

PENGEMBANGAN TES SUMATIF MATA PELAJARAN BIOLOGI SEMESTER GANJIL KELAS IX

Amri

Program Studi Pendidikan Biologi Fakultas Keguruan dan Ilmu Pendidikan, Universitas
Muhammadiyah Parepare, Kampus II Jl. Jln. Ahmad Yani Km. 7 Kota parepare
Telepon: (0421) 22757, E-mail: amryefilein@yahoo.co.id/amribiologi@umpar.ac.id

Syahriani

Jurusan Pendidikan Biologi Fakultas Tarbiyah dan Keguruan,
UIN Alauddin Makassar, Kampus II Jl. H. M. Yasin Limpo No 36 Samata-Gowa,
Sulawesi Selatan 92118, Telepon: (0411) 424835,
E-mail: syahriani.rahman@uin-alauddin.ac.id

Abstrak

Tujuan penelitian ini adalah untuk mengembangkan perangkat tes sumatif yang valid dan reliabel dan perangkat tes sumatif dengan karakteristik internal yang meliputi tingkat kesukaran, daya beda, dan kualitas pengecoh (*distracter*) pada mata pelajaran biologi semester ganjil kelas IX. Jenis Penelitian ini adalah penelitian pengembangan dengan menggunakan pendekatan kualitatif. Model pengembangan yang digunakan model pengembangan 4-D yang terdiri atas tahap *define, design, develop* dan *dessiminate*. Instrumen penelitian yang digunakan yaitu format telaah butir soal dan angket respon siswa terhadap tes yang telah dikembangkan. Pengelolaan data dilakukan secara kualitatif dan kuantitatif. Hasil yang diperoleh pada proses pengembangan pada tahap pertama menghasilkan prototipe 1 perangkat tes kognitif tipe pilihan ganda sebanyak 90 butir. Pengembangan perangkat pada tahap selanjutnya dicapai; 1) valid; (a) validitas isi berdasarkan penilaian ahli >75% dan praktisi dengan sedikit revisi, (b) validitas kriteria pertama dan kedua dari 90 butir tes tersisa 53 tes yang valid dan dikategorikan baik (c) validitas konstruk memerlukan revisi atau peninjauan ulang 2) reliabel; (a) nilai reliabilitas tes pada uji empirik pertama 0,92 dan kedua 0,82 , (3) tingkat kesukaran sedang, (3) daya beda cukup, (4) kualitas pengecoh baik dan (5) respon siswa dan guru terhadap perangkat tes dalam kategori positif.

Kata kunci: Pengembangan tes, Model 4-D, Tes Sumatif

Abstract

The purpose of this research is to develop a valid and reliable summative test instrument and summative test instrument with level of difficulty, distinguishing ability, and distracter quality on biology subject of class IX. Type. This research

is a development research with qualitative approach. The development model used the 4-D development model that were consist of define, design, develop and disseminate stages. Research data were obtained through two criteria, namely (1) validity and reliability data were obtained from expert and practitioner's assessment (2) data of validity, reliability, difficulty, distinguishing ability, and distracter quality were obtained from test result of summative test. Based on limited trial there was a cognitive assessment instrument that consist of 53 numbers of multiple choice question from 90 question numbers from the previous draft with valid and reliable quality categories of easy, moderate and difficult. In terms of difficulty level, easy category was 38.67%, medium category was 58.67% and difficult category was 2.67% and for distinguishing ability there were weak, adequate and good categories. The weak category was 36%, enough category was 56%, and good category was 8%. In terms of distracter effectivity, about 89.33% of the test was effective and 10.67% was not effective.

Key words: *4-D Development, summative test instrument*

PENDAHULUAN

Asesmen merupakan proses pengumpulan informasi selengkap-lengkapny tentang siswa dan kelas untuk tujuan pembuatan keputusan pengajaran. Sedangkan asesmen dalam kaitannya dengan evaluasi adalah proses pengumpulan berbagai data yang dapat memberikan gambaran perkembangan belajar siswa. Gambaran perkembangan belajar siswa perlu diketahui oleh guru agar bisa memastikan bahwa siswa mengalami proses pembelajaran dengan benar. Untuk mengetahui tingkat keberhasilan siswa dalam menguasai bahan pelajaran diperlukan suatu alat ukur yang biasanya berupa tes. Hasil evaluasi tes merupakan salah satu indikator keberhasilan siswa yang dapat dicapai dalam usaha belajarnya.

Hasil evaluasi yang dapat memberikan informasi perbaikan kualitas pendidikan harus menggunakan alat ukur (tes) yang baik, sehingga analisis kualitas tes mutlak dilakukan oleh guru dan merupakan suatu tahap yang harus ditempuh untuk mengetahui derajat kualitas suatu tes, baik tes secara keseluruhan maupun butir soal yang menjadi bagian dari tes. Sebuah tes dapat dikatakan baik atau berkualitas sebagai alat pengukur harus memenuhi persyaratan tes, yaitu memiliki: validitas, reliabilitas, objektivitas, praktibilitas serta ekonomis (Arikunto, 2005).

Praktik evaluasi di sekolah syarat validitas dan reliabilitas sangat sulit dipenuhi, apalagi kalau pembuat tes itu adalah guru atau praktisi pendidikan yang belum memiliki kebiasaan dalam menyusun tes. Seringkali guru acuh dengan kualitas suatu tes. Kualitas tes atau tingkat validitas dan reliabilitas belum menjamin keobjektifannya, sebab hanya diberikan kepada sekelompok peserta didik tertentu saja atau bersifat sektoral karena belum diuji cobakan kepada kelompok besar serta bersifat temporer atau berlaku pada situasi dan kondisi tertentu atau hanya bersifat hafalan semata. Hal yang

hanya dianggap penting bagi guru adalah tersedianya perangkat tes untuk melaksanakan penilaian. Bahkan guru hanya mengambil soal dari buku-buku pelajaran atau dari kumpulan soal belum diketahui tingkat kebaikannya ditinjau dari segi validitas, reliabilitas (Rasyid, 2007).

Salah satu faktor yang mempengaruhi hasil tes tidak valid yaitu, faktor yang berasal dari dalam tes itu sendiri. faktor ini dapat disebabkan oleh arahan tes yang disusun dengan makna tidak jelas sehingga dapat mengurangi validitas tes, kata-kata yang digunakan dalam struktur instrumen evaluasi terlalu sulit, Item-item tes dikonstruksi jelek, tingkat kesulitan item tes tidak tepat dengan materi pembelajaran yang diterima peserta didik, waktu yang dialokasikan tidak tepat, hal ini termasuk kemungkinan terlalu kurang atau terlalu longgar, jumlah item tes terlalu sedikit sehingga tidak mewakili sampel materi pelajaran, dan jawaban masing-masing item evaluasi bisa diprediksi peserta didik (Sukardi, 2008).

Hasil pengamatan peneliti di lapangan ketika melakukan observasi di salah satu sekolah menengah pertama di Kabupaten Enrekang pada pelaksanaan ulangan semester ganjil 2016/2017 menunjukkan bahwa proses penilaian yang dilakukan selama ini semata-mata hanya menekankan pada penguasaan konsep dengan tes subjektif (tes esai). Hal ini didukung oleh penelitian Nuryani (2007) yang mengemukakan bahwa penilaian yang dilakukan selama ini baru mengukur penguasaan materi dengan ranah kognitif tingkat rendah.

Dari hasil wawancara dengan beberapa guru yang mengajar pada Sekolah Menengah Pertama (SMP) di Kabupaten Enrekang. Peneliti memperoleh informasi bahwa setiap soal tes ulangan semester yang dibuat oleh guru harus sesuai dengan silabus yang ada, artinya bahwa setiap soal yang akan diujikan tidak boleh menyimpang dari materi yang telah disampaikan guru kepada siswanya. Namun belum diketahui daya beda dan tingkat kesukaran soal tersebut karena guru diberikan kebebasan untuk menentukan daya beda dan tingkat kesukaran masing-masing soal tes itu sesuai dengan silabus buatan guru.

Berdasarkan fakta-fakta di atas, maka diperlukan perangkat tes yang baik dalam melaksanakan evaluasi di sekolah sehingga peneliti mengembangkan perangkat tes sumatif mata pelajaran biologi semester ganjil kelas IX dengan menggunakan taksonomi Bloom versi revisi dari L.W. Anderson (2001) sebagai acuan. Model pengembangan yang digunakan adalah model 4-D (*define, design, develop, and disseminate*). Model ini cocok digunakan dengan pertimbangan bahwa model 4-D lebih tepat digunakan sebagai dasar untuk mengembangkan perangkat pembelajaran, uraiannya tampak lebih lengkap dan sistematis, pengembangannya melibatkan penilaian ahli, sehingga sebelum dilakukan uji coba di lapangan perangkat pembelajaran telah dilakukan revisi berdasarkan penilaian, saran dan masukan para ahli.

Tujuan dalam penelitian ini adalah untuk mengembangkan perangkat tes sumatif yang valid dan reliabel dan mengembangkan perangkat tes sumatif dengan karakteristik internal yang meliputi tingkat kesukaran, daya beda, dan kualitas pengecoh (*distracter*) pada mata pelajaran biologi semester ganjil kelas IX SMP di Kabupaten Enrekang.

Asesmen secara sederhana dapat diartikan sebagai proses pengukuran dan nonpengukuran untuk memperoleh data karakteristik peserta didik dengan aturan tertentu (Uno & Satria, 2012). Pelaksanaan asesmen pembelajaran guru dihadapkan pada beberapa istilah yang sering dikacaukan pengertiannya atau bahkan sering pula digunakan secara bersama, yaitu istilah pengukuran, tes, penilaian dan evaluasi. Secara konseptual istilah-istilah ini berbeda satu sama lain, tetapi mempunyai hubungan yang sangat erat.

Pengukuran merupakan sekumpulan aturan untuk menetapkan suatu bilangan yang mewakili objek, sifat atau karakteristik, atribut atau tingkah laku. Pengukuran sebagai suatu prosedur pemberian angka (kuantifikasi) terhadap atribut atau variabel sepanjang garis kontinum. (Kusaeri & Suprananto, 2012).

Penilaian adalah suatu prosedur sistematis dan mencakup kegiatan mengumpulkan, menganalisis, serta menginterpretasikan informasi yang dapat digunakan untuk membuat kesimpulan tentang karakteristik seseorang atau objek. Secara khusus untuk dunia pendidikan, Kusaeri & Suprananto (2012), mendefinisikan penilaian sebagai suatu proses yang sistematis dan mencakup kegiatan mengumpulkan, menganalisis, serta menginterpretasikan informasi untuk menentukan seberapa jauh seorang siswa atau sekelompok siswa mencapai tujuan pembelajaran yang telah ditetapkan, baik aspek pengetahuan, sikap maupun keterampilan.

Evaluasi adalah proses pemberian makna atau ketetapan kualitas hasil pengukuran dengan cara membandingkan angka hasil pengukuran tersebut dengan kriteria tertentu. Kriteria sebagai pembanding dari proses pengukuran atau dapat pula ditetapkan sesudah pelaksanaan pengukuran. Kriteria ini dapat berupa proses/kemampuan rata-rata unjuk kerja kelompok dan berbagai patokan yang lain. Kriteria yang berupa batas kriteria minimal yang telah ditetapkan sebelum pengukuran dan bersifat mutlak disebut dengan penilaian acuan patokan atau penilaian acuan kriteria, sedangkan kriteria ditentukan setelah kegiatan pengukuran dilakukan dan didasarkan pada keadaan kelompok dan bersifat relatif disebut dengan Penilaian Acuan Norma (Uno & Satria, 2012).

Sebuah tes yang dapat dikatakan baik sebagai alat pengukur harus memenuhi persyaratan tes yaitu memiliki: validitas, reliabilitas, objektivitas, praktikabilitas serta ekonomis. Validitas didefinisikan sebagai ukuran seberapa cermat suatu tes melakukan fungsi ukurnya. Sebuah tes biasanya hanya menghasilkan ukuran yang valid untuk satu tujuan ukuran tertentu, pernyataan valid harus diiringi oleh keterangan yang menunjuk kepada tujuan yaitu valid mengukur apa, valid bagi siapa. Karena itu, suatu tes yang sangat valid guna pengambilan suatu keputusan sangat tidak berguna dalam pengambilan keputusan yang lain (Mardapi, 2005).

Menurut Arikunto (2010), kata reliabilitas dalam bahasa Indonesia diambil dari kata *reliability* dalam bahasa Inggris, berasal dari kata "*reliable*" yang artinya dapat dipercaya. Penggunaan istilah "reliabilitas" dan *reliable* sering dikacaukan. "Reabilitas" merupakan kata benda, sedangkan "*reliable*" merupakan kata sifat atau keadaan. Seseorang dapat dipercaya jika selalu bicara *ajeg*, tidak berubah-ubah pembicaraannya

dari waktu ke waktu. Demikian pula halnya sebuah tes. Tes tersebut dikatakan dapat dipercaya jika memberikan hasil yang tetap apabila diteskan berkali-kali. Sebuah tes dikatakan reliabel apabila hasil-hasil tes tersebut menunjukkan ketetapan. Dengan kata lain, jika kepada peserta didik diberikan tes yang sama pada waktu yang berlainan, maka setiap peserta didik akan tetap berada dalam urutan (ranking) yang sama dalam kelompoknya. Reliabilitas adalah tingkat atau derajat konsistensi dari suatu instrumen. Reliabilitas dapat diukur dari tiga kriteria, yaitu *stability*, *dependability* dan *predictability*.

Menurut Nurhadi & Suwardi (2010), secara umum tes objektif dibagi empat bentuk yaitu pilihan ganda, benar salah, menjodohkan, dan jawaban singkat. Tes pilihan ganda merupakan bentuk tes yang menyajikan pokok persoalan yang harus dijawab peserta didik dengan memilih jawaban yang benar. Tes pilihan ganda terdiri atas dua bagian yaitu: bagian pertama berupa pembawa pokok persoalan dan bagian kedua berupa daftar pilihan jawaban biasanya terdiri atas empat sampai lima jawaban.

Keunggulan soal bentuk pilihan ganda di antaranya adalah dapat mengukur kemampuan/perilaku secara objektif, sedangkan untuk soal uraian di antaranya adalah dapat mengukur kemampuan mengorganisasikan gagasan dan menyatakan jawabannya menurut kata-kata atau kalimat sendiri. Kelemahan soal bentuk pilihan ganda di antaranya adalah sulit menyusun pengecohnya, sedangkan untuk soal uraian di antaranya adalah sulit menyusun pedoman penskorannya (Depdiknas, 2004).

Tingkat kesukaran soal adalah angka yang menunjukkan proporsi siswa dalam menjawab betul suatu soal. Makin besar tingkat kesukaran berarti soal itu makin mudah, demikian juga sebaliknya yaitu makin rendah tingkat kesukaran berarti soal itu makin sukar. Selanjutnya daya beda soal adalah kemampuan sesuatu soal yang dapat membedakan antara peserta didik yang pandai (berkemampuan tinggi) dengan peserta didik yang bodoh (berkemampuan rendah). Soal yang baik adalah soal yang dapat membedakan peserta didik yang pintar dan peserta didik yang tidak pintar (Purwanto, 2010).

Distracter bertujuan untuk mengecoh mereka yang kurang mampu (tidak tahu) untuk dibedakan dengan yang mampu (lebih tahu). *Distracter* sudah berfungsi dengan baik jika sudah dipilih oleh lebih dari 5% pengikut tes ($p > 5\%$) dan jika kurang atau sama dengan 5% ($p \leq 5\%$), berarti distraktor tidak berfungsi dengan baik. Tujuan utama dari pemasangan *distracter* pada setiap butir item soal pada contoh di atas agar dari sekian banyak testee yang mengikuti tes hasil belajar ada yang tertarik atau terangsang untuk memilihnya sebab mereka menyangka bahwa *distracter* yang mereka pilih merupakan jawaban yang betul (Sukardi, 2008).

Menurut Mardapi (2008) ada sembilan langkah yang harus ditempuh dalam mengembangkan tes hasil belajar. Kesembilan langkah tersebut adalah: (1) pengembangan spesifikasi tes; (2) menulis soal tes; (3) menelaah soal tes; (4) melakukan uji coba tes; (5) menganalisis butir soal tes; (6), memperbaiki tes; (7) merakit tes; (8) melaksanakan tes; dan (9) menafsirkan hasil tes.

Pengembangan program pengajaran adalah rumusan-rumusan tentang apa yang dilakukan guru dan siswa dalam proses pembelajaran untuk mencapai tujuan, sebelum kegiatan pengajaran sesungguhnya dilaksanakan. Menurut Oemar Hamalik (2008), perangkat yang harus disiapkan dalam perencanaan pembelajaran antara lain: (1) memahami kurikulum; (2) menguasai bahan ajar; (3) menyusun program pengajaran; (4) melaksanakan program pengajaran; (5) menilai program pengajaran dan hasil proses belajar mengajar yang telah dilaksanakan.

METODOLOGI PENELITIAN

Jenis penelitian ini adalah penelitian pengembangan (*research and development*). Model pengembangan yang digunakan adalah model 4-D (*Define, Design, Develop, Dessiminate*) dari Thiagarajan and Semmel yang diadaptasi menjadi 4-P (Pendefinisian, Perancangan, Pengembangan, Penyebaran). Model ini digunakan untuk menghasilkan produk tertentu, dan menguji kevalidan, kepraktisan dan keefektifan produk tersebut. Dengan menggunakan model pengembangan 4-D. Perangkat penilaian yang dikembangkan adalah perangkat tes sumatif semester ganjil untuk siswa kelas IX SMP di Kabupaten Enrekang.

Tahap ujicoba berlokasi di SMPN 1 Enrekang dengan melakukan uji coba terbatas di kelas reguler semester ganjil tahun pelajaran 2016/2017, uji coba yang lebih luas dilaksanakan di beberapa sekolah menengah pertama di Kabupaten Enrekang. Tahap penyebaran dilakukan melalui forum MGMP se-Kabupaten Enrekang. Model atau rancangan penelitian yang digunakan adalah model pengembangan dari Thiagarajan, Semmel and Semmel yaitu model 4-D (*Define, Design, Develop, and Disseminate*) diadaptasikan menjadi model 4-P (Pendefinisian, Perancangan, Pengembangan, dan Penyebaran) (Trianto, 2007).

Penelitian ini dilaksanakan dalam beberapa tahap: 1). Tahap pendefinisian adalah menganalisis karakteristik siswa dan kurikulum KTSP untuk mengkaji standar kompetensi, kompetensi dasar, indikator, dan materi yang dijadikan bahan pembelajaran di kelas dalam penelitian. 2). Tahap Perancangan adalah menentukan metode dan format/tipe tes kemudian menyusun kisi-kisi tes dengan tipe yang dipilih dengan pokok bahasan. 3). Tahap pengembangan adalah menelaah soal tes, telaah soal untuk memperbaiki soal yang dalam pembuatannya terdapat kekurangan atau kesalahan. Telaah soal dilakukan ahli/validator kemudian melakukan revisi atau tidak berdasarkan hasil validator ahli. Selanjutnya digunakan untuk uji coba yang digunakan sebagai sarana untuk memperoleh data empirik tentang kebaikan soal yang telah disusun. Melalui uji coba dapat diperoleh data tentang reabilitas, dan validitas, tingkat kesukaran, daya beda dan *distracter* soal yang menjadi dasar untuk memperbaiki atau merevisi soal. 4). Tahap penyebaran atau tahap *disseminate*. Tahap ini merupakan tahap penggunaan yang dikembangkan pada skala yang lebih luas. Penyebaran perangkat tes pada penelitian ini dilakukan melalui Forum Musyawarah Guru Mata Pelajaran Biologi tingkat SMP.

Data validitas perangkat tes kognitif diperoleh melalui validitas ahli, data reliabilitas dan karakteristik internal siswa yang meliputi tingkat kesukaran, daya beda, dan *distractor* (pengecoh) diperoleh melalui uji coba pada siswa kelas IX SMPN 1 Enrekang.

Instrumen penelitian yang digunakan yaitu format telaah butir soal untuk ahli pendidikan biologi, format telaah butir soal praktisi, dan angket respon siswa terhadap tes yang telah dikembangkan. Format telaah butir soal mencakup aspek materi, konstruksi, dan bahasa yang digunakan. Instrumen ini sebelum digunakan telah divalidasi oleh ahli. Pengelolaan data dilakukan secara kualitatif dan kuantitatif. Validitas, reliabilitas, tingkat kesukaran dan daya beda dan efektivitas pengecoh merupakan komponen utama dalam pengembangan instrumen penilaian pada penelitian ini. Analisis kualitatif dilakukan oleh validator dan guru serta siswa yang dijadikan test terhadap instrumen penilaian berdasarkan komentar dan saran yang diberikan saat validasi.

HASIL PENELITIAN DAN PEMBAHASAN

Pengembangan perangkat tes sumatif dalam penelitian ini mengacu kepada pengembangan tes kognitif yang dilakukan secara sistematis dengan menggunakan langkah-langkah model 4-D (*define, design, develop, disseminate*) dari Thiagarajan, Semmel and Semmel yang diadaptasi menjadi 4-P (Pendefinisian, Perancangan, Pengembangan, dan Penyebaran). Ada enam langkah yang telah ditempuh dalam mengembangkan tes ini. Keenam langkah tersebut adalah (1) pengembangan spesifikasi tes, (2) menulis soal tes, (3) menelaah soal tes, (4) melakukan uji coba tes, (5) menganalisis butir soal tes, (6), merakit tes.

Penyusunan spesifikasi tes dimulai dengan menetapkan spesifikasi tes yang berisi uraian yang menunjukkan keseluruhan karakteristik yang harus dimiliki tes. Penyusunan spesifikasi tes mencakup: (1) menentukan tujuan tes, yaitu untuk tes formatif dan sumatif (2) menyusun kisi-kisi tes. Kisi-kisi soal memuat standar kompetensi dan kompetensi dasar, indikator, daftar pokok bahasan dan, subpokok bahasan yang diujikan, langkah ini merupakan tahap pendefinisian atau *define*.

Penulisan soal merupakan langkah menjabarkan indikator menjadi pertanyaan-pertanyaan yang memiliki karakteristik sesuai dengan kisi-kisi soal yang telah dibuat, butir soal ini merupakan draf awal, langkah ini merupakan tahap *design*.

Prototipe 1 kemudian divalidasi oleh dua validator ahli, kemudian dilakukan revisi. Dengan validasi butir soal ini diharapkan dapat semakin memperbaiki kualitas soal yang disusun. Hasil analisis validasi ahli menunjukkan bahwa tes yang dikembangkan dikategorikan sangat valid berdasarkan kriteria yang ditetapkan dengan nilai 1,00. Kriteria sangat valid yang telah ditetapkan menurut konsep validitas isi yang menyatakan bahwa jika koefisien validitas ($>75\%$) atau 0,75 maka dapat dinyatakan koefisien validitas isi yang dihasilkan adalah valid. Karena validator ahli mengevaluasi butir tes menggunakan skala empat, maka penilaian dari setiap penilai pada setiap butir

dapat dikotomiskan, menjadi relevansi lemah (nilai 1 atau 2) melawan relevansi kuat (nilai 3 atau 4), berdasarkan nilai yang diberikan oleh para penilai ahli semua butir soal berada pada nilai tiga atau empat yang dapat dikotomiskan pada relevansi kuat. Perangkat tes yang telah direvisi dan divalidasi diujicobakan di SMPN 1, SMPN 2 dan SMPN 5 Enrekang dengan pertimbangan bahwa sekolah-sekolah tersebut merupakan Sekolah Standar Nasional (SSN). Uji coba digunakan sebagai sarana untuk memperoleh data empirik tentang kebaikan soal yang telah disusun.

Melalui uji coba yang dilakukan diperoleh data tentang validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas pengecoh. Berdasarkan hasil uji coba tersebut dilakukan analisis butir soal. Untuk analisis reabilitas, tingkat kesukaran, daya beda dan kualitas pengecoh dapat dilakukan dengan menggunakan program *Microsoft office Excell 2007* dan SPSS 18. Dasar penggunaan bantuan software tersebut sesuai dengan pendapat Widyoko (2010), bahwa selain menghitung secara manual dengan rumus yang telah ditentukan ada beberapa cara (software) yang dapat digunakan untuk melakukan analisis butir tes dengan mudah dan praktis yang menggunakan software yang khusus dibuat untuk tujuan tersebut seperti ANATES, bisa juga menggunakan software lain seperti SPSS, ITEMAN, Lisrel, dan *microsoft office excel*.

Analisis butir soal dapat dilakukan dengan dua cara sebagaimana yang diungkapkan Lababa (2008) yaitu analisis secara kualitatif teoritik dapat dilakukan sebelum soal-soal diberikan kepada peserta tes. Analisis ini dilakukan dengan cara menelaah kesesuaian dengan kemampuan dasar dan indikator yang hendak diukur serta apakah butir-butir soal tersebut telah memenuhi persyaratan dari aspek materi, konstruksi dan bahasa. Analisis secara kuantitatif dapat dilakukan dengan menghitung validitas, reliabilitas, tingkat kesukaran soal, daya beda, dan pengecoh (*distractor*).

Untuk analisis validitas dan reliabilitas isi berdasarkan penilaian ahli berada pada kategori valid. Analisis validitas isi untuk mengetahui tingkat efisiensinya yaitu relevansi kedua pakar dengan menggunakan teknik yang dikembangkan oleh Gregory, berupa koefisien validator isi menunjukkan angka 1,00 yang berarti sangat tinggi, sesuai dengan ketentuan yang ditetapkan bahwa kriteria 0,9-1,0 adalah sangat tinggi. Analisis reliabilitas dengan menggunakan rumus *percentage of agreements* menunjukkan angka 0,9 yang berarti reliabel. Hal ini sesuai dengan pendapat Nurdin (2007), bahwa instrumen penilaian dikatakan reliabel jika nilai reliabilitasnya lebih besar dari 0,75, sehingga layak untuk digunakan.

Uji empirik yang dilakukan terdiri atas dua tahap, yaitu tahap pertama yaitu uji coba terbatas dengan 35 sampel dan uji tahap kedua secara lebih luas dengan banyaknya responden 150. Hal ini dilakukan untuk melihat butir-butir soal mana yang tidak layak digunakan atau layak digunakan berdasarkan kriteria analisisnya.

Hasil analisis validitas butir pada uji empirik pertama dengan menggunakan rumus korelasi pearson melalui *software* SPSS 18 menunjukkan bahwa terdapat 15 soal dari 90 soal yang diujikan yang tidak valid/drop sehingga tersisa 75 soal yang valid yang akan diujikan secara lebih luas dengan jumlah sampel yang lebih banyak. Hasil analisis reliabilitas pada uji empirik pertama dengan menggunakan rumus K-R 20 dari

Kuder Richardson (KR) melalui *software* SPSS 18. Hasil analisisnya adalah 0,92, dimana kriteria derajat realibilitas suatu tes mengacu pada kriteria yang dibuat oleh Nurdin (2007) yang menyatakan bahwa 0,92 termasuk dalam rentang $0,80 < R \leq 1,00$ dengan konsistensi sangat tinggi.

Hasil analisis reliabilitas pada uji empirik kedua diperoleh melalui cara yang sama dalam pengerjaan uji empirik pertama, tetapi dalam hal ini perbedaan jumlah responden menyebabkan perbedaan koefisien korelasi pearson pada taraf signifikannya yang menunjukkan bahwa terdapat 22 soal dari 75 soal yang diujikan yang tidak valid/drop, sehingga tersisa 53 soal yang valid, sedangkan hasil analisis reliabilitasnya adalah 0,82 dimana kriteria derajat realibilitas suatu tes mengacu pada kriteria yang dibuat oleh Nurdin (2007) yang menyatakan bahwa 0,83 termasuk dalam rentang $0,80 < R \leq 1,00$ dengan konsistensi sangat tinggi.

Untuk analisis tingkat kesukaran butir soal menunjukkan bahwa soal dengan kategori mudah sebanyak 38,67%, kategori sedang sebanyak 58,67% dan kategori sukar sebanyak 2,67%. Kategori tersebut sesuai dengan pendapat Sukiman (2012) bahwa untuk tes sejenis formatif, proporsi tingkat kesukarannya adalah 25% kategori mudah, 50% kategori sedang dan 25% kategori sukar. Proporsi logis antara soal mudah, sedang dan sukar adalah 2 : 6 : 2 atau 20% mudah, 60% sedang, dan 20% soal sukar. Namun dalam penelitian ini tidak sejalan dengan standar proporsinya, khususnya untuk soal yang sukar tidak mencapai 20%, tetapi proporsi soal yang mudah dan sedang sejalan dengan proporsi 20% soal mudah dan 60% soal sedang, sehingga dapat disimpulkan bahwa soal-soal yang dikembangkan dalam penelitian ini berdistribusi normal.

Analisis daya beda butir menunjukkan bahwa soal dengan kategori daya beda lemah sebanyak 36%, soal dengan kategori daya beda cukup sebanyak 56% dan soal dengan daya beda baik sebanyak 8%. Soal yang berdaya beda lemah merupakan soal yang mudah, soal yang mudah juga perlu dimasukkan ke dalam sebuah tes. Jika soal-soal dalam tes tidak ada yang mudah, maka siswa yang berkemampuan rendah akan merasa putus asa karena soal yang diberikan di luar kemampuannya untuk menjawab. Menurut Ivanty, dkk (2013), soal yang dapat diterima adalah soal yang memiliki daya beda baik sekali, baik dan cukup, sedangkan soal yang direvisi/diteliti adalah soal yang tidak memiliki daya beda, dan berdaya beda lemah.

Menurut Sukiman (2012) semakin tinggi indeks daya pembeda soal berarti semakin mampu soal yang bersangkutan membedakan peserta didik yang pandai dengan peserta didik yang kurang pandai.

Soal yang baik adalah soal yang dapat dijawab dengan benar oleh peserta didik (*testee*) yang pandai atau menguasai materi yang diteskan dan tidak dapat dijawab benar oleh peserta didik yang tidak pandai atau tidak menguasai materi yang diteskan. Soal yang tidak baik adalah soal yang ketika digunakan muncul tiga kemungkinan berikut: 1) peserta didik yang pandai dan yang tidak pandai sama-sama mampu menjawab dengan benar, 2) peserta didik yang pandai dan yang tidak pandai sama-sama tidak dapat menjawab dengan benar, atau 3) peserta didik yang pandai tidak dapat menjawab

dengan benar, sebaliknya peserta didik yang tidak pandai justru dapat menjawab dengan benar.

Untuk analisis pengecoh (*distractor*), menunjukkan bahwa banyaknya soal yang memiliki pengecoh yang baik berfungsi efektif sebanyak 89,33%, sedangkan soal yang pengecohnya tidak berfungsi dengan efektif sebanyak 10,67%. Soal yang memiliki option tidak efektif (tidak ada satu pun siswa yang memilih) berarti option tersebut direvisi, tidak dibuang. Pendapat tersebut sesuai dengan pendapat Arikunto (2010) 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 (Purwanto, 2010).

Analisis butir soal tes menentukan soal yang baik, kurang baik, dan tidak baik. Soal yang dinyatakan baik bila soal tersebut tidak terlalu mudah dan atau tidak terlalu sukar. Soal yang terlalu mudah tidak menantang atau merangsang peserta didik untuk berusaha dengan serius memberikan jawaban. Sebaliknya soal yang terlalu sukar akan menyebabkan peserta didik kehilangan semangat untuk berusaha memecahkannya karena diluar jangkauan kemampuannya. Dalam menganalisis butir soal tes dilakukan dalam tiga tahapan yaitu: pertama, menentukan tingkat kesukaran soal; kedua, menentukan daya pembeda soal; dan ketiga, mengetahui pola jawaban soal (Widyoko, 2010).

Keputusan suatu item soal layak digunakan, perlu direvisi atau ditolak berdasarkan pada kriteria keputusan untuk penilaian item soal oleh Ivanti, dkk. (2013) sebagai berikut: 1) Item soal diterima, apabila karakteristik item soal memenuhi semua kriteria. Item soal yang terlalu sukar atau mudah, tetapi memiliki daya beda dan distribusi pengecoh item yang memenuhi kriteria, butir soal tersebut dapat diterima atau dipilih. 2) Item soal direvisi, apabila salah satu atau lebih dari ketiga kriteria karakteristik item soal tidak diterima. 3) Item soal ditolak, jika item soal memiliki karakteristik yang tidak memenuhi semua kriteria.

Hasil analisis tes menunjukkan bahwa dari 90 butir soal yang dikembangkan, setelah dilakukan uji coba terbatas (empirik), maka diperoleh 53 butir soal yang valid dianggap sudah baik dan tidak perlu di revisi, tetapi untuk butir soal nomor 52 dan 71 berdasarkan kesimpulan hasil analisis termasuk dalam kategori tidak digunakan, namun dalam hal ini untuk penggunaan selanjutnya dapat dilakukan revisi terhadap butir soal tersebut baik secara isi maupun konstruk butir soal tersebut. Langkah-langkah di atas merupakan tahapan pengembangan tes sumatif atau tahap *development*.

Langkah selanjutnya yaitu memperbaiki tes atau soal sesuai dengan saran-saran dari guru maupun validator ahli serta berdasarkan hasil analisis uji coba terhadap tes yang masih belum memenuhi kriteria sebagai tes yang baik. Selanjutnya merakit tes yaitu menyusun kembali soal yang telah diperbaiki/direvisi setelah uji coba, dengan demikian dihasilkan sebuah perangkat tes sumatif yang dapat digunakan untuk mengukur kemampuan siswa dan telah diuji kualitasnya. Setelah tes sudah dirakit dengan baik langkah selanjutnya adalah penyebaran perangkat tes. Penyebaran

perangkat tes pada penelitian ini dilakukan melalui Forum Musyawarah Guru Mata Pelajaran Biologi tingkat SMP. Langkah ini merupakan tahap penyebaran atau *dessiminate*.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan penelitian yang telah dikemukakan, serta dihubungkan dengan pertanyaan penelitian, maka dapat disimpulkan beberapa hal pokok yang berkaitan dengan pengembangan tes sumatif mata pelajaran biologi semester ganjil kelas IX SMP di Kabupaten Enrekang sebagai berikut: 1). Proses pengembangan perangkat menggunakan model pengembangan perangkat Thiagarajan yang terdiri dari empat tahap yaitu (1) pendefinisian (*define*), (2) perancangan (*design*), (3) pengembangan (*develop*), dan (4) penyebaran (*dessiminate*). Proses pengembangan pada dua tahap pertama menghasilkan prototipe 1 perangkat tes kognitif tipe pilihan ganda sebanyak 90 butir. Pengembangan perangkat pada tahap selanjutnya dicapai; 1) valid; (a) validitas isi berdasarkan penilaian ahli >75% dan praktisi dengan sedikit revisi, (b) validitas kriteria pertama dan kedua dari 90 butir tes tersisa 53 tes yang valid dan dikategorikan baik (c) validitas konstruk memerlukan revisi atau peninjauan ulang 2) reliabel; (a) nilai reliabilitas tes pada uji empirik pertama 0,92 dan kedua 0,82 , (3) tingkat kesukaran sedang, (3) daya beda cukup, (4) kualitas pengecoh baik dan (5) respon siswa dan guru terhadap perangkat tes dalam kategori positif.

Berdasarkan hasil dan temuan yang diperoleh dalam penelitian ini, dikemukakan beberapa saran sebagai berikut: 1). kepada guru, dalam pembuatan tes yang akan diberikan kepada siswa diharapkan benar-benar tes telah teruji validitas dan reliabilitasnya. 2). untuk menambah kompetensi guru yang ada di masing-masing sekolah, pihak sekolah diharapkan dapat memfasilitasi guru-guru seperti menghadirkan para pakar yang berkompeten, workshop mandiri atau pelatihan tentang evaluasi/penilaian.

DAFTAR PUSTAKA

- Anderson. *Et al.* (2001). *A Taxonomy For Learning, Teaching, And Asuming: A revision of Bloom's Of Educational Objectives*. New York: Longman.
- Arikunto, S. (2005). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arikunto, S. (2010). *Dasar-dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: Bumi Aksara.
- Departemen Pendidikan Nasional. (2004). *Teknik Evaluasi*. Jakarta: Diknas.
- Hamalik, O. (2008). *Perencanaan Pengajaran Berdasarkan Pendekatan Sistem*. Jakarta: PT Bumi Aksara.
- Ivanty, D.W.N, Aminah, N.S, Ekawati, E.Y. (2013). Penyusunan Instrumen Tes Tengah Semester Genap Fisika X SMA Untuk Kelas X SMA. *Jurnal Pendidikan*

- Fisika, (Online)*, Vol. 1, No. 1, ISSN: 2338-0691, Diakses 1 Juli 2017.
- Kusaeri & Suprananto, (2012). *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha ilmu.
- Lababa, J. (2008). Analisis Butir Soal dengan Teori Tes Klasik: Sebuah Pengantar. *Jurnal Iqra', (online)*, Vol. 5 Januari – Juni 2008. Diakses 1 Juli 2017.
- Mardapi. (2005). *Pengembangan Sistem Penilaian Berbasis Kompetensi*. Yogyakarta: HEPI.
- Mardapi. (2008). *Teknik Penyusunan Instrumen Tes dan Non Tes*. Yogyakarta: Mitra Cendekia.
- Nuridin. (2007). Model Pembelajaran Matematik yang Menumbuhkan Kemampuan Metakognitif Untuk Menguasai Bahan Ajar. *Disertasi*. Tidak diterbitkan. Surabaya: Program Pascasarjana universitas Negeri Surabaya.
- Nurhadi & Suwardi. (2010). *Evaluasi Pembelajaran yang Efektif dan Menyenangkan*. Jakarta: Multi Kreasi.
- Nuryani, R. (2007). *Asessmen dalam Pembelajaran Sains*. Bandung: Univesitas Pendidikan Indonesia (UPI).
- Purwanto. N. M. (2010). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosda Karya.
- Rasyid & Mansur. (2007). *Penilaian Hasil Belajar*. Bandung: Wacana Prima.
- Sukardi. (2008). *Evaluasi Pendidikan Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara.
- Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Madani
- Trianto. (2007). *Model Pembelajaran Inovatif Berorientasi Konstruktivistik*. Surabaya: Pustaka Ilmu.
- Uno, Hamzah B. & Satria Koni. (2012). *Asessmen Pembelajaran*. Jakarta: PT. Bumi Aksara.
- Widyoko P.E. (2010). *Evaluasi Program Pengajaran*. Yogyakarta: Pustaka Pelajar.