

**PENGEMBANGAN INSTRUMEN PENILAIAN AKHIR SEMESTER (PAS)
MATA PELAJARAN BIOLOGI PADA KELAS X DI MAN SIDOARJO**

**THE DEVELOPING INSTRUMENT OF ASSESSMENT FINAL EXAM (PAS)
ON BIOLOGY'S COURSE GRADE Xth IN MAN SIDOARJO**

Zahro' Suwitaningsih

Pendidikan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
Jalan Ketintang, Gedung C3, Lt. 2 Surabaya 60231
e-mail: zahrosuwitaningsih@gmail.com

Sifak Indana

Pendidikan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
Jalan Ketintang, Gedung C3, Lt. 2 Surabaya 60231
e-mail: sifakindana@unesa.ac.id

Abstrak

Asesmen diberikan oleh guru dalam bentuk soal dan berfungsi sebagai alat ukur untuk mengetahui kemampuan yang dimiliki peserta didik. Penelitian ini bertujuan untuk mendeskripsikan instrumen pengembangan soal Penilaian Akhir Semester (PAS) Gasal. Instrumen diujicobakan pada 30 peserta didik kelas X MAN Sidoarjo. Penelitian ini dilakukan pada bulan Maret-Mei 2018. Teknik analisis data secara deskriptif. Metode pengumpulan data yang digunakan adalah metode wawancara, validasi, dan tes. Analisis kualitas soal dianalisis berdasarkan validitas teoritis dan empiris. Validitas teoritis ditinjau dari aspek materi, konstruksi, dan bahasa menunjukkan 99,5% dinyatakan sangat layak. Validitas empiris menunjukkan bahwa (1) reliabilitas memiliki nilai sebesar 0,81 kategori sangat tinggi; (2) tingkat kesukaran soal sebesar 33% mudah, 53% sedang, dan 15% sukar; (3) indeks daya beda yaitu 20% cukup, 57% baik, dan 23% baik sekali; dan (4) distraktor (pengecoh) 7% berfungsi cukup, 17% berfungsi baik, dan 76 % berfungsi sangat baik. Secara umum, sebanyak 23 butir soal diterima dan 7 butir direvisi. Simpulan dari penelitian ini ialah instrumen soal Penilaian Akhir Semester (PAS) Gasal valid dan reliabel mengukur kemampuan yang dimiliki peserta didik.

Kata Kunci: Asesmen, Penilaian Akhir Semester gasal Biologi, validitas.

Abstract

Assesment those given by teacher in item test and as a means of measure to know ability owned by students. This research bent on be for describe the developing instrument of assessment problem final (PAS) of Odd Semester. This instrument tested in 30 participant of learner's class X MAN Sidoarjo. The research conducted during March-May 2018. Data analyzed descriptive. Data collecting method that used by was interview, validation, and test. The developing instrument has been revised analyzed theoretically and empiric. Theoretical validation were evaluated from aspect of content, construction, and language shows 99.5% expressed very feasible. Empiric validation indicates that (1) reliability haves value as high as 0.81 very high categories; (2) level of test difficulty as high as 33% easy, 53% , and 15% difficult; (3) index of discrimination that is 7% poor, 17% enough, 20% enough, 53% good, and 23% very good; and (4) distractors 6% poor, 17% deficient, 30% enough, 33% good, and 17% jolly good. In general, 23 item were accepted and 7 item were revised. Conclusion from this research was that test instrument assessment problem final (PAS) of Odd Semester validity and reliability measures ability owned by students.

Keywords: Assessment, problem final odd semester of Biology lesson, validity.

PENDAHULUAN

Guru dituntut untuk memiliki empat kompetensi, meliputi kompetensi kepribadian, sosial, profesional, dan pedagogik. Adapun dalam memenuhi tuntutan kompetensi

tersebut, khususnya kompetensi pedagogik, guru harus memiliki pengetahuan luas, menggunakan teknologi, membuat Rencana Pelaksanaan Pembelajaran (RPP), dan mampu melakukan evaluasi.

Evaluasi ialah komponen dari sistem pendidikan yang berkontribusi pada pengelolaan pembelajaran dari awal sampai akhir dalam kelas (Naqvi *et al.*, 2010). Kegiatan evaluasi harus mempunyai teknik atau alat ukur supaya pelaksanaannya terarah (Rahmani, 2015). Cara guru dalam melakukan evaluasi dengan memberi asesmen pada peserta didik. Asesmen tersebut diberikan dalam bentuk soal-soal yang berfungsi sebagai alat ukur dan bertujuan untuk mengukur kemampuan peserta didik, sejauh mana mereka dapat memahami apa yang telah dipelajari (Hasanah *et al.*, 2016). Asesmen memiliki prinsip-prinsip, antara lain ekonomis, akurat, dan mendorong kualitas sistem pembelajaran (Mardapi, 2017). Asesmen dibedakan menjadi dua, yakni tes dan nontes. Berdasarkan kemampuan yang diujikan kepada peserta didik macam-macam tes, meliputi tes bakat dan minat, tes kepribadian, tes kecerdasan, tes terstandarisasi, dan tes buatan guru (Naqvi *et al.*, 2010).

Tes terstandarisasi merupakan tes yang telah dianalisis berdasarkan validitas dan reliabilitasnya, sehingga tes bersifat mengukur secara valid (tepat) dan reliabel (konsisten). Tes standardisasi dibuat oleh suatu badan eksternal dan akan diuji cobakan selama pengembangan pada sampel besar sesuai kebutuhan (Iliya, 2014). Tes standardisasi digunakan pada *Program for International Students Assessment (PISA)*, *Trends in Mathematics and Science Study (TIMSS)*, dan Ujian Nasional (UN). Sementara, tes buatan guru ialah tes yang digunakan untuk mengukur hasil pembelajaran dan pelaksanaannya terbatas pada suatu sekolah. Tes ini dibuat oleh seorang guru yang mengampu mata pelajaran tertentu.

Praktik di lapangan, kualitas soal-soal buatan guru di Indonesia masih rendah (Rofiah *et al.*, 2013). Guru-guru tidak semua mengikuti pelatihan pembuatan soal yang baik sesuai dengan tata cara penulisan yang baku (Sulistiawan, 2016). Hasil wawancara pada Maret 2018 yang dilakukan terhadap guru Biologi di MAN Sidoarjo, memaparkan bahwa soal ujian buatan guru berkisar pada kategori C1, C2, dan C3. Hal tersebut mengakibatkan kemampuan berpikir peserta didik menjadi rendah karena terbatas hafalan. Data nilai rerata UN mata pelajaran Biologi yang diperoleh di MAN Sidoarjo menunjukkan bahwa mengalami penurunan, pada tahun 2015 sebesar 459,82 menjadi 445,00 di tahun 2016. Penurunan hasil rerata UN tersebut diakibatkan oleh peserta didik kurang terbiasa dalam mengerjakan soal-soal. Oleh karena itu, guru diharapkan mampu membiasakan untuk melatih soal-soal yang mengandung kemampuan berpikir tingkat tinggi. Soal yang mengandung kemampuan berpikir tingkat tinggi dalam konteks penilaian tidak hanya mengingat kembali atau menyampaikan kembali suatu

informasi tetapi merujuk pada mengaitkan informasi satu ke informasi lain, menggunakan informasi supaya dapat memecahkan masalah, dan mengolah informasi secara kritis (Widana, 2017).

Pembuatan instrumen soal idealnya harus ditelaah atau dianalisis terlebih dahulu sebelum diberikan kepada peserta didik, sehingga diketahui bagaimana kualitas instrumen (Hasanah *et al.*, 2016). Hal ini berkebalikan dengan fakta lapangan bahwa penelaahan instrumen dilakukan setelah tes dilaksanakan dan untuk mengetahui kemampuan peserta didik telah mencapai nilai Kriteria Ketuntasan Minimal (KKM) atau belum. Tujuan dari analisis instrumen adalah mengkaji banyak pertanyaan supaya diperoleh informasi mengenai kualitas pertanyaan yang memadai (Sudjana, 2013). Analisis soal penting untuk memperoleh informasi mengenai apakah soal perlu diperbaiki dan dibuang karena dinilai tidak efektif dalam mengukur kemampuan peserta didik terkait ketercapaian memahami materi atau konsep (Masruroh, 2012). Analisis kualitas soal dapat dilakukan secara kualitatif dan kuantitatif. Analisis kualitatif butir soal ditinjau dari aspek materi, bahasa, dan konstruksi, sedangkan analisis kuantitatif ditinjau dari reliabilitas, indeks daya beda, indeks kesukaran soal, dan distraktor (Retnawati, 2016).

Berdasarkan uraian tersebut tujuan penelitian ini ialah untuk mendeskripsikan validitas teoritis dan empiris instrumen pengembangan butir soal Penilaian Akhir Semester (PAS) Gasal pada kelas X buatan Guru Biologi di MAN Sidoarjo tahun ajaran 2017/2018.

METODE

Penelitian ini ialah penelitian deskriptif dengan metode kualitatif dan kuantitatif, karena penelitian ini diperoleh dengan mempelajari masalah, proses, tata cara yang sedang terjadi di masyarakat. Penelitian dilakukan selama Bulan Maret s.d Mei 2018. Sasaran penelitian ini yakni 30 butir soal PAS buatan guru Biologi dalam bentuk pilihan ganda, kunci jawaban, dan lembar jawaban peserta tes. Subjek penelitian sebanyak 30 peserta tes kelas X MIA II di MAN Sidoarjo.

Teknik pengumpulan data berupa wawancara, telaah, dan tes. Teknis analisis data secara kualitatif menggunakan lembar telaah yang ditinjau dari aspek materi, konstruksi, dan bahasa oleh dua orang dosen Biologi. Analisis kuantitatif menggunakan program komputer ITEMAN 3.00, meliputi reliabilitas, indeks daya beda, tingkat kesukaran soal, dan distraktor.

Reliabilitas didefinisikan sebagai konsistensi (keajegan) alat ukur. Hasil analisis nilai reliabilitas diamati berdasarkan nilai *Alpha* ITEMAN 3.00. Indeks daya beda ialah nilai yang terdapat pada butir tes yang

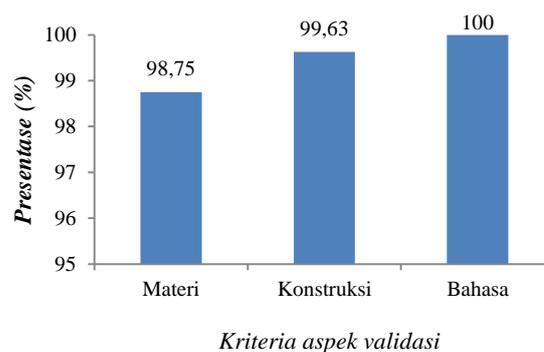
berfungsi untuk membedakan siswa kelompok atas dengan kelompok bawah. Hasil analisis indeks daya beda diamati dari nilai *biserial* analisis ITEMAN 3.00. Indeks kesukaran soal berkaitan dengan proporsi siswa menjawab benar (Handayani *et al.*, 2015). Hasil analisis indeks kesukaran soal diamati dari nilai *proporsi correct* analisis ITEMAN 3.00. Distraktor merupakan alternatif jawaban yang terdapat dalam soal pilihan ganda. Hasil analisis distraktor diamati dari nilai *proporsi endorsing* analisis ITEMAN 3.00. Nilai *prop. endorsing* dihitung sebagai distraktor jika $\geq 0,050$ artinya distraktor/pengecoh dipilih oleh 5% peserta tes.

HASIL DAN PEMBAHASAN

Instrumen soal PAS yang dikembangkan berasal dari soal buatan guru Biologi kelas X di MAN Sidoarjo. Soal-soal tersebut dianalisis secara kualitatif ditinjau berdasarkan validitas teoritis. Setelah dilakukan analisis diketahui bahwa soal PAS tersebut tidak memiliki kisi-kisi soal yang dibuat setelah dilaksanakan PAS. Kisi-kisi soal tersebut digunakan hanya untuk mengetahui peserta didik yang telah memenuhi KKM atau belum.

Hal tersebut mengakibatkan jumlah butir soal pada setiap indikator pembelajaran tidak representatif, sehingga mempengaruhi proporsi cakupan materi. Selain itu, tingkatan kognitif soal PAS buatan guru didominasi oleh C1, C2, dan C3 dengan nilai presentase sebesar 76,6%, sedangkan C4 dan C5 sebesar 23,3%. Instrumen soal buatan guru harus mengacu pada indikator pembelajaran yang tersusun dalam Kompetensi Dasar (KD) sesuai dalam RPP pada saat pembelajaran di kelas (Hasanah *et al.*, 2016). Dengan demikian, pengembangan instrumen soal PAS perlu dilakukan guna memperoleh suatu alat ukur yang dapat mengukur kemampuan peserta didik secara akurat.

Pengembangan instrumen soal PAS menghasilkan produk berupa kisi-kisi soal PAS dan butir-butir soal berbasis *Higher Order Thinking* (HOT). Sebelum diujicobakan secara terbatas, dilakukan telaah oleh dua orang ahli yang ditinjau dari validitas teoritis. Instrumen soal yang telah disusun tidak dapat dinyatakan secara langsung bahwa setiap butir-butir soal baik, maka dari itu diperlukan penelaahan terhadap instrumen soal (Rahmani *et al.*, 2015). Hasil rekapitulasi validasi teoritis pengembangan instrumen soal PAS dapat dilihat pada Gambar 1.



Gambar 1. Hasil rekapitulasi validasi pengembangan instrumen butir soal Penilaian Akhir Semester (PAS) Gasal di MAN Sidoarjo.

Berdasarkan hasil validitas teoritis butir soal Penilaian Akhir Semester (PAS) yang ditinjau dari aspek materi, konstruksi, dan bahasa didapatkan presentase rata-rata keseluruhan sebesar 99,5%. Hasil keseluruhan dinyatakan sangat layak.

Aspek materi diperoleh hasil presentase sebesar 98,75% termasuk dalam kriteria sangat layak. Hal tersebut menunjukkan bahwa pengembangan instrumen sudah terpenuhi dalam semua kriteria yang ingin dinilai. Instrumen tersebut dikatakan sangat layak salah satunya karena indikator pembelajaran dan butir soal telah sesuai. Aspek materi berkaitan dengan keilmuan dan tingkatan berpikir (Retnawati, 2016).

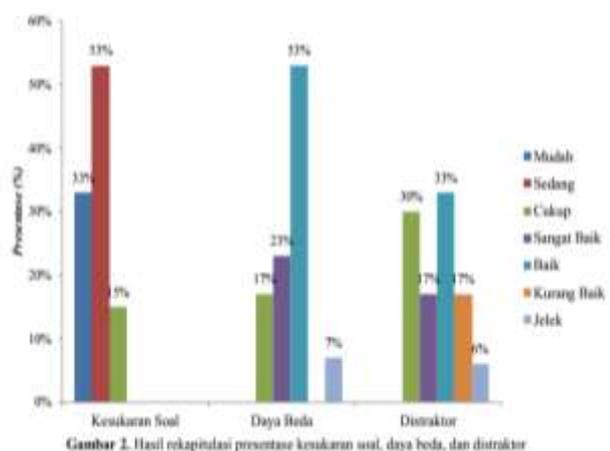
Aspek konstruksi diperoleh hasil presentase sebesar 99,63% termasuk dalam kriteria sangat layak. Hasil tersebut menunjukkan bahwa instrumen butir soal dalam beberapa aspek, seperti terdapat gambar/grafik/tabel, butir soal memiliki rumusan pertanyaan dan jawaban sesuai keperluan, jawaban tersusun atas panjang kalimat relatif sama, dan tidak merujuk jawaban benar disusun secara baik dan proporsional. Aspek konstruksi berkaitan dengan tehnik penulisan butir soal (Mardapi, 2017).

Aspek bahasa diperoleh hasil presentase sebesar 100% termasuk dalam kriteria sangat layak. Aspek bahasa berhubungan dengan kejelasan hal yang menyusun rumusan pertanyaan dalam butir soal (Mardapi, 2017). Instrumen soal yang telah dinyatakan valid secara teoritis menunjukkan bahwa aspek materi, konstruksi, dan bahasa telah terpenuhi (Rachma dan Ratnasari, 2015). Validitas dinyatakan tinggi berarti dapat mendeskripsikan hasil belajar secara tepat (Kunandar, 2013).

Pengembangan instrumen soal PAS sebanyak 30 butir soal, setelah ditelaah secara teoritis diujicobakan secara terbatas dan dilakukan analisis secara empiris. Hasil ujicoba, dianalisis dari segi validitas empiris butir soal menggunakan rogram ITEMAN 3.00. Analisis butir soal meliputi reliabilitas, indeks daya beda, tingkat kesukaran soal, dan distraktor.

Hasil analisis reliabilitas tes ditentukan berdasarkan nilai *Alpha*, yakni sebesar 0,81. Nilai tersebut berada pada interval 0,80-1,00 termasuk dalam kategori sangat tinggi. Hal ini berarti bahwa instrumen perbaikan butir tes Penilaian Akhir Semester (PAS) yang mengacu pada *Higher Order Thinking* (HOT) mempunyai tingkat *keajegan* (konsistensi) yang sangat tinggi. Jika dilakukan tes berulang-ulang pada kelompok yang sama tetapi dalam waktu berbeda. Reliabilitas soal berhubungan dengan konsistensi soal untuk mengukur prestasi belajar peserta didik (Masruroh, 2012). Ada beberapa faktor yang mempengaruhi nilai reliabilitas yaitu secara langsung maupun tak langsung. Faktor secara langsung, meliputi waktu pelaksanaan, panjang instrumen soal, indeks kesukaran soal, penyebaran nilai, dan objektivitas penskoran, sedangkan faktor tidak langsung, yakni petunjuk pelaksanaan jelas, pengawasan, dan kondisi lingkungan (Retnawati, 2016).

Hasil analisis presentase tingkat kesukaran, indeks daya beda, dan distraktor soal disajikan pada Gambar 2.



Gambar 2. Hasil rekapitulasi presentase kesukaran soal, daya beda, dan distraktor

Berdasarkan Gambar 2 diperoleh presentase tingkat kesukaran soal 15% kategori sulit, 53% kategori sedang, dan 33% kategori mudah. Butir soal kategori sulit terdapat pada materi mengenai peranan virus, ciri-ciri Cyanobacteria, peranan bakteri, dan struktur sel bakteri. Butir soal ketegori sedang diantaranya terdapat pada materi cabang ilmu biologi, kegiatan manusia sebagai ancaman keanekaragaman hayati, dasar klasifikasi, dasar sistem klasifikasi, kunci identifikasi, struktur tubuh virus, peranan virus bagi kehidupan, bagian-bagian sel bakteri, cara reproduksi bakteri, struktur tubuh ganggang, dan peranan protozoa. Adapun butir soal kategori mudah diantaranya terdapat pada materi mengurutkan tingkatan organisasi kehidupan, komponen karya tulis ilmiah, persebaran flora dan fauna di Indonesia, tingkatan takson makhluk hidup, serta penulisan nama ilmiah.

Tingkat kesukaran soal ditentukan menggunakan cara penskoran dikotomi, yaitu dengan menghitung jumlah yang menjawab benar diberi nilai satu (1) dan tidak menjawab atau salah diberi nilai nol (0). Apabila didapatkan tingkat kesukaran soal interval nilai 0,00-0,30 maka butir tersebut sukar, interval 0,30 – 0,70 butir soal sedang, dan interval 0,70-1,00 butir soal mudah (Arikunto, 2009). Proporsi indeks kesukaran soal yang ideal antara mudah, sedang, dan sukar mempunyai perbandingan 30%:50%:20% (Kunandar, 2013). Indeks kesukaran soal dijadikan salah satu parameter untuk menganalisis suatu tes karena berguna dalam menentukan kemampuan yang dimiliki antar peserta didik (Retnawati, 2016).

Berdasarkan hasil perhitungan indeks daya beda diperoleh indeks daya beda 23% termasuk kategori baik sekali. Butir soal yang memiliki indeks daya beda kategori baik sekali, terdapat pada contoh soal mengenai dasar menentukan klasifikasi tumbuhan tingkat tinggi, artikel peranan virus, gambar struktur bagian-bagian bakteri, menentukan cara reproduksi bakteri, mengidentifikasi struktur *Clamydomonas*, dan ciri umum Protista.

Presentase indeks daya beda 53% termasuk kategori baik. Butir soal yang memiliki indeks daya beda kategori baik, terdapat pada contoh soal tentang mengidentifikasi cabang biologi dan tingkatan organisasi kehidupan, komponen karya tulis ilmiah, tingkat keanekaragaman hayati dan persebaran flora dan fauna, upaya pelestarian hayati, dasar klasifikasi makhluk hidup, kunci identifikasi, peranan virus dan bakteri, struktur sel bakteri, dan ciri-ciri Cyanobacteria.

Presentase indeks daya beda 17% termasuk kategori cukup. Butir soal yang memiliki indeks daya beda kategori cukup, terdapat pada contoh soal tentang mengurutkan tingkatan organisasi kehidupan, menentukan persebaran zoogeografis Indonesia, menentukan tingkatan takson makhluk hidup, struktur tubuh virus pada saat kondisi ekstraseluler, dan menentukan dasar pengelompokan Protozoa.

Presentase indeks daya beda 7% termasuk kategori jelek. Butir soal yang memiliki indeks daya beda kategori jelek, terdapat pada contoh soal tentang mengidentifikasi tahapan replikasi virus dan peranan virus terkait penyebab penyakit AIDS. Indeks daya beda pada interpretasi kriteria jelek dapat dikatakan bahwa butir soalnya tidak dapat membedakan kemampuan peserta didik yang tinggi dan peserta didik yang bawah. Selain itu, nilai indeks daya beda menunjukkan angka dibawah nol dan negatif. Faktor tersebut dikarenakan peserta didik tidak memahami secara teliti pernyataan dalam soal sehingga peserta tes hanya menebak jawaban. Beberapa faktor yang mempengaruhi nilai indeks daya beda, antara lain tabel/gambar dalam

butir soal tidak efisien sehingga siswa menebak jawaban dan waktu pengerjaan soal terbatas sehingga peserta tes terburu-buru dalam mengerjakan soal (Putri dan Raharjo, 2017).

Berdasarkan pada Gambar 2, menunjukkan bahwa keefektifan distraktor sebesar 17% semua distraktor berfungsi dengan sangat baik, 33% tiga distraktor berfungsi dengan baik, 30% dua distraktor berfungsi dengan cukup, 17% satu distraktor yang berfungsi, dan 6% distraktor tidak berfungsi. Distraktor ditentukan dengan dari hasil perhitungan *proporsi endorsing* menggunakan ITEMAN 3.00. Jika nilai dari *proporsi endorsing* lebih dari sama dengan 0,050 berarti distraktor tersebut dipilih sebanyak 5% oleh peserta tes (Rofiah, 2013). Soal yang memiliki item distraktor kurang baik dan jelek diduga tidak dipilih oleh peserta tes karena alternatif jawaban kurang homogen (Masruroh, 2012).

Berdasarkan analisis menggunakan ITEMAN 3.00 ditinjau dari segi empiris, yakni reliabilitas, indeks kesukaran soal, daya beda, dan distraktor menunjukkan bahwa sebanyak 23 butir diterima dan 7 butir direvisi. Kebijakan dalam mengambil keputusan berdasarkan kriteria berikut: (a) butir soal diterima jika semua kriteria (indeks kesukaran soal, daya beda, dan distraktor berfungsi) terpenuhi. Contohnya, butir soal dinyatakan sukar/mudah tetapi daya beda dan persebaran distraktor memenuhi kriteria butir diterima, (b) butir soal diperbaiki/revisi jika salah satu atau lebih kriteria butir tidak terpenuhi, (3) butir soal ditolak jika soal tidak sesuai terhadap semua kriteria (Wardany *et al.*, 2015).

Instrumen perbaikan soal terhadap PAS gasal buatan guru Biologi dikembangkan untuk mengukur kemampuan berpikir tingkat tinggi (*Higher Order Thinking/HOTS*) peserta didik. Widana (2017) memaparkan bahwa dalam menyusun butir soal HOTS terdapat tiga karakteristik yang direkomendasikan sebagai pedoman, yaitu dapat mengukur kemampuan HOTS, permasalahan yang kontekstual dan menggunakan macam-macam bentuk soal. Kemampuan berpikir tingkat tinggi (HOTS) diperlukan supaya mencapai kompetensi yang dituntut dalam kehidupan abad 21. Peserta didik diharapkan mampu untuk mengkorelasi, memanipulasi, dan mentransfer pengalaman dan pengetahuan yang diperoleh secara berpikir kritis serta kreatif supaya mampu memecahkan masalah (Rofiah *et al.*, 2013). Dikembangkan beberapa kriteria dalam berpikir kritis, seperti mengusulkan permasalahan, membuat perencanaan, menyampaikan argumen dan menanggapi, membuat keputusan, serta memperbaiki konsep (Sunaryo, 2014).

Instrumen perbaikan soal terhadap PAS gasal buatan guru Biologi berhasil dikembangkan soal HOTS

berdasarkan indikator yang ingin dicapai dalam pembelajaran. Pada butir soal tersebut disajikan stimulus yang sesuai dengan isu-isu yang sedang terjadi, misalnya kesehatan, sains, teknologi, dan perkebunan. Isu-isu tersebut dapat diimplementasi dalam bentuk cuplikan artikel, gambar/diagram, dan tabel. Demikian, membiasakan peserta didik supaya dapat mengembangkan kemampuan berpikir kritis dan kreatif.

SIMPULAN

Berdasarkan hasil analisis data, dapat disimpulkan bahwa pengembangan instrumen soal PAS dianalisis berdasarkan validitas teoritis yang ditinjau dari aspek materi, konstruksi, dan bahasa diperoleh hasil sebesar 99,5% termasuk dalam kriteria sangat layak. Adapun analisis validitas empiris, meliputi reliabilitas, indeks daya beda, tingkat kesukaran soal, dan distraktor. Nilai reliabilitas sebesar 0,81 kategori sangat tinggi; tingkat kesukaran soal 33% mudah, 53% sedang, dan 15% sulit; indeks daya beda 23% baik sekali, 53% baik, 17% cukup, dan 7% jelek; distraktor 17% sangat baik, 33% baik, 30% cukup, 17% kurang baik, dan 6% jelek. Butir soal dinyatakan diterima sebanyak 23 butir dan 7 butir direvisi.

SARAN

Guru harus membuat kisi-kisi soal sebelum diujicobakan ke peserta didik dan butir soal jenjang kognitif tinggi perlu ditambahkan dalam butir soal, supaya peserta didik terbiasa untuk berpikir tingkat tinggi dalam menyelesaikan masalah.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada Prof. Dr. Endang Susantini, M.Pd. dan Dra. Rinie Pratiwi Puspitawati, M.Si. selaku desain pengujian serta validator pada penelitian ini, serta Asnani, S.Pd., selaku guru mata pelajaran Biologi. Semua siswa kelas X MIA 2 MAN Sidoarjo atas kerjasama selama pengambilan data.

DAFTAR PUSTAKA

- Arikunto, S. 2009. *Dasar-dasar Evaluasi Pendidikan (edisi revisi)*. Jakarta: Bumi Aksara.
- Handayani, A. P., Muhandjito., dan Sumarjono. 2015. Pengembangan Instrumen Tes Pilihan Ganda Distraktor Bermakna untuk Mengidentifikasi Karakteristik Konsepsi Fisika Siswa Kelas XI SMA Negeri 2 Malang. hlm: 1-13.
- Hasanah, Annisah L., Subali, B., dan Mariyan, S. 2016. Analisis Item Ujian Akhir Semester (UAS) Genap Mata Pelajaran Biologi Kelas X Tahun Ajaran

- 2014/2015. Di SMA yang Mengimplementasikan Kurikulum 2013 di Kabupaten Sleman. *Jurnal Pendidikan Biologi*. Vol. 5 (4); 7-26.
- Iliya, Amos. 2014. Formative and Summative Assessment in Educational Enterprise. *Journal of Educational and Practice*. Vol. 5 (20); 111-117.
- Kunandar. 2013. *Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik berdasarkan Kurikulum 2013)*. Jakarta: Raja Grafindo Persada.
- Putri, B. A. Y., dan Raharjo. 2017. Pengembangan Instrumen Evaluasi Higher Order Thinking (HOT) Berbasis Computer Based Test (CBT) pada Materi Sistem Indera Siswa Kelas XI SMA. *Jurnal Bioedu Unesa*. Vol. 6 (3): 353-359.
- Rachma, N. A dan Ratnasari, Evie. 2015. Pengembangan Tes Elektronik (E-Test) Berbasis Komputer pada Materi Bioteknologi di SMA Negeri 1 Surabaya. *Jurnal Bioedu Unesa*. Vol. 4 (3). page 1018-1022.
- Rahmani, Mita. 2015. Analisis Kualitas Butir Soal Buatan Guru Biologi Kelas X SMA Negeri 1 Tanah Pinoh. *Artikel Penelitian Pendidikan*. Pontianak: Universitas Tanjungpura.
- Retnawati, Heri. 2016. *Validitas, Reliabilitas, dan Karakteristik Butir (Panduan untuk Peneliti, Mahasiswa, dan Psikometrian)*. Yogyakarta: Parama Publishing.
- Rofiah, Emi., Aminah, N. S., dan Ekawati, E. Y. 2013. Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika pada Siswa SMP. *Jurnal Pendidikan Fisika*. Vol. 1(2). halaman 17-22.
- Sudargo, F., dan Soesy, A. S. 2010. Kemampuan Pedagogik Calon Guru dalam Meningkatkan Kemampuan Berpikir Kritis dan Keterampilan Proses Siswa melalui Pembelajaran Berbasis Praktikum. *Jurnal Pengajaran MIPA*. Vol. 16 (1): 4-12.
- Mardapi, D. 2017. *Pengukuran, Penilaian, dan Evaluasi Pendidikan (edisi 2)*. Yogyakarta: Parama Publishing.
- Masruroh., Rudyatmi, Ely., dan Ridlo, Saiful. 2012. Analisis Soal Ulangan Semester Gasal Biologi Kelas X di Kecamatan Petanahan Kebumen. *Unnes Journal of Biology Education*. Vol. 1 (2); 116-121.
- Naqvi, S. I. H., Hashmi, Muhammad Aamir., Hussain, Akhtar., 2010. Validation of Objective-Type Test in Biology at Secondary School Level. *Procedia Sosial and Behavioral Sciences*. 2; 3909-3913.
- Sudaryono., Margono, G., dan Rahayu, W. 2013. Pengembangan Instrumen Penelitian Pendidikan. Yogyakarta: Graha Ilmu.
- Sudjana, N. 2013. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.
- Sulistiawan, C. H. 2016. Kualitas Soal Ujian Sekolah Matematika Program IPA dan Kontribusinya terhadap Hasil Ujian Nasional. *Jurnal Penelitian dan Evaluasi Pendidikan*. Vol. 20 (1); 1-10
- Sunaryo, Yono. 2014. Model Pembelajaran Berbasis Masalah untuk Meningkatkan Kemampuan Berpikir Kritis dan Kreatif Matematika Siswa di Kota Tasikmalaya. *Jurnal Pendidikan dan Keguruan*. Vol. 1 (2)
- Wardany, K., Sajidan, dan Ramli, R. 2015. Penyusunan Instrumen Tes *Higher Order Thinking Skill* pada Materi Ekosistem SMA Kelas X. *Seminar Nasional XII Pendidikan Biologi FKIP UNS*. Hlm. 538-543.
- Widana, I Wayan. 2017. *Modul Penyusunan Soal Higher Order Thinking Skill (HOTS)*. Jakarta: Direktorat Pembinaan SMA, Direktorat Jenderal Pendidikan Dasar dan Menengah, Departemen Pendidikan dan Kebudayaan.