

INSTRUMEN EVALUASI BERBASIS *INQUIRY LAB* UNTUK MENINGKATKAN KEMAMPUAN MENGANALISIS PADA MATERI SISTEM GERAK KELAS XI IPA

Jesicha Muryantono Putri Pradana¹, Sajidan², Sugiyarto³

¹ Program Studi Magister Pendidikan Sains FKIP Universitas Sebelas Maret
Surakarta, 57126, Indonesia
jesticapradana05@yahoo.com

² Program Studi Magister Pendidikan Sains FKIP Universitas Sebelas Maret
Surakarta, 57126, Indonesia
sajidan@fkip.uns.ac.id

³ Program Studi Magister Pendidikan Sains FKIP Universitas Sebelas Maret
Surakarta, 57126, Indonesia
sugiyarto67@staff.uns.ac.id

Abstrak

Tujuan penelitian yaitu: 1) mengetahui karakteristik instrumen evaluasi berbasis *Inquiry Lab* untuk meningkatkan kemampuan menganalisis, 2) mengkaji kelayakan instrumen evaluasi berbasis *Inquiry Lab* untuk meningkatkan kemampuan menganalisis, 3) menguji keefektifan instrumen evaluasi berbasis *Inquiry Lab* untuk meningkatkan kemampuan menganalisis pada materi sistem gerak di SMA negeri 1 Ngawi. Pengembangan instrumen evaluasi berbasis *Inquiry Lab* mengacu pada 9 langkah model *research and development* (R&D) dari Borg and Gall yaitu: 1) penelitian dan pengumpulan data, 2) perencanaan, 3) pengembangan produk, 4) uji coba produk awal, 5) revisi produk I, 6) uji coba lapangan, 7) revisi produk II, 8) uji coba lapangan operasional, 9) revisi produk akhir. Analisis hasil penelitian menggunakan deskriptif kuantitatif. Hasil penelitian meliputi: 1) karakteristik instrumen evaluasi berbasis *Inquiry Lab* untuk meningkatkan kemampuan menganalisis yaitu instrumen evaluasi merupakan bagian dari HOT yang mengacu pada ketercapaian indikator sintak *Inquiry Lab* meliputi observasi, manipulasi, generalisasi, verifikasi dan aplikasi; instrumen evaluasi sesuai Taksonomi Bloom khususnya pada kemampuan menganalisis (C4) dengan indikator membedakan, mengorganisasi dan mendekonstruksi; bentuk instrumen evaluasi adalah *multiple choice* dengan menggunakan dimensi pengetahuan meliputi faktual, konseptual dan prosedural, 2) kelayakan instrumen evaluasi berbasis *Inquiry Lab* untuk meningkatkan kemampuan menganalisis mempunyai tingkat validitas isi dan validitas konstruk yang “baik”; validitas butir soal dengan interpretasi minimal “cukup”; tingkat kesukaran soal dengan proporsi 60% sedang dan 40% sukar; daya pembeda soal dengan interpretasi minimal “cukup”; kepraktisan soal yang “baik”, 3) hasil uji lapangan instrumen evaluasi ditunjukkan dengan rerata hasil kemampuan menganalisis siswa kelas model *Inquiry Lab* (79,36), lebih tinggi dibanding dengan kelas *existing learning* (64,75) dan dapat disimpulkan bahwa instrumen evaluasi berbasis *Inquiry Lab* dinyatakan efektif serta dapat digunakan sebagai alat evaluasi di sekolah.

Kata Kunci : evaluasi, *Inquiry Lab*, *research and development*, kemampuan menganalisis

Pendahuluan

Pada era pengetahuan, modal intelektual, khususnya kecakapan berpikir tingkat tinggi (*higher order thinking*) merupakan kebutuhan sebagai tenaga kerja yang handal (Galbreath, 1999). Untuk memasuki dunia kerja pada abad 21,

keterampilan belajar abad 21 mempunyai tujuh keterampilan yaitu: 1) berpikir kritis dan pemecahan masalah; 2) kreatifitas dan inovasi; 3) kolaborasi, kerjasama tim, 4) pemahaman lintas budaya ; 5) komunikasi, informasi, literatur media; 6) computer dan ICT; dan 7) karir dan belajar kemandirian (Fadel, 2009). Sekolah dituntut untuk merubah

pendekatan pembelajaran yang berpusat pada guru menjadi pendekatan pembelajaran yang berpusat pada siswa, hal ini sesuai dengan tuntutan dunia masa depan yang harus memiliki kecakapan berpikir dan belajar. Kecakapan tersebut diantaranya adalah kecakapan memecahkan masalah, berpikir kritis, kolaborasi dan kecakapan berkomunikasi.

Pentingnya penguasaan kemampuan berpikir tingkat tinggi ada dalam poin Standar Kompetensi kelulusan Sekolah Menengah yaitu siswa dapat membangun dan menerapkan informasi atau pengetahuan secara logis, kritis, kreatif dan inovatif; menunjukkan kemampuan berpikir logis, kritis, kreatif dan inovatif dalam pengambilan keputusan serta menunjukkan kemampuan menganalisis dan memecahkan masalah kompleks (Permendiknas No 23 Tahun 2006). Pembelajaran biologi merupakan pembelajaran yang melakukan kegiatan penyelidikan dan eksperimen sebagai salah satu bagian dari kerja ilmiah. Kerja ilmiah menekankan siswa untuk dapat berpikir kreatif, kritis, analitis dan divergen (BSNP, 2006).

Salah satu upaya yang dilakukan pemerintah adalah mengadakan perubahan terhadap kurikulum yaitu dengan memperbaiki kurikulum 1994 dengan mengembangkan kurikulum 2004 dan Kurikulum tingkat satuan Pendidikan (KTSP). KTSP merupakan kurikulum operasional yang disusun dan dilaksanakan oleh masing-masing satuan pendidikan/sekolah (Muslich, 2007). Pemberlakuan KTSP, menuntut siswa untuk memiliki kompetensi khusus dalam semua mata pelajaran setelah proses pembelajaran. Kompetensi merupakan kemampuan berpikir, bertindak dan bersikap secara konsisten sebagai perwujudan dari pengetahuan, ketrampilan dan nilai (BSNP, 2006).

Idealnya tes formatif yang dilaksanakan oleh guru yaitu 80% harus mencakup kemampuan berpikir tingkat tinggi (Standar Penilaian BAN, 2012). Tes formatif yang digunakan oleh guru adalah tes tulis dalam bentuk pilihan ganda. Soal pilihan ganda digunakan karena penilaian yang mudah, tetapi siswa dalam menjawab secara

untung-untungan atau menebak. Soal pilihan ganda kurang mengukur kemampuan kognitif (Purwanto,2010). Hasil analisis di sekolah mendapat kesimpulan bahwa guru membutuhkan instrumen evaluasi yang mampu mengukur kemampuan berpikir tingkat tinggi terutama pada level C4 yaitu kemampuan menganalisis.

Fakta secara rasional berdasarkan survei TIMMS (*Trends in International Mathematics and Science Study*) yang diadakan oleh IEA (*International Association for the Evaluation of Educational Achievement*). Hasil TIMSS 2011 pada bidang sains menunjukkan indonesia memperoleh nilai 397 dimana nilai ini berada di bawah nilai rata-rata internasional yaitu 500 (Litbang Kemendikbud, 2013). Berdasarkan data prosentase rata-rata jawaban benar untuk konten sains dan domain kognitif khususnya biologi, prosentase jawaban benar pada soal mengingat selalu tinggi dibandingkan dengan prosentase jawaban benar pada soal memahami, mengaplikasi dan menganalisis.

Kurangnya kualitas pembelajaran berdampak pada pencapaian hasil belajar yang rendah. Salah satu pengukuran hasil belajar adalah Ujian Nasional (UN). Hasil UAN Biologi SMA Negeri 1 Ngawi tahun 2013/2014 menunjukkan pada materi Struktur dan Fungsi Organ terutama sistem gerak rata-rata nilai yang diperoleh yaitu 51,36 untuk tingkat nasional; 64,04 untuk tingkat propinsi; 57,91 untuk tingkat kota; dan 64,45 untuk tingkat sekolah. Hasil ulangan harian siswa menunjukkan nilai rata-rata 64,50, dimana nilai ini berada di bawah KKM yaitu 78,00. (Permendiknas, 2007).

Fakta di lapangan menunjukkan rendahnya penguasaan sains berhubungan erat dengan kualitas pembelajaran. Berdasarkan analisis delapan Standar Nasional Pendidikan (SNP) yang meliputi standar isi, standar proses, standar kompetensi kelulusan, standar pendidik dan tenaga kependidikan serta standar penilaian (Badan Akreditasi Nasional, 2012) menunjukkan bahwa GAP pada Standar proses sebesar 1,85% dan Standar penilaian sebesar 1,39%. Adanya GAP ini dapat disimpulkan bahwa kurang maksimalnya

penyusunan dan pelaksanaan pada tiap-tiap indikator.

Analisis soal Ulangan Harian menunjukkan bahwa soal-soal yang digunakan kurang melatih berpikir tingkat tinggi (HOT). Soal kategori C4 hanya 4% dan soal yang digunakan cenderung pada kategori C1 yaitu 38%, C2 yaitu 40% dan C3 yaitu 8%. Data ini menunjukkan kurang maksimalnya pelatihan kemampuan berpikir tinggi bagi siswa. Siswa lebih banyak melatih kemampuan mengingat dan memahami, mengaplikasi. Kemampuan menganalisis kurang dilatihkan bahkan kemampuan berpikir C5 (evaluasi) dan C6 (mencipta) tidak muncul dalam indikator soal (tidak dilatihkan). Setelah dilakukan tes untuk mengetahui kemampuan menganalisis siswa, diperoleh data bahwa dari 27 siswa yang mempunyai kemampuan menganalisis (C4) dengan baik hanya 33% dan 67% siswa mempunyai kemampuan menganalisis kurang baik. Siswa yang memiliki kemampuan berpikir tingkat tinggi tidak hanya dikembangkan dalam pembelajaran tetapi juga didukung dengan instrumen evaluasi. Menurut Arikunto, 2007 menyatakan bahwa evaluasi dapat digunakan dalam mengukur keberhasilan pencapaian indikator pembelajaran yang dilakukan. Indikator pembelajaran yang berupa kognitif, psikomotorik dan afektif. Instrumen evaluasi yang digunakan dalam penelitian didasarkan pada indikator kognitif.

Pembelajaran inquiry merupakan rangkaian kegiatan pembelajaran yang menekankan pada proses berpikir kritis dan analitis untuk mencari dan menemukan sendiri jawaban dari suatu masalah yang dipertanyakan. Wenning (2010) membagi model pembelajaran inquiry menjadi enam tahapan yaitu: *discovery learning*; *interactive demonstration*; *inquiry lesson*; *inquiry laboratory*; *real world application*; dan *hypothetical inquiry*. Hasil survey di SMA Negeri 1 Ngawi bahwa perolehan data saat tes kemampuan inkuiri yaitu 90,62% siswa menguasai inquiry level 1 (*discovery learning*) dengan nilai rata-rata nilai 80,40; siswa menguasai inquiry level 2 (*interactive demonstration*) dengan nilai rata-rata 84,03;

siswa menguasai inquiry level 3 (*inquiry lesson*) dengan rata-rata nilai 86,29, dan siswa kurang menguasai inquiry level 4 (*Inquiry Lab*) dengan nilai rata-rata 67,62. Kemampuan inkuiri siswa pada level 1, level 2, level 3 termasuk kategori baik, sedangkan pada level 4 kemampuan inquiry siswa dalam kategori kurang baik (Robert, 2000).

Kegiatan laboratorium merupakan komponen yang penting dalam pembelajaran IPA. Menurut Lazarowitz & Tamir (1994), kurikulum berbasis inquiry mengalokasikan waktunya sekitar 50% untuk kegiatan laboratorium. Pembelajaran berbasis inquiry laboratorium menekankan pada aktivitas dalam membantu siswa belajar dan memahami proses dan keterampilan berpikir layaknya ilmuwan dan memahami karakteristik penelitian ilmiah (Wenning, 2010 & Khan, 2011).

Materi sistem gerak merupakan materi biologi yang abstrak dan memiliki bahasan luas dan hafalan istilah yang cukup banyak, maka dibutuhkan pembelajaran yang bermakna. Pembelajaran yang bermakna adalah pembelajaran yang banyak melibatkan aktivitas siswa di dalam kelas termasuk dalam kemampuan berpikir tingkat tinggi yaitu kemampuan menganalisis (Burris & Garton 2006). Dengan demikian materi dapat dipahami langsung oleh siswa bukan lagi hafalan yang didapatkan dari penjelasan guru.

Berdasarkan latar belakang, maka diperlukan pengembangan instrumen evaluasi dalam melatih kemampuan menganalisis. Instrumen evaluasi hasil dari pengembangan diharapkan dapat memperbaiki instrumen evaluasi yang sebelumnya digunakan.

Metode Penelitian

Model penelitian dan pengembangan (*Research & Development*) yang diterapkan mengacu pada tahapan menurut Borg & Gall (1983). Pendekatan penelitian dan pengembangan merupakan penelitian yang berorientasi untuk mengembangkan dan memvalidasi produk-produk yang digunakan dalam penelitian.

Langkah-langkah dalam *Research & Development* (R&D) menurut Borg & Gall

(1983) sebagai berikut: 1) *Research and information collecting* (penelitian dan pengumpulan data); 2) *Planning* (perencanaan); 3) *Develop preliminary form product* (pengembangan produk awal); 4) *Preliminary field test* (uji coba lapangan awal); 5) *Main product revision* (revisi produk I); 6) *Main field testing* (uji coba lapangan utama); 7) *Operational product revision* (revisi produk II); 8) *Operational field testing* (uji lapangan operasional); 9) *Final product revision* (revisi akhir); 10) *Dissemination and implementation* (penyebaran dan implementasi). Dalam penelitian ini dapat menggunakan sampel yang tidak terlalu besar dan diperbolehkan mengembangkan sampai tahapan tertentu sesuai dengan kebutuhan peneliti (Emzir, 2012). Pada penelitian ini hanya dilakukan sembilan tahapan dari sepuluh tahapan sebagai keterbatasan penelitian.

Penelitian pengembangan instrumen evaluasi ini dilakukan di Sekolah Menengah Atas Negeri 1 Ngawi, Sampel yang digunakan adalah siswa kelas XI semester ganjil Tahun Pelajaran 2015/2016. Instrumen pengambilan data yang digunakan meliputi angket untuk analisis kebutuhan, lembar *ceklist* 8 SNP, bank soal, dokumentasi, silabus dan RPP, lembar *ceklist* penilaian produk dan kepraktisan. Data analisis kebutuhan dianalisis dengan statistik deskriptif. Hasil angket dideskripsikan untuk menganalisis kebutuhan pengembangan. Hasil analisis digunakan untuk mempertimbangkan kebutuhan pengembangan. Data penilaian ahli dan guru senior terhadap soal dianalisis dengan teknik deskriptif presentase (Purwanto, 2010). Analisis data dilakukan dengan cara menghitung skor yang dicapai dari keseluruhan aspek yang dinilai kemudian menghitungnya dengan rumus sebagai berikut:

$$N = \frac{K}{Nk} \times 100 \%$$

Keterangan:

N : presentase kelayakan aspek

K : skor hasil pengumpulan data

Nk : skor maksimal (skor kriteria tertinggi x jumlah aspek x jumlah validator)

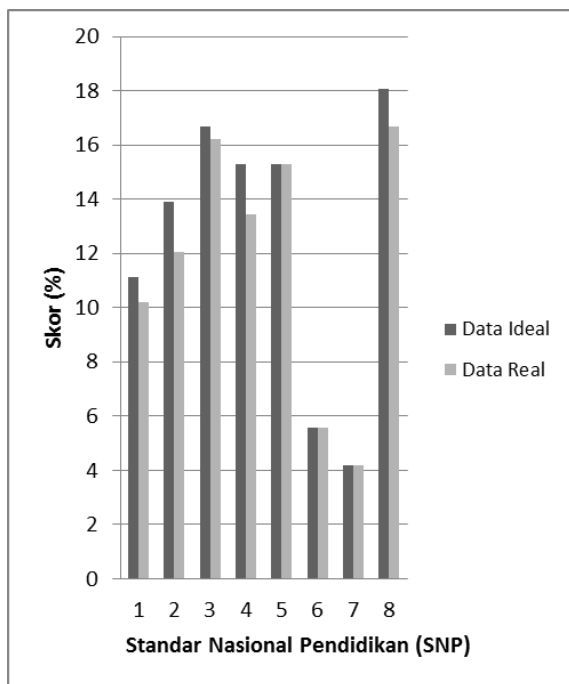
Tabel 1 Kriteria Interpretasi Skor Validasi Ahli

Interval Kriteria	Kriteria	Konversi
$86\% \leq N < 100\%$	Sangat Baik	A
$72\% \leq N < 85\%$	Baik	B
$58\% \leq N < 71\%$	Cukup	C
$44\% \leq N < 57\%$	Kurang	D
$N < 44\%$	Sangat Kurang	E

Validasi dilakukan oleh ahli evaluasi, validasi materi, praktisi pendidikan dan guru bidang studi. Pengumpulan data menggunakan data kualitatif dan data kuantitatif. Data kualitatif diperoleh dari guru bidang studi. Data kuantitatif diperoleh dari pengujian soal meliputi validitas, reliabilitas, daya pembeda dan kepraktisan soal. Teknik yang digunakan dalam penelitian ini meliputi analisis soal-soal tes terkait sistem gerak yang ada pada ulangan harian dan bank soal yang digunakan di sekolah. Penelitian deskriptif kualitatif dan kuantitatif dengan *post test only control group desain* (Sugiyono, 2011). Sampel yang digunakan dalam riset ini adalah 27 siswa pada uji lapangan dan 54 siswa pada uji operasional. Data hasil uji coba dianalisis menggunakan program *Quest* (Rost, 1988).

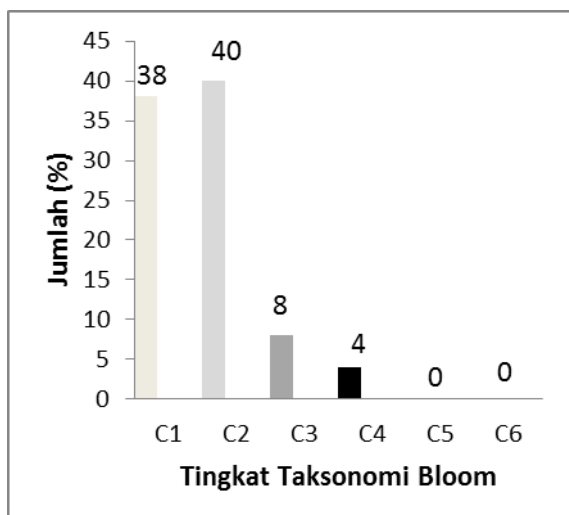
Hasil Penelitian dan Pembahasan

Data yang diperoleh dalam penelitian dan pengembangan antara lain data analisis kebutuhan, data validasi ahli dan praktisi, data hasil uji coba lapangan dan data uji operasional. Data analisis kebutuhan meliputi tingkat pemenuhan standar nasional pendidikan (SNP) di SMA Negeri 1 Ngawi, analisis bank soal yang digunakan guru. Analisis SNP mengenai 8 Standar Pendidikan Nasional digunakan untuk mengetahui standar yang dicapai oleh SMA Negeri 1 Ngawi. Hasil SNP disajikan pada Gambar 1



Gambar 1. Skor Implementasi 8 SNP di SMA Negeri 1 Ngawi

Hasil 8 SNP menunjukkan adanya GAP pada standar proses adalah 1,85% dan standar penilaian adalah 1,39%. Analisis bank soal yang digunakan guru dilakukan untuk mengetahui presentase penggunaan tingkatan Taksonomi Bloom. Hasil temuan bank soal guru disajikan pada gambar 2



Gambar 2 Presentase Soal-Soal Sesuai Taksonomi Bloom di SMA Negeri 1 Ngawi

Soal yang digunakan di sekolah sebagian masih belum meningkatkan

kemampuan menganalisis(C4). Padahal idelanya 80% soal yang digunakan mencakup C4-C6 (BAN, 2006). Kemampuan menganalisis harus dapat diukur dengan instrumen yang jelas, valid dan terkoordinasi sehingga hasilnya dapat dipercaya. Pengembangan instrumen evaluasi berbasis *Inquiry Lab* belum banyak dilakukan oleh guru. Penilaian formatif yang dipadu dengan basis model dapat mendorong siswa untuk meningkatkan kemampuan menganalisis (Ozlem, 2009).

Pengembangan Instrumen evaluasi berbasis *Inquiry Lab* didasarkan pada teori kognitif dari Piaget. Implikasi teori Piaget adalah instrumen disesuaikan dengan tingkat perkembangan kognitif siswa yaitu siswa SMA agar tidak sulit dipahami. Teori belajar yang mendukung instrumen evaluasi yang lain adalah teori berpikir “John Dewey: Implikasi dari teori John Dewey adalah soal yang diberikan dapat merangsang siswa dalam meningkatkan kemampuan berpikir terutama kemampuan menganalisis, hal ini sejalan dengan teori konstruktivisme. Implikasi teori konstruktivisme adalah didalam soal mengandung masalah yang harus dipecahkan oleh siswa, sehingga siswa tidak hanya sekedar menghafal. Siswa harus memiliki ketrampilan dalam mengkaitkan pengetahuan lama dan pengetahuan baru. Indikator dalam pengembangan instrumen evaluasi berbasis *Inquiry Lab* diadapatasi oleh Anderson & Krathwohl (2001).

Hasil penilaian indikator materi dengan kategori sangat baik meliputi konsep materi soal benar; cakupan materi sesuai tingkatan siswa; istilah yang digunakan jelas; materi soal mudah dipahami; dan materi soal ditulis sistematis, runtut dan alur logika jelas. Penilaian oleh ahli materi bertujuan untuk mengkaji validitas isi dari instrumen evaluasi pengembangan. Rata-rata hasil penilaian indikator materi termasuk dalam kategori “sangat baik”. Perbaikan dilakukan sesuai saran meliputi gambar yang diperjelas dan kebenaran konsep untuk dipertegas.

Hasil penilaian indikator konstruk dengan kategori sangat baik meliputi letak pilihan jawaban benar ditentukan; wacana,

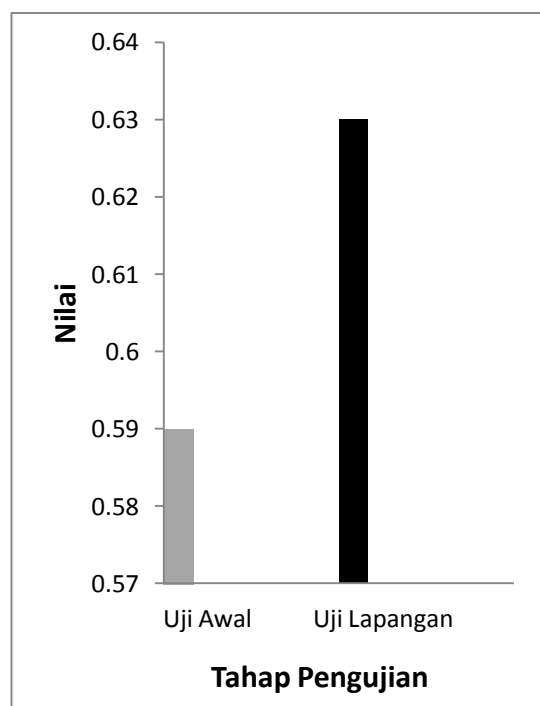
gambar atau grafik berfungsi; rumusan kalimat komunikatif; kalimat menggunakan bahasa yang baik dan benar; rumusan kalimat tidak mengandung penafsiran ganda; menggunakan bahasa yang umum; rubrik penilaian benar; dan penskoran objektif. Hasil penilaian indikator konstruk dengan kategori baik meliputi butir soal sesuai indikator; butir soal sesuai dengan materi; isi materi sesuai; soal hanya mengandung satu jawaban benar; pokok soal dirumuskan dengan jelas; pokok soal merupakan kalimat yang diperlukan saja; pilihan jawaban homogen; panjang alternatif pilihan jawaban sama; pokok soal tidak menunjuk ke arah jawaban; tidak ada kalimat “semua jawaban benar” atau “semua jawaban salah”; pengecoh berfungsi; pokok soal tidak mengandung pernyataan negatif ganda; antara butir soal tidak tergantung satu sama lain; dan rumusan soal tidak menyinggung perasaan. Rata-rata hasil penilaian indikator konstruk termasuk dalam kategori “baik”. Perbaikan telah dilakukan sesuai saran antara lain gambar yang diperjelas, tata cara penulisan yang benar dan kesesuaian taksonomi bloom.

Hasil penilaian indikator kelayakan oleh praktisi pendidikan dengan kategori sangat baik meliputi perintah mengerjakan soal jelas; istilah yang digunakan jelas; tidak ada kesalahan tata tulis, ejaan dan tanda baca; kunci jawaban benar; dan penskoran objektif. Hasil penilaian indikator kelayakan dengan kategori baik meliputi soal sesuai dengan KD; soal sesuai dengan indikator; soal dapat mengukur kemampuan menganalisis; maksud pertanyaan jelas; dan waktu siswa cukup untuk mengerjakan soal. Rata-rata hasil penilaian indikator kelayakan termasuk dalam kategori “baik”. Perbaikan telah dilakukan sesuai saran dari guru pengguna meliputi penulisan jawaban untuk tidak mengulang soal dan tata cara penulisan benar.

Menurut Mardapi (2008) menyebutkan waktu yang dibutuhkan untuk mengerjakan tes bentuk pilihan ganda adalah 2-3 menit untuk setiap butir tes pilihan ganda sekitar 20-30 soal, yang berarti setiap butir soal dikerjakan selama 3-4 menit. Alokasi waktu mengerjakan soal tergantung banyaknya soal dan bentuk soal (Arikunto, 2007). Berdasarkan kajian

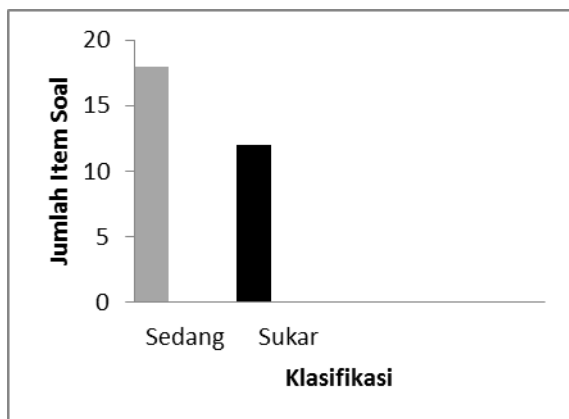
yang dilakukan maka waktu yang diberikan dalam mengerjakan soal berbasis *Inquiry Lab* adalah 60 menit untuk 30 soal yang diberikan, sehingga masing-masing butir soal dikerjakan selama 2 menit.

Uji coba instrumen evaluasi dilakukan dua kali yaitu uji coba awal dan uji coba lapangan. Hasil uji coba awal mendapatkan reliabilitas 0,59 dan uji coba lapangan mendapatkan reliabilitas 0,63. Hasil reliabilitas disajikan pada Gambar 3.



Gambar 3 Hasil Uji Coba Instrumen Evaluasi

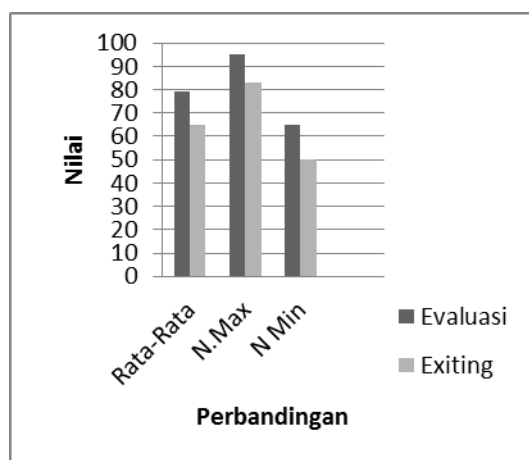
Nilai reliabilitas soal adalah 0,63 dengan kategori “baik”. Uji coba lapangan digunakan untuk mendapatkan revisi dan dilakukan perbaikan selanjutnya. Uji coba lapangan bertujuan untuk mendapatkan produk instrumen penilaian yang baik meliputi, valid, reliabel serta praktis (Arikunto, 2007)



Gambar 4 Hasil Analisis Uji Item Soal Evaluasi

Hasil uji coba lapangan mendapatkan kesimpulan bahwa penilaian kemampuan berpikir tingkat tinggi siswa terutama kemampuan menganalisis memiliki validitas dan reliabilitas “baik”, memiliki taraf kesukaran soal dari 30 soal dengan proporsi 60% sedang dan 40% sukar untuk soal pilihan ganda, memiliki daya pembeda soal dengan interpretasi minimal “cukup” dan memiliki kepraktisan soal evaluasi yang “baik”. Hasil kepraktisan soal dengan kategori sangat baik meliputi waktu penyusunan tidak lebih dari 1 bulan; dan pelaksanaan tes mudah. Hasil kepraktisan soal dengan kategori baik meliputi biaya penyusunan tes; penyusunan tes dapat dilakukan oleh guru biologi; penilaian tes mudah; mengolah hasil tes mudah; dan waktu pelaksanaan di sekolah cukup. Rata-rata hasil penilaian kepraktisan soal yang dilakukan oleh guru bidang studi termasuk dalam kriteria baik.

Kepraktisan instrumen evaluasi yang baik adalah instrumen evaluasi dapat digunakan kembali oleh guru untuk mengukur tujuan pembelajaran pada suatu saat nanti (Purwanto, 2010). Produk pengembangan instrumen evaluasi berbasis *Inquiry Lab* yang valid dan reliabel selanjutnya dilakukan uji coba operasional atau keefektifan untuk mengetahui keterlaksanaan kemampuan menganalisis. Pengujian dilakukan pada dua kelas yang setara. Kedua kelas diberikan perlakuan yang sama, namun pada akhir pembelajaran satu kelas diuji dengan instrumen evaluasi berbasis *Inquiry Lab*. Hasil perbandingan nilai kemampuan menganalisis dari kedua kelas disajikan pada Gambar 5.



Gambar 5 Nilai Kemampuan Menganalisis

Perolehan nilai pada kelas model *Inquiry Lab* memiliki rata-rata 79,36 sedangkan kelas *existing learning* memiliki rata-rata 64,75. Kelas evaluasi memperoleh nilai di atas KKM. Perbedaan nilai antar dua kelas dipengaruhi oleh beberapa faktor yaitu faktor internal dan faktor eksternal. Faktor internal antara lain kecerdasan, bakat, kematangan dan kesipaan siswa dalam menerima instrumen evaluasi. Faktor eksternal antara lain lingkungan sekolah atau lingkungan di dalam kelas (Slameto, 2003).

Instrumen evaluasi bertujuan untuk mendapatkan alternatif instrumen formatif. Penilaian formatif adalah kegiatan penilaian yang bertujuan dengan mencari umpan balik dan hasilnya dapat digunakan untuk memperbaiki proses pembelajaran. Manfaat instrumen evaluasi formatif adalah untuk mengetahui sejauh mana bahan yang diajarkan dapat diterima oleh siswa, penguatan bagi siswa untuk menghasilkan skor yang diharapkan dan memberikan motivasi untuk meningkatkan penguasaan materi (Purwanto, 2010). Perbandingan instrumen evaluasi berbasis *Inquiry Lab* dengan hasil peneliti lain adalah untuk mengetahui kelayakan atau kriteria instrumen yang baik. Instrumen Evaluasi berbasis *Inquiry Lab* yang dibandingkan dengan hasil instrumen peneliti lain. Perbandingan pada masing-masing peneliti mengkaji kualitas instrumen meliputi tingkat kesukaran, daya pembeda dan reliabilitas.

Hasil perbandingan kualitas instrumen pada tingkat kesukaran kategori mudah menurut (Mufida, 2013) adalah 15%; (Walid, 2015) adalah 1,6%; (Kusumawardany, 2015) adalah 1,93%; dan (Jesicha, 2015) tidak ditemukan soal dengan tingkat kesukaran kategori mudah. Tingkat kesukaran kategori sedang menurut (Mufida, 2013) adalah 80%; (Walid, 2015) adalah 58%; (Kusumawardany, 2015) adalah 37%; dan (Jesicha, 2015) adalah 60%. Tingkat kesukaran kategori sukar menurut (Mufida, 2013) adalah 5%; (Walid, 2015) adalah 40%; (Kusumawardany, 2015) adalah 0,56%; dan (Jesicha, 2015) adalah 40%.

Kualitas instrumen evaluasi pada daya pembeda kategori kurang tidak ditemukan pada masing-masing hasil peneliti. Daya pembeda kategori cukup menurut (Mufida, 2013) adalah 50%; (Walid, 2015) adalah 37,6%; (Kusumawardany, 2015) adalah 33,87%; dan (Jesicha, 2015) adalah 34%. Daya pembeda kategori baik menurut (Mufida, 2013) adalah 50%; (Walid, 2015) adalah 54%; (Kusumawardany, 2015) adalah 4,63%; dan (Jesicha, 2015) adalah 66%. Nilai reliabilitas menurut menurut (Mufida, 2013) adalah 0,84; (Walid, 2015) adalah 0,70; (Kusumawardany, 2015) adalah 0,65; dan (Jesicha, 2015) adalah 0,63.

Perbedaan hasil reliabilitas, disebabkan karena masing-masing peneliti menggunakan bentuk soal dan jumlah soal yang berbeda. Reliabilitas pada masing-masing peneliti termasuk dalam kategori “baik”. Menurut Arikunto (2007) untuk kriteria instrumen yang baik adalah memiliki validitas, reliabilitas, dan daya pembeda minimal “Cukup”.

Kesimpulan dan Rekomendasi

Kesimpulan dari penelitian dan pengembangan instrumen evaluasi antara lain :

1. Karakteristik instrumen evaluasi berbasis *Inquiry Lab* untuk meningkatkan kemampuan menganalisis yaitu instrumen evaluasi merupakan bagian dari HOT yang mengacu pada ketercapaian indikator sintak *Inquiry Lab* meliputi observasi, manipulasi, generalisasi, verifikasi dan aplikasi; instrumen evaluasi sesuai Taksonomi

Bloom khususnya pada kemampuan menganalisis (C4) dari (Anderson & Krathwohl, 2001) dengan indikator membedakan, mengorganisasi dan mendekonstruksi; bentuk instrumen evaluasi adalah *multiple choice* dengan menggunakan dimensi pengetahuan meliputi faktual 30%, konseptual 46% dan prosedural 24%.

2. Kelayakan dari hasil pengembangan instrumen evaluasi berbasis *Inquiry Lab* untuk meningkatkan kemampuan menganalisis telah melalui validitas isi dan konstruk yang dilakukan oleh validasi ahli evaluasi, ahli materi. Validitas isi oleh ahli materi dinilai sangat baik, validitas konstruk oleh ahli evaluasi dinilai baik. Validitas pada butir soal dengan interpretasi minimal cukup. Tingkat kesukaran soal dengan proporsi 60% sedang dan 40% sukar. Daya pembeda pada soal dengan interpretasi minimal cukup dan tingkat kepraktisan soal dengan interpretasi baik.
3. Instrumen evaluasi berbasis *Inquiry Lab* efektif untuk meningkatkan kemampuan menganalisis. Efektifitas instrumen evaluasi ditunjukkan adanya perbedaan nilai kemampuan menganalisis antara kelas model *Inquiry Lab* (79,36) dan kelas *existing learning* 64,75.

Rekomendasi untuk penelitian dan pengembangan instrumen evaluasi antara lain :

1. Guru dalam memberikan materi pada KD lain sebaiknya menggunakan model yang dapat meningkatkan kemampuan menganalisis.
2. Alokasi waktu dalam mengerjakan soal evaluasi berbasis *Inquiry lab* tidak kurang dari 60 menit.
3. Sebelum mengerjakan soal evaluasi berbasis *Inquiry Lab* diharapkan siswa sebelumnya telah mendapatkan materi pada KD 3.1 Menjelaskan keterkaitan antara struktur, fungsi dan proses serta kelainan penyakit yang terjadi pada sistem gerak manusia.
4. Instrumen evaluasi yang terkait dengan model pembelajaran yang dapat meningkatkan kemampuan menganalisis

- perlu dikembangkan lebih lanjut menggunakan soal bertingkat.
5. Perlu adanya pengkajian tentang dampak dari hasil pengembangan instrumen evaluasi berbasis *Inquiry Lab*.

Daftar Pustaka

- Arikunto, S. 2007. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta : PT. Bumi Aksara
- Anderson, L. W dan D.R Krathwohl. 2001. *A Taxonomy for Learning Teaching and Assesing*. New York: Longman
- Badan Akreditasi Nasional. 2013.<http://ban.pt.kemdiknas.go.id> (diakses tgl 7 Maret 2014).
- Borg, W. R and Gall, M. D. 1983. *Education Research and introductoin* . New York and Longman
- BSNP. 2006. *Panduan Penyusunan Kurikulum Tingkat Satuan Pendidikan. Jenjang Pendidikan Dasar dan Menengah*. Jakarta: Badan Standar Nasional Pendidikan.
- Burris, S & Garton, B.L. 2006. An Investigation of The Critical Thingking Ability of Secondary Agriculture Students. *Journal of Southern Agricultural Education Research*. 56(1): 18-19
- Djemari Mardapi. 2008. *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendekia yogyakarta Press.
- Emzir. 2012. *Metodologi Penelitian Pendidikan: Kualitatif dan Kuantitatif*. Jakarta : Raja Grafindo Persada.
- Fadel, T. 2009. *21 ST Century Skills*. United States: Jossey-Bass A Wiley Imprint.
- Galbreath, J. 1999. *Preparing the 21th century worker : The link between computer-based technology and future skill sets*. Educational Technology Nopember-Desember 1999.
- Lazarowitz, R. & P. Tamir. 1994. *Research on Using Laboratory Instruction in Science*. Handbook of Research on Science Teaching and Learning. Edited by: D. L. Gabel. New York: Macmillan Publishing Company
- Kemendikbud Badan Penelitian dan Pengembangan. 2013. *Survei Internasional TIMSS*. Acceced on Desember 2014; <http://litbang.kemdikbud.go.id;internet>
- Khan, M.S. 2011. Effect Of Inquiry Method On Achievemment Of Student In Chemistry At Secondary Level. *Gilgit : International Journal of Academic Research*, 3(1): 13-20
- Muslich, M. 2007. *KTSP Dasar Pemahaman dan Pengembangan*. Jakarta: Bumi Aksara
- 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
- Ozlem Isik. 2009. The Importance of Group Work in Mathematics. *Procedia Social and behavioral Sciences*. Vol.1, 2363-2365
- Permendiknas No 20 Tahun 2007. *Standar Penilaian Pendidikan*. Jakarta: Depdiknas.
- Purwanto, N. 2010. *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: PT Remaja Rosdakarya.
- Rost, (1988). Rating Scale Analysis With Latents Class Models. *Psychometrika*, 53(3), 327-348.
- Robert, R. 2000. Inquiry Based Experiments in The Introductory Physics Laboratory. *Journal of Education Conference*. 2(4): 18-31
- Slameto. 2003. *Belajar dan Faktor-faktor yang Mempengaruhinya*. Jakarta: PT. Rineka Cipta.

- Walid, A. 2015. Pengembangan Penilaian Untuk Mengukur Kemampuan Berpikir Tingkat Tinggi Siswa di Lima SMA Negeri di Surakarta Materi Sistem Reproduksi. Universitas Sebelas Maret: Surakarta.
- Wardany, K. 2015. Pengembangan Penilaian Untuk Mengukur *Higher Order Thinking Skill* Siswa Pada Materi Ekosistem dan Lingkungan. Universitas Sebelas Maret: Surakarta.
- Wenning, C. J .2010. Levels of Inquiry: Using Inquiry Spectrum Learning Sequence to Teach Science. *Journal of Physics teacher Education*. 5(1): 5-19.