

## PENGEMBANGAN TES KOGNITIF BERBASIS REVISI TAKSONOMI BLOOM PADA MATERI SISTEM REPRODUKSI UNTUK SISWA SMA

**Ainul Uyuni Taufiq**

Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar,  
Kampus II Jl. H. M. Yasin Limpo No 36 Samata-Gowa, Sulawesi Selatan (92118)  
Telepon: 085242728022 / E-mail: ainul.uyuni@uin-alauddin.ac.id

### Abstrak

Penelitian ini bertujuan untuk mengetahui kualitas perangkat tes yang dikembangkan. Kualitas perangkat tes dari segi validitas, reliabilitas, tingkat kesukaran, daya beda, dan pengecoh atau distraktor.

Penelitian ini menggunakan pendekatan kualitatif dengan metode penelitian pengembangan. Pengembangan perangkat menerapkan prosedur pengembangan perangkat model 4-D yang terdiri atas tahap *define, design, develop* dan *dessiminate*.

Data penelitian diperoleh melalui 2 kriteria yaitu (1) data validitas diperoleh melalui penilaian ahli, (2) data validitas butir soal, reliabilitas, tingkat kesukaran, daya beda, dan pengecoh atau distractor diperoleh melalui hasil uji coba tes kognitif. Data validitas yang diperoleh dianalisis secara matematis dan dikategorikan berdasarkan standar yang telah ditetapkan, sedangkan data validitas butir soal, reliabilitas, tingkat kesukaran, daya beda, dan pengecoh atau distractor dianalisis dengan menggunakan bantuan Microsoft Excel 2007.

Hasil validasi menunjukkan bahwa tes kognitif memiliki kategori sangat valid dengan nilai 3,54 berdasarkan penilaian validator ahli. Selanjutnya dilakukan uji coba untuk mengetahui kualitas tes kognitif. Dari segi validitas butir soal, kategori valid (92% untuk tipe A, 92% untuk tipe B), tidak valid (8% untuk tipe A, 8% untuk tipe B). Dari segi reliabilitas, termasuk dalam kategori reliabel dengan nilai 0,71 untuk soal tipe A dan 0,74 untuk soal tipe B. Dari segi tingkat kesukaran, kategori mudah (30% untuk tipe A, 22% untuk tipe B), sedang (48% untuk tipe A, 65% untuk tipe B), sukar (22% untuk tipe A, 13% untuk tipe B). Dari segi daya beda, kategori lemah (9% tipe A, 4% tipe B), cukup (35% tipe A, 31% tipe B), baik (52% tipe A, 61% tipe B), baik sekali (4% tipe A, 4% tipe B). Dari segi pengecoh, kategori efektif (87% tipe A, 91% tipe B), tidak efektif (13% tipe A, 9% tipe B).

**Kata kunci:** Tes Kognitif, Analisis Butir Soal, Sistem Reproduksi.

### Abstract

*This study aims to determine the quality of the test device was developed. In terms of the quality of test validity, reliability, difficulty level, determinant indeks, and effectiveness of distractors.*

*This study used a qualitative approach to the development of research methods. Development of procedures to implement the software development model 4-D*

*which consists of phases define, design, develop and disseminate.*

*The data were obtained through two criteria: (1) the validity of the data obtained through experts assessment, (2) item validity data, reliability, difficulty level, determinant index, and effectiveness of distractors obtained through the test results of cognitive tests. Validity of the data obtained was analyzed mathematically and categorized based on the standards set, while item validity data, reliability, difficulty level, determinant index, and effectiveness of distractors analyzed with the help of Microsoft Excel 2007.*

*Validation results showed that cognitive tests have very valid category with a value of 3.54 based on experts assessment. The next conducted trials to determine the quality of the cognitive tests. In terms of item validity, valid category (92% for type A, 92% for type B), not valid (8% for type A, 8% for type B). In terms of reliability, including in the category of reliable with values about 0,71 for type A and 0,74 for type B. In terms of difficulty level, easy category (30% for type A, 22% for type B), medium category (48% for type A, 65% for type B), difficult category (22% for type A, 13% for type B). Determinant index, poor category (9% type A, 4% type B), satisfactory (35% type A, 31% type B), good (52% type A, 61% type B), excellent (4% type A, 4% type B). Effectiveness of distractors, effective category (87% type A, 91% type B), not effective category (13% type A, 9% type B)*

**Keywords:** *Cognitive test, analysis of items, reproduction system*

## PENDAHULUAN

Pendidikan adalah kegiatan sadar tujuan yang mengandung tanggung jawab. Menurut konsep pertanggungjawaban dalam pendidikan (*accountability*), siapa saja yang disertai tugas untuk mendidik harus dapat mempertanggungjawabkan tugasnya. Tanggung jawab itulah yang mengharuskan pendidikan mempersiapkan materi yang akan diajarkan, memilih metode dan alat bantu yang akan digunakan dalam mengajar, membuat alat evaluasi untuk mengetahui apakah pelajaran yang disajikan itu berhasil diketahui oleh siswa. Untuk dapat mempertanggungjawabkan suatu tugas, perlu adanya alat yang dapat memberi petunjuk bahwa tugas itu berhasil atau tidak. Di sinilah letak betapa pentingnya evaluasi dalam proses belajar mengajar.

Peranan evaluasi dalam pendidikan sangat penting. Oleh sebab itu, seluk beluknya harus mendapat perhatian sepenuhnya. Untuk mengetahui kemampuan yang dimiliki oleh siswa kita harus memiliki informasi mengenai siswa, yang diperoleh melalui evaluasi, yang dilakukan oleh guru. Evaluasi memberikan informasi mengenai hasil belajar yang telah dimiliki oleh siswa. Dengan informasi tersebut guru dapat menentukan apakah tujuan yang telah ditetapkan itu telah tercapai atau belum (Sahabuddin, 2007).

Fungsi penting bagi siswa dalam evaluasi belajar adalah umpan balik kepada siswa tersebut untuk mengetahui hasil belajar dari proses pembelajaran yang dilakukan. Pengetahuan dan pemahaman pada pencapaian hasil belajar siswa akan membantu guru

untuk mengadakan refleksi guna memperbaiki kinerjanya di masa yang akan datang; sebab informasi tersebut adalah sangat penting untuk merencanakan pembelajaran selanjutnya. Masukan dari evaluasi hasil belajar mungkin akan merubah berbagai metode dan strategi pembelajaran yang biasa dilakukan atau mungkin perlu tidaknya penambahan media pembelajaran dilakukan. Umpan balik dari evaluasi pembelajaran minimal memberikan beberapa fungsi sebagai berikut: (a) guru dan siswa memperoleh pengetahuan tentang seberapa besar suatu tujuan pembelajaran atau suatu kompetensi sudah dikuasai; dan (b) guru mengetahui efektifitas dari program pembelajaran yang dilaksanakan (Sudaryono, 2012).

Syarat-syarat yang harus dipenuhi oleh guru agar mereka dapat mengevaluasi hasil belajar siswa dengan baik, antara lain: (a) menguasai dan memahami berbagai peraturan perundang-undangan yang di dalamnya terdapat pasal-pasal mengenai evaluasi hasil belajar; dan (b) menguasai teori evaluasi pembelajaran.

Mengingat kompleksnya proses penilaian, guru perlu memiliki pengetahuan dan keterampilan yang memadai. Dalam tahap persiapan terdapat beberapa kegiatan, antara lain penyusunan tabel spesifikasi yang didalamnya terdapat sasaran penilaian, teknik penilaian serta jumlah instrumen yang diperlukan. Pada tahap pelaksanaan, dilakukan pemakaian instrument untuk menemukan respon peserta didik terhadap instrument tersebut sebagai bentuk hasil belajar, selanjutnya dilakukan penelitian terhadap data yang telah dikumpulkan dan dianalisis untuk membuat tafsiran tentang kualitas prestasi belajar peserta didik, baik dengan acuan criteria maupun dengan acuan kelompok.

Untuk melaksanakan evaluasi hasil mengajar dan belajar itu, seorang guru dapat menggunakan dua macam tes, yakni tes yang telah distandarkan (*standardized test*) dan tes buatan guru sendiri (*teacher-made test*). Yang dimaksud dengan *standardized test* ialah tes yang telah mengalami proses standardisasi, yakni proses validasi dan keandalan (*reliability*) sehingga tes tersebut benar-benar valid dan andal untuk suatu tujuan dan bagi suatu kelompok tertentu (Purwanto, 2004).

Berdasarkan hasil observasi di lapangan yaitu di beberapa sekolah SMA yang ada di Makassar menunjukkan bahwa guru di sekolah melakukan evaluasi, khususnya pada materi sistem reproduksi kelas XI semester genap menggunakan alat yang belum memenuhi kriteria sebagai alat tes yang baik, sehingga kemampuan siswa tidak dapat terukur dengan baik oleh guru. Selain itu, tes yang digunakan tidak diuji cobakan terlebih dahulu sehingga syarat validitas dan reliabilitas tidak terpenuhi, demikian halnya dengan analisis butir soal yang meliputi: tingkat kesukaran, daya beda, dan kualitas pengecoh tidak dilakukan. Tes yang dipakai jika ditinjau dengan menggunakan revisi taksonomi Bloom hanya berada pada kisaran C1 sampai C3. Pada materi sistem reproduksi siswa dituntut untuk memiliki kompetensi kognitif sampai C6, hal inilah yang sulit dilakukan oleh guru untuk membuat soal tes hingga tingkatan kognitif C6 apalagi jika soalnya berbentuk tes objektif (pilihan ganda). Oleh karena itu, peneliti tertarik untuk mengembangkan perangkat tes kognitif berorientasi revisi Taksonomi

Bloom pada materi sistem reproduksi, dimana pengembangan ini diharapkan mampu memenuhi kriteria tes yang baik sehingga pelaksanaan evaluasi di sekolah juga dapat menggunakan alat evaluasi yang baik.

### **Rumusan Masalah**

Berdasarkan latar belakang yang telah dibahas sebelumnya, maka rumusan masalah dalam penelitian ini adalah : (1) Bagaimana kualitas tes kognitif yang berbasis revisi taksonomi Bloom ditinjau dari segi validitas dan reliabilitas pada materi sistem reproduksi untuk siswa SMA?; (2) Bagaimana kualitas tes kognitif yang berbasis revisi taksonomi Bloom ditinjau dari segi karakteristik internal yang meliputi tingkat kesukaran, daya beda, dan kualitas pengecoh (*distractor*) pada materi sistem reproduksi untuk siswa SMA?

### **METODOLOGI PENELITIAN**

Jenis penelitian ini adalah penelitian pengembangan (*development research*). Dengan menggunakan model pengembangan 4-D (*Define, Design, Develop, Disseminate*) dari Thiagarajan, Semmel and Semmel yang diadaptasi menjadi 4-P (Pendefinisian, Perancangan, Pengembangan, Penyebaran). Tes kognitif yang dikembangkan adalah tes kognitif materi sistem reproduksi berbasis revisi taksonomi Bloom untuk siswa kelas XI SMA di Kota Makassar. Lokasi uji coba di SMA Negeri 2 Makassar dan SMA Negeri 14 Makassar.

Model pengembangan perangkat 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) menurut Trianto (2009). Penelitian ini dilaksanakan dalam beberapa tahap sebagai berikut:

1. Tahap pendefinisian adalah menganalisis karakteristik siswa, pada penelitian ini yang dikaji adalah pengelompokan umur siswa untuk menentukan tingkat kognitif soal yang disusun yang berdasar pada teori Piaget, mengkaji standar kompetensi, kompetensi dasar, indikator, tujuan pembelajaran yang harus dicapai dalam pembelajaran di kelas;
2. Tahap Perancangan adalah menentukan metode dan format/tipe tes yaitu soal yang berbasis revisi taksonomi Bloom, tipe tes berbentuk pilihan ganda kemudian menyusun kisi-kisi tes. Selanjutnya penulisan soal yang merupakan langkah penjabaran indikator menjadi pertanyaan-pertanyaan yang memiliki karakteristik sesuai dengan kisi-kisi soal yang telah dibuat. Soal yang terdapat dalam tes ini ada yang dibuat sendiri oleh penulis, modifikasi dari buku, dan mengambil soal dari buku. Tes pada tahap ini merupakan draft awal.
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 reliabilitas, dan validitas, tingkat kesukaran, daya beda dan *distractor* 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, yaitu mensosialisasikan tes kognitif pada forum MGMP (Musyawarah Guru Mata Pelajaran).

Data validitas perangkat tes kognitif diperoleh melalui validasi ahli, data reliabilitas dan karakteristik butir soal meliputi tingkat kesukaran, daya beda, dan *distractor* (pengecoh) diperoleh melalui uji coba pada siswa kelas XI SMA Negeri 2 Makassar dan SMA Negeri 14 Makassar dengan jumlah sampel sebanyak 291 orang siswa.

Instrumen penelitian yang digunakan yaitu format telaah butir soal untuk ahli pendidikan, format telaah butir soal ahli konten, 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.

Tes yang telah dibuat (draf awal) akan divalidasi oleh validator ahli kemudian direvisi, selanjutnya akan divalidasi oleh praktisi sebelum uji coba terbatas. Penilaian validator terhadap tes hasil belajar mencakup tiga aspek yaitu: materi, konstruksi, dan bahasa.

Soal dalam tes kognitif ini dianalisis validitas butirnya untuk setiap item soal. Analisis dilakukan dengan mengkorelasikan skor butir soal setiap nomor yang dicapai oleh setiap siswa dengan skor total seluruh soal yang dicapai oleh setiap siswa. Skor butir soal menjadi variabel X dan skor total seluruh soal menjadi variabel Y. Selanjutnya dihitung reliabilitasnya dengan menggunakan program *Microsoft Excel 2007*. Soal yang dihitung reliabilitasnya adalah soal yang valid berdasarkan perhitungan analisis validitas butir soal untuk setiap item, sedangkan soal yang tidak valid tidak dihitung reliabilitasnya. Rumus yang digunakan untuk menghitung reliabilitas adalah rumus K-R20. Uji tingkat kesukaran soal, daya pembeda, dan *distractor* dianalisis dengan program *Microsoft Excel 2007*.

## HASIL PENELITIAN DAN PEMBAHASAN

Tes kognitif yang telah dikembangkan dalam penelitian ini mengikuti prosedur pengembangan perangkat model 4-D (*define, design, develop, disseminate*) dari Thiagarajan, Semmel and Semmel yang diadaptasi menjadi 4-P (Pendefenisian, Perancangan, Pengembangan, Penyebaran). Adapun tahap-tahap pengembangan yang dilakukan selama penelitian diuraikan sebagai berikut.

## 1. Tahap pendefinisian (*define*)

Tahap pendefinisian diawali dengan analisis SKL (Standar Kompetensi Lulusan), SK (Standar Kompetensi), KD (Kompetensi Dasar), indikator, tujuan pembelajaran, materi untuk penyusunan soal, dan analisis kemampuan kognitif siswa.

**SKL (Standar Kompetensi Lulusan):** Menjelaskan sistem reproduksi manusia dan proses pembentukan sel kelamin.

**SK (Standar Kompetensi):** Menjelaskan struktur dan fungsi organ manusia dan hewan tertentu, kelainan/penyakit yang mungkin terjadi serta implikasinya pada salingtemas

**KD (Kompetensi Dasar):** Menjelaskan keterkaitan antara struktur, fungsi, dan proses yang meliputi pembentukan sel kelamin, ovulasi, menstruasi, fertilisasi, kehamilan, dan pemberian ASI, serta kelainan/penyakit yang dapat terjadi pada sistem reproduksi manusia.

Indikator pencapaian kompetensi dasar:

- 1) Menjelaskan struktur dan fungsi organ reproduksi pada manusia.
- 2) Menjelaskan proses gametogenesis (spermatogenesis dan oogenesis).
- 3) Menjelaskan hormon-hormon yang terlibat dalam sistem reproduksi manusia.
- 4) Menjelaskan mekanisme terjadinya menstruasi
- 5) Menjelaskan proses ovulasi
- 6) Menjelaskan proses terjadinya fertilisasi, kehamilan, dan persalinan
- 7) Menjelaskan pentingnya pemberian ASI
- 8) Menjelaskan alat kontrasepsi.
- 9) Menjelaskan kelainan/penyakit pada sistem reproduksi manusia.

Pada penelitian ini, dikembangkan 50 tujuan pembelajaran yang mengacu pada sembilan indikator yang disebutkan di atas. Dari tujuan pembelajaran tersebut kemudian disusun menjadi soal/tes kognitif. Tujuan pembelajaran yang telah dikembangkan disusun dengan mempertimbangkan indikator yang dibuat oleh guru di sekolah.

## 2. Tahap perancangan (*design*)

Pada tahap perancangan, peneliti menetapkan jenis soal yang dikembangkan adalah soal pilihan ganda (*multiple choice*) dengan pertimbangan bahwa soal pilihan ganda cocok untuk digunakan jika peserta tes banyak. Hal ini dilakukan mengingat bahwa jumlah sampel uji coba dalam penelitian ini berjumlah 291 orang. Metode penilaian yang dilakukan adalah untuk jawaban benar skornya 1 (satu) dan untuk jawaban salah skornya 0 (nol). Jumlah soal yang disusun sebanyak 50 butir soal (dibuat sendiri oleh penulis sebanyak 18 soal, modifikasi dari buku sebanyak 21 soal, mengambil dari buku sebanyak 11 soal) kemudian soal terbagi ke dalam 2 (dua) bagian yaitu bagian A sebanyak 25 butir soal (dibuat sendiri oleh penulis sebanyak 7 soal yaitu soal nomor 5, 7, 9, 13, 22, 23, 24, modifikasi dari buku sebanyak 11 soal yaitu soal nomor 1, 3, 6, 10, 12, 15, 16, 17, 18, 20, 21, mengambil dari buku sebanyak 7 soal yaitu

soal nomor 2, 4, 8, 11, 14, 19, 25) dan bagian B sebanyak 25 butir soal (dibuat sendiri oleh penulis sebanyak 11 soal yaitu soal nomor 1, 2, 6, 10, 14, 15, 17, 18, 21, 22, 23, modifikasi dari buku sebanyak 10 soal yaitu soal nomor 4, 5, 8, 11, 12, 13, 16, 19, 20, 25, mengambil dari buku sebanyak 4 soal yaitu soal nomor 3, 7, 9, 24). Hal ini dilakukan dengan pertimbangan bahwa soal yang disusun adalah tipe soal untuk ulangan formatif (ulangan harian). Untuk ulangan harian, 50 butir soal terlalu banyak dan sangat melelahkan bagi siswa.

### 3. Tahap pengembangan (*develop*)

Tahap pengembangan dilakukan dengan cara menelaah soal-soal yang telah disusun dengan tujuan untuk memperbaiki soal-soal yang masih terdapat kekurangan/kesalahan. Telaah soal dilakukan oleh validator ahli yang terdiri dari 1 orang ahli pendidikan bidang evaluasi dan 1 orang ahli konten/materi reproduksi. Telaah soal juga dilakukan oleh praktisi (guru) di sekolah. Selanjutnya dilakukan uji coba terbatas di 2 (dua) sekolah yaitu SMA Negeri 2 Makassar dan SMA Negeri 14 Makassar dengan jumlah sampel 291 orang. Uji coba dilakukan untuk menguji kualitas tes yang telah dikembangkan. Melalui uji coba dapat diperoleh data tentang validitas butir setiap item soal, reliabilitas, tingkat kesukaran, daya beda, dan *distractor* (pengecoh) soal yang menjadi dasar untuk memperbaiki atau merevisi soal.

#### a. Hasil analisis validasi ahli

Analisis validitas soal dilakukan berdasarkan hasil penilaian validator ahli evaluasi pendidikan dan ahli konten/materi reproduksi. Hasil penilaian kedua validator ahli tersebut dapat dilihat pada Tabel 1:

**Tabel 1. Rekap Hasil Analisis Validasi Ahli**

No	Aspek yang dinilai	Hasil penilaian		Rerata Kriteria	Keterangan
		Validator 1	Validator 2		
1.	Materi	3,50	3,75	3,63	Sangat Valid
2.	Konstruksi	3,75	3,20	3,48	Valid
3.	Bahasa	3,67	3,33	3,50	Sangat Valid
Rerata Aspek				3,54	Sangat Valid

Sumber: Data Peneliti

Sajian data pada Tabel 4.3 menunjukkan bahwa tes kognitif yang dikembangkan berada pada kategori sangat valid dengan nilai 3,54. Hal ini sesuai dengan ketentuan yang ditetapkan yaitu hasil penilaian validator yang berada pada rentang skor  $3,5 \leq M \leq 4$  termasuk ke dalam kategori sangat valid. Dengan demikian kedua validator ahli memutuskan tes kognitif yang dikembangkan dinyatakan sangat valid dan layak digunakan untuk uji coba dengan revisi sedikit.

#### b. Hasil analisis validitas angket respon siswa dan guru

Analisis validitas soal dilakukan berdasarkan hasil penilaian dua validator ahli. Hasil penilaian kedua validator ahli tersebut dapat dilihat pada Tabel 2:

**Tabel 2. Rekap Hasil Analisis Validitas Angket Respon Siswa dan Guru Berdasarkan Penilaian Ahli**

No	Aspek yang dinilai	Hasil penilaian		Rerata Kriteria	Keterangan
		Validator 1	Validator 2		
1.	Petunjuk	4,00	3,50	3,75	Sangat Valid
2.	Bahasa	3,50	3,25	3,38	Valid
3.	Isi	3,20	3,00	3,10	Valid
Rerata Aspek				3,41	Valid

Sumber: Data Peneliti

Berdasarkan Tabel 2 terlihat bahwa angket respon siswa dan guru terhadap tes kognitif yang dikembangkan berada pada kategori valid dengan nilai 3,41. Hal ini sesuai dengan ketentuan yang ditetapkan yaitu hasil penilaian validator yang berada pada rentang skor  $2,5 \leq M \leq 3,5$  termasuk ke dalam kategori valid. Dengan demikian kedua validator ahli memutuskan angket respon siswa dan guru terhadap tes kognitif yang dikembangkan dinyatakan valid dan layak digunakan dengan revisi sedikit.

c. Hasil analisis uji coba tes kognitif

Uji coba terbatas dalam penelitian ini dilakukan di dua sekolah yaitu SMA Negeri 2 Makassar dengan jumlah sampel 216 orang yang terbagi ke dalam tujuh kelas XI IPA dan SMA Negeri 14 Makassar dengan jumlah sampel sebanyak 75 orang yang terbagi ke dalam dua kelas XI IPA. Total keseluruhan sampel adalah 291 orang. Soal yang diberikan terbagi ke dalam dua bagian yaitu soal tipe A sebanyak 25 butir dan soal tipe B sebanyak 25 butir dengan tingkat kesukaran yang sama. Waktu untuk mengerjakan soal tersebut adalah 45 menit untuk masing-masing tipe.

2) Hasil analisis validitas butir tes kognitif

Analisis validitas butir soal dilakukan dengan bantuan *software Microsoft Excel* 2007. Nilai siswa yang diperoleh pada saat pelaksanaan uji coba diinput ke program Microsoft Excel kemudian dihitung berdasarkan rumus yang telah ditetapkan. Rumus yang digunakan untuk menghitung validitas butir soal adalah rumus korelasi product moment. Rumus tersebut kemudian diinput/dimasukkan ke dalam rumus/formula yang ada pada program *Microsoft Excel*. Adapun hasil perhitungan reliabilitas soal dapat dilihat pada Tabel 3:

**Tabel 3. Rekap Hasil Analisis Validitas Butir Soal**

Kategori	Tipe A		Tipe B	
	Nomor Soal	Jumlah	Nomor soal	Jumlah
Valid	1, 2, 3, 4, 5, 6, 7, 8, 9, 10,	23 (92%)	3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	23 (92%)
	12, 13, 14, 15, 16, 17, 18,		13, 14, 15, 16, 17, 18, 19,	
	19, 20, 21, 22, 23, 24		20, 21, 22, 23, 24, 25	
Tidak valid	11, 25	2 (8%)	1, 2	2 (8%)

Sumber: Data Peneliti

Sajian data hasil analisis validitas butir soal pada Tabel 3 menunjukkan bahwa



kedua tipe soal memiliki jumlah soal yang sama untuk setiap kategori. Untuk soal tipe A terdapat 23 soal yang valid, 2 soal yang tidak valid, untuk soal tipe B juga terdapat 23 soal yang valid, 2 soal yang tidak valid. Pengkategorian tersebut sesuai dengan ketentuan yang ditetapkan yaitu dengan cara melihat harga  $r$  hitung dan kemudian dikonsultasikan dengan harga  $r$  tabel *product moment* (pada penelitian ini harga  $r$  tabel yang digunakan adalah 0,176) dengan kriteria apabila harga  $r$  hitung sama dengan atau lebih besar dengan harga  $r$  tabel berarti soal dinyatakan valid.

3) Hasil analisis reliabilitas tes kognitif

Analisis reliabilitas soal dilakukan dengan bantuan software *Microsoft Excel* 2007. Nilai siswa yang diperoleh pada saat pelaksanaan uji coba diinput ke program *Microsoft Excel* kemudian dihitung berdasarkan rumus yang telah ditetapkan. Soal yang dihitung reliabilitasnya adalah soal yang valid berdasarkan perhitungan analisis validitas butir soal untuk setiap item (soal tipe A sebanyak 23 soal, soal tipe B sebanyak 23 soal), sedangkan soal yang tidak valid tidak dihitung reliabilitasnya. Rumus yang digunakan untuk menghitung reliabilitas adalah rumus K-R 20 dari *Kuder-Richardson* (KR). Rumus tersebut kemudian diinput / dimasukkan ke dalam rumus/formula yang ada pada program *Microsoft Excel*. Adapun hasil perhitungan reliabilitas soal dapat dilihat pada Tabel 4:

**Tabel 4. Rekap Hasil Analisis Reliabilitas Soal**

No	Tipe Tes	Nilai	Kategori
1.	A	0,71	Reliabel
2.	B	0,74	Reliabel

Sumber: Data Peneliti

Hasil analisis reliabilitas pada Tabel 4 menunjukkan bahwa kedua tipe soal berada pada kategori reliabel dengan nilai 0,71 untuk soal tipe A dan 0,74 untuk soal tipe B. Pengkategorian tersebut sesuai dengan ketentuan yang ditetapkan yaitu *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.

4) Hasil analisis tingkat kesukaran tes kognitif

Analisis tingkat kesukaran soal dilakukan dengan bantuan *software Microsoft Excel* 2007. Nilai siswa yang diperoleh pada saat pelaksanaan uji coba diinput ke program *Microsoft Excel* kemudian dihitung berdasarkan rumus yang telah ditetapkan. Rumus yang digunakan untuk menghitung tingkat kesukaran adalah banyaknya siswa yang menjawab benar butir soal dibagi dengan banyaknya siswa yang mengikuti tes. Rumus tersebut kemudian diinput/ dimasukkan ke dalam rumus/formula yang ada pada program *Microsoft Excel*. Selanjutnya untuk pengkategorian dihitung dengan cara mengkategorikan nilai indeks tingkat kesukaran soal. Adapun hasil perhitungan tingkat kesukaran soal dapat dilihat pada Tabel 5:

**Tabel 5. Rekap Hasil Analisis Tingkat Kesukaran Soal Tipe A dan Tipe B**

Kategori	Tipe A		Tipe B	
	Nomor Soal	Jumlah	Nomor soal	Jumlah
Mudah	1, 2, 3, 5, 6, 17, 21,	7 (30%)	4, 5, 6, 20, 21	5 (22%)
Sedang	4, 7, 9, 10, 12, 14, 15, 16, 18, 20, 22	11 (48%)	3, 7, 8, 9, 10, 12, 15, 16, 17, 18, 19, 22, 23, 24, 25	15 (65%)
Sukar	8, 13, 19, 23, 24	5 (22%)	11, 13, 14	3 (13%)

Informasi yang tersaji pada Tabel 5 menunjukkan bahwa kedua tipe soal berada pada kategori mudah, sedang, sukar dengan proporsi jumlah soal berbeda, namun perbedaan jumlah tersebut tidak terlalu berarti (tidak menyolok). Dengan demikian soal yang dikembangkan tergolong homogen dari segi tingkat kesukaran. Pengkategorian tersebut sesuai dengan 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.

#### 5) Hasil analisis daya beda tes kognitif

Analisis daya beda soal dilakukan dengan bantuan software Microsoft Excel 2007. Nilai siswa yang diperoleh pada saat pelaksanaan uji coba diinput ke program Microsoft Excel kemudian dihitung berdasarkan rumus yang telah ditetapkan. Rumus yang digunakan untuk menghitung daya beda adalah jumlah jawaban benar siswa kelompok atas dikurangi jumlah jawaban benar siswa kelompok bawah dibagi dengan setengah jumlah siswa yang mengikuti tes. Rumus tersebut kemudian diinput/dimasukkan ke dalam rumus/ formula yang ada pada program Microsoft Excel. Selanjutnya untuk pengkategorian dihitung dengan cara mengkategorikan nilai indeks daya beda soal. Data hasil analisis daya beda pada Tabel 6 menunjukkan bahwa kedua tipe soal berada pada kategori lemah, cukup, baik, baik sekali dengan proporsi jumlah soal berbeda, namun perbedaan jumlah tersebut tidak terlalu berarti (tidak menyolok). Secara keseluruhan soal diterima jika nilai IDB  $\geq 0,20$  dan direvisi/diteliti jika nilai IDB  $< 0,20$  (Ivanty, dkk, 2013). Dengan demikian soal yang dikembangkan rata-rata memiliki daya pembeda yang baik. Pengkategorian tersebut sesuai dengan ketentuan yang ditetapkan yaitu apabila IDB (Indeks Daya Beda)  $< 0,00$  (negatif) soal tidak memiliki daya beda, IDB  $< 0,20$  soal berdaya beda lemah, IDB 0,20-0,39 soal berdaya beda cukup, IDB 0,40-0,69 soal berdaya beda baik, IDB 0,70-1,00 soal berdaya beda baik sekali. Adapun hasil perhitungan daya beda soal dapat dilihat pada Tabel 6:

**Tabel 6 Rekap Hasil Analisis Daya Beda Soal Tipe A dan Tipe B**

Kategori	Tipe A		Tipe B	
	Nomor Soal	Jumlah	Nomor soal	Jumlah
Tidak ada daya beda	-	0 (0%)	-	0 (0%)
Lemah	1, 5	2 (9%)	6	1 (4%)
Cukup	2, 3, 10, 13, 17, 19, 23, 24	8 (35%)	5, 11, 13, 14, 17, 20, 23	7 (31%)

Baik	4, 6, 7, 8, 12, 14, 15, 16, 18, 20, 21, 22	12 (52%)	4, 7, 8, 9, 10, 12, 15, 16, 18, 19, 21, 22, 24, 25	14 (61%)
Baik sekali	9	1 (4%)	3	1 (4%)

Sumber: Data Peneliti

#### 4. Tahap penyebaran (*disseminate*)

Tahap penyebaran dilakukan dengan cara mensosialisasikan tes kognitif yang dikembangkan kepada guru-guru yang tergabung dalam forum MGMP (Musyawarah Guru Mata Pelajaran) biologi Kota Makassar. Penyebaran ini dilakukan agar tes yang dikembangkan dapat digunakan secara lebih luas untuk mengukur kemampuan siswa dalam menyerap materi pelajaran. Selain itu, penyebaran juga dilakukan dengan harapan bahwa guru-guru dapat memberikan saran maupun kritikan konstruktif agar tes kognitif yang dikembangkan dapat mengukur kemampuan siswa dengan baik. Tahap penyebaran ini dilaksanakan di SMA Negeri 5 Makassar yang merupakan sekretariat MGMP biologi Kota Makassar.

#### Pembahasan Hasil Penelitian

Pengembangan tes kognitif dilakukan dengan mengikuti langkah-langkah pengembangan tes yang dikemukakan oleh Thiagarajan. Langkah pengembangan tes yang dilakukan dalam penelitian ini merujuk pada langkah pengembangan perangkat model 4-D (*define, design, develop, dessiminate*) dari Thiagarajan, Semmel and Semmel yang diadaptasi menjadi 4-P (Pendefenisian, Perancangan, Pengembangan, Penyebaran). Adapun langkah-langkah pengembangan yang dilakukan selama penelitian yaitu: (1) menentukan tujuan dan kawasan tes, (2) menguraikan materi dan batasan perilaku yang akan diukur, (3) menyusun kisi-kisi, (4) memilih bentuk tes, (5) menentukan panjang tes, (6) menulis soal tes, (7) menelaah soal tes, (8) melakukan uji coba tes, (9) menganalisis butir soal, (10) memperbaiki tes, dan (11) merakit tes (Sukiman, 2012).

Menentukan tujuan dan kawasan tes yaitu tes formatif, jenis tes ini merupakan tes untuk mengukur tingkat pemahaman siswa setiap selesai pembelajaran untuk satu kompetensi dasar, kompetensi dasar yang diujikan dalam penelitian ini adalah KD 3.7 Menjelaskan keterkaitan antara struktur, fungsi, dan proses yang meliputi pembentukan sel kelamin, ovulasi, menstruasi, fertilisasi, kehamilan dan pemberian ASI, serta kelainan/penyakit yang dapat terjadi pada sistem reproduksi manusia. Menguraikan materi dan batasan perilaku yang akan diukur, yaitu menentukan topik-topik materi yang akan dibuatkan soal yang ada pada KD 3.7. Topik-topik materi yang ada pada KD 3.7 adalah Organ reproduksi pada pria dan wanita, Pembentukan sel kelamin (spermatogenesis dan oogenesis), Menstruasi, ovulasi, fertilisasi, kehamilan, pemberian ASI, kontrasepsi, dan gangguan/penyakit pada sistem reproduksi manusia. Topik materi tersebut semuanya mengacu pada kompetensi dasar. Menyusun kisi-kisi, yaitu menentukan proporsi soal untuk setiap materi, tingkatan kognitif dan aspek pengetahuan

yang ada pada revisi taksonomi Bloom.

Memilih bentuk tes, yaitu bentuk tes objektif pilihan ganda. Peneliti memilih bentuk pilihan ganda dikarenakan jumlah sampel uji coba yang banyak yakni 291 orang. Menentukan panjang tes, yaitu 50 soal. Selanjutnya 50 soal ini dibagi menjadi dua tipe yaitu tipe A dan tipe B dengan pertimbangan jenis tes merupakan tes formatif (ulangan harian yang hanya mengukur satu kompetensi dasar). Menulis soal tes, yaitu menulis 50 soal pilihan ganda dengan option jawabannya masing-masing. Menelaah soal tes, yaitu soal ditelaah oleh dua orang validator ahli, yaitu ahli pendidikan bidang evaluasi dan ahli konten/materi bidang reproduksi serta praktisi/guru di sekolah. Hasil validasi ahli menunjukkan bahwa tes kognitif yang dikembangkan berada pada kategori sangat valid dengan nilai 3,54. Pengkategorian tersebut sesuai dengan ketentuan yang ditetapkan yaitu hasil penilaian validator yang berada pada rentang skor  $3,5 \leq M \leq 4$  termasuk ke dalam kategori sangat valid (Nurdin, 2007).

Melakukan uji coba tes, yaitu tes kognitif di uji cobakan secara terbatas di dua sekolah yakni SMA Negeri 2 Makassar dan SMA Negeri 14 Makassar dengan jumlah sampel 291 orang. Menganalisis butir soal, ada 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*). Pada penelitian ini analisis butir soal telah dilakukan sesuai dengan yang dilakukan dalam penelitian Lababa (2008), yaitu setelah uji coba dilakukan analisis validitas butir, reliabilitas, tingkat kesukaran, daya beda, pengecoh (*distractor*), validitas empirik dengan menggunakan bantuan software Microsoft Excel 2007 dan SPSS 17. Dasar penggunaan bantuan software tersebut sesuai dengan pendapat Kasim (2009) dan Widodo (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 yaitu menggunakan software yang khusus dibuat untuk tujuan tersebut seperti ANATES, bisa juga menggunakan software lain seperti SPSS, Lisrel, dan microsoft office Excel.

Untuk analisis validitas butir soal menunjukkan bahwa kedua tipe soal memiliki jumlah soal yang sama untuk setiap kategori. Untuk soal tipe A terdapat 23 soal yang valid, 2 soal yang tidak valid, untuk soal tipe B juga terdapat 23 soal yang valid, 2 soal yang tidak valid. Pengkategorian tersebut sesuai dengan ketentuan yang ditetapkan yaitu dengan cara melihat harga  $r$  hitung dan kemudian dikonsultasikan dengan harga  $r$  tabel product moment (pada penelitian ini harga  $r$  tabel yang digunakan adalah 0,176) dengan kriteria apabila harga  $r$  hitung sama dengan atau lebih besar dengan harga  $r$  tabel berarti soal dinyatakan valid (Sukiman, 2012).

Untuk analisis reliabilitas menunjukkan bahwa kedua tipe soal berada pada kategori reliabel dengan nilai 0,71 untuk soal tipe A dan 0,74 untuk soal tipe B. Pengkategorian tersebut sesuai dengan ketentuan yang ditetapkan yaitu 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 (Sukiman, 2012).

Untuk analisis tingkat kesukaran, menunjukkan bahwa kedua tipe soal berada pada kategori mudah (30% untuk tipe A, 22% untuk tipe B), sedang (48% untuk tipe A, 65% untuk tipe B), sukar (22% untuk tipe A, 13% untuk tipe B). Kategori tingkat kesukaran tersebut sejalan dengan pendapat (Arikunto, 2010) bahwa soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya. Kategori tersebut juga 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 (Adiputra, 2012). Hasil penelitian Adiputra (2012) menunjukkan bahwa proporsi tingkat kesukaran soal ulangan IPS terpadu adalah 23% soal mudah, 62% soal sedang, dan 15% soal sukar, dimana hasil penelitian ini menunjukkan hasil yang mendekati proporsional. Hasil analisis menunjukkan persentase untuk setiap kategori tidak sama dengan pendapat ahli, namun perbedaan tersebut masih dalam kisaran yang dapat ditoleransi. Dengan demikian soal yang dikembangkan tergolong homogen dan hampir mendekati jumlah persentasenya dengan yang dikemukakan oleh Sukiman dan Adiputra. Pengkategorian tingkat kesukaran sesuai dengan 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 (Sukiman, 2012).

Untuk analisis daya beda, menunjukkan bahwa kedua tipe soal berada pada kategori lemah (9% tipe A, 4% tipe B), cukup (35% tipe A, 31% tipe B), baik (52% tipe A, 61% tipe B), baik sekali (4% tipe A, 4% tipe B) dengan proporsi jumlah soal berbeda, namun perbedaan jumlah tersebut masih dalam kisaran yang dapat ditoleransi. Dengan demikian soal yang dikembangkan tergolong homogen. 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. Hasil penelitian Ivanty, dkk (2013) menunjukkan bahwa 65 % soal diterima dan 35%

soal direvisi/diteliti. Untuk analisis daya beda, hasil penelitian ini menunjukkan bahwa untuk tipe A 84% soal diterima, 16% direvisi dan tipe B 88% diterima, 12% direvisi. Pengkategorian daya beda sesuai dengan ketentuan yang ditetapkan yaitu apabila IDB (Indeks Daya Beda)  $< 0,00$  (negatif) soal tidak memiliki daya beda, IDB  $< 0,20$  soal berdaya beda lemah, IDB  $0,20-0,39$  soal berdaya beda cukup, IDB  $0,40-0,69$  soal berdaya beda baik, IDB  $0,70-1,00$  soal berdaya beda baik sekali (Sukiman, 2012).

Untuk analisis pengecoh (distractor), menunjukkan bahwa kedua tipe soal berada pada kategori efektif (87% tipe A, 91% tipe B), tidak efektif (13% tipe A, 9% tipe B). 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, 2013).

Untuk hasil analisis validitas empirik dianalisis dengan menggunakan analisis faktor. Analisis faktor digunakan untuk mengetahui apakah perangkat tes yang dikembangkan sesuai dengan hasil validasi ahli ketika diujicobakan dilapangan. Hasil analisis uji validitas empirik khususnya pada tabel *Anti-image Correlation* pada nilai diagonal yang berpangkat a, jika nilainya lebih besar dari 0,500 maka komponen tersebut dinyatakan memenuhi, artinya hasil uji coba sesuai dengan hasil validasi ahli (Indahati, 2012). Pada soal tipe A, soal C4 tidak memenuhi, namun hal ini tidak menunjukkan perbedaan yang berarti karena hampir mendekati 0,500 yaitu 0,428, pada soal tipe B semua soal C1 – soal C6 memenuhi kriteria. Untuk dimensi pengetahuan soal tipe A dan tipe B semuanya memenuhi kriteria.

Dari segi validitas dan reliabilitas, 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 (Sukiman 2012).

Dari segi karakteristik butir soal, keputusan suatu item soal layak digunakan, perlu direvisi atau ditolak didasarkan pada kriteria keputusan untuk penilaian item soal oleh Ivanty, 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 penelitian Ivanty, dkk (2013) menunjukkan bahwa 60% soal diterima dan 40% direvisi. Dalam penelitian ini, 76% soal diterima, 24 % ditolak. Soal yang diterima merupakan soal yang telah memenuhi semua kriteria soal yang baik dalam analisis butir soal dan memenuhi semua materi dan indikator yang mengacu pada kompetensi dasar.

Memperbaiki tes, yaitu memperbaiki 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. Merakit tes yaitu menyusun kembali soal yang telah diperbaiki/direvisi setelah uji coba, dengan demikian dihasilkan sebuah tes kognitif yang dapat digunakan untuk mengukur kemampuan siswa dan telah diuji kualitasnya. Selanjutnya dilakukan tahap Tahap penyebaran dilakukan dengan cara mensosialisasikan tes kognitif yang dikembangkan kepada guru-guru yang tergabung dalam forum MGMP (Musyawarah Guru Mata Pelajaran) biologi Kota Makassar yang dilaksanakan di SMA Negeri 5 Makassar yang merupakan sekretariat MGMP biologi Kota Makassar. Penyebaran ini dilakukan agar tes yang dikembangkan dapat digunakan secara lebih luas untuk mengukur kemampuan siswa dalam menyerap materi pelajaran.

## KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dikemukakan, serta mengacu pada rumusan masalah penelitian maka dapat disimpulkan beberapa hal sebagai berikut:

1. Tes kognitif yang telah dikembangkan termasuk dalam kategori sangat valid dengan nilai 3,54 berdasarkan penilaian validator ahli. Untuk analisis validitas butir: kategori valid (92% untuk tipe A, 92% untuk tipe B), tidak valid (8% untuk tipe A, 8% untuk tipe B) serta termasuk dalam kategori reliabel dengan nilai 0,71 untuk soal tipe A dan 0,74 untuk soal tipe B.
2. Tes kognitif yang telah dikembangkan memiliki kualitas:
  - a. Tingkat kesukaran: kategori mudah (30% untuk tipe A, 22% untuk tipe B), sedang (48% untuk tipe A, 65% untuk tipe B), sukar (22% untuk tipe A, 13% untuk tipe B)
  - b. Daya beda: kategori lemah (9% tipe A, 4% tipe B), cukup (35% tipe A, 31% tipe B), baik (52% tipe A, 61% tipe B), baik sekali (4% tipe A, 4% tipe B)
  - c. Pengecoh (distractor): kategori efektif (87% tipe A, 91% tipe B), tidak efektif (13% tipe A, 9% tipe B).

## DAFTAR PUSTAKA

- Adiputra, I.B.R. (2012). Analisis Butir Soal Tes Ulangan Akhir Semester IPS Terpadu Buatan MGMP IPS Kabupaten Gianyar Kelas VII Semester 1 Tahun Pelajaran 2011 – 2012, (online), ([http://www.google.co.id/url?sa=t&rct=j&q=jurnal%2Banalisis%2Bpengecoh%2Fdistraktor%2Bsoal&source=web&cd=1&cad=rja&ved=0CCUQFjAA&url=http%3A%2F%2Fpasca.undiksha.ac.id%2Fjournal%2Findex.php%2Fjurnal\\_ep%2Farticle%2Fdownload%2F370%2F164&ei=cZ25UfOSFcPJrAfRvYDoBw&usg=AFQjCNEXksN465JfmbOHnSRa7ZVCprTLsA&bvm=bv.47883778,d.bmk](http://www.google.co.id/url?sa=t&rct=j&q=jurnal%2Banalisis%2Bpengecoh%2Fdistraktor%2Bsoal&source=web&cd=1&cad=rja&ved=0CCUQFjAA&url=http%3A%2F%2Fpasca.undiksha.ac.id%2Fjournal%2Findex.php%2Fjurnal_ep%2Farticle%2Fdownload%2F370%2F164&ei=cZ25UfOSFcPJrAfRvYDoBw&usg=AFQjCNEXksN465JfmbOHnSRa7ZVCprTLsA&bvm=bv.47883778,d.bmk)), Diakses 13 Juni 2013).
- Arikunto, S. (2010). Dasar-Dasar Evaluasi Pendidikan (Edisi Revisi). Jakarta: PT. Bumi Aksara.
- Indahati. (2012). Pengembangan Tes Diagnostik Biologi Kelas XI IPA SMA di Kabupaten Luwu. Tesis tidak diterbitkan. Makassar: Program Pascasarjana Universitas Negeri Makassar.
- 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, ([http://www.google.co.id/url?sa=t&rct=j&q=jurnal%2Banalisis%2Bdaya%2Bbeda%2Bsoal&source=web&cd=13&cad=rja&ved=0CC0QFjACOAO&url=http%3A%2F%2Fjurnal.fkip.uns.ac.id%2Findex.php%2Fpfisika%2Farticle%2Fdownload%2F1789%2F1282&ei=Vpm5UYeRKMejrQf00YCoAQ&usg=AFQjCNFgxflix-ztQr\\_G6D3uuXpeFBq3Mow&bvm=bv.47883778,d.bmk](http://www.google.co.id/url?sa=t&rct=j&q=jurnal%2Banalisis%2Bdaya%2Bbeda%2Bsoal&source=web&cd=13&cad=rja&ved=0CC0QFjACOAO&url=http%3A%2F%2Fjurnal.fkip.uns.ac.id%2Findex.php%2Fpfisika%2Farticle%2Fdownload%2F1789%2F1282&ei=Vpm5UYeRKMejrQf00YCoAQ&usg=AFQjCNFgxflix-ztQr_G6D3uuXpeFBq3Mow&bvm=bv.47883778,d.bmk)), Diakses 13 Juni 2013).
- Kasim, J. (2009). Apakah Analisis Butir Soal Perlu?. Jurnal Medik, (online), No. 3, September-Desember 2009, (<http://bdkjakarta.kemenag.go.id/file/media/3234forumapakahanalisisbutirsoalperlu.pdf>), Diakses 13 Juni 2013).
- Lababa, J. (2008). Analisis Butir Soal Dengan Teori Tes Klasik: Sebuah Pengantar. Jurnal Iqra', (online), Vol. 5 Januari – Juni 2008, (<http://www.google.co.id/url?sa=t&rct=j&q=jurnal%2Banalisis%2Bpengecoh%2Fdistraktor%2Bsoal&source=web&cd=25&cad=rja&ved=0CDgQFjAEOBQ&url=http%3A%2F%2Fblog.tp.ac.id%2Fwpcontent%2Fuploads%2F4328%2Fdownload-03-jun2936.pdf&ei=bqK5UaHGKYb8rAfJiYDICw&usg=AFQjCNGajET02n-MerIZ5wZqgvcYjaD5iRw&bvm=bv.47883778,d.bmk>), Diakses 13 Juni 2013).
- Nurdin. (2007). Model Pembelajaran Matematik yang Menumbuhkan Kemampuan Metakognitif Untuk Menguasai Bahan Ajar. Disertasi tidak diterbitkan. Surabaya: Program Pascasarjana Universitas Negeri Surabaya.



- Purwanto, N. (2004). Prinsip-Prinsip dan Teknik Evaluasi Pengajaran. Bandung: PT. Remaja Rosdakarya.
- Purwanto. (2013). Evaluasi Hasil Belajar. Yogyakarta: Pustaka Pelajar.
- Sahabuddin. (2007). Mengajar dan Belajar. Makassar: Badan Penerbit Universitas Negeri Makassar.
- Sudaryono. (2012). Dasar-dasar Evaluasi Pembelajaran. Yogyakarta: Graha Ilmu.
- Sukiman. (2012). Pengembangan Sistem Evaluasi. Yogyakarta: Insan Madani.
- Trianto. (2009). Mendesain Model Pembelajaran Inovatif-Progresif. Jakarta: Kencana Prenada Media Group.
- Widodo. (2010). Analisis Butir Soal Tes. Jurnal Pendidikan Penabur, (Online), No.14, Tahun ke-9, (<http://www.bpkpenabur.or.id/files/Hal.%2058-67%20Analisis%20Butir%20Soal.pdf>, diakses 13 Juni 2013).