

Mulianti dkk

Pengembangan Instrumen Penilaian
Autentik pada Materi Sistem Ekskresi di
Kelas XI SMA Negeri 1 Tellusiattinge

**PENGEMBANGAN INSTRUMEN PENILAIAN AUTENTIK PADA
MATERI SISTEM ESKRESI DI KELAS XI SMA NEGERI 1
TELLUSIATTINGE****DEVELOPMENT OF INSTRUMENTS IN THE MATERIALS AUTHENTIC
ASSESSMENT EXCRETION SYSTEM IN CLASS XI SMA Negeri 1
TELLUSIATTINGE****Mulianti¹, Yusminah Hala², A. Mushawwir Taiyeb³****¹Guru SMA Negeri 1 Tellusiattinge****^{2,3}Dosen Program Pascasarjana Universitas Negeri Makassar****Abstract**

This research is the development model adopted Borg and Gall conducted in SMA Negeri 1 Tellusiattinge District of Bone. Sampel study were 30 students of SMA Negeri 1 Tellusiattinge. This research was implemented in the second semester of the academic year 2015/2016. The results showed that the authentic assessment instruments that have been developed on the material excretory system is very valid in the category with a value of 3.78 for the attitude, skill and 3.64 to 3.58 for the cognitive tests. To test the reliability of the instrument, the assessment and cognitive tests that are in the category of reliable value 0.760 and 0.723, while the skills assessment instruments that are in the category of unreliability with a value of 0.633. Cognitive tests that have been developed have quality levels of difficulty: easy 33% category, the category was 45%, 23% difficult category; different power: category 3% excellent, 53% good category, a category quite 20%, 18% weak category, and the category no different power 8%; quality humbug: effective category 35% and 65% category ineffective.

Keywords: *Authentic, Assessment Instruments, Excretion System***Abstrak**

Penelitian ini merupakan penelitian pengembangan yang mengadopsi model Borg dan Gall yang dilaksanakan di SMA Negeri 1 Tellusiattinge Kabupaten Bone. Sampel penelitian adalah 30 siswa SMA Negeri 1 Tellusiattinge. Penelitian ini dilaksanakan pada semester genap tahun pelajaran 2015/2016. Hasil penelitian menunjukkan bahwa instrumen penilaian autentik yang telah dikembangkan pada materi sistem ekskresi berada pada kategori sangat valid dengan nilai 3,78 untuk sikap, 3,64 untuk keterampilan dan 3,58 untuk tes kognitif. Untuk uji reliabilitas instrumen penilaian sikap dan tes kognitif berada pada kategori reliabel dengan nilai 0,760 dan 0,723, sedangkan instrumen penilaian keterampilan berada pada kategori tidak reliabel dengan nilai 0,633. Tes kognitif yang telah dikembangkan memiliki kualitas tingkat kesukaran: kategori mudah 33%, kategori sedang 45%, kategori sukar 23%; daya beda: kategori baik sekali 3%, kategori baik 53%, kategori cukup 20%, kategori lemah 18%, dan kategori tidak ada daya beda 8%; kualitas pengecoh: kategori efektif 35% dan kategori tidak efektif 65%.

Kata Kunci : *Autentik, Instrumen penilaian, Sistem ekskresi***PENDAHULUAN**

Penilaian adalah proses pengumpulan informasi/ data tentang capaian pembelajaran peserta didik dalam aspek sikap, aspek pengetahuan, dan aspek keterampilan yang dilakukan secara terencana dan sistematis untuk memantau proses, kemajuan belajar, dan perbaikan hasil belajar dan evaluasi hasil belajar. Pencapaian hasil belajar peserta didik mencakup ujian nasional, ujian sekolah,

ujian mutu tingkat kompetensi, ujian tingkat kompetensi, ulangan akhir semester, ulangan tengah semester, ulangan harian, penilaian berbasis portofolio, penilaian diri dan penilaian autentik [1].

Asesmen dan keputusan-keputusan evaluatif yang dihasilkan haruslah akurat sehingga mampu mencegah kekeliruan pemahaman dan komunikasi oleh pihak-pihak terkait. Oleh karenanya diperlukan berbagai jenis asesmen yang secara bersama-sama

akan menghasilkan informasi evaluatif yang lengkap dan akurat. Beberapa tipe dan kategori asesmen diantaranya asesmen formal dan informal, asesmen kuantitatif dan kualitatif, evaluasi formatif dan sumatif, asesmen baku dan tak baku, asesmen rujukan norma dan asesmen rujukan kriteria, asesmen obyektif dan subyektif, asesmen tes dan asesmen alternatif [2].

Penilaian autentik adalah istilah yang diciptakan untuk menjelaskan berbagai metode penilaian alternatif yang memungkinkan siswa dapat mendemonstrasikan kemampuannya dalam menyelesaikan tugas-tugas dan menyelesaikan masalah. Sekaligus, mengekspresikan pengetahuan dan keterampilannya dengan cara mensimulasikan situasi yang dapat ditemui di dalam dunia nyata (di luar lingkungan sekolah). Dalam hal ini adalah simulasi yang dapat mengekspresikan prestasi (*performance*) peserta didik yang ditemui di dalam praktik dunia nyata.

Secara paradigmatis penilaian autentik memerlukan perwujudan pembelajaran autentik (*authentic instruction*) dan belajar autentik (*authentic learning*). Hal ini diyakini bahwa penilaian autentik lebih mampu memberikan informasi kemampuan peserta didik secara holistik dan valid. Penilaian autentik merupakan pendekatan dan instrumen penilaian yang memberikan kesempatan luas peserta didik untuk menerapkan pengetahuan, keterampilan dan sikap yang sudah dimilikinya dalam bentuk tugas: membaca dan meringkasnya, eksperimen, mengamati, survei, proyek, makalah, diskusi kelas dan membuat karangan [3].

Penilaian autentik memiliki relevansi kuat terhadap pendekatan ilmiah dalam pembelajaran sesuai dengan tuntutan Kurikulum 2013. Asesmen semacam ini mampu menggambarkan peningkatan hasil belajar peserta didik, baik dalam rangka mengobservasi, menalar, mencoba, membangun jejaring. Penilaian autentik cenderung fokus pada tugas-tugas kompleks atau kontekstual, memungkinkan peserta didik untuk menunjukkan kompetensi mereka dalam pengaturan yang lebih autentik. Karenanya, asesmen autentik sangat relevan dengan pendekatan tematik terpadu dalam pembelajaran, khususnya jenjang sekolah dasar atau untuk mata pelajaran yang sesuai.

Penilaian autentik bukanlah hal baru bagi guru sejak adanya pendekatan *Contekstual Teaching Learning* (CTL) pada Kurikulum Tingkat Satuan Pendidikan (KTSP) mereka sudah mengenali penilaian ini. Namun, pengaplikasiannya ditegaskan dalam penerapan kurikulum 2013. Meskipun sudah terbilang lama, namun masih ada guru yang belum memahami penilaian autentik. Guru masih dibayang-bayangi dengan penilaian tradisional. Akhirnya pembelajaran yang tercipta bukanlah pembelajaran yang autentik sehingga penilaian yang dihasilkan pun belum autentik.

SMA Negeri 1 Tellusiattinge merupakan salah satu sekolah yang telah menerapkan penilaian autentik. Kenyataan yang ada di sekolah ini, guru masih belum memahami penilaian autentik, sehingga belum mampu mengaplikasikan penerapannya dalam pembelajaran. Penilaian autentik akan menilai secara holistik kemampuan peserta didik dalam proses pembelajaran.

Penilaian autentik dapat dilihat dari tiga ranah yaitu sikap, keterampilan dan pengetahuan. Permasalahan yang ditemukan dalam ranah sikap adalah penilaian yang dilakukan difokuskan pada sikap peserta didik yang menonjol baik dan menonjol buruk dalam proses pembelajaran. Hal ini menjadikan guru tidak fokus dalam menilai karena dalam kompetensi inti ada dua hal yang mesti dinilai dalam ranah sikap yakni sikap spiritual dan sikap sosial. Jika hanya meninjau sikap yang menonjol baik dan menonjol buruk maka dikhawatirkan ada sikap yang tidak ternilai. Olehnya itu, Kompetensi Inti 1 (KI 1) dan Kompetensi Inti 2 (KI 2) mestinya diekstraksi untuk disesuaikan dengan materi yang akan dibahas dalam beberapa indikator sikap, sebab tidak semua sikap akan dinilai sekaligus dalam waktu yang bersamaan.

Penilaian kognitif dapat diketahui melalui jbaran indikator dari Kompetensi Inti 3 dan Kompetensi Dasarnya. Indikator yang dijabarkan dari setiap kompetensi dasar mengacu pada kata operasional dari revisi Taksonomi Bloom. Indikator pencapaian kompetensi dapat dilihat dari rencana pembelajaran setiap guru. Permasalahan yang terkait dengan kognitif adalah guru belum mampu menerjemahkan dengan baik keinginan dari kompetensi dasar yang akan dicapai. Biasanya guru menyusun indikator berdasarkan buku pegangan bukan menganalisis kurikulum. Permasalahan kedua

adalah dalam pembuatan tes kognitif tidak mengacu pada indikator pencapaian kompetensi, sehingga soal tes yang dibuat oleh guru tidak mengacu lagi pada kompetensi dasar. Akhirnya guru tidak mampu menilai kemampuan peserta didik secara holistik melalui tes tertulis, ditambah lagi dengan tes yang belum tervalidasi.

Penilaian keterampilan mengacu pada Kompetensi Inti 4 (KI 4) yang selanjutnya diturunkan pada kompetensi dasar dan indikator pencapaian kompetensi. Permasalahan dalam penilaian ini hampir sama dengan penilaian kognitif yakni guru belum terbiasa untuk menganalisis kurikulum pada saat akan menyusun indikator pencapaian kompetensi sehingga penilaian tidak sesuai dengan tuntutan kurikulum. Selain itu, dalam penilaian keterampilan nilai dapat diperoleh melalui kinerja, proyek maupun portofolio. Kenyataan yang ada guru masih sekedar menempelkan nilai pada hasil karya peserta didik tanpa menilai prosesnya. Guru bahkan tidak membuat rubrik penilaian sebagai acuan dalam melakukan penilaian.

Sehubungan dengan uraian tentang penerapan penilaian autentik di sekolah, maka seorang guru harus cerdas dan kreatif dalam merancang sebuah instrumen penilaian autentik yang didasari oleh pembelajaran yang bersifat autentik pula. Olehnya itu, diperlukan sebuah instrumen penilaian autentik yang mampu memberikan solusi untuk permasalahan tersebut. Akhir-akhir ini telah banyak dilakukan penelitian pengembangan (*Research and Development*) yang bertujuan untuk menghasilkan suatu produk atau merevisi suatu produk yang ada yang hasilnya dapat dipertanggungjawabkan. Penelitian pengembangan ini akan melalui prosedur perancangan, pengembangan, validasi, uji coba sampai menghasilkan suatu produk yang valid dan reliabel. Permasalahan terkait penilaian autentik yang terjadi di SMA Negeri 1 Tellusiattinge berada pada cakupan instrumen penilaian yang belum baik. Sebenarnya sudah ada dalam panduan penilaian yang dirilis oleh Direktorat Jenderal Pendidikan Dasar dan Menengah, namun dalam penerapannya mesti disesuaikan dengan kondisi elemen sekolah di daerah setempat. Olehnya itu, melalui penelitian pengembangan ini akan

dikembangkan sebuah instrumen penilaian autentik.

Menelaah beberapa hal yang berkaitan dengan latar belakang, maka rumusan masalah dalam penelitian ini adalah sebagai berikut: (1) Bagaimana kualitas instrumen penilaian autentik ditinjau dari segi validitas dan reliabilitas pada materi sistem ekskresi di kelas XI SMA Negeri 1 Tellusiattinge? (2) Bagaimana kualitas tes kognitif ditinjau dari karakteristik internal yang meliputi tingkat kesukaran, daya beda, dan kualitas pengecoh (*distractor*) pada materi sistem ekskresi di kelas XI SMA Negeri 1 Tellusiattinge?

Adapun tujuan yang ingin dicapai dari penelitian ini antara lain: (1) Untuk menguji kualitas instrumen penilaian autentik ditinjau dari segi validitas dan reliabilitas pada materi sistem ekskresi di kelas XI SMA Negeri 1 Tellusiattinge. (2) Untuk menguji kualitas tes kognitif ditinjau dari karakteristik internal yang meliputi tingkat kesukaran, daya beda, dan kualitas pengecoh (*distractor*) pada materi sistem ekskresi di kelas XI SMA Negeri 1 Tellusiattinge.

METODE PENELITIAN

Penelitian ini merupakan penelitian kualitatif dengan model penelitian pengembangan (*development research*). Penelitian mengadopsi model Borg and Gall yang mencakup tahap merancang, mengembangkan, validasi, dan uji coba dan penyebaran. Penelitian ini mengembangkan instrumen penilaian autentik pada materi sistem ekskresi. Penilaian autentik mencakup penilaian kompetensi sikap (jurnal observasi sikap), penilaian keterampilan (unjuk kerja). Penilaian pengetahuan (tes hasil belajar) hanya mengukur pengaruh atau hasil dari penilaian sikap dan keterampilan. Penelitian pengembangan instrumen penilaian autentik ini, dilakukan beberapa penyesuaian dalam desainnya. Desain penelitian ini meliputi proses merancang, mengembangkan, validasi, uji coba dan penyebaran.

Data validitas instrumen penilaian autentik diperoleh melalui validasi ahli, data reliabilitas dan karakteristik butir soal diperoleh melalui uji coba pada peserta didik S Kelas XI SMA Negeri 1 Tellusiattinge. Peserta didik berjumlah 30 orang. Penelitian ini dilaksanakan pada tanggal 28 Maret – 17 Mei 2016.

Instrumen penilaian yang digunakan yaitu format jurnal observasi sikap. Penilaian keterampilan terdiri dari instrumen penilaian

unjuk kerja. Masing-masing penilaian dilengkapi rubrik. Penilaian pengetahuan dibuat dalam bentuk tes tertulis pilihan ganda berjumlah 40 nomor.

Teknik analisis data yaitu, tes yang telah dibuat (draft awal) akan divalidasi oleh validator ahli kemudian direvisi, selanjutnya akan divalidasi oleh praktisi sebelum uji coba terbatas. Yang terdiri atas: (1) Analisis kevalidan format instrumen penilaian non tes: Instrumen penilaian sikap dan keterampilan merupakan instrumen non tes. Instrumen yang baik adalah instrumen yang valid dan reliabel. Instrumen yang valid harus mempunyai validitas internal dan eksternal. Validitas internal untuk instrumen non tes berupa validasi konstruk. Setelah pengujian konstruk dari ahli dan berdasarkan pengalaman empiris di lapangan selesai, maka diteruskan dengan uji coba instrumen. Jumlah anggota sampel yang digunakan sekitar 30 orang. Setelah data ditabulasikan, maka pengujian validitas konstruksi dilakukan dengan analisis faktor, yaitu dengan mengkorelasikan antar skor item instrumen dalam salah satu faktor, dan mengkorelasikan skor faktor dengan skor total [4]. (2) Analisis Kevalidan Format Instrumen Penilaian Kognitif, penilaian validator terhadap instrumen penilaian autentik mencakup tiga aspek yaitu: isi, konstruksi, dan bahasa, adapun analisis instrumen penilaian autentik pada materi sistem ekskresi untuk kelas XI MIA 2. (3) Analisis format instrumen tes: analisis format instrumen tes dilakukan untuk memperoleh tes yang valid dan reliabel. (4) Analisis data angket respon peserta didik: Data respon peserta didik yang diperoleh melalui angket dianalisis menggunakan statistik deskriptif dalam bentuk persentase, Persentase dari setiap respon peserta didik. (5) Analisis data angket respon guru: Data respon guru yang diperoleh melalui angket dianalisis menggunakan statistik deskriptif dalam bentuk persentase, persentase dari setiap respon peserta didik.

HASIL PENELITIAN

1. Hasil Penelitian

a. Deskripsi Hasil Analisis Data

Penelitian ini merupakan penelitian kualitatif dengan model penelitian pengembangan (*development research*). Penelitian mengadopsi model Borg and

Gall yang mencakup tahap merancang, mengembangkan, validasi, dan uji coba. Penelitian ini mengembangkan instrumen penilaian autentik yang mencakup penilaian kompetensi sikap (jurnal observasi sikap), penilaian keterampilan (unjuk kerja), dan penilaian kognitif (tes hasil belajar).

1) Tahap Validasi

Tahap validasi merupakan tahapan telaah instrumen penilaian autentik oleh ahli. Validasi yang dilakukan oleh pakar adalah validasi konstruk (isi). Instrumen penilaian autentik ini divalidasi oleh dua pakar yaitu Prof. Dr. Ir. Yusminah Hala, M.S., dan Prof. Dr. Danial, M.Si.

Hasil analisis validitas soal diperoleh berdasarkan hasil penilaian ahli yang dilakukan oleh validator ahli pendidikan, validator ahli konten. Hasil rekapitulasi analisis validitas kedua ahli di atas dapat dilihat pada Tabel 1 untuk penilaian sikap, Tabel 2 untuk penilaian keterampilan dan Tabel 3 untuk penilaian kognitif (tes hasil belajar).

Berdasarkan Tabel 1, tampak bahwa instrumen penilaian sikap melalui metode observasi yang dikembangkan dikategorikan sangat valid dengan nilai 3,78. Hal tersebut sudah sesuai dengan standar yang telah ditetapkan yaitu $3,5 \leq x \leq 4,0$ dikategorikan sangat valid [5]. Instrumen penilaian sikap yang sudah valid ini layak untuk diujicobakan pada subyek penelitian dengan revisi sedikit.

Tabel 1. Rekap Hasil Analisis Validitas Instrumen Penilaian Sikap berdasarkan Penilaian Ahli

Aspek yang Dinilai	Rerata Aspek		Rerata Aspek	Keterangan
	Agreement	Disagreement		
Materi	3	0	4,00	SV
Konstruksi	3	0	3,67	SV
Bahasa	3	0	3,67	SV
Rerata Total Aspek			3,78	SV

Berdasarkan Tabel 2, tampak bahwa instrumen penilaian keterampilan melalui penilaian unjuk kerja yang dikembangkan dikategorikan sangat valid dengan nilai 3,64. Hal tersebut sudah sesuai dengan standar yang telah ditetapkan yaitu $3,5 \leq x \leq 4,0$ dikategorikan sangat valid [5]. Instrumen penilaian sikap yang sudah valid ini layak untuk diujicobakan pada subyek penelitian dengan revisi sedikit.

Tabel 2. Rekap Hasil Analisis Validitas Instrumen Penilaian Keterampilan berdasarkan Penilaian Ahli

Aspek yang Dinilai	Rerata Aspek		Rerata Aspek	Keterangan
	Agreement	Disagreement		
Materi	2	0	3,75	SV
Konstruksi	3	0	3,50	V
Bahasa	3	0	3,67	SV
Rerata Total Aspek			3,64	SV

Tabel 3. Rekap Hasil Analisis Validitas Instrumen Tes Hasil Belajar berdasarkan Penilaian Ahli

Aspek yang Dinilai	Rerata Aspek		Rerata Aspek	Keterangan
	Agreement	Disagreement		
Materi	4	0	3,75	SV
Konstruksi	7	0	3,5	SV
Bahasa	3	0	3,5	SV
Rerata Total Aspek			3,58	SV

Berdasarkan Tabel 3, tampak bahwa instrumen tes hasil belajar yang dikembangkan dikategorikan sangat valid dengan nilai 3,58. Hal tersebut sudah sesuai dengan standar yang telah ditetapkan yaitu $3,5 \leq x \leq 4,0$ dikategorikan sangat valid (Nurdin, 2007 dalam Adnan, 2014). Tes hasil belajar yang sudah valid ini layak untuk diujicobakan pada subyek penelitian. Jumlah butir soal ada 40 dengan tipe pilihan ganda, peserta didik diberikan tenggang waktu untuk mengerjakannya selama 90 menit.

2) Tahap Uji Coba

Setelah semua instrumen telah di validasi dan dinyatakan valid dan layak, maka dilanjutkan dengan uji coba pada subjek penelitian. Uji coba dilaksanakan pada kelas XI MIA 2 yang berjumlah 30 orang. Penilaian sikap dilakukan selama 4 kali pertemuan dengan tiga orang observer, masing-masing observer menilai 10 orang. Observer dalam penilaian sikap antara lain Herliyani, S.Pd (Pengajar Biologi), Asrida Agriani, S.Pd., M.Pd. (Pengajar Biologi) dan Suparman, S.Pd. (Wakasek Urusan Kurikulum). Penilaian Keterampilan dalam hal ini menilai unjuk kerja peserta didik dalam pengujian kandungan glukosa pada urine dibantu oleh dua orang observer yaitu Herliyani, S.Pd (Pengajar Biologi), Asrida Agriani, S.Pd., M.Pd. (Pengajar Biologi).

Penilaian untuk tes hasil belajar dilakukan pada akhir pertemuan.

Hasil uji coba ini melahirkan data penilaian sikap, penilaian keterampilan dan karakteristik tes hasil belajar yang telah dikembangkan. Pengembangan instrumen penilaian ini terdiri dari instrumen non tes (penilaian sikap dan keterampilan) dan instrumen tes (tes hasil belajar).

Instrumen non tes yang digunakan untuk mengukur sikap cukup memenuhi validitas kontrak. Setelah pengujian konstruk dari ahli dan berdasarkan pengalaman empiris di lapangan selesai, maka diteruskan dengan uji coba instrumen. Setelah data ditabulasikan, maka pengujian validitas konstruksi dilakukan dengan analisis faktor, yaitu dengan mengkorelasikan antar skor item instrumen dalam suatu faktor, dan mengkorelasikan skor faktor dengan skor total. Adapun data hasil analisis faktor diperoleh dari output program SPSS 20 untuk penilaian sikap dan keterampilan sebagai berikut:

a. Penilaian Sikap

1) KMO (Kaiser-Meyer-Olkin) dan Bartlett's Test

Tabel 4. KMO dan Bartlett's Test Hasil Analisis Instrumen Penilaian Sikap

Nilai	Kaiser-Meyer-Olkin	MSA
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.729
Bartlett's Test of Sphericity	Approx. Chi-Square	354.191
	Df	136
	Sig.	.000

(KMO-MSA) sebesar 0,729 sudah menunjukkan bahwa hasil yang baik (> 0.60). Dengan demikian, syarat analisis faktor pertama dapat dipenuhi sehingga dapat dilanjutkan untuk pengujian selanjutnya. Kedua, nilai Bartlett's Test of Sphericity sebesar 354,191 pada signifikansi 0,000. Nilai ini menandakan bahwa faktor pembentuk variabel sudah baik.

2) Variance Explained

Berdasarkan hasil uji total variance explained diketahui bahwa 4 (empat) faktor yang digunakan mampu menjelaskan 73% dari keseluruhan konstruk sikap. Dapat disimpulkan bahwa melalui analisis faktor berhasil dibuktikan bahwa 17 (tujuh belas) item yang digunakan untuk mengukur variabel sikap dikelompokkan menjadi 4

faktor yaitu yaitu faktor 1 yang menjelaskan 48% varian, faktor 2 menjelaskan 9% varian, faktor 3 yang menjelaskan 8% varian, dan faktor 4 yang menjelaskan 7% varian konstrak.

3) *Rotated Komponent Matrix*

Hasil *Rotated Component Matrix* menunjukkan bahwa item indikator 8, indikator 11, indikator 17, indikator 10, indikator 16, indikator 5, indikator 12, indikator 2 mengelompok pada faktor 1; indikator 6, indikator 9, indikator 14 mengelompok pada faktor 2; item indikator 3, indikator 7, indikator 1 mengelompok pada faktor 3; dan item indikator 15, indikator 4, indikator 13 mengelompok pada faktor 4. Dari hasil ini jelas dapat disimpulkan bahwa konstrak sikap merupakan konstrak multidimensi yang terdiri dari 4 (empat) faktor.

4) Hasil Perhitungan Pengujian Validitas Konstrak dan Reliabilitas

Untuk mengetahui kualitas butir dari setiap indikator pada instrumen penilaian sikap valid atau tidak, dapat diketahui dengan cara mengkorelasikan antara skor butir dengan total. Bila harga korelasi dibawah 0,05 (taraf signifikan 5%), maka butir instrumen tersebut valid, sebaliknya jika diatas 0,05, maka butir instrumen tersebut tidak valid. Untuk uji reliabilitas instrumen penilaian sikap, jika diatas angka 0,70 berarti instrumen tersebut reliabel, jika dibawah angka 0,70 berarti instrumen tersebut tidak reliabel. Berdasarkan hasil perhitungan pengujian validitas konstrak instrumen penilaian sikap, maka semua indikator dinyatakan valid dengan nilai korelasi pada setiap indikator berada lebih kecil dari taraf signifikan 0,05. Instrumen penilaian sikap termasuk kategori reliabel dengan nilai 0,760.

b. Penilaian Keterampilan

1) *Variance Explained*

Berdasarkan hasil uji total variance explained pada instrumen penilaian keterampilan diketahui bahwa 4 (empat) faktor yang digunakan mampu menjelaskan 74% dari keseluruhan konstrak keterampilan. Dapat disimpulkan bahwa melalui analisis faktor berhasil dibuktikan bahwa 12 (tujuh belas) item yang digunakan untuk mengukur variabel keterampilan dikelompokkan menjadi 4 faktor yaitu yaitu faktor 1 yang menjelaskan 24% varian, faktor 2 menjelaskan 21% varian, faktor 3 yang

menjelaskan 17% varian, dan faktor 4 yang menjelaskan 11% varian konstrak.

2) *Rotated Component Matrix*

Hasil *Rotated Component Matrix* menunjukkan bahwa item indikator 1, indikator 7, indikator 11, indikator 2 mengelompok pada faktor 1; indikator 10, indikator 4, indikator 8 mengelompok pada faktor 2; item indikator 5 dan indikator 12, mengelompok pada faktor 3; dan item indikator 3, indikator 9, indikator mengelompok pada faktor 4.

3) Hasil Perhitungan Pengujian Validitas Konstrak dan Reliabilitas

Untuk mengetahui kualitas butir dari setiap indikator pada instrumen penilaian keterampilan valid atau tidak, dapat diketahui dengan cara mengkorelasikan antara skor butir dengan total. Bila harga korelasi dibawah 0,05 (taraf signifikan 5%), maka butir instrumen tersebut valid, sebaliknya jika diatas 0,05, maka butir instrumen tersebut tidak valid. Untuk uji reliabilitas instrumen penilaian keterampilan, jika diatas angka 0,70 berarti instrumen tersebut reliabel, jika dibawah angka 0,70 berarti instrumen tersebut tidak reliabel.

Tabel 5. Rekapitulasi Perhitungan Pengujian Validitas Konstrak Instrumen Penilaian Keterampilan

Kategori	Jumlah	Persentase (%)	Nomor Indikator
Valid	7	58	1, 2, 4, 7, 8, 10, 11
Tidak Valid	5	42	3, 5, 6, 9, 12
Jumlah	12	100	

Tabel 6. Hasil Perhitungan Pengujian Reliabilitas Instrumen Penilaian Keterampilan

Cronbach's Alpha	N of Items
,633	12

c. Penilaian Kognitif

1) Hasil analisis uji coba tes hasil belajar

Uji coba dilakukan pada kelas XI MIA 2 SMA Negeri 1 Tellusiattinge dengan jumlah peserta didik 30 orang. Waktu untuk mengerjakan soal tersebut adalah 90 menit. Data mengenai jawaban dan skor yang diperoleh peserta didik selengkapnya dapat dilihat pada lampiran.

2) Hasil analisis validitas butir tes kognitif Analisis validitas butir soal dilakukan dengan menggunakan SPSS 20. Berdasarkan Tabel 1.6 diperoleh hasil analisis untuk mengetahui nilai validitas butir soal bahwa dari 40 item tes hasil belajar yang telah diujikan, terdapat 25 item tes yang diinterpretasikan valid yaitu item tes nomor 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 20, 23, 24, 27, 28, 29, 31, 34, 35, 36, 37, dan 40. Jumlah item tes yang tidak valid berjumlah 15 yaitu item tes nomor 1, 10, 15, 16, 17, 18, 19, 21, 22, 25, 26, 30, 32, 33, dan 38. Adapun rekapitulasi hasil analisis validitas butir soal dapat dilihat pada Tabel 7.

Tabel 7. Rekapitulasi Hasil Analisis Validitas Butir Soal

Kategori	Jumlah	Persentase (%)
Valid	25	63%
Invalid	15	38%
Total	40	100%

3) Hasil analisis reliabilitas butir tes kognitif

Analisis reliabilitas butir soal dilakukan dengan menggunakan SPSS 20. Adapun hasil perhitungan validitas butir soal dapat dilihat pada Tabel 8

Tabel 8. Rekap Hasil Analisis Reliabilitas Soal

Cronbach's Alpha	N of Items
,723	40

Bertitik tolak dari hasil analisis yang diperlihatkan pada Tabel 8 menunjukkan bahwa tes hasil belajar yang dikembangkan berada pada kategori reliabel yang tinggi dengan nilai 0,723. Hal ini sesuai dengan ketentuan yang ditetapkan yaitu apabila R (nilai reliabilitas) sama dengan atau lebih besar daripada 0,70 berarti soal tersebut telah memiliki reliabilitas yang tinggi (= reliabel). Apabila R (nilai reliabilitas) lebih kecil daripada 0,70 berarti tes hasil belajar tersebut belum memiliki reliabilitas yang tinggi (= unreliabel / tidak reliabel).

4) Hasil analisis karakteristik butir tes hasil belajar kognitif

Analisis karakteristik butir tes hasil belajar kognitif yang meliputi tingkat kesukaran, daya beda, dan keefektifan pengecoh juga diperoleh melalui hasil analisis dengan menggunakan SPSS 20.

a) Tingkat kesukaran

Hasil analisis tingkat kesukaran butir soal diperoleh dari analisis nilai peserta didik dengan menggunakan SPSS 20. Analisis tingkat kesukaran juga dapat dilakukan dengan menggunakan rumus yaitu banyaknya peserta didik yang menjawab benar butir soal dibagi dengan banyaknya peserta didik yang mengikuti tes.

Mencermati hasil analisis pada 9 menunjukkan bahwa tipe tes hasil belajar yang dikembangkan secara umum berkategori sedang sebanyak 45%, kategori mudah sebanyak 33% dan kategori sukar sebanyak 23%. Pengkategorian tersebut didasarkan atas ketentuan yang ditetapkan yaitu apabila ITK (Indeks Tingkat Kesukaran) 0,00-0,30 soal tergolong sukar, ITK 0,31-0,70 soal tergolong sedang, ITK 0,71-1,00 soal tergolong mudah [6].

Tabel 9. Rekapitulasi Hasil Analisis Tingkat Kesukaran Soal

Kategori	Jumlah Soal	Persentase (%)	Nomor Butir Soal
Sukar	9	23%	15, 18, 19, 20, 21, 26, 33, 38, 40
Sedang	18	45%	1, 2, 4, 5, 6, 9, 12, 16, 22, 25, 27, 28, 29, 30, 31, 34, 35, 36,
Mudah	13	33%	3, 7, 8, 10, 11, 13, 14, 17, 23, 24, 32, 37, 39
Total	40	100	40

b) Daya Pembeda

Analisis daya pembeda soal dilakukan melalui program SPSS 20. Adapun hasil perhitungan daya beda soal dapat dilihat pada Tabel 10

Tabel 10. Rekapitulasi Hasil Analisis Daya Beda Soal

Kategori	Jumlah Soal	Persentase	Nomor Soal
Tidak ada daya beda	3	8%	1, 15, 26
Daya beda lemah	7	18%	10, 17, 18, 19, 25, 33, 38
Daya beda cukup	8	20%	16, 21, 22, 23, 30, 32, 35, 40
Daya beda baik	21	53%	2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 20, 24, 28, 29, 31, 34, 36, 37, 39
Daya beda baik sekali	1	3%	27

Informasi dari Tabel 10 menunjukkan bahwa tipe soal yang berada pada kategori baik sekali 3%, kategori daya beda baik 53%, kategori daya beda cukup 20%, kategori daya beda lemah 18%, dan tidak ada daya beda 8%.

c) Kualitas pengecoh

Analisis pengecoh dilakukan dengan cara menghitung jumlah peserta didik yang memilih option jawaban pada setiap soal. Berdasarkan hasil perhitungan dari Tabel 1.10 menunjukkan bahwa tipe soal yang dikembangkan memiliki butir yang optionnya tidak dipilih oleh peserta didik sehingga dikatakan tidak efektif. Perhitungan tersebut telah dilakukan dengan mengacu pada kriteria bahwa pengecoh dikatakan berfungsi efektif apabila ada peserta didik yang terkecoh memilih. Selain itu, ada juga option yang tergolong tidak efektif karena banyaknya peserta didik yang memilihnya termasuk peserta didik yang tergolong kelompok atas.

Tabel 11. Rekapitulasi Hasil Analisis Kualitas Pengecoh

Kategori	Jumlah	Persentase	Nomor Item
Efektif	14	35%	5, 6, 7, 15,16,17,18, 20, 21, 22, 27, 30, 35, 40
			Tidak efektif

2. Pembahasan Penelitian

Hasil pengembangan dari setiap kompetensi kemudian di validasi oleh ahli. Hasil validasi untuk instrumen penilaian sikap, dalam hal ini jurnal observasi sikap berada pada kategori sangat valid yaitu 3,78. Instrumen penilaian sikap telah direvisi berdasarkan saran validator dengan menambahkan checklist waktu pada setiap indikator sikap disetiap pertemuan. Hasil validasi oleh ahli untuk instrumen penilaian keterampilan, dalam hal ini unjuk kerja praktikum berada pada kategori sangat valid yaitu 3,64. Instrumen penilaian keterampilan mengalami revisi berdasarkan saran validator untuk mengubah kata “tepat” menjadi kata yang lebih operasional dan mudah diukur. Hasil validasi oleh ahli untuk instrumen penilaian pengetahuan,

dalam hal ini tes hasil belajar berada pada kategori sangat valid yaitu 3,58. Instrumen penilaian kognitif tes hasil belajar telah mengalami revisi atas saran validator untuk menambahkan alokasi waktu pada identitas soal. Instrumen yang telah dinyatakan valid dan layak, kemudian diujikan pada kelas XI MIA 2 SMA Negeri 1 Tellusiattinge yang berjumlah 30 orang. Instrumen penelitian yang digunakan berupa instrumen non tes dan tes. Instrumen non tes terdiri dari instrumen penilaian sikap dan instrumen penilaian keterampilan. Setelah instrumen penilaian sikap dan keterampilan di validasi konstruk oleh ahli kemudian hasilnya diuji cobakan, lalu data hasil observasi dan penilaian unjuk kerja masing-masing ditabulasikan lalu dilanjutkan dengan analisis faktor. Analisis faktor memungkinkan peneliti untuk 1) menguji ketepatan model (*goodness of fit test*) faktor yang terbentuk dari item alat ukur 2) menguji kesetaraan unit pengukuran antar item, 3) menguji reliabilitas item-item pada tiap faktor yang diukur, 4) menguji adanya invarian item pada populasi [7].

Hasil analisis faktor (uji total varians) pada jurnal observasi untuk instrumen penilaian sikap diketahui bahwa ada 4 (empat) faktor yang digunakan mampu menjelaskan 73% dari keseluruhan konstruk sikap. Melalui analisis faktor menunjukkan hasil bahwa 17 (tujuh belas) item yang digunakan untuk mengukur variabel sikap dikelompokkan menjadi 4 faktor yaitu faktor 1 (indikator 8, 11, 17, 10, 16, 5, 12 dan 5) yang menjelaskan 48% varian, faktor 2 (indikator 6, 9, 14) menjelaskan 9% varian, faktor 3 (indikator 3,7,1) yang menjelaskan 8% varian, dan faktor 4 (indikator 15, 4, 13) yang menjelaskan 7% varian. Indikator penilaian sikap dapat diamati pada lampiran.

Hasil analisis faktor (uji total varians) pada penilaian unjuk kerja untuk instrumen penilaian keterampilan diketahui bahwa 4 (empat) faktor yang digunakan mampu menjelaskan 74% dari keseluruhan konstruk keterampilan. Melalui analisis faktor menunjukkan hasil bahwa 12 (tujuh belas) item yang digunakan untuk mengukur variabel keterampilan dikelompokkan menjadi 4 faktor yaitu faktor 1 (indikator 1,7,11 dan 2) yang menjelaskan 24% varian, faktor 2 (indikator 10,4, dan 8) menjelaskan 21% varian, faktor 3 (indikator 5 dan 12) yang menjelaskan 17% varian, dan faktor 4 (indikator 3 dan 9) yang menjelaskan 11% varian konstruk. Indikator penilaian

keterampilan dapat diamati pada lampiran. Berdasarkan hasil analisis faktor baik pada penilaian sikap maupun penilaian keterampilan dapat disimpulkan bahwa konstruk sikap dan keterampilan merupakan konstruk multidimensi yang terdiri dari 4 (empat) faktor.

Selain instrumen non-tes pada penelitian ini juga mengembangkan instrumen tes berupa tes kognitif dalam bentuk pilihan ganda. Hasil tes yang telah diujicobakan kemudian dianalisis untuk mengetahui validitas dan reliabilitas dari butir tes kognitif karakteristik butir soal yang meliputi tingkat kesukaran soal, daya pembeda, dan kualitas pengecoh (pola distribusi jawaban). Validitas tes kognitif yaitu validitas isi untuk mengetahui kesesuaian soal yang disusun dengan materi yang ada di kompetensi dasar. Reliabilitas diperlukan dalam penilaian tes, sebab tes yang memiliki reliabilitas tinggi akan memiliki daya pembeda yang tinggi pula [8]. Hal tersebut dapat dilihat pada hasil analisis validitas butir soal tes kognitif pada penelitian ini, persentase soal valid yaitu 63% dan soal invalid 38%, artinya 63% soal yang disusun sudah sesuai dengan materi yang terdapat pada kompetensi dasar 3.9. Reliabilitas dari soal tes kognitif ini dengan nilai 0,723. Hal ini sesuai dengan ketentuan yang ditetapkan yaitu apabila R (nilai reliabilitas) sama dengan atau lebih besar daripada 0,70 berarti soal tersebut telah memiliki reliabilitas yang tinggi (= reliabel). Apabila R (nilai reliabilitas) lebih kecil daripada 0,70 berarti tes hasil belajar tersebut belum memiliki reliabilitas yang tinggi [9].

Selain analisis validitas dan reliabilitas soal, juga dilakukan analisis karakteristik butir soal. Berdasarkan hasil analisis program SPSS 20 diperoleh tingkat kesukaran soal umumnya berada pada kategori sedang yaitu 45%, 33% tergolong mudah dan 23% tergolong sukar. Persentase tersebut khususnya pada kategori soal mudah memenuhi persyaratan tes tertulis pilihan ganda seperti yang diungkapkan oleh [6], sedangkan soal sedang 50% dan soal sukar 20%. Hasil ini sejalan dengan hasil penelitian yang dilakukan oleh Sudirman (2013) analisis tingkat kesukaran yang telah dilakukan, dapat diketahui bahwa tes hasil belajar yang dikembangkan umumnya berada pada kategori sedang dengan proporsi soal yang

tingkat kesukarannya kategori mudah sebanyak 17%, kategori sedang sebanyak 60%, dan yang berkategori sukar sebanyak 23%. Soal tes kognitif yang dikembangkan untuk karakteristik butir soal yang ditinjau dari tingkat kesukaran soal secara umum belum sesuai teori tetapi tidak berbeda secara signifikan sehingga masih dapat ditoleransi. Dengan demikian, soal yang dikembangkan tergolong homogen dan hampir mendekati jumlah persentasenya dengan yang dikemukakan oleh ahli. Bermutu atau tidaknya soal tes hasil belajar dapat diketahui dari derajat kesukaran yang dimiliki oleh masing-masing butir soal tersebut.

Karakteristik butir soal selanjutnya yang dianalisis adalah daya pembeda. Daya pembeda pada soal tes kognitif kategori baik sekali 3%, kategori baik 53%, kategori cukup 20%, kategori lemah 18%, dan kategori tidak ada daya beda 8%. Soal yang perlu perbaikan dan ditolak adalah soal yang tidak memiliki daya beda, dan soal yang diterima dan perlu perbaikan merupakan soal yang memiliki daya beda cukup. Soal yang tidak memiliki daya beda termasuk soal yang tergolong mudah sehingga peserta didik yang tergolong dalam kelompok rendah juga mampu untuk menjawabnya. Setiap soal memiliki daya pembeda yang memadai, artinya tiap butir dalam tes itu dapat membedakan peserta didik yang belajar atau menguasai materi dan peserta didik yang belum belajar atau belum menguasai materi.

Untuk analisis pengecoh atau distactor soal efektif sebanyak 35% dan soal yang tidak efektif sebanyak 65%. Soal yang memiliki option tidak efektif (tidak ada satu pun siswa yang memilih) berarti option tersebut direvisi, tidak dibuang. Pendapat tersebut sesuai dengan pendapat [8] bahwa menulis soal adalah suatu pekerjaan yang sulit, sehingga apabila masih dapat diperbaiki, sebaiknya diperbaiki saja, tidak dibuang. Perhitungan analisis pengecoh telah dilakukan dengan mengacu pada kriteria bahwa pengecoh dikatakan berfungsi efektif apabila paling tidak ada siswa yang terkecoh memilih [11].

Hasil validasi untuk instrumen tes dan non-tes Berdasarkan validitas dan reliabilitas dari instrumen penilaian sikap maka dinyatakan layak untuk digunakan karena semua indikator valid dan instrumen tersebut sudah reliabel. Untuk instrumen penilaian keterampilan masih ada indikator yang belum valid dan belum reliabel, sehingga sebelum

digunakan masih membutuhkan revisi. Hasil validasi yang telah diujicobakan kemudian dilakukan analisis faktor yang menghasilkan data bahwa kontrak sikap dan keterampilan merupakan konstruk multidimensi yang terdiri dari empat faktor.

Untuk instrumen tes kognitif, keputusan suatu item soal layak digunakan, perlu direvisi atau ditolak didasarkan pada kriteria keputusan bahwa soal tersebut telah valid dan reliabel berdasarkan ketentuan sebagai berikut: 1) apabila nilai validitas sama dengan atau lebih besar dengan nilai r tabel product moment (pada penelitian ini harga r tabel yang digunakan adalah 0,176) berarti soal dinyatakan valid. Jika nilai validitas lebih kecil dari nilai r tabel product moment (pada penelitian ini harga r tabel yang digunakan adalah 0,176) berarti soal dinyatakan tidak valid. 2) apabila R (nilai reliabilitas) sama dengan atau lebih besar daripada 0,70 berarti soal tersebut reliabel. Apabila R (nilai reliabilitas) lebih kecil daripada 0,70 berarti tes hasil belajar tersebut tidak reliabel [9].

Dari segi karakteristik butir soal, keputusan suatu item soal layak digunakan, perlu direvisi atau ditolak didasarkan pada kriteria keputusan untuk penilaian item soal adalah memiliki validitas yang tinggi, memiliki reliabilitas yang tinggi, tiap butir soal memiliki daya pembeda yang memadai, tingkat kesukaran tes berdasar kelompok yang akan dites (kira-kira 30% soal mudah, 50% soal sedang, 20% soal sulit), mudah diadministrasikan [6].

Memperbaiki instrumen autentik, yaitu memperbaiki instrumen penilaian sikap, keterampilan, maupun kognitif sesuai dengan saran-saran dari guru maupun validator ahli serta berdasarkan hasil analisis uji coba terhadap instrumen yang masih belum memenuhi kriteria sebagai instrumen yang baik. Merakit instrumen autentik yaitu menyusun kembali instrumen penilaian sikap, keterampilan, dan kognitif yang telah diperbaiki/direvisi setelah uji coba, dengan demikian dihasilkan sebuah instrumen autentik yang dapat digunakan untuk mengukur kemampuan siswa dan telah diuji kualitasnya. Selanjutnya dilakukan tahap penyebaran dengan cara mensosialisasikan instrumen penilaian autentik yang dikembangkan kepada guru-guru yang tergabung dalam forum MGMP (Musyawarah Guru Mata Pelajaran) Biologi Kabupaten Bone yang dilaksanakan

di SMA Negeri 1 Watampone yang merupakan sekretariat MGMP Biologi Kabupaten Bone. Penyebaran ini dilakukan agar instrumen penilaian autentik yang dikembangkan dapat digunakan secara lebih luas untuk mengukur kemampuan peserta didik dalam menyerap materi pelajaran.

KESIMPULAN

Berdasarkan hasil dan pembahasan maka diperoleh kesimpulan sebagai berikut :

1. Instrumen penilaian autentik yang telah dikembangkan pada materi sistem ekskresi berada pada kategori sangat valid dengan nilai 3,78 untuk sikap, 3,64 untuk keterampilan dan 3,58 untuk tes kognitif. Untuk uji reliabilitas instrumen penilaian sikap dan tes kognitif berada pada kategori reliabel dengan nilai 0,760 dan 0,723, sedangkan instrumen penilaian keterampilan berada pada kategori tidak reliabel dengan nilai 0,633.
2. Tes kognitif yang telah dikembangkan memiliki kualitas yaitu tingkat kesukaran: kategori mudah 33%, kategori sedang 45%, kategori sukar 23%; daya beda: kategori baik sekali 3%, kategori baik 53%, kategori cukup 20%, kategori lemah 18%, dan kategori tidak ada daya beda 8%; dan kualitas pengecoh: kategori efektif 35% dan kategori tidak efektif 65%.

DAFTAR PUSTAKA

- [1]. Direktorat Jenderal Pendidikan Dasar dan Menengah. 2015. *Panduan Penilaian untuk Sekolah Menengah Atas*. Jakarta: Direktorat Jenderal Pendidikan Dasar dan Menengah Kementerian Pendidikan dan Kebudayaan.
- [2]. Meta. 2010. *Jenis-jenis Penilaian*. <http://www.manggamudaku.net/2010/11/jenis-jenis-penilaian.html>. Diakses pada tanggal 7 Januari 2016.
- [3]. Mohammad, S. 2015. *Materi Pelatihan Guru Implementasi Kurikulum 2013 tahun 2015 Mata Pelajaran Biologi*. Jakarta: Kementerian Pendidikan dan kebudayaan.
- [4]. Sugiyono. 2014. *Metode Penelitian Pendidikan*. Bandung: Alfabeta.
- [5]. Adnan. 2014. *Model Pembelajaran Biologi Konstruktivistik berbasis TIK*

untuk meningkatkan Motivasi dan Kemampuan Kognitif Siswa SMP. Disertasi. Tidak diterbitkan. Makassar: Universitas Negeri Makassar.

- [6]. Kunandar. 2013. *Penilaian Autentik*. Jakarta: PT. Raja Grafindo Perkasa.
- [7]. Widhiarso, Wahyu. 2009. *Handout Mata Kuliah Psikometri Fakultas Psikologi UGM*. www.distrodoc.com/2125-prosedur-analisis-faktor-dengan-menggunakan-komputer. Diakses pada tanggal 09 Juni 2016.
- [8]. Arikunto, S. 2002. *Prosedur Penelitian*. Jakarta: PT. Rineka Cipta.
- [9]. Sukiman. 2012. *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Madani.
- [10]. Sudirman, A. 2013. *Pengembangan Tes Hasil Belajar Kognitif Berorientasi Higher Order Thinking Skills (HOTS) pada Materi Sistem Ekskresi di SMA*. Tesis. Tidak diterbitkan. Makassar: Universitas Negeri Makassar.
- [11]. Purwanto. 2013. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.