumuskan soal pilihan ganda jawaban tunggal dan asesmen kinerja berdasarkan indikator dari KD 3.4, 3.5, 3.6, 4.4, 4.5, dan 4.6.

2. Metode

1. Rancangan Penelitian

Model pengembangan instrumen penilaian pada penelitian ini memodifikasi model pengembangan ADDIE yang dikembangkan oleh Branch (2009) yang terdiri dari 5 tahap yaitu langkah sebagai berikut: (1) *Analyze*, (2) *Design*, (3) *Develop*, (4) *Implement*, dan (5) *Evaluate*. Model ADDIE dipilih karena pada model tersebut memberikan kesempatan pada pengembang untuk melakukan revisi (evaluasi) berkali-kali selama hal tersebut dibutuhkan. Revisi yang dapat dilaksanakan terus menerus ini memberikan kesempatan kepada pengembang agar menghasilkan instrumen evaluasi yang lebih baik.

Instrumen evaluasi yang dikembangkan dalam penelitian ini adalah 3 asesmen kinerja, 34 soal pilihan ganda jawaban tunggal dengan 5 pilihan jawaban dan 5 soal uraian. Pada pengembangan instrumen evaluasi asesmen kinerja dilakukan revisi berdasarkan masukan dan saran dari validator ahli evaluasi yaitu pada fase pengembangan (*develop*). Pengembangan instrumen evaluasi soal pilihan ganda jawaban tunggal dilakukan revisi berdasarkan hasil validasi dari ahli evaluasi, ahli materi dan ahli lapangan pada fase pengembangan serta revisi dilakukan berdasarkan hasil uji coba kelompok kecil dan uji coba kelompok besar pada fase pengembangan dan implementasi.

2. Analisis Data

2.1. Analisis Validasi Logis

Analisis validitas logis dilakukan oleh 3 ahli, yaitu ahli evaluasi, ahli materi, dan ahli lapangan (guru). Hasil validasi logis yang dilakukan oleh ahli evaluasi, ahli materi dan ahli lapangan (guru) dihitung persentasenya. Persentase yang dihitung adalah jumlah jawaban ya dan tidak pada setiap aspek yang diukur. Selanjutnya angka persentase pada seluruh aspek dihitung reratanya. Hasil persentase dibandingkan dengan kriteria kelayakan validitas pada Tabel 1.

Tabel 1. Kriteria Kelayakan Validitas

Kriteria Kelayakan	Tingkat Kelayakan
80%-100%	Layak/ valid
60%-79%	Cukup layak/ cukup valid
50%-59%	Kurang layak/ kurang valid
0%-49%	Tidak layak/ tidak valid

(Arikunto, 2012)

Berdasarkan hasil analisis dapat diketahui soal mana yang perlu diperbaiki kualitasnya. Perbaikan soal disesuaikan berdasarkan saran dari validator. Analisis data dilakukan secara kualitatif.

2.2. Analisis Validasi Empiris

Analisis validasi empiris dilakukan setelah uji coba kelompok kecil dan uji coba kelompok besar, meliputi tingkat kesukaran, daya beda, reliabilitas, dan analisis pengecoh. Analisis dilakukan dengan menggunakan program ITEMAN. Sedangkan analisis pengecoh dilakukan secara manual menggunakan *Microsoft office excel*. Data analisis validasi empiris dilakukan secara kuantitatif.

2.2.1. Tingkat Kesukaran

Berdasarkan hasil tingkat kesukaran butir soal dapat diketahui proporsi soal sulit, sedang, dan mudah. Penelitian ini menggunakan proporsi soal sukar, sedang, dan mudah dengan rasio 3: 4: 3 (Sudjana, 2012). Jika ditemukan proporsi yang tidak sesuai dengan rasio 3: 4: 3 maka dilakukan revisi terhadap soal sehingga rasio akan terpenuhi. Klasifikasi tingkat kesukaran terdapat pada Tabel 2.

Tabel 2 Klasifikasi Tingkat Kesukaran

No	Rentang	Kriteria
1.	0.00 - 0.30	Soal sukar
2.	0.30 - 0.70	Soal sedang
3.	0.70 - 1.00	Soal mudah

(Arikunto, 2009)

2.2.2. Daya Beda

Daya beda atau daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang kurang pandai (berkemampuan rendah) (Arikunto, 2009). Artinya, bila soal tersebut diberikan kepada anak yang mampu, hasilnya menunjukkan prestasi yang tinggi, dan bila diberikan pada siswa yang lemah, hasilnya rendah. Klasifikasi daya beda terdapat pada Tabel 3.

Tabel 3. Klasifikasi Daya Pembeda

No	Rentang	Kriteria
1.	0.00 - 0.20	Jelek
2.	0.20 - 0.40	Cukup
3.	0.40 - 0.70	Baik
4.	0.70 - 1.00	Baik sekali
5.	Negatif	Tidak baik

(Arikunto, 2009)

2.2.3. Reliabilitas

Reliabilitas berhubungan dengan masalah kepercayaan. Suatu tes dapat dikatakan mempunyai taraf kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap. Instrumen baik adalah instrumen yang dapat dengan ajeg memberikan data yang sesuai dengan kenyataan. Ajeg atau tetap tidak selalu harus sama tetapi mengikuti perubahan secara ajeg (Arikunto, 2009). Artinya suatu tes yang sama diberikan kepada siswa berlainan

waktu maka siswa tersebut akan memperoleh hal yang sama. Klasifikasi reliabilitas terdapat pada Tabel 4.

Tabel 4. Klasifikasi Reliabilitas

No	Rentang	Kriteria
1.	0,81-1,00	reliabilitas sangat tinggi
2.	0,61-0,80	reliabilitas tinggi
3.	0,41-0,60	reliabilitas sedang
4.	0,21-0,40	reliabilitas rendah
5.	0,00-0,20	reliabilitas sangat rendah (tidak reliabel)

(Arikunto, 2009)

2.2.4. Analisis Pengecoh

Menurut Sunarti dan Rahmawati (2014) penentuan apakah suatu item memiliki efektivitas pengecoh yang baik atau tidak sebagai berikut.

1. Pada alternatif jawaban betul, kelompok atas harus memilih lebih banyak jawaban betul.
2. Pada alternatif jawaban pengecoh, kelompok bawah harus memilih lebih banyak dari pada kelompok atas.
3. Semua alternatif jawaban harus terpilih minimal 5% pengikut tes.
4. Omit (jawaban kosong tidak menjawab) baik jika dipilih tidak lebih dari 10% pengikut tes.
5. Jika syarat 1 sampai dengan 4 tidak terpenuhi, maka pengecoh dikatakan tidak efektif

3. Hasil dan Pembahasan

Hasil penilaian validasi oleh validator dalam pengembangan instrumen evaluasi diperoleh saran dan masukan dari validator ahli evaluasi antara lain: (1) Pilihan jawaban untuk soal nomor 6, 9, dan 10 perlu diperbaiki agar lebih homogen dan logis serta panjang rumusan pilihan jawabannya relatif sama. (2) soal sebaiknya dilengkapi dengan tingkatan kognitif. (3) rubrik penilaian pada asesmen produk poster perlu diperbaiki karena kurang jelas. Saran dan masukan dari validator yaitu ahli materi antara lain: (1) Gambar pada soal nomor 5, 13 dan 31 kurang jelas sehingga perlu perbaikan, (2) Gambar untuk soal nomor 18 pilihan gambar 5 kurang jelas sehingga perlu perbaikan, (3) Soal nomor 23 kata protozoa diganti dengan protista, (4) Soal nomor 24 sebaiknya pokok soalnya diganti, dan (5) Soal nomor 32 kata organisme diganti dengan protista. Saran dan masukan dari validator ahli lapangan (guru) antara lain materi dan kata kerja operasional pada soal nomor 2, 3, 5, 9, 13, 22, 24, 28, 33, dan 34 belum sesuai dengan indikator dan dibuat lebih spesifik. Saran dan masukan dari validator digunakan untuk merevisi instrumen evaluasi yang dikembangkan agar dapat dilanjutkan pada tahap uji coba kelompok kecil.

Sebelum diuji cobakan instrumen evaluasi yang dikembangkan selanjutnya akan dianalisis validasi logis (validitas isi dan konstruk). Berdasarkan hasil analisis dapat disimpulkan bahwa butir soal sudah mengacu pada semua materi yang telah ada artinya secara validitas isi instrumen tersebut sudah valid. Berdasarkan hasil analisis validasi konstruk dapat disimpulkan bahwa kata kerja dalam soal sesuai dengan kata kerja pada

indikator soal. Hal ini berarti soal telah memiliki validitas konstruk yang baik. Menurut Arikunto (2009) suatu tes dapat dikatakan baik jika memenuhi persyaratan validitas, reliabilitas. Berdasarkan hasil analisis validitas logis instrumen evaluasi sebanyak 34 butir soal sudah valid karena telah memenuhi validitas isi maupun konstruk. Setelah uji coba kelompok kecil, dilakukan analisis validitas empiris menggunakan ITEMAN. Berdasarkan analisis tingkat kesukaran instrumen evaluasi dapat dilihat pada Tabel 5.

Tabel 5. Persentase Analisis Tingkat Kesukaran Uji Coba Kelompok Kecil

Kategori	Jumlah	Persentase	Nomor Soal
Mudah	13	38,23%	1, 2, 4, 5, 6, 8, 9, 15, 18, 22, 24, 26, 34
Sedang	14	41,18%	3, 7, 10, 12, 14, 16, 17, 19, 20, 21, 27, 28, 29, 33
Sukar	7	20,59%	11, 13, 23, 25, 30, 31, 32
Total	34	100%	34

Berdasarkan tabel 5 diantara soal mudah, sedang dan sukar, butir soal pada kategori sedang memiliki jumlah dan persentase terbanyak yaitu 14 soal dengan persentase 41,18% dari total 34 soal, sehingga soal memiliki kategori baik. Rasio soal yang dihasilkan pada uji coba kelompok kecil soal mudah: sedang: sukar yaitu 4:4:2. Hal ini belum sesuai dengan rasio 3:4:3 (Sudjana, 2012). Untuk mencapai proporsi 3:4:3 dilakukan revisi soal dengan mengubah beberapa soal sedang menjadi soal mudah dan sukar.

Hasil analisis daya beda instrumen evaluasi dapat dilihat pada Tabel 6.

Tabel 6. Persentase Analisis Daya Beda Uji Coba Kelompok Kecil

Kategori	Jumlah	Persentase
Negatif	-	-
Jelek	5	14,70%
Cukup	4	11,76%
Baik	18	52,95%
Baik sekali	7	20,59%
Total	34	100%

Berdasarkan tabel 6 hasil persentase terbanyak yaitu kategori baik dengan jumlah 18 soal dan persentase 52,95%, dari total 34 soal dengan demikian soal mempunyai daya beda sangat baik. Artinya soal dapat membedakan siswa yang berkemampuan tinggi dan berkemampuan rendah (Arikunto, 2009). Berdasarkan analisis pengecoh instrumen evaluasi dapat dilihat pada Tabel 7.

Tabel 7. Persentase Analisis Pengecoh Uji Coba Kelompok Kecil

Kategori	Jumlah	Persentase
Berfungsi	14	41,17%
Tidak Berfungsi	20	58,83%
Total	34	100%

Berdasarkan tabel 7 analisis pengecoh pada uji coba kelompok kecil menghasilkan pilihan jawaban kategori berfungsi 14 soal dengan persentase sebesar 41,17% dengan

demikian soal mempunyai analisis pengecoh yang cukup (Sudjana, 2012). Setelah uji coba kelompok besar, dilakukan analisis validitas empiris menggunakan ITEMAN. Berdasarkan analisis tingkat kesukaran instrumen evaluasi berdasarkan hasil analisis uji coba kelompok besar dapat dilihat pada Tabel 8.

Tabel 8. Persentase Analisis Tingkat Kesukaran Uji Coba Kelompok Besar

Kategori	Jumlah	Persentase	Nomor Soal
Mudah	12	35,30%	1, 2, 4, 5, 6, 9, 12, 18, 19, 24, 29, 34
Sedang	13	38,23%	3, 8, 10, 14, 16, 17, 20, 21, 23, 27, 28, 30, 31
Sukar	9	26,47%	7, 11, 13, 15, 22, 25, 26, 32, 33
Total	34	100%	34

Berdasarkan tabel 8 diantara soal mudah, sedang dan sukar, butir soal pada kategori sedang memiliki jumlah dan persentase terbanyak yaitu 13 soal dengan persentase 38,23% dari total 34 soal sehingga soal memiliki kategori baik. Rasio soal menunjukkan rasio soal mudah: sedang: sukar yaitu 4:4:2 Hal ini masih belum sesuai dengan rasio 3:4:3. Untuk mencapai proporsi soal yang seimbang yaitu 3:4:3 dilakukan revisi soal dengan memperhatikan hasil data tingkat kesukaran dan daya beda soal (Sudjana, 2012). Hasil analisis daya beda uji coba kelompok besar terdapat pada Tabel 9.

Tabel 9. Persentase Analisis Daya Beda Uji Coba Kelompok Besar

Kategori	Jumlah	Persentase
Negatif	-	-
Jelek	4	11,76%
Cukup	6	17,64%
Baik	14	41,17%
Baik sekali	10	29,43%
Total	34	100%

Berdasarkan tabel 9 hasil persentase terbanyak yaitu 14 soal kategori baik dengan persentase 41,17%, dari total 34 soal dengan demikian soal mempunyai daya beda sangat baik. Artinya soal dapat membedakan siswa yang berkemampuan tinggi dan berkemampuan rendah (Arikunto, 2009). Berdasarkan analisis pengecoh instrumen evaluasi hasil analisis uji coba kelompok besar dapat dilihat pada Tabel 10.

Tabel 10. Persentase Analisis Pengecoh Uji Coba Kelompok Besar

Kategori	Jumlah	Persentase
Berfungsi	28	82,36%
Tidak Berfungsi	6	17,64%
Total	34	100%

Berdasarkan Tabel 10 analisis pengecoh pada uji coba kelompok besar menghasilkan pilihan jawaban yang berfungsi 28 soal dengan persentase sebesar 82,36% dengan demikian mempunyai analisis pengecoh yang baik. Pengecoh yang baik jika pilihan jawaban pengecoh dalam rentang 75%-100% (Sudijono, 2011). Revisi terhadap pengecoh dilakukan setelah uji coba kelompok besar. Hasil analisis reliabilitas menggunakan program ITEMAN

menunjukkan bahwa soal pilihan ganda jawaban tunggal yang dikembangkan memiliki nilai reliabilitas sebesar 0,855 artinya instrumen tes soal pilihan ganda mempunyai tingkat reliabilitas yang sangat tinggi. Spesifikasi dari produk yang dikembangkan terdapat pada Tabel 10.

Tabel 11. Spesifikasi Produk yang Dikembangkan

No	Identifikasi produk	Penjelasan
1.	Jenis	Instrumen Evaluasi
2.	Bentuk soal	Pilihan ganda jawaban tunggal dan asesmen kinerja
3.	Nama	Penilaian kognitif dan psikomotor
4.	Tujuan	Untuk mengetahui kompetensi siswa
5.	Ranah Kognitif	C1 (mengingat), C2(memahami/ mengerti), C3 (mengaplikasi-kan) , C4 (menganalisis)
6.	Jumlah butir soal	34
7.	Karakteristik Validitas butir soal Tingkat kesukaran Daya beda Analisis pengecoh Reliabilitas	Validitas konstruk sesuai dengan tujuan pembelajaran, validitas isi sesuai dengan materi Tingkat kesukaran meliputi soal yang sukar 26,47% (9 soal), soal sedang 38,23% (13 soal) dan soal mudah 35,30% (12 soal) Daya beda didominasi kategori baik sebesar 41,17% dengan rentangan 0,40-0,70 Analisis pengecoh yang berfungsi sebesar 82,36% sedangkan yang tidak berfungsi 17,64% Reliabilitas sebesar 0,855 yang berarti sangat tinggi

Instrumen evaluasi yang dikembangkan adalah instrumen evaluasi dengan materi virus, bakteri dan protista kelas X semester ganjil. Instrumen evaluasi ranah psikomotor yang dikembangkan berbentuk asesmen kinerja, sedangkan instrumen evaluasi kognitif yang dikembangkan berbentuk soal pilihan ganda jawaban tunggal dengan jumlah 34 butir. Mempunyai validitas konstruk sesuai dengan tujuan pembelajaran, validitas isi sesuai dengan materi tingkat kesukaran soal.

4. Metode

Berdasarkan hasil pembahasan pengembangan instrumen evaluasi didapatkan kesimpulan sebagai berikut. (1) Dirumuskan soal pilihan ganda jawaban tunggal yang terdiri dari 5 pilihan jawaban sejumlah 34 soal dari Kompetensi Dasar 3.4, 3.5, dan 3.6. (2) Dirumuskan 3 asesmen kinerja yang tersiri dari asesmen produk, asesmen laporan dan asesmen presentasi dari Kompetensi Dasar 4.4, 4.5, dan 4.6. Berdasarkan hasil pembahasan, saran yang dapat diberikan untuk peneliti lain sebagai berikut. Saran yang dapat diberikan yaitu instrumen ini bisa digunakan oleh guru pelajaran Biologi namun perlu mengkaji pada beberapa soal karena perbaikan soal harus dilakukan berulang-ulang agar mendapatkan soal yang valid.

Daftar Rujukan

- Anas, S. (2011). Pengantar evaluasi pendidikan. *Jakarta: Rajawali Pers.*
- Arikunto, S. (2012). Dasar-dasar evaluasi pendidikan edisi 2. *Jakarta: Bumi Aksara.*
- Arikunto, S. (1992). *Prosedur penelitian: Suatu pendekatan praktik.* Rineka cipta.
- Branch, R. M. (2009). *Instructional design: The ADDIE approach* (Vol. 722). Springer Science & Business Media.
- Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2016). Permendikbud Nomor 24 tahun 2016.
- Noverina, S., & Wiyono, K. (2014). Pengembangan Rubrik Penilaian Keterampilan dan Sikap Ilmiah Mata Pelajaran Fisika Kurikulum 2013 di Kelas X Sekolah Menengah Atas. *Jurnal Inovasi Dan Pembelajaran Fisika, 1*(2), 145-151.
- Nomor, P. (23). tahun 2016a tentang Standar Penilaian. *Jakarta: BSNP.*
- Sudjana, N. (2014). Penilaian hasil proses belajar mengajar.
- Sunarti, S. R. (2014). Penilaian Dalam Kurikulum 2013. *Yogyakarta: CV Andi Offset.*
- Sundari, S. (2014). Model pengembangan asesmen kinerja (performance assessment) mata pelajaran IPA berbasis nilai karakter di SMP kota Ternate Maluku Utara. *Jurnal Edubio Tropika, 2*(1).
- Widyantoro, D. (2009). Pengembangan soal tes pilihan ganda kompetensi sistem starter dan pengisian program keahlian teknik mekanik otomotif Kelas XII. *Jurnal Pendidikan Teknik Mesin, 9*(1).