

Penyusunan Instrumen Tes *Higher Order Thinking Skill* Pada Materi Ekosistem SMA Kelas X

Kusuma Wardany^{1*}, Sajidan², Murni Ramli²

¹Program Studi Pendidikan Sains Pascasarjana Fakultas Keguruan dan Ilmu Pendidikan
Universitas Sebelas Maret Surakarta, Indonesia

² Pendidikan Biologi FKIP Universitas Sebelas Maret Surakarta, Indonesia

* Email: kusuma.wardany@gmail.com

Abstract: Penelitian ini bertujuan untuk menyusun instrumen tes *Higher Order Thinking Skill* pada materi ekosistem yang akan diujikan pada siswa SMA kelas X. Instrumen disusun berdasarkan langkah Riset Pengembangan Borg & Gall. Langkah pertama adalah menganalisis semua bentuk tes yang ada pada Ujian Nasional, Ulangan formatif dan sumatif, dan tes yang ada di dalam buku Biologi SMA dengan sampel SMA di Surakarta. Hasil analisis menunjukkan bahwa sebagian besar tes mencerminkan Kemampuan Berpikir Tingkat Rendah (LOTS). Langkah kedua adalah penyusunan draft Tes HOTS pada materi ekosistem yang terdiri dari 20 butir soal pilihan ganda yang akan dikerjakan dalam waktu 60 menit. Draft tes divalidasi oleh validator berkualifikasi dan guru. Instrumen tes selanjutnya diujikan pada sejumlah kecil siswa dalam Uji Coba Pertama. Hasil uji coba kesatu dianalisis secara kuantitatif menggunakan program MicroCat ITEMAN versi 3.00 yang digunakan untuk menentukan reliabilitas, taraf kesukaran, daya beda serta efektivitas distraktor. Berdasarkan hasil Uji Coba Tahap I diketahui bahwa instrumen tes *Higher Order Thinking Skill* pada siswa SMA kelas X untuk materi Ekosistem yang menunjukkan 45 % item diterima atau valid, 40% item direvisi dan 15% item ditolak. Proses revisi draft instrumen tes dan Uji Coba Tahap II dalam proses penyelesaian..

Keywords: HOTS, assesmen, ekosistem, Biology SMA

1. PENDAHULUAN

Assesmen atau penilaian merupakan istilah umum yang mencakup keseluruhan prosedur yang digunakan untuk mendapatkan informasi tentang hasil belajar siswa (pengamatan, peringkat, pengujian menggunakan kertas dan pensil) dan membuat penilaian mengenai proses pembelajaran (Gronlund & Linn, 1995). Dalam dunia pendidikan, penilaian diartikan sebagai prosedur yang digunakan untuk mendapatkan informasi untuk mengukur taraf pengetahuan dan keterampilan subjek didik yang hasilnya akan digunakan untuk keperluan evaluasi (Reynolds, Livingston, Willson, 2010).

Hasil pemetaan pendidikan melalui pemenuhan delapan (8) Standar Nasional Pendidikan (SNP) berbagai Sekolah Menengah Atas (SMA) Negeri di Surakarta menunjukkan bahwa capaian pemenuhan standar proses (standar 2) dan standar penilaian (standar 8) menunjukkan adanya gap yang terbesar antara skor ideal dan skor riil, yaitu sebesar 3,70 %. Hal ini terjadi karena ketercapaian kompetensi pembelajaran di SMA Surakarta melalui aspek kognitif belum dapat meningkatkan *Higher Order Thinking Skill* siswa (Sajidan, Sugiharto & Prasetyani, 2013).

Higher Order Thinking Skill merupakan suatu keterampilan berpikir yang tidak hanya membutuhkan keterampilan mengingat, tetapi membutuhkan keterampilan lain yang lebih tinggi. Indikator untuk mengukur *Higher Order Thinking Skill* meliputi keterampilan menganalisa (C4), mengevaluasi (C5), dan menciptakan (C6) (Anderson & Krathwohl, 2001). *Higher Order Thinking Skill* sebagai keterampilan berpikir yang terjadi ketika seseorang mengambil informasi baru dan informasi yang sudah tersimpan dalam ingatannya, selanjutnya menghubungkan informasi tersebut dan menyampaikannya untuk mencapai tujuan atau jawaban yang dibutuhkan (Lewis & Smith, 1993).

Hal ini sejalan dengan karakteristik *skills* masyarakat abad ke-21 yang dipublikasikan oleh *Partnership of 21st Century Skill* mengidentifikasi bahwa pebelajar pada abad ke-21 harus mampu mengembangkan keterampilan kompetitif yang diperlukan pada abad ke-21 yang berfokus pada pengembangan *Higher Order Thinking Skill*, seperti: berpikir kritis (*critical thinking*), pemecahan masalah (*problem solving*), keterampilan berkomunikasi (*communication skills*), melek TIK, teknologi



informasi dan komunikasi (*ICT, information and Communication Technology*), melek informasi (*information literacy*), dan melek media (*media literacy*) (Basuki & Haryanto, 2012).

Hasil Observasi SMA di Surakarta yang dipilih secara acak, yaitu di SMAN 2, SMAN 3, SMAN 4, SMAN 6, dan SMAN 7 melalui Ulangan dan Ujian seperti Ujian Nasional, Ujian Akhir Semester, Ujian Tengah Semester, Ujian Sekolah, Ulangan Harian, maupun dari buku paket yang guru dan siswa gunakan pada materi ekosistem dan lingkungan menunjukkan bahwa soal dan pertanyaan masih dalam ranah kognitif yang rendah (*lower order thinking skills*). Rendahnya presentase soal *Higher Order Thinking Skills* menjadi indikator rendahnya kognitif siswa di sekolah.

Penilaian yang mengukur *Higher Order Thinking Skill* dapat menggunakan bentuk tes subjektif dan tes objektif. Tes subjektif merupakan bentuk tes bentuk esai. Tes esai adalah suatu bentuk uraian dengan mempergunakan bahasa sendiri. Dalam tes bentuk esai siswa dituntut untuk berpikir tentang dan mempergunakan apa yang diketahui yang berkenaan dengan pertanyaan yang harus dijawab. Tes objektif merupakan bentuk tes yang terdiri dari tes jawaban benar-salah (*true false*), pilihan ganda (*multiple choice*), isian (*completion*), dan penjodohan (*matching*) (Suwandi, 2009).

Untuk menyatakan hasil belajar baik atau buruk, berhasil atau gagal, sukses atau tidaknya sesuatu, maka data harus benar-benar dapat dipercaya/akurat agar ketetapan yang di ambil tidak salah. Jika salah datanya salah pula hasil penilaiannya dan akibatnya salah pula keputusannya. Oleh karena itu diperlukan adanya alat tes yang baik agar data dapat akurat (Subali, 2010).

Berdasarkan hal tersebut, masalah dalam penelitian ini adalah bagaimana menyusun instrumen tes *Higher Order Thinking Skill* pada materi ekosistem yang akan diujikan pada siswa SMA kelas X.

Adapun tujuan dari penelitian ini adalah untuk menyusun instrumen tes *Higher Order Thinking Skill* pada materi ekosistem yang akan diujikan pada siswa SMA kelas X.

2. METODE PENELITIAN

Instrumen disusun berdasarkan langkah Riset Pengembangan Borg & Gall (1983). Teknik analisis data yang digunakan yaitu analisis data kualitatif dan kuantitatif. Analisis secara kualitatif dilakukan melalui penelaah untuk mengetahui validitas isi instrument tes yaitu kesesuaian antara soal-soal dalam tes dengan indikator yang telah disusun sebelumnya. Analisis secara kuantitatif dilakukan dengan pendekatan teori tes klasik yang dibantu oleh program MicroCat versi 3.00. Beberapa aspek yang dianalisis secara kuantitatif yaitu reliabilitas, tingkat kesukaran, butir soal, daya pembeda butir soal serta efisiensi pengecoh.

2.1 Reliabilitas

Reliabilitas seringkali disebut dengan derajat konsistensi (keajegan). Sebuah alat ukur mempunyai reliabilitas tinggi, artinya meskipun pengukuran dilakukan berulang-ulang dengan alat ukur tersebut menghasilkan informasi yang sama atau mendekati sama (Purwanto, 2010). Reliabilitas instrument tes juga ditinjau dari hasil analisis program ITEMAN versi 3.00 yaitu nilai *alpha*.

2.2 Tingkat Kesukaran

Tingkat kesukaran suatu pokok uji atau soal (dilambangkan dengan P) adalah proporsi dari keseluruhan siswa yang menjawab benar pada pokok uji atau soal.

Bilangan yang menunjukkan sukar dan mudahnya suatu soal disebut indeks kesukaran (p). Besarnya indeks kesukaran antara 0,00 sampai dengan 1,00. Soal dengan indeks kesukaran 0,0 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1,0 menunjukkan bahwa soal itu terlalu mudah. Analisis terhadap tingkat kesukaran butir soal penelitian ditinjau dari nilai Prop. Correct (p) hasil analisis ITEMAN versi 3.00. Berdasarkan nilai p tersebut dapat diketahui butir soal yang termasuk kriteria mudah, sedang atau sulit. Butir soal yang baik dan dapat diterima adalah butir soal yang termasuk kriteria sedang, sedangkan butir yang termasuk kriteria mudah atau sulit dianggap sebagai soal yang tidak baik dan perlu direvisi.

2.3 Daya Pembeda

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah. menjawab benar pada soal yang dianalisis. Pada hasil analisis menggunakan program ITEMAN versi 3.00. Daya beda butir soal dilihat dari nilai Biser, dan Point Biser. Semakin tinggi nilai biserial maka semakin tinggi kemampuan butir soal tersebut untuk membedakan siswa yang pandai dan kurang pandai.

3. HASIL DAN PEMBAHASAN

Sebelum instrumen diujikan ke lapangan, instrumen ditelaah secara kualitatif. Telaah instrumen secara teoritis atau deskriptif dilakukan untuk melihat keterbacaan instrumen dan untuk validasi isi. Telaah instrumen secara deskriptif dilakukan dengan peninjauan aspek materi, konstruksi dan bahasa. Dalam penelitian ini, telaah instrumen secara deskriptif dilakukan oleh ahli instrument evaluasi, dosen ahli materi, ahli bahasa dan guru praktisi. Hasil telaah deskriptif ditemukan beberapa soal yang belum sesuai dengan kriteria sehingga harus direvisi.

Analisis kualitatif dilakukan untuk mereview butir soal dari aspek materi, kontruksi, dan bahasa sehingga diketahui validitas instrument tes.

Tabel 1. Hasil Penilaian Aspek Materi *HOTS* dari Ahli Materi (Materi Ekosistem)

No	Aspek yang dinilai	Skor (%)	Konversi	Kriteria
1.	Kebenaran	98,08	A	Sangat
2.	Materi Soal	98	A	Baik
3.	Kedalaman Materi Keterbacaan	100	A	Sangat Baik
Rata-Rata		98,72		

Sumber: Lembar Penilaian Ahli Materi

Tabel menunjukkan skor rata-rata keseluruhan aspek materi adalah 98,72% dan kategori termasuk "A".

Tabel 2. Hasil Penilaian Aspek Materi *HOTS* dari Guru Praktisi 1 (Materi Ekosistem)

No	Aspek yang dinilai	Skor (%)	Konversi	Kriteria
1.	Isi Soal	100	A	Sangat Baik
2.	Keterbacaan	75	B	Baik
3.	Soal	75	B	Baik
4.	Penggunaan	100	A	Sangat Baik
5.	Bahasa Rubrik Penilaian Manajemen Waktu	75	B	Baik
Rata-Rata		85		

Sumber: Lembar Penilaian Guru Praktisi

Tabel 2 menunjukkan skor rata-rata keseluruhan aspek materi adalah 85 % dan kategori termasuk "B" (Baik)

Tabel 3. Hasil Penilaian Aspek Materi *HOTS* dari Guru Praktisi 2 (Materi Ekosistem)

No	Aspek yang dinilai	Skor (%)	Konversi	Kriteria
1.	Isi Soal	88,78	A	Sangat Baik
2.	Keterbacaan	99,36	A	Sangat Baik
3.	Soal	95,67	A	Sangat Baik
4.	Penggunaan	99,04	A	Sangat Baik
5.	Bahasa Rubrik Penilaian Manajemen Waktu	100	A	Sangat Baik
Rata-Rata		96,57		

Sumber: Lembar Penilaian Guru Praktisi

Tabel 3 menunjukkan skor rata-rata keseluruhan aspek materi adalah 96,57 % dan kategori termasuk "A" (Sangat Baik).

Tabel 4. Hasil Penilaian Aspek Bahasa dari Ahli Bahasa (Materi Ekosistem)

No	Aspek yang dinilai	Skor (%)	Konversi	Kriteria
1.	Bahasa Indonesia yang baik dan benar	100	A	Sangat Baik
2.	Peristilahan	97,78	A	Sangat Baik
3.	Kejelasan Bahasa Kesesuaian Bahasa	93,33	A	Sangat Baik
4.	Kejelasan Bahasa Kesesuaian Bahasa	91,67	A	Sangat Baik
Rata-Rata		95,69		

Sumber: Lembar Penilaian Ahli Bahasa

Tabel 4 menunjukkan skor rata-rata keseluruhan aspek materi adalah 95,69 % dan kategori termasuk "A" (Sangat Baik)

Tabel 5. Hasil Penilaian Aspek Materi *HOTS* dari Ahli Instrumen (Materi Ekosistem)

No	Aspek yang dinilai	Skor (%)	Konversi	Kriteria
1.	Ranah Materi	87,75	A	Sangat Baik
2.	Ranah Materi	60,27	C	Baik
3.	Ranah Kontruksi	75	B	Cukup
4.	Ranah Bahasa Rubrik Penilaian	100	A	Sangat Baik
Rata-Rata		80,95		

Sumber: Lembar Penilaian Guru Praktisi

Tabel 5 menunjukkan skor rata-rata keseluruhan aspek materi adalah 80,95% dan kategori termasuk "B" (Baik)

Setelah instrumen tes ditelaah secara deskriptif, selanjutnya instrumen tersebut diujicobakan pada siswa kelas X di SMA 2 Karanganyar. Uji coba I dilaksanakan pada Rabu, 16 Mei 2012. Analisis kuantitatif dilakukan dengan menggunakan program MicroCat ITEMAN versi 3.00. Program ITEMAN versi 3.00 secara otomatis menganalisis tingkat kesukaran, daya beda, reabilitas soal serta beberapa data statistik lainnya. Hasil analisis instrumen tes secara keseluruhan pada uji coba I dapat dilihat pada tabel dibawah.

Hasil analisis butir tes ditinjau dari tingkat kesukaran, daya beda, serta efektifitas distraktor pada program MicroCat ITEMAN versi 3.00 disajikan lebih rinci pada Tabel 2, Tabel 3, dan Tabel 4.



Tabel 6. Tingkat Kesukaran Item Tes Uji Coba I

kategori	Tes	
	Nomor item	Jumlah
Sulit ($P < 0,3$)	1, 2, 5, 6, 7, 9, 10, 11, 12, 14, 16, 17, 20	13 (65%)
Sedang ($0,3 < P < 0,7$)	3, 4, 13, 18, 19	5 (25%)
Mudah ($P > 0,7$)	8, 15	2 (10%)
Jumlah		20

Berdasarkan Tabel 6, untuk tes terdapat 65% item yang termasuk kategori sulit, 25% item termasuk kategori sedang dan terdapat 10% mudah. Secara keseluruhan 13 butir soal yang tingkat kesukarannya sulit memerlukan revisi, begitu pula untuk untuk 2 butir soal yang tingkat kesukarannya mudah juga memerlukan revisi. Dan untuk kategori sedang sebanyak 5 butir soal dapat diterima.

Tabel 7. Daya Beda Item Tes Uji Coba I

Kategori	Tes tipe A		Keterangan
	Nomor item	Jumlah	
Rendah Sekali ($< 0,0$)	1, 5, 13, 14, 15	5 (25%)	Ditolak (25%)
Rendah ($0,00-0,20$)	9	1 (5%)	Direvisi (5%)
Sedang ($0,20-0,40$)	3, 8, 18, 19	4 (20%)	
Tinggi ($0,40-0,70$)	6, 10, 16, 17, 18	5 (25%)	Diterima (70%)
Tinggi Sekali ($0,70-1,00$)	2, 4, 7, 11, 20	5 (25%)	
Jumlah		20	

Berdasarkan Tabel 7 tingkat daya beda, terdapat 25% item dengan kategori rendah sekali dan menunjukkan sebanyak 5 butir soal perlu direvisi. Begitu pula sebesar 5% item dengan 1 butir soal yang perlu direvisi. Terdapat 70% item dengan kategori 20 % item sedang sebanyak 4 butir soal, 25% item dengan kategori tinggi sebanyak 5 butir soal dan 25% item dengan kategori tinggi sekali sebanyak 5 butir sial dapat diterima.

Tabel 8. Efektivitas Distraktor Item Tes Uji Coba I

Distraktor yang berfungsi	Tes tipe A		Keterangan
	Nomor item	Jumlah	
3	13, 15, 16, 18, 19	5 (25%)	Baik
2	4, 6, 7, 8, 9, 12, 17	7 (35%)	Baik
1	2, 3, 10, 11, 20	5 (25%)	Tidak baik
0	1, 5, 14	3 (15%)	Tidak baik
Jumlah		20	

Pada Tabel 8 dapat dilihat bahwa paket tes memiliki 60% item yang semua distraktor berfungsi baik, serta 8% item distraktor yang tidak berfungsi. Secara keseluruhan berdasarkan efektifitas distraktor sebanyak 8 butir soal dengan 40% item perlu direvisi.

Berdasarkan hasil analisis untuk tingkat kesukaran, daya beda dan efektifitas distraktor menggunakan program ITEMAN versi 3.00 maka diperoleh kesimpulan penerimaan item tes pada uji coba instrumen tes yang disajikan pada Tabel 5.

Tabel 9. Kesimpulan

Kategori	No.Item	Jumlah
Diterima	4, 6, 7, 8, 12, 16, 17, 18, 19	9 (45%)
Direvisi	2, 3, 9, 10, 11, 13, 15, 20	8 (40%)
Ditolak	1, 5, 14	3 (15%)
Jumlah		20

Hasil analisis uji tes Higher Order Thinking Skill pada materi ekosistem pada siswa SMA kelas X menggunakan program ITEMAN versi 3.00 yang dirangkum pada tabel 5 menunjukkan bahwa 45% item dengan 9 butir soal diterima, 40% item dengan 8 butir soal perlu direvisi, dan 15% item dengan 3 butir ditolak.

Berdasarkan hasil uji validitas isi serta analisis butir tes yaitu tingkat kesukaran, daya beda dan efektifitas distraktor pada penyusunan instrumen tes kemampuan berpikir tingkat tinggi diperoleh hasil bahwa indikator – indikator yang dikembangkan dalam bentuk item tes telah memenuhi syarat sebagai instrumen tes yang baik walaupun ada beberapa butir soal yang perlu perbaikan.

Dengan demikian, indikator kemampuan berpikir tingkat tinggi yang digunakan telah memenuhi syarat sebagai indikator yang baik untuk dikembangkan lagi dalam bentuk item tes guna menguji kemampuan

berpikir tingkat tinggi siswa. Item tes yang disusun dari indikator pada aspek ini sebagian besar memiliki pokok soal berupa pernyataan-pernyataan berisi masalah untuk memberi stimulus pada siswa agar kritis dalam menyelesaikan soal tersebut.

Soal yang telah diujicobakan dan dianalisis selanjutnya dipilih untuk digunakan pada uji coba II. Pemilihan soal didasarkan pada hasil analisis ITEMAN yang meliputi peninjauan taraf kesukaran item, daya pembeda dan efektifitas distraktor. Dari 20 soal yang telah direvisi untuk uji coba II. Pengambilan keputusan didasarkan pada kriteria sebagai berikut : (1) item soal diterima apabila karakteristik item soal memenuhi semua kriteria. Item soal terlalu skar atau terlalu mudah tetapi memiliki daya pembeda dan distribusi pengecoh yang memenuhi kriteria maka butir soal tersebut dapat diterima atau dipilih, (2) item soal direvisi, apabila salah satu atau lebih dari ketiga kriteria karakteristik item soal tidak memenuhi syarat, (3) item soal ditolak, jika item soal memiliki karakteristik yang tidak memenuhi semua kriteria (Yusliana, 2000:187).

Soal termasuk dalam kategori mudah mengindikasikan bahwa soal tersebut mampu dikerjakan oleh peserta tes.

Soal termasuk dalam kategori sedang menunjukkan bahwa soal tersebut dapat dijawab oleh kelompok atas dan kelompok bawah dengan proporsi yang tepat artinya peserta tes kelas atas yang mampu menjawab jumlahnya lebih banyak jika dibandingkan dengan peserta tes yang mampu menjawab dari kelas bawah.

Sedangkan soal yang dapat dikategorikan sulit soal yang memiliki tingkat kesukaran tinggi menunjukkan bahwa untuk menjawab soal tersebut diperlukan tingkat kemampuan yang tinggi sehingga hanya peserta tes dari kelas atas yang mampu menjawab soal dengan tepat.

Daya pembeda soal poor dan ditolak merupakan kategori daya pembeda yang jelek, soal dikatakan memiliki daya pembeda jelek artinya soal – soal tersebut tidak mampu membedakan peserta tes kelas atas dan kelas bawah. Soal yang termasuk kategori poor dan ditolak sebaiknya direvisi terlebih dahulu dengan mengkaji ulang kemungkinan penyebab rendahnya daya pembeda soal tersebut. Hal – hal yang dapat menjadi sebab rendahnya nilai daya pembeda misalnya, pilihan jawaban yang membingungkan, pernyataan soal yang kurang jelas atau mungkin peserta tes yang tidak memahami konsep sehingga menganggap soal terlalu sulit sehingga siswa memilih jawaban secara acak.

4. KESIMPULAN

Berdasarkan hasil penelitian dari penyusunan instrumen tes *Higher Order Thinking Skill* pada siswa kelas X SMA, peneliti menyimpulkan sebagai berikut:

1. Hasil analisis uji tes Higher Order Thinking Skill pada materi ekosistem pada siswa SMA kelas X menggunakan program ITEMAN versi 3.00 yang dirangkum pada tabel 5 menunjukkan bahwa 45% item dengan 9 butir soal diterima. 40% item dengan 8 butir soal perlu direvisi, dan 15% item dengan 3 butir ditolak.
2. Indikator kemampuan berpikir tingkat tinggi yang digunakan telah memenuhi syarat sebagai indikator yang baik untuk dikembangkan lagi dalam bentuk item tes guna menguji kemampuan berpikir tingkat tinggi siswa.

Berdasarkan hasil penelitian dari penyusunan instrument tes di atas maka peneliti dapat memberikan saran sebagai berikut:

1. Penyusunan instrumen tes yang memuat *Higher Order Thinking Skill* memerlukan kreativitas dan ketelitian yang tinggi sehingga agar dapat menghasilkan instrumen tes yang baik diperlukan kegiatan diskusi dan sosialisasi dengan pelaku bidang pendidikan serta pihak yang lebih ahli baik dari segi materi maupun evaluasi.
2. Pemahaman terhadap tahap-tahap penyusunan instrumen tes sangat diperlukan untuk menghasilkan instrumen tes yang berkualitas.
3. Tampilan instrumen tes sebaiknya dibuat lebih menarik agar siswa tidak bosan selama mengerjakan tes.

5. DAFTAR PUSTAKA

- Anderson, W. L & Krathwohl, R, D. (2001). *A Taxonomy for Learning Teaching and Assesing A Revision of Bloom's Taxonomy of Educational Objectives*. USA: Addison Wesley Longman.
- Arikunto, Suharsimi. (2007). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- Basuki, Ismet, & Hariyanto. (2012). *Assesmen Penelitian*. Bandung: PT Remaja Rosdakarya Offset
- Borg & Gall. (1983). *Educational Research, An Introduction (4th Ed)*. Newyork and London. Longman Inc.
- BSNP. (2007). *Peraturan Menteri Pendidikan Nasional Republik Indonesia No 42 Tahun 2007 tentang Standar Proses*. Jakarta: BSNP
- Campbell, Neil A., Reece, Jane B., & Mitchell, Lawrence G. (2004). *Biology Edisi Kelima Jilid 3*. Erlangga: Jakarta
- Gronlund, M. S & Linn, J. (1995). *Measurement And Assesment In Teaching*. Prentice-Hall, Pearson Education Upper Saddle River, New Jersey



- Hosnan. (2014). *Pendekatan Saintifik dan Kontekstual Dalam Pembelajaran Abad 21*. Ghalia Indonesia: Bogor
- Kemendikbud. (2013). *Permendikbud No. 81A tentang Implementasi Kurikulum*. Jakarta: Kementerian Pendidikan dan Kebudayaan
- King, J.F., Goodson, L., & Rohani, F. (2010). *Higher Order Thinkin Skills, Definition, Teaching Strategis, Assesment. A Publication of The Educational Services Program. www.Cala.fsu.edu*
- Lewis, A., & Smith, D. (2001). Defining Higher Order Thingking. *Theory Into Practice*, XXXII (3): 131-137
- Lissa. (2012). *Pengembangan Instrumen Penilaian Keterampilan Berpikir Tingkat Tinggi Materi Sistem Respirasi dan Ekskresi*. Universitas Negeri Semarang: Semarang
- Nofiana, M. (2013). *Pengembangan Instrumen Evaluasi Two-Tier Multiple Choice Question Untuk Mengukur Keterampilan Berpikir Tingkat Tinggi Pada Materi Kingdom Plantae*. Universitas Sebelas Maret: Surakarta
- Partnership for 21st century Skill. (2009). *21st Century Skills Map*. <http://science.nsta.org/ps/Final21stCenturyMapScience.pdf>. 1/03/2015
- Peraturan Menteri Pendidikan Nasional Republik Indonesia. (2007). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 20 Tahun 2007 Tentang Standar Penilaian Pendidikan untuk Satuan Pendidikan Dasar dan Menengah*. Jakarta: Depdiknas
- Poerwanti, E. (2001). *Evaluasi Pembelajaran, Modul Akta Mengajar*. UMM Press.
- Purwanto. (2010). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar
- Reynold, L., W. (2010). *Measurement and Assesment in Education*. Pearson Eucation, Inc., Upper Saddle River, New Jersey 07458. Pearson.
- Rochmah, T. F. (2013). *Penerapan Model Pembelajaran Inkuiri Untuk Meningkatkan Hasil Belajar Biologi Siswa Kelas X-G MAN Malang 1*. Universitas Negeri Malang: Malang
- Rohayati, I. (2013). *Penyusunan Instrumen Formatif Fisika SMP*. Universitas Sebelas Maret: Surakarta
- Sajidan, Sugiharto, B., & Prasetyani, N.M. (2013). *Penerapan Dan Pengembangan Mutu Pendidikan Melalui Lesson Studi Pada Mata Pelajaran Biologi Berbasis Pemetaan Ketuntasan UN Biologi*, Laporan Hibah Guru Besar. LPPM: UNS
- Subali, B. (2010). *Penilaian, Evaluasi dan Remediasi Pembelajaran Biologi*. UNY: Yogyakarta
- Sudjana, N. (2009). *Penilain Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya
- Sukiman. (2012). *Pengembangan Media Pembelajaran*. Yogyakarta: Pendajogja
- Suwandi, S. (2009). *Model Asesmen Dalam Pembelajaran*. Mata Padi Presindo: Surakarta
- Elvin, Y. (2010). *Pemanfaatan Telaah Kualitatif dan Kuantitatif Sebagai Prasyarat Penyelesaian Item Tes Pilihan Ganda Dalam Penyelesaian tugas Akhir (Skripsi) di LPTK*. Akademika II (2), 181 – 187

Saran :

Putri Agustina, M.Pd
(Universitas Muhammadiyah Surakarta /UMS)

Mohon diperbaiki dari kata-kata reabilitas yang benar adalah reliabilitas.

Alangkah baiknya apabila hasil penelitian kualitatif serta ditambahkan saran dari semua validator (ahli materi, ahli instrument, ahli praktisi)