



Analisis kualitas butir soal instrumen *Assessment Diagnostic* untuk mendeteksi miskonsepsi siswa SMA pada materi virus

Prima Mitha Puspitasari Setyaningrum ^{a, 1}, Murni Ramli ^{b, 2*}, Yudi Rinanto ^{b, 3}

^a Magister Pendidikan Sains, Pascasarjana, Universitas Sebelas Maret, Jl. Ir. Sutami 36A, Surakarta 57126, Indonesia

^b Pendidikan Biologi, Fakultas Keguruan dan Ilmu Pendidikan Biologi, Universitas Sebelas Maret Jl Ir. Sutami 36A, Surakarta 57126, Indonesia

¹ pprimamitha@yahoo.com; ² mramlim04@fkip.uns.ac.id; ³ yudi.rinanto@gmail.com

*korespondensi penulis

Informasi artikel

Sejarah artikel:
Diterima 18 Agustus 2018
Revisi 19 November 2018
Dipublikasikan 22 Desember 2018

Kata kunci:

Analisis butir soal
Assessment Diagnostic
Kolom Bukti Fakta
Structure Communication Grid

ABSTRAK

Evaluasi merupakan kegiatan, untuk melihat sejauh mana keberhasilan pembelajaran telah dilaksanakan. Kegiatan evaluasi memerlukan alat sebagai pengumpul data yang disebut instrumen. Sebuah instrumen dikatakan baik apabila memenuhi beberapa kriteria. Salah satunya dengan melakukan analisis butir soal. Penelitian ini bertujuan untuk mengetahui validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal pada instrumen *diagnostic* materi virus. Responden penelitian berjumlah 346 siswa kelas X, XI, dan XII SMA Negeri dan Swasta di Kota Magelang. Instrumen yang diberikan berupa *Assessment Diagnostic* pada materi virus, yang terdiri dari soal Kolom Bukti Fakta (KBF), *essay*, dan *Structure Communication Grid* (SCG). Penelitian ini merupakan penelitian kuantitatif untuk mengetahui kualitas instrumen. Hasil penelitian menunjukkan persentase validitas soal KBF dan SCG 63,2% , soal *essay* 100%, semua soal reliabel, tingkat kesukaran soal mudah KBF dan SCG 38,6%, soal sedang 37,1%, dan 24,3% soal sukar. Sedangkan kesukaran soal *essay* 20% pada kategori mudah dan 80% pada kategori sedang. Daya pembeda soal KBF dan SCG sebesar 62,2% kategori jelek, 31,5% kategori cukup, 6,3% kategori baik. Daya pembeda soal *essay* 60% soal pada kategori cukup dan 40% soal kategori baik. Berdasarkan hasil analisis, dilakukan perbaikan butir soal. Instrumen pada materi virus yang sedang dikembangkan, merupakan yang pertama di Indonesia.

ABSTRACT

Key word:

Analysis of items
Assessment Diagnostic
Column Evidence of Facts
Structure Communication Grid

Quality analysis of item in the *Assessment Diagnostic* instrumen to detect misconceptions in high school students on virus material. Evaluation is an activity to see the extent to which the success of learning has been carried out. Evaluation activities require tools as data collectors called instruments. An instrument is good, if it meets several criteria. One of them is by analyzing questions' item. This study was aimed to determine the validity, reliability, level of difficulty, and differentiation of problems in the diagnostic instruments of virus material. Respondents in this study are 346 students of class X, XI, XII of public and private High Schools in Magelang City. The instrumen provided is in the form of a Diagnostic Assessment on virus material, which consists of the Facts Proof Column (KBF), essay and Structure Communication Grid (SCG). This research is a quantitative study which determine the quality of the instrumen. The results showed the percentage of the validity of KBF and SCG questions was 63,2%, essay questions were 100%, all questions are reliable, the level of difficulty of easy questions of KBF and SCG is 38,6%, moderate questions are 37,1%, and 24,3% are difficult questions. While the difficulty of the essay is 20% in the easy category and 80% in the medium category. The differentiation of the KBF and SCG questions is 62,2% in the bad category, 31,5% in the sufficient category, 6,3% in the good category. The differentiation of essay questions 60% in adequate category and 40% in good categories. Based on the analysis, further improvements were made to the items. Instrumens in the virus material which are being developed are the first in Indonesia.

Copyright © 2018 Universitas Ahmad Dahlan

Pendahuluan

Proses pembelajaran merupakan sistem yang meliputi tiga komponen pokok, yaitu merancang, melaksanakan, dan melakukan evaluasi kegiatan pembelajaran. Ketiga komponen tersebut saling berkaitan dan berinteraksi dalam mencapai tujuan pembelajaran. Melalui kegiatan evaluasi, kemajuan program pembelajaran yang telah dilaksanakan dapat diketahui (Borualogo, Kusdiyati, Susandari, & Sirodj, 2017). Proses evaluasi tidak hanya mengukur sejauh mana tujuan tercapai, namun menentukan keputusan lanjutan dalam kegiatan pembelajaran.

Evaluasi sebagai kegiatan terstruktur, menyeluruh, dan berkelanjutan dalam upaya penjaminan dan pengendalian mutu pembelajaran. Sehubungan dengan evaluasi pembelajaran, maka sangat penting untuk dilakukan penilaian pembelajaran. Clark (2015) telah berhasil melakukan identifikasi tiga tujuan penilaian yang berbeda, namun berkaitan satu sama lain, yaitu penilaian pembelajaran (*Assessment of Learning*), penilaian sebagai pembelajaran (*Assessment as Learning*), dan penilaian untuk pembelajaran (*Assessment for Learning*). *Assessment Diagnostic* yang dikembangkan dalam penelitian ini merupakan bagian dari *AFL*, yang merupakan penilaian formatif dan bertujuan untuk memperbaiki pembelajaran berikutnya. Penyusunan *Assessment Diagnostic* dikhususkan pada materi virus, yang belum pernah dikembangkan di Indonesia.

Evaluasi berkaitan erat dengan kegiatan pengumpulan data, dimana diperlukan alat yang disebut instrumen. Pengumpulan data dalam kegiatan evaluasi pendidikan, berupa tes atau non tes (Purwanto, 2011). Teknik tes digunakan untuk mengetahui hasil belajar kognitif, sedangkan teknik non tes untuk mengukur hasil belajar afektif (aspek sikap) dan psikomotorik (aspek keterampilan). Penggunaan alat sebagai teknik penilaian, bertujuan agar pelaksanaan pembelajaran menjadi terarah.

Hasil wawancara terhadap guru biologi dan siswa SMA di Kota Magelang pada November 2017, diketahui bahwa instrumen penilaian yang digunakan seringkali diperoleh dari buku paket atau soal kumpulan ujian. Bentuk soal yang diujikan kepada siswa yaitu *multiple choice* (pilihan ganda) dan *essay* (uraian). Soal berbentuk pilihan ganda dan *essay* hanya menguji aspek ingatan yang akan membuat siswa untuk belajar dengan cara menghafal materi bukan memahami materi. Soal-soal yang menguji aspek ingatan termasuk dalam ranah kognitif level bawah yaitu C1 (hafalan/ingatan) dan C2 (pemahaman). Akibatnya siswa tidak terbiasa untuk berpikir tingkat tinggi, yang berdampak pada ketidakmampuan siswa dalam menghubungkan

pengetahuan yang dimilikinya ketika dihadapkan pada soal yang menuntut analisis maupun evaluasi dalam pembelajaran.

Biologi mempelajari kehidupan organisme, lingkungan, dan relasi tentang keduanya. Materi biologi terdiri atas objek yang bersifat konkret dan objek yang bersifat abstrak. Objek konkret teramati melalui fenomena alam yang terjadi. Sedangkan objek yang bersifat abstrak diantaranya: proses metabolisme di dalam tubuh manusia (meliputi: sistem saraf, sistem indra, sistem hormon, sistem koordinasi, dan sebagainya). Sifat objek pada materi biologi sangat beragam, baik ditinjau dari ukuran organisme (makroskopis seperti: animalia dan plantae; mikroskopis seperti: virus, bakteri, protista), keterjangkauan (ekosistem dan komponen penyusun didalamnya), sifat pathologi/penyebab penyakit (virus, bakteri, protista, organisme lain), dan sebagainya (Sudarisman, 2015). Sejalan dengan hal tersebut, Jayawardana (2017) berpendapat bahwa "Proses belajar mengajar pada pembelajaran biologi seringkali dihadapkan pada materi abstrak dan di luar pengalaman siswa sehari-hari, sehingga materi menjadi sulit dipahami siswa". Beberapa materi biologi melibatkan proses-proses yang saling berhubungan, yaitu: sistem digesti (gerakan peristaltik kerongkongan hingga lambung), respirasi (proses inspirasi dan ekspirasi pernapasan dada serta perut), genital (proses fertilisasi hingga terbentuknya zigot), cardiovascular (hubungan antara kerja jantung dan pembuluh darah), dan sebagainya (Jayawardana, 2017).

Ayu, Saputri, & Widyaningrum (2016), menyatakan bahwa "Materi virus merupakan salah satu materi yang sulit untuk dibelajarkan dan sulit untuk dipahamkan kepada siswa". Konsep virus yang diajarkan pada tingkat SMA meliputi: ciri-ciri virus sebagai peralihan benda mati dan makhluk hidup, struktur tubuh virus yang membedakannya dengan benda mati maupun organisme seluler, dasar klasifikasi virus, replikasi virus dalam menginfeksi inangnya, dan peranan virus. Partikel virus berukuran mikroskopis dan sangat sederhana jika dibandingkan bakteri maupun prokariota, virus tidak memiliki komponen penyusun sel, serta tidak bermetabolisme di luar sel inang. Karakteristik partikel virus yang tidak dapat dilihat oleh mata tanpa alat khusus, berakibat pada terbentuknya pemikiran abstrak siswa yang berkembang menjadi miskonsepsi (Güne, Güne, & Hoplan, 2011).

Miskonsepsi adalah konsepsi atau pandangan tentang sains yang berbeda dengan konsep yang sebelumnya telah disepakati oleh para ahli (Kose, Pekel, & Hasenekoglu, 2009). Suratno (2008), menyatakan bahwa sebelum mengikuti pembelajaran

sebenarnya siswa sudah memiliki pemikiran sendiri terhadap suatu konsep, dan mereka akan membawa konsepsi tersebut ke dalam kelas. Apabila konsepsi tersebut salah, maka akan menjadi miskonsepsi yang jika tidak diperbaiki akan terus menjadi kesalahan konsep pada siswa (Gurel, Eryilmaz, & McDermott, 2015). Miskonsepsi mudah muncul dalam konteks menjelaskan fenomena yang ditemui sehari-hari di lingkungan sekitar (Chi, Slotta, & Leeuw, 1994). Sejalan dengan hal tersebut, Aprilia (2016) berpendapat “Pemahaman konsep sebenarnya merupakan jenjang kemampuan berpikir yang setingkat lebih tinggi dari ingatan dan hafalan”.

Kara & Çelikler (2015) menyatakan “Dalam pembelajaran, guru memiliki peran sentral dalam upaya merubah kesalahan konseptual (miskonsepsi) siswa yaitu dengan mengembangkan dan mengevaluasi pembelajaran yang terjadi dalam kelas”. Miskonsepsi dapat dideteksi dengan menemukan penyebab miskonsepsi tersebut melalui teknik tes. Beberapa bentuk tes yang dapat digunakan untuk mendeteksi miskonsepsi yaitu: metode diskusi, praktikum tanya jawab, *Certainty of Response Index* (CRI), tes pilihan ganda disertai alasan terbuka, peta konsep, tes *essay* tertulis, wawancara, dan tes diagnostik (Suparno, 2013).

Terdapat beberapa instrumen diagnostik yang telah dikembangkan pada penelitian-penelitian sebelumnya untuk melacak terjadinya miskonsepsi, yaitu: 1) Tes *certainty of response index* (CRI). Metode CRI dapat mengeksplorasi pemahaman sains, tingkat kepercayaan diri, dan mengidentifikasi miskonsepsi yang terjadi (Mustaqim, Zulfiani, & Herlanti, 2014), 2) Instrumen *two-tier multiple choice question* yang dikembangkan pada materi *Plantae* dapat melatih kemampuan berpikir tingkat tinggi (HOTs) (Nofiana, Sajidan, & Puguh, 2016), 3) Instrumen *three-tier diagnostic test*, yaitu tes dengan tiga tingkatan soal (Syahrul & Setyarsih, 2015). Penelitian pada materi virus yang telah dilakukan Ayu et al. (2016) menyatakan bahwa terdapat miskonsepsi pada buku teks, yang berimbas pada kesalahan siswa dalam memahami konsep pada materi virus. Belum adanya tes diagnostik, terutama dalam bentuk KBF dan SCG menjadi dasar bagi peneliti untuk menyusun suatu instrumen penilaian berupa *Assessment Diagnostic* pada materi virus.

Tes diagnostik merupakan suatu jenis *AFL* atau penilaian untuk pembelajaran yang dapat memberikan umpan balik kepada guru maupun siswa terkait kemajuan pembelajaran dalam mencapai tujuan pembelajaran (Yankovskaya, Levin, & Fuks, 2014). Umpan balik ini akan dipergunakan oleh guru untuk merevisi dan mengembangkan pengajaran berikutnya.

Tes diagnostik yang disusun berupa Kolom Bukti Fakta (KBF), merupakan tes yang diadopsi dari “*Resources Pack*” King College London yang dikembangkan oleh Osborne (2004), sedangkan *Structure Communication Grid* (SCG) diadopsi dari penelitian Johnstone, Bahar, & Hansell (2000). Tes diagnostik KBF mampu mengukur tingkat pemahaman konsep siswa terhadap *key concept* pada pokok bahasan tertentu, yang cenderung dipahami secara salah. Tes SCG digunakan untuk mengetahui tingkat pemahaman siswa tentang konsep yang berupa suatu tahapan atau siklus.

Penggunaan tes diagnostik KBF dan SCG memberikan gambaran kepada guru tentang konsep-konsep yang sulit dipahami oleh siswa, serta mengetahui kelemahan siswa tentang materi yang diajarkan. Hal ini sejalan dengan pemikiran Yankovskaya et al., (2014), “Kelemahan siswa dapat dideteksi dengan menggunakan tes diagnostik”.

Tes diagnostik KBF dan SCG yang dikembangkan pada penelitian ini adalah instrumen untuk mendeteksi miskonsepsi pada materi virus. Tes diagnostik yang dikembangkan perlu diuji coba untuk mengetahui tingkat keefektifan dalam melacak siswa yang kemungkinan mengalami paham konsep, miskonsepsi, dan tidak paham konsep. Selanjutnya dari hasil jawaban siswa, dilakukan analisis kuantitatif butir soal untuk mengetahui seberapa baik instrumen tersebut dapat mengukur miskonsepsi siswa. *Assessment Diagnostic* materi virus yang dikembangkan, diharapkan dapat memberikan gambaran kepada guru mengenai jenis penilaian guna memperbaiki proses pembelajaran di sekolah.

Penelitian mengenai kesulitan pembelajaran pada materi virus sebelumnya telah dilakukan oleh beberapa peneliti. Akan tetapi belum ada penelitian yang mengkaji lebih dalam tentang penggunaan *Assessment for Learning* yang mampu mendeteksi miskonsepsi siswa. Beberapa pertimbangan tersebut, melatarbelakangi peneliti untuk menyusun butir soal instrumen *Assessment Diagnostic* pada materi virus, Penelitian ini bertujuan untuk mengetahui kualitas butir soal instrumen. Selanjutnya hasil analisis butir soal digunakan sebagai salah satu dasar perbaikan instrumen soal sebelum dilakukan uji coba kedua.

Metode

Penelitian ini merupakan penelitian deskriptif kuantitatif untuk mengecek kualitas instrumen. Data yang diperoleh berasal dari jawaban siswa, dianalisis menggunakan aplikasi IBM SPSS *Statistics* versi 21. Tahapan penelitian yang dilakukan, yaitu: 1) Menyusun instrumen soal berupa KBF disertai *essay*, dan SCG untuk mendeteksi miskonsepsi siswa pada

materi virus, 2) Melakukan validasi terkait *content* materi bersama validator ahli, 3) Melakukan survey di SMA di Kota Magelang yang akan dijadikan tempat penelitian, 4) Menentukan jumlah responden yang terlibat berdasarkan penentuan populasi siswa, 5) Melaksanakan pengujian instrumen kepada 346 siswa sebagai responden, 6) Menganalisis hasil uji coba soal menggunakan aplikasi IBM SPSS *Statistic* versi 21.

Sekolah yang menjadi sampel dalam penelitian ditentukan berdasarkan teknik *cluster sampling*. Dari 12 SMA di Kota Magelang, sebanyak 6 sekolah dipilih berdasarkan kriteria jenis sekolah, yaitu sekolah negeri dan swasta. Jumlah responden pada setiap SMA yang terlibat ditentukan berdasarkan teknik *proportionate stratified random sampling*, yaitu sekitar 20 siswa dari setiap kelas untuk setiap *grade* X, XI, dan XII SMA yang telah memperoleh materi virus di sekolah. Siswa di SMA negeri dan swasta di Kota Magelang berjumlah 5045 siswa, sedangkan siswa dari 6 sekolah yang dipilih sebagai responden penelitian berjumlah 3046 siswa. Dari total siswa 6 SMA tersebut, yang terlibat dalam tes miskonsepsi sebanyak 346 siswa.

Jumlah sampel ditentukan berdasarkan formula *Isaac & Michael, (1995)*, dengan rumus:

$$S = \frac{X^2 \cdot N \cdot P(1 - P)}{d^2(N - 1) + X^2P(1 - P)}$$

Keterangan: S = jumlah sampel, N = jumlah populasi, P = proporsi populasi sebagai dasar asumsi pembuatan tabel (ditetapkan P=0.5), d = derajat ketetapan yang direfleksikan oleh kesalahan yang dapat ditoleransi dalam fluktuasi proporsi sampel P (umumnya ditetapkan 0.5), X² = Nilai tabel chi square untuk satu derajat kebebasan relatif level konfidensi yang diinginkan (X² = 3.841 tingkat kepercayaan 0.95).

Kasus: Ciri-Ciri Virus

a) Berikan tanda yang berbeda untuk menjawab masing-masing kotak!
 b) Virus merupakan partikel aseluler yang bersifat parasit karena mampu menyebabkan berbagai macam penyakit dan dapat menyebar diantara organisme. Akan tetapi virus tidak mampu bereproduksi atau melaksanakan aktivitas metabolisme di luar sel inang. Pengelompokan virus masih menjadi perdebatan para ahli. Apakah termasuk benda mati atau makhluk hidup?
 c) Gunakan tanda sesuai dengan instruksi di bawah untuk mengidentifikasi virus sebagai benda mati atau makhluk hidup.

V	Bukti yang mengatakan virus adalah makhluk hidup
X	Bukti yang mengatakan virus adalah benda mati

Lembar soal:

1.	Virus memiliki asam nukleat yang dibungkus selubung protein	11.	Ukuran virus terbesar beberapa nanometer (nyaris tidak tampak dibawah mikroskop cahaya)
2.	Virus tidak memiliki enzim untuk melakukan metabolisme sendiri	12.	Virus tidak memiliki ribosom untuk mensintesis protein
3.	Virus memiliki kemampuan replikasi didalam sel inang	13.	Virus tidak mampu memperbanyak dirinya sendiri
4.	Virus hanya mengandung satu tipe asam nukleat (RNA/DNA), tidak pernah keduanya	14.	Partikel virus diluar sel inang tidak mempunyai kegiatan metabolik yang mandiri
5.	Virus tidak memiliki sel (bersifat aseluler)	15.	Virus dapat menyebabkan penyakit bagi inang
6.	Virus tidak memiliki nukleus	16.	Beberapa jenis virus dilindungi oleh amplop bermembran
7.	Virus terdiri atas partikel yang dapat dikristalkan	17.	Virus terkecil hanya memiliki empat gen dalam genomnya
8.	Keberadaan virus sangat bergantung pada inang yang ditumpanginya	18.	Virus tidak mampu menghasilkan ATP
9.	Virus tidak memiliki dinding peptidoglikan	19.	Virus tidak dapat dibiakkan pada medium nutrisi dalam tabung reaksi atau cawan petri
10.	Setiap tipe virus dapat menginfeksi sel dari ragam inang yang terbatas (host range)	20.	Virus dapat didistribusikan berdasarkan jenis asam nukleat yang dimiliki, ukuran dan morfologinya

1) Dengan mempertimbangan argumen-argumen tersebut, menurut Anda virus merupakan makhluk hidup atau benda mati?
 2) Jelaskan alasan yang melatarbelakangi pilihan Anda!

Gambar 1. Soal Kolom Bukti Fakta (KBF) tentang ciri-ciri virus sebagai peralihan benda mati dan makhluk hidup

Tipe soal KBF yang dikembangkan dalam penelitian ini, yaitu siswa diberikan kasus dengan pernyataan ciri-ciri virus sesuai subkonsep, kemudian siswa menentukan jawaban yang benar serta menyimpulkan konsep yang sesuai dengan jawaban yang dipilih. Tipe soal SCG dalam penelitian ini, yaitu siswa diberikan pernyataan tentang tahapan-tahapan konsep reproduksi virus, konsep yang diberikan terdiri dari konsep benar dan konsep salah, kemudian siswa menentukan konsep yang benar dan mengurutkan tahapan reproduksi virus dengan tepat. Contoh soal KBF pada materi virus, yang telah disusun dan diuji kepada siswa kelas X, XI, dan XII disajikan pada *Gambar 1*.

Data yang dianalisis dalam penelitian ini didasarkan pada lembar jawaban siswa berkaitan dengan instrumen *Assessment Diagnostic* pada materi virus. Instrumen dibuat oleh peneliti dan diujikan kepada siswa SMA di enam sekolah berstatus negeri dan swasta yang ada di Kota Magelang pada tahun pelajaran 2017/2018. Instrumen disajikan dalam bentuk soal KBF dan soal *essay* yang berjumlah lima kasus (terdiri dari 63 kolom pernyataan) serta soal SCG yang berjumlah satu kasus (terdiri dari 20 kolom pernyataan). Sedangkan soal *essay* berjumlah 5 buah (*Gambar 2*).

Hasil analisis data selanjutnya digunakan sebagai dasar pertimbangan tentang diterima atau tidaknya setiap butir soal. Jika hasil analisis memenuhi kriteria (valid, reliabel, proporsi tingkat kesukaran seimbang, daya pembeda baik), maka butir soal dapat diterima. Tetapi jika tidak memenuhi kriteria, maka butir soal tidak dapat diterima atau dapat diterima dengan melakukan perbaikan.

Kasus Replikasi Virus

a) Perhatikan masing-masing uraian pada setiap kotak!

1. Enzim-enzim inang mentranskripsikan genom virus menjadi mRNA virus	2. Terjadi hidrolisis DNA pada sel inang	3. mRNA virus digunakan oleh ribosom inang untuk membuat lebih banyak protein kapsid
4. Terjadi replikasi genom fag tanpa menghancurkan inang	5. Virus menempel dan memasuki sel inang (Adsorpsi)	6. Pembentukan genom fag didalam kapsid
7. Virus mencopot selubung, melepaskan DNA dan protein kapsid (penetrasi)	8. Genom virus dan protein kapsid merakit diri menjadi partikel virus baru yang keluar dari sel	9. Sel inang mengalami kerusakan akibat terjadinya osmosis oleh material genetik virus
10. DNA bakteri berintegrasi dengan kromosom virus menjadi profag	11. Pada siklus reproduksi fag, terjadi kematian sel inang	12. Enzim-enzim inang mereplikasi genom virus (sintesis)

Diatas merupakan reaksi siklus reproduksi virus yang terjadi saat virus menginfeksi inang. Salah satu inang yang dapat tenfeksi oleh virus adalah bakteri. Virus yang menyerang bakteri disebut sebagai bakteriofage atau fag. Gunakan angka dari kotak diatas untuk menjawab pertanyaan berikut ini!

b)
 Q1: Kotak manakah yang berisi tahapan siklus reproduksi virus dalam proses menginfeksi sel inang?
 1) Pilihlah kotak yang sesuai.....
 2) Tulislah pilihan jawaban anda dalam urutan yang logis.....
 Q2: Kotak manakah yang berisi tahapan siklus reproduksi virus fag dalam proses menginfeksi hingga terjadi siklus lisis pada inang?
 1) Pilihlah kotak yang sesuai.....
 2) Tulislah pilihan jawaban anda dalam urutan yang logis.....
 Q3: Kotak manakah yang berisi tahapan siklus reproduksi virus fag dalam proses menginfeksi hingga terjadi siklus lisogenik pada inang?
 1) Pilihlah kotak yang sesuai.....
 2) Tulislah pilihan jawaban anda dalam urutan yang logis.....

Gambar 2. Soal *Structure Communication Grid* (SCG) tentang tahapan replikasi virus dalam menginfeksi inangnya

Hasil dan pembahasan

Hasil analisis data pada penelitian ini meliputi: validitas, reliabilitas, tingkat kesukaran dan daya pembeda.

Validitas

Validitas adalah ketepatan suatu tes dalam mengukur apa yang akan diukur (Purwanto, 2011). Soal memiliki validitas tinggi, jika dapat mengukur kompetensi yang diharapkan. Sebaliknya, soal yang memiliki validitas rendah, apabila tidak bisa mengukur kompetensi yang diharapkan. Butir soal dengan validitas tinggi menunjukkan soal itu secara tepat mampu mengukur kemampuan siswa.

Sebaliknya, butir soal yang bervaliditas rendah mencerminkan perlu diadakannya perbaikan.

Berdasarkan analisis validitas instrumen *Assessment Diagnostic* materi virus, diketahui bahwa validitas soal SCG sebesar 95%. SCG merupakan soal yang disusun oleh peneliti pada subbab reproduksi (replikasi) virus. Sedangkan pada soal KBF, validitas tertinggi terdapat pada subbab ciri-ciri virus sebagai benda mati dan makhluk hidup yaitu sebesar 80%. Validitas terendah pada soal KBF terdapat pada subbab peranan virus dalam kehidupan manusia yaitu sebesar 37,5%. Semua soal *essay* dinyatakan valid dengan rata-rata persentase 100%. Validitas butir soal pada materi virus dapat diamati pada Tabel 1.

Tabel 1. Analisis validitas instrumen *Assessment Diagnostic* pada materi virus

Soal	No. Soal	Kategori		Jumlah Soal	Persentase	
		Valid	Tidak Valid			
KBF	P1	1, 2, 4, 5, 6, 7, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20	✓	-	16	80%
		3, 8, 9, 11	-	✓	4	20%
	P2	3, 5, 9, 11, 12, 13, 15	✓	-	7	46.7%
		1, 2, 4, 6, 7, 8, 10, 14	-	✓	8	53.3%
	P4	1, 2, 6, 7, 9, 10	✓	-	6	60%
		3, 4, 5, 8	-	✓	4	40%
P5	1, 5, 6, 7, 9, 10	✓	-	6	60%	
	2, 3, 4, 8	-	✓	4	40%	
P6	4, 5, 6	✓	-	3	37.5%	
	1, 2, 3, 7, 8	-	✓	5	62.5%	
SCG	P3	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20	✓	-	19	95%
		13	-	✓	1	5%
		1P1 (point a)	✓	-	1	100%
Essay		1P2 (point b)	✓	-	1	100%
		2	✓	-	1	100%
		5	✓	-	1	100%
		6	✓	-	1	100%

Keterangan: P1: Pertanyaan nomor 1 (Kasus 1), P2: Pertanyaan nomor 2 (Kasus 2), P3: Pertanyaan nomor 3 (Kasus 3), P4: Pertanyaan nomor 4 (Kasus 4), P5: Pertanyaan nomor 5 (Kasus 5), P6: Pertanyaan nomor 6 (Kasus 6)

Tabel 1 menunjukkan hasil analisis validitas butir soal. Diketahui bahwa dari rerata persentase seluruh soal berbentuk KBF (P1, P2, P4, P5, dan P6) dan SCG (P3) terdapat 63,2% soal valid dan 36,8% soal tidak valid. Untuk semua soal *essay* dinyatakan 100% valid.

Contoh pernyataan soal yang valid pada subbab ciri-ciri virus dapat dilihat pada Gambar 3.

1.	Virus memiliki asam nukleat yang dibungkus selubung protein
2.	Virus tidak memiliki enzim untuk melakukan metabolisme sendiri

Gambar 3. Pernyataan soal valid

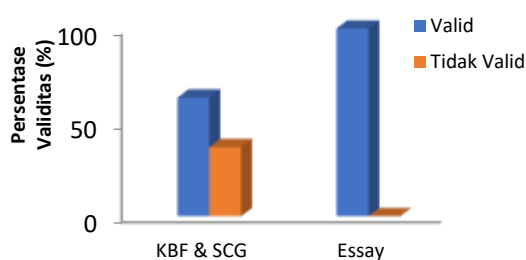
Sedangkan contoh pernyataan soal yang tidak valid dapat dilihat pada Gambar 4.

3.	Virus memiliki kemampuan replikasi didalam sel inang
8.	Keberadaan virus sangat bergantung pada inang yang ditumpangangi

Gambar 4. Pernyataan soal tidak valid

Uji validitas diawali dengan melakukan perhitungan terhadap korelasi *product moment*. Selanjutnya hasil dari perhitungan digunakan sebagai dasar untuk uji reliabilitas, dimana item soal yang tidak valid harus dihilangkan dan direvisi terlebih dulu. Validitas soal berkaitan dengan hasil reliabilitas. Semakin tinggi indeks reliabilitas (keajegan / ketepatan) suatu tes, berarti tinggi pula tingkat validitasnya (Donald, Lucy, & Asghar, 1982). Sejalan dengan hal tersebut, Amalia & Widayati (2012) menyatakan bahwa "Reliabilitas sebagai penyokong terbentuknya validitas butir soal".

Persentase validitas instrumen *Assessment Diagnostic* dapat dilihat pada Gambar 5.



Gambar 5. Validitas Butir Soal KBF, SCG, dan *Essay*

Dari gambar diatas, dapat dilihat bahwa tingkat validitas instrumen assersmen pada uji coba pertama mendekati 70% untuk soal KBF dan SCG. Sedangkan soal *essay* 100% valid.

Uji validitas berguna sebagai pemetaan butir-butir tes mana yang menyebabkan soal memiliki kriteria jelek, jika validitasnya rendah. Borualogo et al. (2017), berpendapat bahwa “Butir soal dikatakan valid apabila memberikan dukungan besar terhadap skor total”. Skor pada butir soal mengakibatkan skor total menjadi tinggi atau rendah. Jika skor pada butir soal memiliki kesejajaran dengan skor total, maka soal dikatakan memiliki validitas tinggi (Nofiana, 2015).

Arifin (2011) mengategorikan faktor-faktor yang berpengaruh terhadap hasil validitas, yaitu: jawaban siswa, instrumen evaluasi, dan administrasi evaluasi serta penskoran. Dalam pelaksanaannya siswa masih merasa asing dengan bentuk soal yang disediakan dan masih bingung dengan cara mengisi jawaban pada lembar jawab, meskipun pada lembar soal telah disediakan petunjuk cara pengisian jawaban. Siswa juga belum terbiasa untuk memahami maksud dari setiap pertanyaan dan cenderung menjawab sebatas dia memahami pertanyaan yang diberikan dan jarang bertanya. Siswa juga kesulitan untuk menyelesaikan semua soal dalam waktu 90 menit. Faktor jawaban siswa, diantaranya: kecenderungan siswa dalam menjawab soal secara cepat, dan penggunaan gaya bahasa tertentu dalam menjawab soal menjadi penentu valid atau tidaknya suatu butir soal.

Reliabilitas

Reliabilitas diartikan sebagai konsistensi dari suatu instrumen. Suatu tes dikatakan reliabel jika tetap menunjukkan hasil yang sama apabila diujikan pada kelompok yang sama dalam kesempatan berbeda (Maenani & Oktova, 2015). Hasil analisis reliabilitas instrumen *Assessment Diagnostic* pada materi virus disajikan pada Tabel 2.

Tabel 2. Analisis reliabilitas instrumen *Assessment Diagnostic* pada materi virus

Soal	Nilai (<i>Cronbach's Alpha</i>)	N of items	Keterangan
KBF & SCG	0.759	57	Reliabel
<i>Essay</i>	0.618	5	Reliabel

Berdasarkan Tabel 2, Analisis reliabilitas instrumen *Assessment Diagnostic* pada materi virus menunjukkan bahwa semua soal yang diujikan reliabel. Reliabilitas instrumen dibuktikan dengan nilai *Cronbach's Alpha* (Koefisien Alpha). *Cronbach's Alpha* merupakan suatu teknik yang digunakan untuk menguji stabilitas internal dari suatu tes, yang penerapannya bukan hanya digunakan pada tes dengan dua pilihan saja, tetapi meliputi reliabilitas skala pengukuran sikap dengan tiga, lima, atau tujuh pilihan (Arifin, 2011).

Cronbach's Alpha merupakan pengukuran keandalan indikator yang digunakan dalam kuesioner penelitian. Alasan penggunaan uji *Cronbach's Alpha*, yaitu: 1) Teknik ini paling sering digunakan pada pengujian keandalan kuesioner, 2) Mampu mendeteksi indikator-indikator yang tidak konsisten. Nilai *Cronbach's Alpha* tertera pada Tabel 3.

Tabel 3. Kriteria nilai *Cronbach's Alpha*

Nilai <i>Cronbach's Alpha</i>	Kriteria Reliabilitas
0.0 – 0.2	Sangat Rendah
>0.2 – 0.4	Rendah
>0.4 – 0.6	Cukup Tinggi
>0.6 – 0.8	Tinggi
>0.8 – 1.0	Sangat Tinggi

Hasil analisis reliabilitas yang baik pada penelitian ini, berkaitan erat dengan hasil validitas instrumen yang telah terukur cukup baik. Hal ini sejalan dengan pernyataan Purwanto (2011), bahwa “Tinggi rendahnya validitas menunjukkan tinggi rendahnya reliabilitas. Soal yang terdiri dari banyak butir soal akan lebih valid daripada yang memiliki sedikit butir, secara otomatis soal yang terdiri dari banyak butir akan lebih reliabel”.

Tingkat Kesukaran

Tingkat kesukaran adalah pengukuran seberapa besar derajat kesukaran suatu soal (Arifin, 2011). Perhitungan tingkat kesukaran soal menggunakan formula:

$$TK = \frac{WL+WH}{nL+nH} \times 100\%$$

Keterangan:

- WL = Jumlah siswa yang menjawab salah dari kelompok bawah
- WH = Jumlah siswa yang menjawab salah dari kelompok atas
- nL = Jumlah kelompok bawah
- nH = Jumlah kelompok atas

Berdasarkan analisis tingkat kesukaran butir soal berbentuk KBF dan SCG, diketahui bahwa

sebanyak 38,6% soal termasuk pada kategori mudah, 37,1% pada kategori sedang, dan 24,3% soal pada kategori sukar. Sedangkan soal *essay* terdiri atas

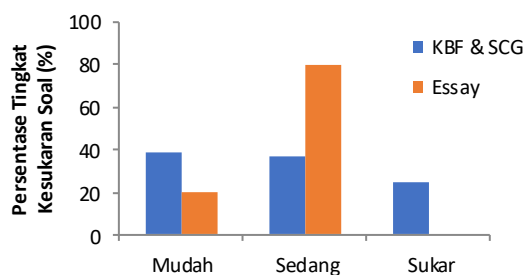
20% soal mudah dan 80% soal sedang. Hasil analisis tingkat kesukaran disajikan pada [Tabel 4](#).

Tabel 4. Analisis tingkat kesukaran instrumen *Assessment Diagnostic* pada materi virus

Kategori Tingkat Kesukaran	Soal KBF & SCG			Soal <i>Essay</i>			
	No. Soal	Jumlah Soal	Persentase	No. Soal	Jumlah Soal	Persentase	
Mudah $P < 0,3$	P1	1, 2, 3, 5, 6, 8, 10, 11, 12, 13, 15, 16, 18, 20	14	70%			
	P2	2, 7, 8, 13,	4	26.67%			
	P3	1	1	5%	6	1	20%
	P4	3, 4, 5, 8,	4	40%			
	P5	1, 3, 7, 9	4	40%			
	P6	1, 4, 6, 8	4	50%			
Sedang $0,3 < P < 0,7$	P1	4, 7, 14, 17, 19,	5	25%			
	P2	3, 4, 5, 6, 9, 10, 11, 14, 15	9	60%			
	P3	2, 4, 7, 8, 14, 15,	6	30%	1P1, 1P2, 2,	4	80%
	P4	7, 10	2	20%	5		
	P5	2, 5, 6, 8, 10	5	50%			
	P6	3, 5, 7	3	37.5%			
Sukar $P > 0,7$	P1	9	1	5%			
	P2	1, 12	2	13.33%			
	P3	3, 5, 6, 9, 10, 11, 12, 13, 16, 17, 18, 19, 20	13	65%	-	-	-
	P4	1, 2, 6, 9,	4	40%			
	P5	4	1	10%			
	P6	2	2	12.5%			

Berdasarkan [Tabel 4](#), analisis tingkat kesukaran butir soal KBF dan SCG sebanyak 38,6% soal termasuk pada kategori mudah, 37,1% soal pada kategori sedang, dan 24,3% soal pada kategori sukar. Sedangkan soal *essay* terdiri atas 20% soal pada kategori mudah dan 80% soal pada kategori sedang. Pada soal *essay*, persentase tingkat kesukaran soal didominasi pada kriteria sedang yaitu 80%. Hal ini menunjukkan bahwa sebagian besar soal berbentuk *essay*, dapat dijawab oleh siswa. Namun soal *essay* belum memenuhi keseimbangan proporsi tingkat kesukaran.

Tingkat kesukaran soal dapat dikelompokkan ke dalam 3 kategori, yaitu 1) Soal dengan kategori sukar memiliki nilai $P < 0,3$, 2) Soal sedang memiliki nilai $P = 0,3 - 0,7$, dan 3) Soal mudah memiliki nilai $P > 0,7$. Persentase tingkat kesukaran soal dapat diamati pada [Gambar 6](#).



Gambar 6. Tingkat Kesukaran Butir Soal KBF, SCG, dan *Essay*

Berdasarkan [Gambar 6](#), persentase soal pada kategori sedang untuk soal KBF, SCG, dan *essay* mendominasi pada aspek tingkat kesukaran instrumen. Soal dikatakan baik apabila tidak terlalu sukar dan tidak terlalu mudah. Butir soal yang terlalu mudah atau terlalu sukar boleh digunakan, tetapi dengan memperhatikan proporsinya.

Naskah soal sebaiknya memiliki proporsi tingkat kesukaran berimbang, yaitu: 25% soal sukar, 50% soal sedang, dan 25% soal mudah ([Amalia & Widayati, 2012](#)). Untuk memperoleh hasil belajar yang baik, proporsi tingkat kesukaran soal harus tersebar secara normal. Terdapat tiga perhitungan perbandingan soal sukar, sedang, mudah dalam mencapai proporsi normal, yaitu: 1) 25% : 50% : 25%, atau 2) 20% : 60% : 20%, atau 3) 15% : 70% : 15% ([Arifin, 2011](#)).

Menurut [Maenani & Oktova \(2015\)](#), meminta *judgment*/penilaian dari guru merupakan salah satu upaya untuk memperoleh komposisi soal dengan kriteria yang sesuai. Akan tetapi hal ini juga memiliki kelemahan, sebab belum tentu butir soal yang dianggap guru sebagai soal yang sulit juga dirasakan sulit oleh siswa. Guru mengklasifikasikan butir soal pada kriteria tingkat kesukaran tertentu (mudah, sedang, dan sukar) berdasarkan intuisinya. Sehingga sangat sulit menentukan kriteria suatu butir soal, sebelum siswa melakukan tes.

Beberapa hal yang harus diperhatikan dalam penyusunan soal berkaitan dengan tingkat kesukaran,

yaitu: 1) Soal yang terlalu sukar atau mudah tidak memberikan informasi yang berguna bagi sebagian besar siswa karena syarat distribusi jawaban tidak terpenuhi, 2) Soal masih memenuhi syarat untuk diterima jika menunjukkan jawaban yang merata, logis, dan daya bedanya negatif (kecuali kunci jawaban), 3) Jika daya pembeda memenuhi kriteria, maka soal yang terlalu mudah atau sukar tersebut dapat diterima, 4) Jika daya pembeda belum memenuhi kriteria, maka soal tersebut harus direvisi dan diuji coba lagi (Arifin, 2011).

Daya Pembeda

Daya pembeda adalah kemampuan soal untuk membedakan siswa pandai dan siswa yang kurang pandai). Analisis daya pembeda menggunakan formula:

$$DP = \frac{(WL - WH)}{n}$$

Keterangan:

DP = Daya pembeda

WL = Jumlah siswa yang gagal dari kelompok bawah

WH = Jumlah siswa yang gagal dari kelompok atas

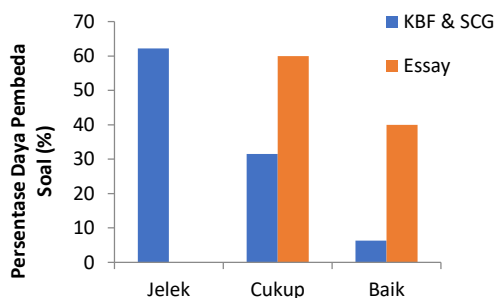
n = 27% x jumlah responden

Berdasarkan hasil analisis daya pembeda soal, persentase instrumen *Assessment Diagnostic* pada materi virus terdiri atas 62,2% soal dengan kategori jelek, 31,5% soal dengan kategori cukup, dan 6,3% soal dengan kategori baik pada soal berbentuk KBF dan SCG. Sedangkan pada soal *essay* terdiri atas 60% soal dengan kategori daya pembeda cukup dan 40% soal dengan kategori daya pembeda baik. Analisis daya pembeda butir soal *Assessment Diagnostic* pada materi virus dapat dilihat pada Tabel 5.

Tabel 5. Analisis daya pembeda instrumen *Assessment Diagnostic* pada materi virus

Kategori Daya Beda	Soal KBF & SCG				Soal Essay		
	No. Soal	Jumlah Soal	Persentase	No. Soal	Jumlah Soal	Persentase	
Jelek <0.20	P1	1, 3, 8, 9, 10, 11, 13, 15, 20	9	45%			
	P2	1-14	14	93.3%			
	P3	6, 9, 13, 16, 18, 20	6	30%			
	P4	1, 3, 4, 5, 8, 9	6	60%	-	-	-
	P5	1, 2, 3, 4, 5, 7, 8	7	70%			
	P6	1, 2, 3, 6, 7, 8	6	75%			
Cukup 0.2-0.4	P1	2, 4, 5, 6, 7, 12, 14, 16, 17, 18, 19	11	55%			
	P2	15	1	6.67%	1P1,		
	P3	1, 2, 3, 4, 5, 11, 12, 17, 19	9	45%	1P2,	3	60%
	P4	2, 6, 7, 10	4	40%	5		
	P5	6, 9, 10	3	30%			
	P6	4	1	12.5%			
Baik >0.40	P3	7, 8, 10, 14, 15	5	25%	2, 6	2	40%
	P6	5	1	12.5%			

Tabel 5 menunjukkan 62,2% soal termasuk dalam kategori jelek; 31,5% soal dengan kategori cukup, dan 6,3% soal dengan kategori baik pada KBF dan SCG. Sedangkan pada soal *essay*, 60% soal memiliki daya pembeda cukup dan 40% soal memiliki daya pembeda baik. Persentase daya pembeda instrumen *Assessment Diagnostic* dapat dilihat pada Gambar 7.



Gambar 7. Persentase daya pembeda butir soal KBF, SCG, dan *Essay*

Gambar 7 menunjukkan bahwa instrumen *Assessment Diagnostic* pada aspek daya pembeda memiliki kriteria yang tidak baik. Persentase soal yang jelek lebih besar dari 50% yang berarti sebagian besar soal tidak mempunyai daya pembeda.

Soal mempunyai daya pembeda, jika soal dapat dijawab oleh siswa berkemampuan tinggi tetapi tidak dapat dijawab oleh siswa berkemampuan rendah. Jika soal dapat dijawab oleh semua siswa, baik siswa pintar maupun kurang pintar, maka soal tersebut tidak mempunyai daya beda. Begitu juga jika siswa pintar dan siswa kurang pintar tidak bisa menjawab, berarti soal tersebut tidak baik karena tidak mempunyai daya pembeda.

Butir soal dengan kriteria jelek sebaiknya tidak digunakan, butir soal yang memiliki daya beda cukup dapat diterima dengan perbaikan, dan butir soal yang memiliki daya beda baik dapat diterima tanpa perbaikan.

Untuk meningkatkan daya pembeda, sebaiknya dilakukan ujicoba soal sebelum digunakan. Beberapa

faktor penunjang, agar diperoleh soal dengan daya pembeda baik menurut pendapat Maenani & Oktova (2015), yaitu: mengoptimalkan proses penyusunan dan perakitan soal sesuai dengan kaidah, rumusan soal harus jelas sehingga tidak menimbulkan penafsiran ganda, gambar, grafik,

tabel, dan diagram harus jelas dan berfungsi, soal menggunakan bahasa yang komunikatif.

Berdasarkan hasil analisis butir soal yang telah dilakukan, peneliti menentukan kategorisasi dalam menentukan butir soal mana yang diterima, diterima dengan perbaikan, dan butir soal yang ditolak. Matriks kategorisasi analisis kualitas soal dapat dilihat pada Tabel 6.

Tabel 6. Kategorisasi kualitas butir soal instrumen *Assessment Diagnostic* pada materi virus

Validitas	Reliabilitas	Aspek Tingkat Kesukaran			Daya Pembeda	Kesimpulan Kategorisasi
		Mudah	Sedang	Sukar		
Valid	✓	✓	✓	✓	Baik	Diterima
	✓	✓	✓	✓	Cukup	Diterima
	✓	✓	✓	✓	Jelek	Diterima dengan perbaikan
	✓	✓	✓	✓	Jelek	Diterima dengan perbaikan
Tidak Valid	✓	✓	✓	✓	Baik	Ditolak
	✓	✓	✓	✓	Cukup	Ditolak
	✓	✓	✓	✓	Jelek	Ditolak
	✓	✓	✓	✓	Jelek	Ditolak

Berdasarkan hasil analisis kuantitatif butir soal, diperoleh kesimpulan kategorisasi dimana 40 soal diterima, 23 soal yang memiliki daya pembeda jelek akan diperbaiki, dan 25 butir soal tidak valid dengan daya pembeda jelek akan dibuang. Setelah dilakukan revisi, selanjutnya instrumen akan diuji kembali pada uji coba kedua.

Menurut Amalia & Widayati (2012), soal yang berkualitas adalah soal yang dapat memberikan informasi yang tepat dalam menentukan siswa yang sudah dan belum menguasai materi yang diajarkan. Manfaat melakukan analisis butir soal, yaitu: 1) Memperoleh informasi perbaikan soal dengan mengidentifikasi soal-soal yang kurang baik, 2) Meningkatkan validitas dan reliabilitas soal, dan 3) Meningkatkan keterampilan guru dalam menyusun instrumen.

Hasil analisis kualitas butir soal diukur berdasarkan jawaban siswa. Kemampuan siswa dalam menjawab soal *Assessment Diagnostic* dipengaruhi oleh banyak faktor. Beberapa penyebab kesulitan siswa menurut peneliti diantaranya: 1) Siswa tidak pernah dihadapkan pada bentuk soal KBF dan SCG sehingga merasa asing dengan bentuk soal tersebut, 2) Pemahaman siswa terkait materi virus hanya sebatas pada tahapan mengingat/hafalan, 3) Motivasi, minat, dan perhatian siswa yang rendah pada pelajaran biologi khususnya materi virus akan

menimbulkan ketidaktekunan belajar yang akhirnya berdampak pada menurunnya prestasi belajar siswa (Winarni, 2014).

Simpulan

Validitas soal KBF dan SCG sebesar 63,2% dan soal *essay* 100% valid, semua butir soal reliabel, persentase tingkat kesukaran soal KBF dan SCG yaitu 38,6% mudah, 37,1% sedang, dan 24,3% sukar. Pada soal *essay* 20% soal pada kategori mudah dan 80% soal pada kategori sedang; Persentase daya pembeda soal KBF dan SCG yaitu 62,2% jelek, 31,5% cukup dan 6,3% baik. Pada soal *essay*, 60% soal dengan kategori daya pembeda cukup dan 40% soal dengan kategori baik. Berdasarkan hasil analisis kuantitatif butir soal, 40 soal diterima, 23 soal yang memiliki daya pembeda jelek akan diperbaiki, dan 25 butir soal tidak valid dengan daya pembeda jelek akan dibuang.

Referensi

- Amalia, A. N., & Widayati, A. (2012). Analisis butir soal tes kendali mutu kelas XII SMA mata pelajaran ekonomi akuntansi di kota Yogyakarta tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, 10(1), 1–26. <https://doi.org/10.21831/jpai.v10i1.919>

- Aprilia, N. (2016). Implementasi model pembelajaran reflektif untuk meningkatkan kemampuan pemahaman mahasiswa pendidikan biologi pada mata kuliah strategi pembelajaran di program studi FKIP Universitas Ahmad Dahlan. *JURNAL BIOEDUKATIKA*, 4(1), 27–30. <https://doi.org/10.26555/bioedukatika.v4i1.4739>
- Arifin, Z. (2011). *Evaluasi pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Ayu, D., Saputri, F., & Widyaningrum, T. (2016). Misconceptions analysis on the virus chapter in biology textbooks for high school students grade X, 1(1), 31–37. <https://doi.org/10.15294/ijal.v1i1.7781>
- Borualogo, I. S., Kusdiyati, S., Susandari, S., & Sirodj, D. A. N. (2017). Analisis item soal UTS pedologi semester ganjil 2015-2016. *SCHEMA Journal of Psychological Research*, 3(1), 46–57. Diambil dari <https://ejournal.unisba.ac.id/index.php/schema/article/view/1808/1761>
- Chi, M. T. H., Slotta, J. D., & Leeuw, N. de. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Journal Learning and Instruction*, 4, 27–43. Diambil dari <http://www.public.asu.edu/~mtchi/papers/ChiSlottaLeeuw.pdf>
- Clark, I. (2015). Formative assessment : Translating high-level curriculum principles into classroom practice. *The Curriculum Journal*, 26(1), 91–114. <https://doi.org/10.1080/09585176.2014.990911>
- Donald, A., Lucy, J. C., & Asghar, A. (1982). *Pengantar penelitian dalam pendidikan*. (A. Furchan, Penerj.). Surabaya: Usaha Nasional.
- Güne, M. H., Güne, O., & Hoplan, M. (2011). The using of computer for elimination of misconceptions about photosynthesis. *Procedia - Social and Behavioral Sciences*, 15, 1130–1134. <https://doi.org/10.1016/j.sbspro.2011.03.251>
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 989–1008. <https://doi.org/10.12973/eurasia.2015.1369a>
- Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences* (3 ed.). San Diego: EdITS. [Google Books](https://books.google.com/books)
- Jayawardana, H. B. A. (2017). Paradigma pembelajaran biologi di era digital. *JURNAL BIOEDUKATIKA*, 5(1), 12–17. <https://doi.org/10.26555/bioedukatika.v5i1.5628>
- Johnstone, A. H., Bahar, M., & Hansell, M. H. (2000). Structural communication grids: A valuable assessment and diagnostic tool for science teachers. *Journal of Biological Education*, 34(2), 87–89. <https://doi.org/10.1080/00219266.2000.9655691>
- Kara, F., & Çelikler, D. (2015). Development of achievement test : Validity and reliability study for achievement test on matter changing. *Journal of Education and Practice*, 6(24), 21–26. Diambil dari <https://iiste.org/Journals/index.php/JEP/article/view/25254/25854>
- Kose, E. O., Pekel, O., & Hasenekoglu, I. (2009). Misconceptions and alternative concepts in biology textbooks : Photosynthesis and respiration. *Journal of Science Education*, 10(2), 91–93. Diambil dari <http://chinaxjy.com/downloads/2006-2010.html>
- Maenani, L., & Oktova, R. (2015). Analisis butir soal fisika ulangan umum kenaikan kelas X Madrasah Aliyah Se-Kabupaten Banjarnegara, Jawa Tengah tahun pelajaran 2011/2012. *Jurnal Berkala Fisika Indonesia*, 7(1), 5–11. Diambil dari <http://journal.uad.ac.id/index.php/BFI/article/view/1539/1049>
- Mustaqim, T. A., Zulfiani, Z., & Herlanti, Y. (2014). Identifikasi miskonsepsi siswa dengan menggunakan metode certainty of response index (CRI) pada konsep fotosintesis dan respirasi tumbuhan. *Edusains*, 6(2), 145–152. <https://doi.org/10.15408/es.v6i2.1117>
- Nofiana, M. (2015). Kualitas penulisan butir soal ujian nasional biologi tahun 2014/2015 ditinjau dari aspek teoritik. *Saintifika*, 17(1), 1–13. Diambil dari <https://jurnal.unej.ac.id/index.php/STF/article/view/3127/2495>
- Nofiana, M., Sajidan, S., & Puguh. (2016). Pengembangan instrumen evaluasi higher order thinking skills pada materi kingdom plantae. *Pedagogi Hayati*, 1(1), 46–53. Diambil dari <http://ojs.umrah.ac.id/index.php/pedagogihayati/article/view/37>
- Osborne, J. (2004). *IDEAS: Ideas, evidence and argument in science (IDEAS) project*. London: King's College London.
- Purwanto. (2011). *Evaluasi hasil belajar*. Yogyakarta: Pustaka Pelajar.
- Sudarisman, S. (2015). Memahami hakikat dan karakteristik pembelajaran biologi dalam upaya menjawab tantangan abad 21 serta optimalisasi implementasi kurikulum 2013. *Florea : Jurnal Biologi dan Pembelajarannya*, 2(1), 29–35. <https://doi.org/10.25273/florea.v2i1.403>
- Suparno, P. (2013). *Miskonsepsi dan perubahan konsep dalam pendidikan fisika*. Jakarta: PT Gramedia Widiasarana Indonesia. [Google Books](https://books.google.com/books)

- Suratno, T. (2008). Konstruktivisme, konsepsi alternatif dan perubahan konseptual dalam pendidikan IPA. *Jurnal Pendidikan Dasar*, 10(1), 1–3.
- Syahrul, D. A., & Setyarsih, W. (2015). Identifikasi miskonsepsi dan penyebab miskonsepsi siswa dengan three-tier diagnostic test pada materi dinamika rotasi. *Jurnal Inovasi Pendidikan Fisika*, 4(3), 67–70. Diambil dari <http://jurnalmahasiswa.unesa.ac.id/index.php/inovasi-pendidikan-fisika/article/view/13305/12214>
- Winarni, S. (2014). Pengaruh perhatian guru, motivasi belajar, dan kecerdasan emosional terhadap prestasi belajar biologi siswa SMA Negeri 2 Bantul. *JURNAL BIOEDUKATIKA*, 2(1), 42.
- <https://doi.org/10.26555/bioedukatika.v2i1.4111>
- Yankovskaya, A., Levin, I., & Fuks, I. (2014). Assessment of teaching and learning by mixed diagnostic testing. In *Proceedings of the Frontiers in Mathematics and Science Education Research Conference*. Famagusta. Diambil dari https://www.researchgate.net/profile/Ilya_Levin/publication/262915611_Assessment_of_Teaching_and_Learning_by_Mixed_Diagnostic_Testing/links/58a4923c92851cf0e3894741/Assessment-of-Teaching-and-Learning-by-Mixed-Diagnostic-Testing.pdf