

Pengembangan Tes Diagnostik Pilihan Ganda Dua Tingkat untuk Mendiagnosis Miskonsepsi Siswa dengan Metode MAK (Matriks Analisis Konsep) dan CRI (*Certainty of Response Index*)

Alindah^{ax}, Edy Chandra^a

^a Jurusan Tadris Biologi, IAIN Syekh Nurjati Cirebon, Jawa Barat, 45132, Indonesia

^xCorresponding author: Jl. Perjuangan Bypass Sunyaragi, Cirebon, Jawa Barat, 45132, Indonesia. E-mail addresses: alinlindah@gmail.com

Article history

Received 17 Juli 2021
Received in revised form
22 Agustus 2021
Accepted 11 September 2021

Abstract

Preliminary studies reveal that the development of two-tier multiple-choice diagnostic tests to identify students' misconceptions has not yet been widely applied in schools. This prompted the researcher to explore the topic. The study aims to: (1) conduct a theoretical analysis, (2) perform an empirical analysis, and (3) evaluate the empirical effectiveness of developing a two-tier multiple-choice diagnostic test on the Animalia concept for Grade X students in the Cirebon region. Employing a Research and Development (R&D) design, the study consisted of several stages, including preliminary study, product development, and testing. The test design was guided by Treagust's framework. Participants included Grade X students from SMAN 1 Beber, MAN 1 Kota Cirebon, and MANU Buntet. The findings revealed: (1) two sets of 50-item tests (draft A and B) were constructed and validated by experts in content, construction, and language; (2) item analysis covered validity, reliability, item difficulty, discrimination index, distractor functionality, and guessing factor; and (3) effectiveness testing yielded validity values ranging from 0.20-0.40, reliability scores between 0.60-0.80, moderate difficulty levels ($0.30 < IK < 0.70$), acceptable discrimination power (0.20-0.40), effective distractors, and low guessing rates. The highest incidence of misconceptions was identified in the Pisces subtopic using the Concept Analysis Matrix (MAK) and Certainty of Response Index (CRI) methods. Ultimately, 23 valid test items were produced through this development.

Keywords : R&D, diagnostic test, two tier multiple choice, misconception, MAK, CRI

Abstrak

Berdasarkan studi pendahuluan, pengembangan instrumen tes diagnostik pilihan ganda dua tingkat untuk mengidentifikasi miskonsepsi siswa masih jarang diterapkan di lingkungan sekolah. Hal ini mendorong peneliti untuk mengangkat topik tersebut. Tujuan utama dari penelitian ini meliputi: (1) menganalisis secara teoritis, (2) menganalisis secara empiris, serta (3) mengkaji efektivitas hasil analisis empiris terhadap pengembangan tes diagnostik *two-tier multiple choice* pada materi Animalia kelas X untuk mendeteksi miskonsepsi siswa di wilayah Cirebon. Penelitian ini menggunakan pendekatan *Research and Development* (R&D) yang melibatkan beberapa tahapan, yaitu studi pendahuluan, pembuatan produk, dan uji coba. Pengembangan instrumen ini merujuk pada model yang dikemukakan oleh Treagust. Subjek penelitian mencakup siswa kelas X dari SMAN 1 Beber, MAN 1 Kota Cirebon, dan MANU Buntet. Hasil penelitian menunjukkan bahwa: (1) telah disusun dua versi soal, masing-masing terdiri dari 50 butir soal (draft A dan B), yang divalidasi oleh ahli materi, konstruksi, dan bahasa; (2) analisis butir soal dilakukan melalui uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, efektivitas pengecoh, dan dugaan acak; serta (3) uji efektivitas menunjukkan rentang validitas 0,20-0,40, reliabilitas antara 0,60-0,80, tingkat kesukaran sedang ($0,30 < IK < 0,70$), daya pembeda cukup (0,20-0,40), pengecoh berfungsi dengan baik, dan dugaan acak rendah. Identifikasi miskonsepsi paling banyak ditemukan pada materi Pisces, berdasarkan analisis Matriks Analisis Konsep (MAK) dan *Certainty of Response Index* (CRI). Dari keseluruhan pengembangan, diperoleh sebanyak 23 soal yang dinyatakan valid.

Kata kunci : R&D, tes diagnostik, pilihan ganda dua tingkat, miskonsepsi, MAK, CRI

1. Pendahuluan

Dalam konteks praktik pendidikan di sekolah, siswa memperoleh pengetahuan melalui peran guru sebagai fasilitator utama. Salah satu aspek krusial dalam proses pembelajaran adalah kegiatan evaluasi. Evaluasi pembelajaran berfungsi sebagai sarana untuk mengumpulkan data yang kemudian digunakan sebagai dasar dalam pengambilan keputusan yang tepat (Arikunto, 2011). Selain itu,

evaluasi juga berperan penting dalam mendukung perkembangan potensi siswa, khususnya untuk menilai sejauh mana pencapaian kompetensi mereka terhadap materi yang telah dipelajari selama proses pembelajaran berlangsung.

Berdasarkan hasil observasi di sejumlah sekolah di wilayah Cirebon, diketahui bahwa dalam kegiatan evaluasi pembelajaran, guru umumnya masih menggunakan tes pilihan ganda konvensional yang hanya berfokus pada pencapaian hasil belajar siswa. Jenis tes yang mengukur tingkat pemahaman siswa terhadap konsep yang diajarkan masih jarang diterapkan. Padahal, untuk mengetahui sejauh mana siswa memahami ide-ide yang disampaikan selama pembelajaran, diperlukan instrumen yang lebih mendalam. Oleh karena itu, peneliti menerapkan penggunaan tes pilihan ganda dua tingkat sebagai salah satu bentuk tes diagnostik. Menyatakan bahwa tes diagnostik merupakan alat yang digunakan untuk mengidentifikasi permasalahan belajar yang dialami siswa. Hasil dari tes ini dapat dijadikan dasar untuk merancang pengajaran yang lebih sesuai dengan kemampuan dan kebutuhan siswa. Tujuan utama dari tes diagnostik adalah membantu siswa dalam mengetahui tingkat pemahaman konsep yang telah dipelajari serta mendorong peningkatan hasil belajar mereka.

Dalam penelitian ini, penulis menggunakan model pengembangan tes pilihan ganda dua tingkat yang pertama kali dikembangkan oleh David F. Treagust pada tahun 1988. Struktur dari tes ini terdiri atas dua bagian: tingkat pertama berisi pertanyaan mengenai konsep yang diuji, sedangkan tingkat kedua menyajikan pilihan alasan yang mendasari jawaban pada tingkat pertama, sehingga berfungsi sebagai alat diagnosis pemahaman siswa. Jika dibandingkan dengan bentuk soal lainnya, pilihan ganda dua tingkat memiliki sejumlah keunggulan. Tuysuz (2009) menjelaskan bahwa salah satu kelebihan utama dari tes ini adalah kemampuannya dalam mengurangi kemungkinan jawaban benar akibat tebakan. Dalam pilihan ganda konvensional dengan lima opsi jawaban, peluang siswa menebak dengan benar mencapai 20%, sedangkan pada pilihan ganda dua tingkat, peluang tersebut turun drastis menjadi hanya sekitar 4%. Selain itu, tes ini juga memudahkan guru dalam proses penskoran. Hal ini sejalan dengan pendapat Tan dan Treagust (1999) yang menyatakan bahwa tes diagnostik dua tingkat tidak hanya mudah diterapkan, tetapi juga efisien dalam proses penilaian, sehingga sangat berguna dalam konteks pembelajaran di kelas.

Dalam proses pembelajaran, penting bagi siswa untuk benar-benar memahami setiap konsep yang dipelajari. Ketika terjadi pemahaman yang tidak tepat, hal ini dapat menimbulkan miskonsepsi. Miskonsepsi merupakan struktur kognitif yang menyimpang dari pemahaman ilmiah yang telah diterima secara luas, dan keberadaannya dapat menghambat siswa dalam menyerap pengetahuan baru. Pandangan ini diperkuat oleh Suparno (2005) yang menyatakan bahwa miskonsepsi adalah pemahaman yang tidak sesuai dengan konsep yang dikemukakan oleh para ahli. Miskonsepsi dapat

muncul dalam berbagai bentuk, seperti kesalahan dalam konsep awal atau pemahaman yang keliru mengenai hubungan antarkonsep. Dengan demikian, miskonsepsi merupakan kesalahan dalam memahami suatu konsep yang membuat siswa cenderung menjelaskan materi menggunakan pemikiran mereka sendiri, bukan berdasarkan kerangka ilmiah yang sebenarnya.

Untuk mendiagnosis miskonsepsi yang dialami siswa, digunakan dua pendekatan, yaitu Matriks Analisis Konsep (MAK) dan *Certainty of Response Index* (CRI). MAK merupakan metode yang digunakan untuk mengembangkan analisis konsep menjadi rancangan soal yang mampu mengidentifikasi tiga kategori pemahaman siswa, yakni miskonsepsi, paham konsep, dan tidak paham konsep, sehingga dapat mempermudah proses identifikasi miskonsepsi (Fortuna, 2013). Sementara itu, CRI yang dikembangkan oleh Hasan (1999) merupakan metode yang menilai tingkat kepastian atau keyakinan siswa dalam memberikan jawaban, guna mendeteksi apakah siswa benar-benar memahami konsep atau hanya menebak. Dalam penelitian ini, tingkat kepastian diukur menggunakan skala enam poin, mulai dari 0 hingga 5.

Materi pembelajaran merupakan seperangkat bahan ajar yang dirancang untuk membentuk pengetahuan, keterampilan, dan sikap yang perlu dikuasai siswa guna mencapai standar kompetensi yang telah ditetapkan (Sophie, 2015). Dalam penelitian ini, penulis memilih konsep *Animalia* sebagai fokus pengembangan, yang merupakan bagian dari materi kelas X IPA. Pemilihan konsep ini didasarkan pada hasil wawancara dengan guru dan siswa yang menunjukkan bahwa *Animalia* memiliki cakupan konsep yang cukup luas, disertai dengan banyak istilah asing atau istilah dalam biologi yang cenderung sulit dipahami. Selain itu, siswa sering kali telah memiliki pemahaman awal terhadap konsep tersebut yang belum tentu sesuai dengan konsep ilmiah. Jika pemahaman awal tersebut benar, maka siswa dianggap telah memahami konsep dengan baik; namun, jika pemahamannya keliru, maka hal tersebut mengindikasikan adanya dugaan miskonsepsi.

Penelitian ini dikembangkan dengan pendekatan yang berbeda dari studi sebelumnya, yakni dengan merancang tes diagnostik pilihan ganda dua tingkat untuk mengidentifikasi miskonsepsi siswa menggunakan dua metode, yaitu Matriks Analisis Konsep (MAK) dan *Certainty of Response Index* (CRI), khususnya pada materi *Animalia* di kelas X di wilayah Cirebon. Tujuan dari penelitian ini adalah untuk mengkaji secara teoritis pengembangan tes diagnostik tersebut, serta untuk menelusuri kemungkinan adanya kendala atau potensi ancaman yang dapat memengaruhi keefektifan dan keakuratan instrumen dalam mengidentifikasi miskonsepsi siswa secara tepat.

2. Metode Penelitian

Penelitian ini dilaksanakan selama tiga bulan, terhitung sejak Maret hingga Mei 2018, bertepatan dengan semester genap tahun ajaran 2017–2018. Lokasi penelitian mencakup tiga sekolah di wilayah

Cirebon, yaitu SMAN 1 Beber, MAN 1 Kota Cirebon, dan MANU Buntet. Sampel penelitian terdiri dari beberapa kelompok siswa kelas X IPA, yaitu 30 siswa dari kelas X IPA 2 SMAN 1 Beber, 30 siswa dari kelas X IPA 3 SMAN 1 Kota Cirebon, 26 siswa dari kelas X IPA 3 MAN 1 Kota Cirebon, serta 32 siswa dari kelas X IPA 1 MANU Buntet.

Mengacu pada pendapat Sugiyono (2012), penelitian ini menggunakan metode penelitian dan pengembangan (*Research and Development*) yang mencakup tiga tahapan utama, yaitu studi pendahuluan, pengembangan produk, dan uji coba produk. Model pengembangan tes diagnostik dalam penelitian ini mengikuti peran yang dijelaskan oleh Treagust, yaitu dimulai dari pemilihan konsep yang akan diteliti, pengumpulan informasi terkait miskonsepsi siswa, hingga penyusunan instrumen berupa tes pilihan ganda dua tingkat. Proses analisis data kualitatif dilakukan oleh sejumlah ahli yang terdiri atas ahli konten, ahli konstruksi, dan ahli bahasa. Analisis ini mencakup berbagai aspek, seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, efektivitas pengecoh, serta faktor dugaan. Selain itu, analisis miskonsepsi siswa dilakukan dengan bantuan perangkat lunak *Microsoft Excel*.

3. Hasil dan Pembahasan

3.1 Analisis Validasi Teoritis Pengembangan Tes Diagnostik Pilihan Ganda Dua Tingkat untuk Mendiagnosis Miskonsepsi Siswa

Salah satu bentuk evaluasi yang digunakan dalam pembelajaran adalah tes diagnostik, yang berfungsi untuk mengidentifikasi kelemahan atau kesulitan belajar siswa (Arikunto, 2011). Dalam penelitian ini, penulis mengembangkan tes diagnostik berbentuk pilihan ganda dua tingkat. Instrumen ini terdiri atas dua bagian: tingkat pertama memuat pertanyaan yang menguji pemahaman terhadap konsep tertentu, sementara tingkat kedua menyajikan alasan atau penjelasan dari pilihan jawaban pada tingkat pertama. Dengan struktur tersebut, tes ini berperan sebagai alat diagnosis yang mampu mengungkap sejauh mana siswa memahami konsep yang diajarkan.

Setelah memperoleh informasi mengenai permasalahan penelitian, penulis menyusun peta konsep dan melakukan analisis konsep sebagai langkah awal dalam pengembangan instrumen. Peta konsep disusun untuk mengidentifikasi subkonsep yang akan diajarkan, dengan mengacu pada kompetensi inti dan kompetensi dasar yang berlaku. Menurut Nurhasanah (2014), peta konsep membantu menentukan subkonsep mana yang perlu dianalisis lebih lanjut. Miskonsepsi pada siswa dapat terjadi akibat penetapan konsep yang tidak tepat, sehingga peneliti merujuk pada kamus biologi dan berbagai buku ajar sebagai sumber acuan dalam penyusunan konsep.

Berdasarkan tabel 1, hasil evaluasi terhadap peta konsep yang telah disusun menunjukkan adanya sejumlah saran perbaikan. Beberapa masukan tersebut meliputi pentingnya membuat

tampilan tulisan menjadi lebih komunikatif, serta memperjelas garis dan teks agar tidak terjadi kerancuan antara konsep-konsep yang saling berkaitan. Sebagai contoh, untuk mempertegas perbedaan antara filum dan kelas dalam kingdom Animalia, disarankan agar keduanya diberi garis pembeda yang jelas. Selain itu, beliau juga menyarankan dilakukannya analisis konsep yang lebih mendalam serta penyempurnaan klasifikasi Animalia. Hal ini mencakup pembagian filum ke dalam kelas-kelas secara tepat dan sesuai dengan buku ajar kelas X serta referensi dari buku Campbell edisi semester genap. Penulisan nama ilmiah juga perlu diperbaiki dengan penggunaan huruf miring (*italic*). Oleh karena itu, baik peta konsep maupun analisis konsep perlu direvisi agar dapat digunakan secara optimal dalam penyusunan instrumen tes diagnostik.

Tabel 1. Saran Perbaikan Peta Konsep dan Analisis Konsep Animalia

Kategori	Saran Perbaikan
Peta Konsep	Penulisan lebih komunikatif, garis dan tulisan lebih diperjelas.
Analisis Konsep	Pahami dan perbaiki kembali tentang klasifikasi Animalia

Dalam penyusunan tes diagnostik pilihan ganda dua tingkat, penulis mengacu pada taksonomi Marzano (2007) yang mencakup empat jenjang kognitif, yaitu pengetahuan, pemahaman, analisis, dan pemanfaatan pengetahuan. Setelah dilakukan revisi, peta konsep dan analisis konsep telah diperbarui sesuai dengan kebutuhan pengembangan instrumen. Penulis kemudian merancang empat belas konsep utama dalam materi Animalia yang digunakan untuk mengidentifikasi kemungkinan miskonsepsi siswa. Setelah penyusunan soal, para ahli melakukan analisis teoritis terhadap draft instrumen. Analisis teoritis ini merupakan hasil penelaahan dari para pakar terhadap kualitas butir soal yang telah disusun.

Tabel 2. Hasil Validasi Ahli Terhadap Instrumen Soal

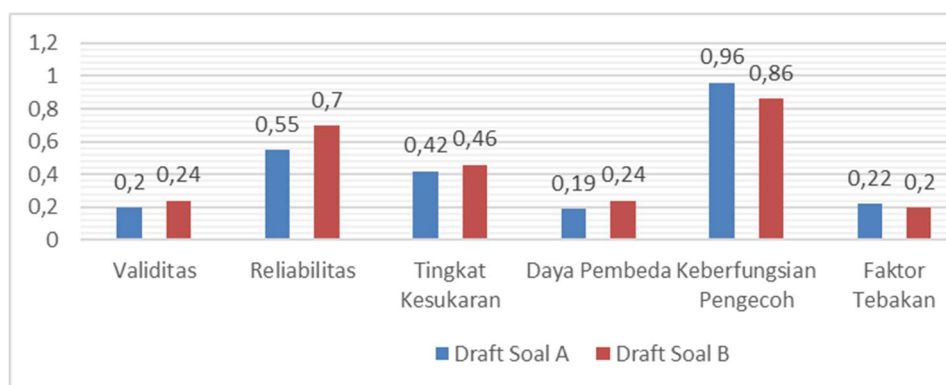
Validasi Ahli	Saran Perbaikan
Ahli Konten	Gunakan bahasa yang efektif dan komunikatif sehingga tidak multi jawaban, dan pahami kembali pertanyaan sehingga jawaban tingkat pertama dan kedua saling berkaitan.
Ahli Konstruksi	Diperbaiki urutan soalnya, dan gambar diperjelas lagi
Ahli Bahasa	Pahami kaidah PUEBI dan terapkan dalam penyusunan instrumen soal, dan banyak soal yang salah dalam tata kalimat sehingga harus diperbaiki susunannya

Berdasarkan hasil validasi, ahli konten menilai bahwa soal-soal yang dikembangkan telah menggunakan bahasa yang efektif, namun menekankan pentingnya keterkaitan antara jawaban pada tingkat pertama dan tingkat kedua, sebagaimana dijelaskan dalam Tabel 2. Beberapa butir soal, seperti nomor 28, 37, dan 43 pada draft soal A, dinilai masih memerlukan penambahan penjelasan agar lebih jelas. Sementara itu, ahli konstruksi memberikan masukan agar kualitas gambar dalam soal diperjelas dan menyarankan perbaikan terhadap urutan penyajian soal. Di sisi lain, ahli bahasa menekankan bahwa soal harus disusun sesuai dengan kaidah Ejaan Bahasa Indonesia (EBI), serta mencatat bahwa beberapa butir soal pada draft A dan B masih membutuhkan revisi. Secara khusus,

soal nomor 3, 18, dan 40 pada draft A, serta nomor 9 dan 17 pada draft B, perlu diperbaiki dari segi struktur dan kejelasan kalimat.

3.2 Analisis Validasi Empiris Pengembangan Tes Diagnostik Pilihan Ganda Dua Tingkat untuk Mendiagnosis Miskonsepsi Siswa

Analisis validitas empiris dilakukan melalui serangkaian tahap uji coba, yang terdiri dari uji coba terbatas, uji coba lapangan tahap 1, dan uji coba lapangan tahap 2. Seluruh rangkaian uji coba dilaksanakan dalam rentang waktu 9 hingga 21 April 2018. Uji coba terbatas melibatkan 30 siswa dari kelas X IPA 2 dan X IPA 3 SMAN 1 Beber. Dalam tahap ini, digunakan sebanyak 100 butir soal yang dibagi ke dalam dua paket, yakni draft soal A dan draft soal B, masing-masing terdiri atas 50 soal. Tujuan dari pelaksanaan uji coba ini adalah untuk menyaring soal-soal yang memenuhi kriteria kualitas berdasarkan aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran, efektivitas pengecoh, serta faktor-faktor lain yang relevan dalam penyusunan tes diagnostik.

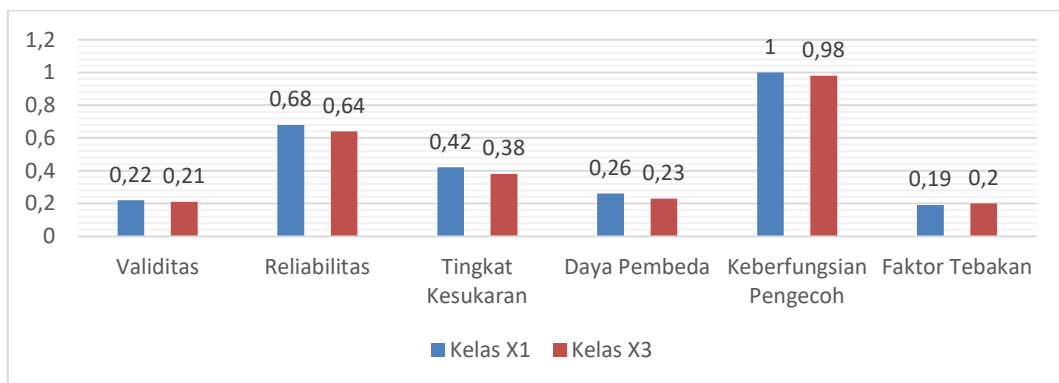


Gambar 1. Karakteristik Butir Soal Draft A dan B Uji Coba Terbatas

Berdasarkan hasil analisis pada gambar 1, karakteristik butir soal dalam draft A pada tahap uji coba terbatas menunjukkan bahwa validitas soal berada pada angka 0,20 yang termasuk dalam kategori sangat rendah. Nilai reliabilitasnya mencapai 0,55 yang tergolong sedang. Tingkat kesukaran soal sebesar 0,42, juga termasuk dalam kategori sedang. Daya pembeda soal menunjukkan nilai 0,19 yang berada pada kategori rendah. Keberfungsian pengecoh mencapai 0,96 yang menunjukkan bahwa pengecoh berfungsi dengan baik. Sementara itu, faktor tebakan berada pada nilai 0,22.

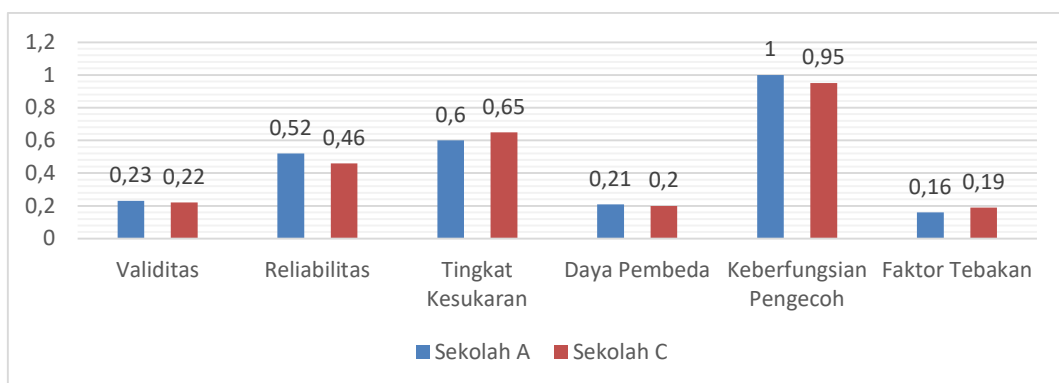
Uji coba lapangan pertama dilaksanakan pada tanggal 23 April hingga 5 Mei 2018. Kegiatan ini melibatkan siswa dari MAN 1 Kota Cirebon, yaitu 26 siswa dari kelas X IPA 1 dan 29 siswa dari kelas X IPA 3. Butir-butir soal yang telah dinyatakan valid pada tahap uji coba terbatas kemudian digunakan kembali dalam uji coba lapangan 1 ini. Adapun hasil rekapitulasi karakteristik soal berdasarkan uji coba lapangan 1 disajikan pada gambar 2.

Berdasarkan hasil yang ditampilkan pada gambar 2, karakteristik butir soal untuk kelas X IPA 1 pada uji coba lapangan 1 menunjukkan nilai validitas sebesar 0,22 yang termasuk dalam kategori rendah. Nilai reliabilitas mencapai 0,68 dan dikategorikan tinggi. Tingkat kesukaran soal berada pada angka 0,42 yang termasuk dalam kategori sedang. Daya pembeda menunjukkan angka 0,26 yang tergolong cukup. Keberfungsian pengecoh memperoleh nilai 1, yang menunjukkan bahwa semua pengecoh berfungsi dengan baik. Sementara itu, faktor tebakan tercatat sebesar 0,19.



Gambar 2. Karakteristik Butir Soal Kelas X IPA 1 dan 3 Uji Coba Lapangan 1

Uji coba lapangan 2 dilakukan pada tanggal 7 Mei - 21 Mei 2018. Subjek dalam uji coba lapangan 2 adalah siswa SMAN 1 Beber kelas X IPA 5 berjumlah 30 Siswa dan siswa MANU Buntet kelas X IPA 1 berjumlah 32 Siswa. Uji coba lapangan 2 yang dilakukan terdiri atas 38 soal. Berikut ini hasil rekapitulasi karakteristik uji coba lapangan 2.



Gambar 3. Karakteristik Butir Soal sekolah A Uji Coba Lapangan 2

Merujuk pada gambar 3, karakteristik butir soal dari sekolah A dalam uji coba lapangan 2 menunjukkan nilai validitas sebesar 0,23 yang tergolong rendah. Reliabilitas soal tercatat sebesar 0,52 dan dikategorikan sedang. Tingkat kesulitan soal berada pada angka 0,60, yang termasuk dalam kategori sedang. Daya pembeda soal menunjukkan nilai 0,21 yang dinilai cukup. Keberfungsian pengecoh bernilai 1, menandakan bahwa semua pengecoh bekerja dengan baik. Sementara itu, faktor tebakan berada pada angka 0,16.

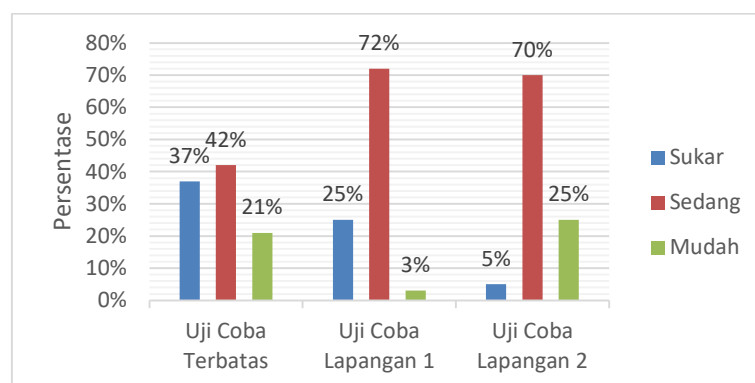
3.3 Efektivitas Pengembangan Tes Diagnostik Pilihan Ganda Dua Tingkat Untuk Mendiagnosis Miskonsepsi Siswa dengan Metode MAK dan CRI

Uji coba terbatas menunjukkan validitas berkategori rendah dengan 0,22. Uji coba lapangan satu juga menunjukkan validitas berkategori rendah dengan 0,22. Ini didasarkan pada Arikunto (2013), yang menyatakan bahwa korelasi antara 0,21 dan 0,40 berkategori rendah. Selain itu, reliabilitas uji coba terbatas sebesar 0,62 berkategori tinggi, sedangkan reliabilitas uji coba lapangan 1 sebesar 0,67 berkategori tinggi, dan reliabilitas uji coba lapangan 2 sebesar 0,49 berkategori sedang. Ini didasarkan pada penjelasan Arikunto bahwa korelasi 0,40-0,59 berkategori sedang, dan korelasi 0,60-0,79 berkategori tinggi. Nilai-nilainya masih dianggap baik meskipun validitas dan reliabilitasnya berubah-ubah. Oleh karena itu, pertanyaan yang dibuat dapat diterima untuk dibahas lebih lanjut.

Tabel 3. Hasil Perbandingan Validitas dan Reliabilitas pada Setiap Uji Coba

Uji Coba	Validitas	Reliabilitas
Uji coba terbatas	0.22	0.62
Uji coba lapangan 1	0.21	0.67
Uji coba lapangan 2	0.22	0.49

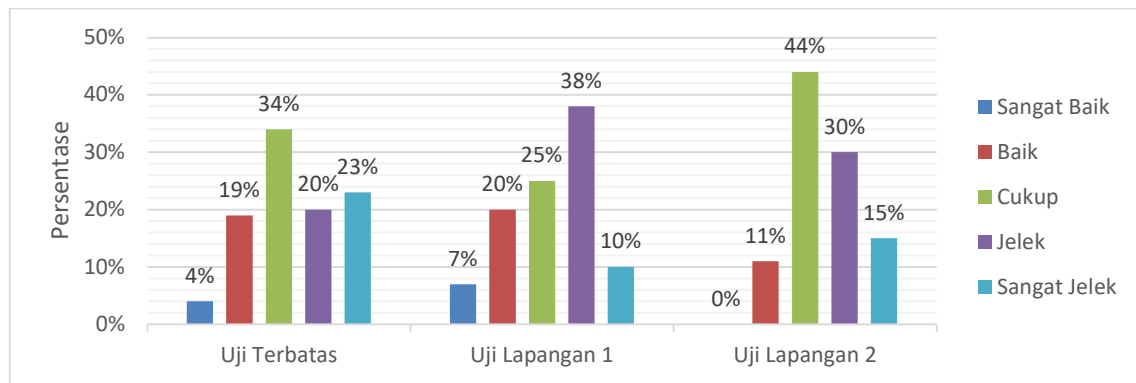
Soal yang baik idealnya memiliki tingkat kesukaran pada kategori sedang, yaitu tidak terlalu sulit maupun terlalu mudah. Soal dengan taraf kesukaran yang seimbang mencerminkan proporsi yang sesuai antara siswa yang mampu menjawab dengan benar dan yang tidak. Setiap tahap uji coba menunjukkan variasi dalam tingkat kesukaran soal. Perbedaan tersebut dapat diamati melalui gambar 4, yang menggambarkan perbandingan tingkat kesukaran antar uji coba.



Gambar 4. Perbandingan Tingkat Kesukaran Soal dalam Setiap Uji Coba

Tingkat kesukaran soal pada setiap tahap uji coba menunjukkan variasi, sebagaimana ditampilkan pada gambar di atas. Uji coba terbatas menunjukkan dominasi soal-soal dalam kategori kesukaran sedang, dan kecenderungan yang sama juga tampak pada uji coba lapangan 1 maupun lapangan 2. Indeks kesukaran yang didominasi oleh kategori sedang ini mengindikasikan bahwa butir-butir soal telah disusun secara proporsional, sehingga layak digunakan dalam proses pembelajaran.

Daya pembeda merupakan kemampuan suatu butir soal untuk membedakan antara siswa yang berkemampuan tinggi dan siswa yang berkemampuan rendah. Nilai daya pembeda ini dinyatakan dalam bentuk indeks diskriminasi (D), yang menunjukkan seberapa kuat suatu soal dapat melakukan pembedaan tersebut. Setiap tahapan uji coba menghasilkan nilai daya pembeda yang bervariasi. Gambar 5 menggambarkan perbandingan daya pembeda antar uji coba yang telah dilakukan.

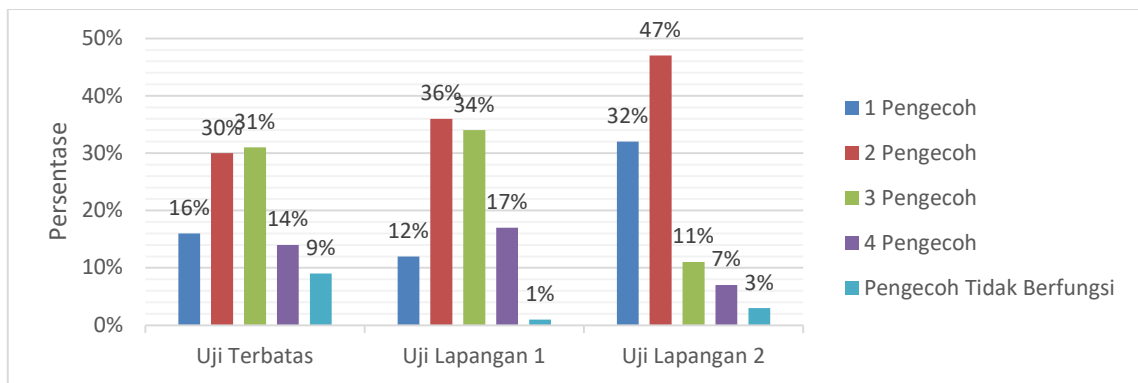


Gambar 5. Perbandingan Daya Pembeda Soal dalam Setiap Uji Coba

Hasil uji coba menunjukkan bahwa nilai daya pembeda pada setiap tahapan pengujian bervariasi, sebagaimana ditampilkan pada grafik di atas. Pada uji coba terbatas, mayoritas butir soal termasuk dalam kategori cukup. Sementara itu, pada uji coba lapangan pertama, sebagian besar soal tergolong dalam kategori jelek. Adapun pada uji coba lapangan kedua, butir soal kembali didominasi oleh kategori cukup. Dengan mempertimbangkan bahwa sebagian besar soal memiliki daya pembeda dalam kategori cukup, maka dapat disimpulkan bahwa soal-soal yang dikembangkan telah disusun secara layak dan memiliki potensi untuk digunakan dalam pembelajaran.

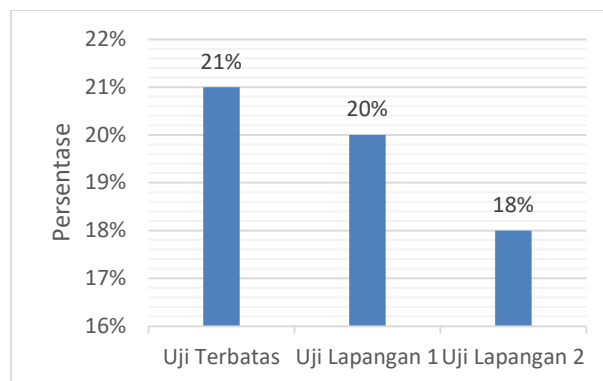
Keberfungsian pengecoh (distractor) berperan dalam membedakan peserta tes dengan kemampuan rendah dan tinggi. Berdasarkan pendapat Sudjono (2005), suatu pengecoh dianggap berfungsi dengan baik apabila dipilih oleh lebih dari 5% peserta tes. Sebaliknya, apabila pengecoh dipilih oleh kurang dari 5% peserta, maka pengecoh tersebut dianggap tidak berfungsi secara optimal. Dalam setiap tahap uji coba, keberfungsian pengecoh pada butir soal menunjukkan variasi yang berbeda. Gambar 6 menyajikan perbandingan keberfungsian pengecoh dalam masing-masing uji coba.

Kualitas pengecoh pada setiap tahap uji coba menunjukkan hasil yang bervariasi, sebagaimana ditampilkan pada grafik di atas. Pada uji coba terbatas, mayoritas soal didominasi oleh pengecoh kategori 3 yang berfungsi dengan baik. Sementara itu, pada uji coba lapangan 1 dan 2, pengecoh yang paling banyak berfungsi dengan baik berasal dari kategori 2. Adapun pengecoh yang tidak berfungsi dengan baik memiliki persentase yang paling rendah di seluruh tahapan uji coba. Berdasarkan distribusi ini, dapat disimpulkan bahwa kualitas soal tergolong cukup baik, karena secara umum didominasi oleh pengecoh yang memenuhi kriteria fungsional.



Gambar 6. Perbandingan Keberfungsian Pengecoh Soal dalam Setiap Uji Coba

Menurut Susetyo (2015), skor yang diperoleh peserta tes bisa saja berasal dari jawaban atas butir-butir soal yang tidak sepenuhnya mereka pahami, yang kemungkinan besar dijawab dengan cara menebak. Jika skor tersebut disebabkan oleh tebakan, maka harus dianggap sebagai hasil tebakan. Oleh karena itu, diperlukan koreksi untuk mengetahui kemampuan peserta yang sesungguhnya. Setiap tahap uji coba menghasilkan faktor tebakan yang berbeda. Gambar 7 menyajikan perbandingan hasil koreksi skor akibat tebakan acak pada masing-masing uji coba.



Gambar 7. Perbandingan Faktor Tebakan dalam Setiap Uji Coba

Gambar 7 menunjukkan bahwa terdapat perbedaan faktor tebakan pada setiap uji coba. Uji coba terbatas memiliki faktor tebakan paling tinggi, yaitu sebesar 21%, yang menunjukkan bahwa lebih banyak peserta mengabaikan atau menjawab secara acak pada tahap ini dibandingkan dengan uji coba lainnya. Sementara itu, uji coba lapangan 1 menunjukkan faktor tebakan sebesar 20%, dan uji coba lapangan 2 sebesar 18%. Perbedaan ini menunjukkan bahwa soal-soal yang diujikan semakin baik dari segi kejelasan dan kemampuan mengukur pemahaman peserta secara aktual, karena semakin rendah faktor tebakan, semakin kecil kemungkinan peserta menjawab secara acak. Dengan demikian, soal-soal yang dikembangkan dapat dikategorikan sebagai soal yang layak dan memiliki kualitas yang cukup baik.

Menurut Yulianti (2011), miskonsepsi terjadi ketika seseorang keliru dalam memahami atau mengaitkan hubungan antar gagasan, baik karena intuisi yang salah maupun pandangan yang keliru.

Sementara itu, Brown mendefinisikan miskonsepsi sebagai penjelasan yang tidak tepat dan tidak sejalan dengan pemahaman ilmiah yang telah disepakati oleh para ahli. Secara lebih spesifik, bentuk-bentuk miskonsepsi dapat mencakup kesalahan dalam memahami suatu konsep, penggunaan konsep yang tidak tepat, generalisasi yang keliru dalam penerapan suatu ide, maupun perbedaan interpretasi terhadap makna dari suatu gagasan.

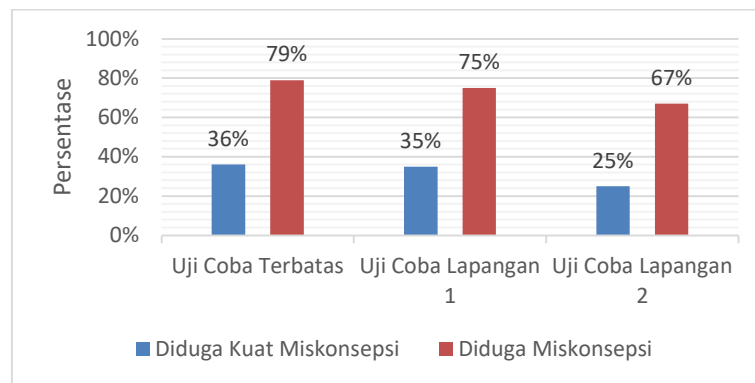
Peneliti merancang tes diagnostik berbentuk pilihan ganda dua tingkat untuk mengidentifikasi miskonsepsi yang dimiliki oleh siswa. Dalam menganalisis miskonsepsi tersebut, peneliti menggunakan dua pendekatan, yaitu metode Matriks Analisis Konsep (MAK) dan *Certainty of Response Index* (CRI). Metode MAK dilakukan dengan merancang butir soal yang pilihannya jawabannya memiliki makna konseptual, seperti mewakili miskonsepsi, ketidaktahuan terhadap konsep, dan pemahaman konsep yang benar. Sementara itu, metode CRI digunakan untuk mengukur tingkat keyakinan atau kepastian siswa dalam memilih jawaban, sehingga memberikan gambaran mengenai tingkat kepercayaan terhadap pengetahuan yang dimilikinya. Hasil analisis miskonsepsi siswa berdasarkan arah soal dan respons siswa pada setiap uji coba ditampilkan dalam tabel 4.

Tabel 4. Hasil Analisis Miskonsepsi Siswa dalam Setiap Uji Coba

Uji Coba	Jumlah Soal	
	Diduga Kuat Miskonsepsi	Diduga Miskonsepsi
Uji Coba Terbatas	506	1.211
Uji Coba Lapangan 1	586	1.212
Uji Coba Lapangan 2	252	844
Uji Coba	Jumlah Siswa	
	Diduga Kuat Miskonsepsi	Diduga Miskonsepsi
Uji Coba Terbatas	574	1.143
Uji Coba Lapangan 1	552	1.246
Uji Coba Lapangan 2	341	735

Berdasarkan tabel 4, terdapat dua pendekatan untuk menganalisis miskonsepsi siswa pada setiap butir soal. Pertama, apabila hasil analisis menggunakan metode MAK dan CRI sama-sama menunjukkan bahwa jawaban siswa salah, maka siswa tersebut mengalami miskonsepsi. Kedua, apabila hanya metode MAK yang menunjukkan kesalahan sedangkan CRI menunjukkan bahwa siswa tidak yakin dengan jawabannya, maka kesalahan tersebut lebih cenderung disebabkan oleh ketidaktahuan, bukan miskonsepsi (Fortuna, 2013). Jenis miskonsepsi siswa kemudian dibedakan ke dalam dua kategori arah, yaitu arah soal dan arah siswa. Hasil analisis menunjukkan bahwa pada uji coba lapangan 1, jumlah miskonsepsi paling banyak ditemukan baik dari arah soal maupun arah siswa. Temuan ini menunjukkan bahwa kombinasi metode MAK dan CRI efektif untuk mengidentifikasi arah miskonsepsi siswa. Hal ini sejalan dengan pendapat Trie (2018) yang

menyatakan bahwa teknik CRI dapat mengidentifikasi berbagai konsep yang mengalami miskonsepsi pada siswa.



Gambar 8. Perbandingan Analisis Miskonsepsi dalam Setiap Uji Coba Soal

Berdasarkan gambar 8, uji coba terbatas menunjukkan persentase tertinggi pada kategori soal dengan indikasi kuat miskonsepsi maupun dugaan miskonsepsi. Jika dibandingkan dengan uji coba lapangan 1 dan 2, persentase soal yang diduga kuat mengandung miskonsepsi mencapai angka tertinggi, yaitu sebesar 36%, sedangkan soal dengan dugaan miskonsepsi mencapai 79%. Secara khusus, soal yang menguji konsep *Pisces* menunjukkan tingkat dugaan miskonsepsi tertinggi, yakni sebesar 79%, yang mengindikasikan bahwa pada konsep tersebut siswa mengalami kesulitan paling besar dalam memahami materi secara ilmiah.

Uji coba lapangan 1 menunjukkan persentase tertinggi pada kategori dugaan miskonsepsi, yakni sebesar 35%, sedangkan pada kategori dugaan miskonsepsi saja tercatat sebesar 18%. Sementara itu, pada uji coba lapangan 2, persentase dugaan miskonsepsi mencapai 25%, dan dugaan miskonsepsi saja sebesar 18%. Dengan demikian, uji coba lapangan 1 lebih banyak mengindikasikan kemungkinan terjadinya miskonsepsi dibandingkan uji coba lapangan 2.

Konsep *Pisces* menjadi yang paling banyak mengalami miskonsepsi, sebagaimana terungkap dari hasil wawancara dengan siswa. Hal ini menunjukkan bahwa siswa telah membawa ide awal yang tidak sesuai dengan pemahaman ilmiah. Temuan ini sejalan dengan pendapat Mustain (2015), yang menyatakan bahwa kemampuan siswa dalam memahami suatu konsep dapat tercermin dari ungkapan mereka, baik secara lisan maupun tertulis, serta dari cara mereka menjelaskan kembali konsep tersebut dengan bahasa sendiri, apakah penjelasan tersebut sesuai atau tidak dengan pandangan ilmiah. Miskonsepsi terjadi ketika guru tidak melakukan klarifikasi terhadap pemahaman yang keliru, sehingga konsep yang salah tersebut tertanam kuat dalam ingatan siswa.

Untuk mengetahui konsep-konsep yang telah dipahami dan yang belum dipahami oleh siswa selama proses pembelajaran, metode MAK dan CRI dapat dimanfaatkan untuk menentukan urutan konsep pembelajaran serta tingkat pencapaian siswa terhadap konsep tersebut. Berdasarkan hasil uji coba lapangan 2, diketahui bahwa jawaban siswa yang termasuk dalam kategori dengan indeks CRI

paling rendah berada pada kategori 2, yakni dengan nilai CRI < 2,5. Jawaban benar dengan CRI rendah menunjukkan bahwa siswa sebenarnya tidak memahami konsep meskipun jawabannya benar, sedangkan jawaban salah dengan CRI rendah juga mengindikasikan ketidaktahuan terhadap konsep. Dengan demikian, siswa yang memberikan jawaban dengan CRI paling rendah dikategorikan ke dalam kategori 2, yaitu kelompok siswa yang belum memahami konsep.

Hasil dari uji coba terbatas, uji coba lapangan 1, dan uji coba lapangan 2 menunjukkan bahwa konsep *Pisces* merupakan konsep yang paling kuat diduga mengalami miskonsepsi, baik pada kategori dugaan kuat miskonsepsi maupun dugaan miskonsepsi saja. Dengan demikian, dapat disimpulkan bahwa hasil tes diagnostik pilihan ganda dua tingkat yang diterapkan di lokasi penelitian pada sekolah yang berbeda menghasilkan temuan yang sejalan dengan hasil wawancara terhadap siswa dan guru mata pelajaran Biologi. Hal ini menunjukkan bahwa tes diagnostik tersebut efektif dalam mengungkap miskonsepsi siswa.

Tes diagnostik pilihan ganda dua tingkat, yang terdiri atas seratus butir soal, telah diuji cobakan pada berbagai subjek dengan latar belakang pendidikan dan tingkat kemampuan yang beragam. Uji coba dilakukan dalam dua tahap, yaitu uji coba terbatas dan uji coba lapangan 1 serta 2. Tes ini dikembangkan dengan menggunakan metode *MAK* (Matriks Analisis Konsep) dan *CRI* (*Certainty of Response Index*), yang mampu mengidentifikasi adanya dugaan kuat miskonsepsi maupun miskonsepsi pada siswa. Berdasarkan hasil analisis, sebanyak 23 soal terpilih dinyatakan layak untuk diuji cobakan lebih lanjut di sekolah lain. Oleh karena itu, evaluasi pembelajaran di institusi pendidikan perlu lebih menekankan pada identifikasi konsep-konsep yang paling sering disalahpahami siswa, agar guru dapat merancang strategi pembelajaran yang lebih tepat dan efektif dalam meningkatkan kualitas pemahaman konsep.

4. Simpulan

Berdasarkan hasil analisis data dan pembahasan, dapat disimpulkan bahwa tes diagnostik pilihan ganda dua tingkat yang dikembangkan memerlukan perbaikan dari segi konten, konstruksi, dan bahasa, seperti penambahan penjelasan rinci, penyempurnaan gambar soal, revisi jenjang kognitif, serta penyesuaian dengan kaidah Ejaan Bahasa Indonesia (EBI). Secara empiris, tes ini memiliki validitas sebesar 0,22, reliabilitas 0,60, daya pembeda 0,34 (kategori cukup), tingkat kesukaran 0,61 (kategori sedang), pengecoh berfungsi baik pada dua opsi jawaban, serta faktor tebakan tergolong rendah. Efektivitas tes diagnostik terbukti tinggi, khususnya dalam mengidentifikasi miskonsepsi menggunakan metode Matriks Analisis Konsep (MAK) dan *Certainty of Response Index* (CRI), dengan temuan utama bahwa konsep *Pisces* merupakan subkonsep yang paling banyak mengandung

dugaan miskonsepsi. Dari seratus soal yang dikembangkan, sebanyak 23 soal dinyatakan valid dan layak diuji cobakan lebih lanjut dalam konteks evaluasi pembelajaran.

Daftar Pustaka

- Arikunto, S. (2011). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Fortuna R, D., Chandra, E & Yulia G, R. (2013). Pengembangan Tes Diagnostik untuk Mengukur Miskonsepsi Siswa pada Pokok Bahasan Sistem Regulasi Manusia untuk Siswa SMA Kelas XI Semester II. *Scientiae Educatia: Jurnal Sains dan Pendidikan Sains*. 2(2).
- Marzano, (2007). *The New Taxonomy of Educational Objectives*. USA: Corwin Press.
- Mustain, I. (2015). Kemampuan Membaca Dan Interpretasi Grafik Dan Data: Studi Kasus Pada Siswa Kelas 8 SMPN. *Scientiae Educatia: Jurnal Sains dan Pendidikan Sains*. 5(2).
- Nurhasanah, R. (2014). *Pengembangan Tes Diagnostik pada Pokok Bahasan Sistem Ekskresi untuk Mendeteksi Miskonsepsi Siswa Kelas XI di Kabupaten Cirebon*. [Skripsi]. Cirebon: IAIN Syekh Nurjati.
- Sophie E, V. Chandra, E & Mulyani, A. (2015). Analisis Kualitas Representasi Visual Buku Biologi SMA Kelas XI Kurikulum 2013 pada Materi Sel. *Scientiae Educatia: Jurnal Sains dan Pendidikan Sains*. 2(2).
- Sudjono, A. (2005). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.
- Sugiyono. (2012). *Metode Penelitian Pendidikan Pendekatan Kualitatif, Kuantitatif dan R&D*. Bandung: Alfabeta.
- Suparno, P. (2005). *Miskonsepsi dan Perubahan Konsep dalam Pendidikan Fisika*. Jakarta: Grasindo.
- Susetyo, B. (2015). *Prosedur Penyusunan dan Analisis Tes*. Bandung: Refika Aditama.
- Trie K, D. Maryanti, S. (2018). *Analysis of Math Teacher Candidates' Misconception on the Dynamic Electricity Concept*. *Scientiae Educatia: Jurnal Sains dan Pendidikan Sains*. 7 (1): 67-78.
- Tuysuz, C. (2009). Development of Two Tier Diagnostic Instrument And Assess Students Understanding In Chemistry. *Scientific Research and Essay*. 4(4). 626-631.
- Yulianti, L. (2011). *Miskonsepsi dan Mediasi Pembelajaran IPA*. Surabaya: Universitas Negeri Surabaya Press.