

Pengembangan Instrumen Penilaian Tematik Berbasis *Higher Order Thinking Skills* untuk Peserta Didik Kelas III Sekolah Dasar di Kota Metro

Nur Aisyah¹, Lilik Sabdaningtyas², Alben Ambarita³

¹ Peneliti, Magister Keguruan Guru Sekolah Dasar, FIP, Universitas Negeri Lampung
^{2,3} Dosen Pembimbing, Magister Keguruan Guru Sekolah Dasar, FIP, Universty Negeri Lampung
Jalan Sumantri Brojonegoro No. 1, Bandar Lampung, Indonesia,
email:nur.aisyah5415@gmail.com
Telep. 087722506610

Abstract: Development of Instrument Thematic Assessment Based On Higher Order Thinking Skills for Grade Iii Elementary Schools Students in Metro City. The aim of this research is to develop and to describe an affective assessment instrument. The type of this research is Research and Development that has been modified for 7 steps. The technique used to compile the data are tests and non-tests techniques. The tools to compile the data are questionnaire and test items. Qualitative and quantitative analysis technique were used to analyze data. The results are: assessment instrument is theoretically and empirically eligible to use. Theoretically means that it fulfilled the construction, linguistic, and material aspects. Empirically means that it has been through validity, reliability, difficulty levels, distinguish, and distractor examinations; assessment instrument has fulfilled Higher Order Thinking Skills characteristics; and assessment instrument is effective as a study evaluation tool because this instrument can distinguish students of low, middle, and high study capabilities.

Keywords: Instrument Assessment, Thematic, Higher Order Thinking Skills.

Abstrak: Pengembangan Instrumen Penilaian Tematik Berbasis *Higher Order Thinking Skills* untuk Peserta Didik Kelas III Sekolah Dasar di Kota Metro Tujuan penelitian ini adalah untuk mengembangkan dan mendeskripsikan instrumen penilaian yang efektif. Jenis penelitian ini adalah *Research and Development* yang dimodifikasi menjadi tujuh langkah. Teknik pengumpulan data menggunakan teknik nontes dan tes. Alat pengumpulan data berupa angket dan soal tes. Teknik analisis data menggunakan analisis kualitatif dan kuantitatif. Hasil penelitian ini, yaitu: instrumen penilaian yang layak digunakan, baik layak secara teoritis maupun empiris. Layak secara teoritis yaitu telah memenuhi aspek konstruksi, bahasa, dan materi. Layak secara empiris yaitu melalui uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan distraktor; instrumen penilaian telah memenuhi karakteristik *Higher Order Thinking Skills*; dan instrumen penilaian efektif digunakan sebagai alat evaluasi hasil belajar karena instrumen penilaian dapat membedakan siswa yang berkemampuan tinggi, sedang, dan rendah.

Kata Kunci: Instrumen Penilaian, Tematik, *Higher Order Thinking Skills*.

PENDAHULUAN

Standar penilaian berdasarkan Peraturan Pemerintah Republik Indonesia No. 32 Tahun 2013 Pasal 1 ayat (12) merupakan standar nasional pendidikan yang berkaitan dengan mekanisme, prosedur, dan instrumen penilaian hasil belajar peserta didik (Kemendikbud, 2013). Penilaian adalah pengumpulan informasi untuk membuat keputusan evaluatif dan digunakan dalam kaitannya dengan tes (Abosalem, 2016: 3). Penilaian adalah alat untuk mengukur sejauh mana peserta didik telah meningkatkan pembelajaran mereka berdasarkan standar Haertel dalam (Mangiante, 2013: 222). Hasil dari penilaian adalah proses awal sebelum membuat keputusan tentang langkah-langkah berikutnya dalam meningkatkan kemampuan untuk memahami belajar peserta didik (Ammie, 2016: 59).

Proses penilaian sangat baik dan efektif apabila prinsip-prinsip berikut diperhatikan, yaitu: 1) adanya pembatasan dengan jelas apa yang diutamakan untuk dinilai. 2) teknik penilaian yang dipilih harus sesuai dengan karakteristik atau kemampuan yang akan diukur. 3) penilaian secara komprehensif membutuhkan berbagai jenis teknik penilaian. 4) sebaiknya disadari bahwa teknik penilaian yang digunakan memiliki keterbatasan. 5) penilaian harus dapat digunakan untuk meningkatkan prestasi belajar peserta didik (Gronlund & Linn, 1990: 6-8).

Kemampuan berpikir tingkat tinggi didefinisikan sebagai penggunaan pikiran secara luas untuk menemukan tantangan baru (Heong, et.al., 2011: 122). Berpikir tingkat tinggi adalah proses yang melibatkan operasi-operasi mental seperti klasifikasi, induksi, deduksi, dan penalaran (Sastrawati, et.al., 2011: 6).

Higher Order Thinking Skills menurut King, Goodson, dan Rohani (2010: 1-2) meliputi berpikir kritis, logis, reflektif, metakognitif, dan kreatif.

Permasalahan yang terjadi di sekolah, soal-soal cenderung lebih banyak menguji aspek ingatan yang kurang melatih keterampilan berpikir tingkat tinggi peserta didik, kemampuan berpikir anak Indonesia secara ilmiah dianggap masih rendah dilihat dari hasil survei TIMSS, salah satu faktor penyebabnya antara lain karena peserta didik di Indonesia kurang terlatih dalam menyelesaikan soal-soal yang mengukur *Higher Order Thinking Skills* dan masalah yang dihadapi oleh guru adalah kemampuan guru dalam mengembangkan instrumen penilaian *Higher Order Thinking Skills* masih kurang dan belum tersedianya instrumen penilaian yang didesain khusus untuk melatih *Higher Order Thinking Skills*, sehingga perlu dikembangkan instrumen penilaian *Higher Order Thinking Skills* (Budiman & Jailani, 2014: 142).

Permendikbud No. 81 Tahun 2013 tentang implementasi kurikulum menyebutkan bahwa kebutuhan kompetisi masa depan diperlukan *Higher Order Thinking Skills* secara kritis, keterampilan komunikasi, dan kreatif (Budiman & Jailani, 2014: 141). Hal ini sejalan dengan pendapat (Rofiah, et.al, 2013: 18) mengenai karakteristik *skills* masyarakat abad ke-21 yang dipublikasikan oleh *Partnership of 21st Century Skill* mengidentifikasi bahwa pembelajar pada abad ke-21 harus mampu mengembangkan keterampilan kompetitif yang diperlukan pada abad ke-21 yang berfokus pada *Higher Order Thinking Skills*.

Pentingnya kemampuan berpikir tingkat tinggi diungkapkan oleh Peter

(2012: 39) agar dapat bersaing dalam dunia kerja dan kehidupan pribadi, seperti: memberikan alasan secara tepat (*reasoning*), berpikir secara kritis (*critical thinking*), berpikir secara kreatif (*thinking creatively*), membuat keputusan (*decisions making*), dan menyelesaikan masalah (*problem solving*).

Menghadapi perubahan yang sangat cepat, hendaknya peserta didik perlu diberikan bekal tentang bagaimana belajar dan bagaimana berpikir agar mereka memiliki kesiapan dan kemampuan untuk memasuki dunia kerja sesuai dengan persyaratan yang ditentukan (Wening, 2015: 1046).

Berdasarkan hal tersebut, Kurikulum 2013 diharapkan dapat membantu siswa untuk meningkatkan *Higher Order Thinking Skills*. Berpikir tingkat tinggi dapat mendorong peserta didik untuk berpikir secara luas dan mendalam tentang materi pelajaran. Peserta didik yang cenderung berpikir tingkat rendah perlu dilatih sejak dasar, agar pada saat memasuki jenjang pendidikan berikutnya peserta didik tidak merasa takut jika dihadapkan pada persoalan yang lebih rumit. Oleh karena itu, guru harus kreatif dan mampu membuat pertanyaan yang dapat melatih kemampuan berpikir tingkat tinggi.

Hasil pra survey berupa angket mengenai instrumen penilaian dengan sasaran 30 orang guru kelas di SD Kota Metro yang menerapkan Kurikulum 2013 yaitu, terdapat 83% guru tidak membuat instrumen penilaian sendiri, 90% dalam proses pembuatan instrumen penilaian guru tidak membuat indikator dari KI dan KD yang memuat dimensi dan aspek kognitif, 90% guru tidak membuat kisi-kisi instrumen penilaian dan mengalami kesulitan untuk membuat

instrumen penilaian, 100% soal UTS dan UAS yang diberikan tidak terintegrasi, 100% guru tidak mempunyai dan tidak tahu cara membuat instrumen penilaian tematik berbasis *Higher Order Thinking Skills*, 90% guru menyusun butir soal cenderung *Lower Order Thinking Skills* pada ranah kognitif C1-C3 dan terdapat pada jenis pengetahuan dimensi faktual dan konseptual, 100% guru tidak pernah memperhatikan kualitas butir soal dan tidak pernah melakukan analisis soal, dan 100% guru membutuhkan instrumen penilaian yang valid dan reliabel

Peneliti membuat instrumen penilaian berupa tes objektif berbentuk uraian dan tes subjektif berbentuk pilihan ganda. Karena pada dasarnya semua soal bentuk pilihan ganda dan uraian dapat dikembangkan menjadi soal-soal yang *Higher Order Thinking Skills*.

Hal ini sejalan dengan pendapat (Abdullah, 2016:176-177) tes yang paling sering digunakan untuk menilai pengetahuan peserta didik adalah tes pilihan ganda dan tes uraian. Tes PG dapat digunakan untuk menguji penguasaan materi dengan cakupan yang banyak dan lebih mudah penskorannya. Oleh sebab itu, tes pilihan ganda digunakan dalam UN dan di tingkat internasional, misalnya: TIMSS, PISA, dan PIRLS.

Penelitian ini sejalan dengan Rubin & Manikya (2015: 37) yaitu penelitian tersebut berusaha untuk menggali potensi menggunakan pilihan ganda dengan alasan untuk mengajarkan pemecahan masalah dalam berpikir tingkat tinggi. Soal uraian menuntut peserta didik untuk berpikir tingkat tinggi.

Mohammed, et.al. (2015: 14) aspek kognitif berdasarkan Taksonomi Bloom revisi dibedakan atas enam

level, tiga level pertama merupakan *Lower Order Thinking Skills*, sedangkan tiga level berikutnya *Higher Order Thinking Skills* yaitu aspek mengingat (C1), memahami (C2), menerapkan (C3), menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Sejalan dengan (Anderson & Karthwol, 2001) indikator untuk mengukur *Higher Order Thinking Skills* meliputi aspek C4-C6.

Menurut Gunawan & Anggarini (2008: 30-31) dimensi pengetahuan merupakan dimensi tersendiri dalam Taksonomi Bloom revisi. Pada dimensi ini dipaparkan empat jenis kategori pengetahuan. Dimensi pengetahuan dikemukakan oleh Anderson, L.W. & Karthwol, D.R. (2001: 38-62), yaitu pengetahuan faktual, pengetahuan konseptual, pengetahuan prosedural, dan pengetahuan metakognitif.

Kategori berpikir tingkat tinggi menurut Brookhart (2010: 14-15) meliputi beberapa aspek, yaitu: analisis, evaluasi, kreasi; penalaran yang logis atau logika beralasan; keputusan dan berpikir kritis; pemecahan masalah; kreatifitas dan berpikir kreatif. Krulik & Rudnick (1999: 138) kemampuan berpikir terdiri atas empat tingkat, yaitu: menghafal (*recall thinking*), berpikir dasar (*basic thinking*), berpikir kritis (*critical thinking*) dan berpikir kreatif (*creative thinking*).

Devi dalam Rochmah & Asih (2015: 29) ada beberapa cara yang dapat dijadikan pedoman oleh para penulis soal untuk menulis butir soal yang menuntut berpikir tingkat tinggi, yakni: setiap pertanyaan diberikan dasar pertanyaan (stimulus) dan soal mengukur kemampuan berpikir kritis. Agar butir soal yang ditulis dapat menuntut berpikir tingkat tinggi, maka

setiap butir soal selalu diberikan dasar pertanyaan (stimulus) yang berbentuk sumber/bahan bacaan seperti: teks bacaan, paragraf, teks drama, penggalan novel/cerita/dongeng, puisi, kasus, gambar, grafik, foto, rumus, tabel, daftar kata/symbol, contoh, peta, film/rekaman suara.

Indikator keterampilan berpikir kritis menurut Robert H. Ennis dalam Costa (1985: 54-57) adalah: 1) memfokuskan pada pertanyaan, 2) menganalisis argumen, 3) mempertimbangkan yang dapat dipercaya, 4) mempertimbangkan laporan observasi, 5) membandingkan kesimpulan, 6) menentukan kesimpulan, 7) mempertimbangkan kemampuan induksi, 8) menilai, 9) mendefinisikan konsep, 10) mendefinisikan asumsi, dan 11) mendeskripsikan. Sedangkan menurut Torrance dalam Filsaime (2008: 20) ada empat karakteristik berpikir kreatif, sebagai sebuah proses yang melibatkan unsur-unsur orisinalitas, kelancaran, fleksibilitas dan elaborasi.

Instrumen penilaian berbasis *Higher Order Thinking Skills* dalam penelitian hanya memuat satu tema kelas III, yaitu Perubahan di Alam. Pengembangan instrumen penilaian yang peneliti kembangkan adalah pembelajaran tematik. Berdasarkan hal tersebut, peneliti mengembangkan KD menjadi indikator-indikator yang berpikir tingkat tinggi pada, yang meliputi aspek C4-C6. Menjembatani adanya permasalahan tersebut, perlunya adanya penelitian tentang instrumen penilaian tematik berbasis *Higher Order Thinking Skills* untuk siswa kelas III SD di Kota Metro.

Berdasarkan uraian di atas, tujuan penelitian ini yaitu: mendeskripsikan terpenuhinya kelayakan instrumen penilaian tematik berbasis *Higher Order Thinking Skills* untuk peserta

didik kelas III SDN Kota Metro, mendeskripsikan terpenuhinya karakteristik *Higher Order Thinking Skills* pada instrumen penilaian tematik, untuk peserta didik kelas III SDN Kota Metro, dan mengukur keefektifan instrumen penilaian tematik berbasis *Higher Order Thinking Skills* untuk peserta didik kelas III SDN Kota Metro.

METODE PENELITIAN

Jenis penelitian ini adalah penelitian dan pengembangan atau *Research and Development (R&D)*. Produk yang dikembangkan adalah instrumen penilaian *Higher Order Thinking Skills*. Untuk mendapatkan prototipe pengembangan, pada penelitian ini dilakukan adaptasi dari model pengembangan Borg & Gall dalam (Sugiyono, 2014: 298).

Berdasarkan 10 langkah penelitian, pengembangan ini diadopsi menjadi tujuh langkah, yaitu: tahap potensi dan masalah, pengumpulan data, desain produk, validasi desain, revisi desain, uji coba lapangan, dan revisi penyempurnaan produk.

Subjek penelitian ini berjumlah 119 siswa kelas III di empat SDN ini kota Metro tahun ajaran 2016/2017, yaitu SDN 1 Metro Pusat, SDN 1 Metro Utara, SDN 1 Metro Timur, dan SDN 1 Metro Barat, terdiri dari: 12 orang untuk uji coba kelompok, 32 orang untuk uji coba kelas, dan 80 orang untuk uji lapangan.

Teknik analisis data yang digunakan yaitu analisis data kualitatif dan kuantitatif. Analisis secara kualitatif dilakukan melalui penelaah untuk mengetahui kelayakan teoritik melalui validitas isi instrumen tes, baik dari segi materi, konstruksi, bahasa, dan kaidah penulisan soal. Selain mengetahui kelayakan teoritik,

analisis data kualitatif dalam penelitian ini adalah untuk mengetahui pemenuhan karakteristik *Higher Order Thinking Skills* pada instrumen penilaian. Data ini termasuk data kualitatif berupa kritik, saran, dan tanggapan dari validator dianalisis secara deskriptif. Analisis secara kuantitatif dalam penelitian ini dibagi menjadi tiga, yaitu:

1. Uji Kelayakan Instrumen

Uji kelayakan ini terbagi menjadi dua, yaitu kelayakan teoritik dan kelayakan empiris. Data kuantitatif dalam menguji kelayakan teoritik didapatkan dari jumlah skor dalam lembar validasi produk oleh para ahli yang diperoleh dari angket.

Angket ini menggunakan skala Guttman, maka skor untuk jawaban Ya adalah 1 dan skor untuk jawaban Tidak adalah 0. Soal dikatakan layak secara teoritik (jika rata-rata jumlah setiap skor butir soal yang diberikan semua ahli $\geq 50\%$). Soal dikatakan tidak layak secara teoritik (jika rata-rata jumlah setiap skor butir soal yang diberikan semua ahli $< 50\%$).

Kelayakan empiris digunakan untuk menganalisis butir soal dengan menggunakan program ANATES. Analisis butir soal digunakan untuk pengujian terhadap kualitas soal yang diujicobakan pada uji coba kelas. Diantaranya adalah dengan menguji validitas, reliabilitas, daya pembeda, indeks kesukaran dan distraktor.

2. Uji Karakteristik HOTS

Data kuantitatif dalam menilai terpenuhinya karakteristik *Higher Order Thinking Skills* pada instrumen penilaian tematik ini didapatkan dari jumlah poin atau skor dalam lembar validasi produk oleh para ahli. Angket ini menggunakan skala Guttman, maka skor untuk jawaban Ya adalah 1, dan skor untuk jawaban Tidak adalah 0. Soal yang memenuhi karakteristik

Higher Order Thinking Skills dikatakan terpenuhi (jika rata-rata skor jawaban $\geq 50\%$). Soal yang memenuhi karakteristik *Higher Order Thinking Skills* dikatakan tidak terpenuhi (jika rata-rata skor jawaban $< 50\%$).

Tabel 1. Indikator Penilaian Berbasis HOTS

Karakteristik HOTS	Kriteria	Ket.
Dasar Pertanyaan (Stimulus)	a. gambar b. tabel c. grafik d. simbol e. foto f. contoh g. rumus h. penggalan kasus i. persamaan j. diagram k. teks	Setiap aspek dikatakan terpenuhi, jika terdapat salah satu indikator pada masing-masing instrumen penilaian tersebut
Kemampuan berpikir kritis	a. memberikan penjelasan sederhana b. membangun keterampilan dasar c. menyimpulkan d. membuat penjelasan lanjut e. mengatur strategi dan taktik	
Kemampuan berpikir kreatif	a. kelancaran b. fleksibilitas c. orisinalitas d. elaborasi	

3. Uji Efektivitas Instrumen

Uji efektivitas instrumen penilaian ini dilakukan untuk mengetahui perbedaan rata-rata yang signifikan antara kelompok tinggi, sedang, dan rendah pada uji lapangan terhadap produk melalui uji ANAVA dengan menggunakan IBM SPSS 20,0.

HASIL DAN PEMBAHASAN

1. Kelayakan Instrumen Penilaian

a. Kelayakan Teoritik

Tabel 2. Hasil Rekapitulasi Penilaian Kelayakan Bentuk PG

No.	Validasi	Persentase	Kriteria
1	Ahli Evaluasi	100%	Layak
2	Ahli Materi	100%	Layak
3	Ahli Bahasa	90,2%	Layak
Rata-rata		96,7%	Layak

Tabel 2 menunjukkan persentase rata-rata keseluruhan rekapitulasi

kelayakan bentuk PG oleh ahli adalah 96,7% dengan kriteria layak.

Tabel 3. Hasil Rekapitulasi Penilaian Kelayakan Bentuk Uraian

No.	Validasi	Persentase	Kriteria
1	Ahli Evaluasi	100%	Layak
2	Ahli Materi	100%	Layak
3	Ahli Bahasa	96%	Layak
Rata-rata		98,7%	Layak

Tabel 3 menunjukkan persentase rata-rata keseluruhan rekapitulasi kelayakan bentuk uraian oleh ahli adalah 98,7% dengan kriteria layak. Uji kelayakan instrumen penilaian secara teoritik dalam penelitian ini menggunakan angket yang sudah divalidasi oleh ahli, memenuhi aspek materi, konstruksi, dan bahasa. Berdasarkan hasil penelitian uji kelayakan, instrumen penilaian ini sudah memenuhi kaidah-kaidah secara teoritik.

Hal ini sejalan dengan pendapat (Abdullah, 2016: 186) untuk mendapatkan tes yang berkualitas, maka penulis harus memperhatikan kaidah-kaidah penulisan soal,. Penelitian ini sependapat dengan penelitian yang dilakukan oleh (Wardany, et al., 2015: 538) telaah instrumen secara teoritis atau deskriptif dilakukan untuk melihat keterbacaan instrumen dan untuk validasi isi.

b. Kelayakan Empiris

Hasil analisis secara empiris meliputi analisis validitas, reliabilitas, tingkat kesukaran, daya pembeda soal, fungsi distraktor, validitas dan reliabilitas.

1) Validitas

Tabel 4. Analisis Validitas Butir Bentuk PG Pada Uji Kelompok Kecil

No.	Kriteria	Nomor Soal	Jml	%
1	Tidak Valid	2, 3, 4, 5, 16, 34	6	17
2	Valid	1, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 35.	29	83
Jumlah			35	100

Berdasarkan Tabel 4, hasil analisis uji validitas soal PG pada uji kelompok, terdapat 83% soal yang valid. Sedangkan, soal yang tidak valid dilakukan revisi, untuk diujicobakan kembali pada uji kelas.

Tabel 5. Analisis Validitas Butir Bentuk PG Pada Uji Kelas

No.	Kriteria	Nomor Soal	Jml	%
1	Tidak Valid	-	0	0
2	Valid	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,35.	35	100
Jumlah			35	100

Berdasarkan Tabel 5, hasil analisis uji validitas soal PG pada uji kelas, 100% soal dinyatakan valid.

Tabel 6. Analisis Validitas Butir Bentuk Uraian Pada Uji Kelompok Kecil

No.	Kriteria	No. Soal	Jml	%
1	Tidak Valid	1	20	20
2	Valid	2, 3, 4, 5	80	80
Jumlah			5	100

Berdasarkan Tabel 6, hasil analisis uji validitas soal uraian pada uji kelompok kecil, terdapat 80% soal yang valid. Sedangkan, soal yang tidak valid dilakukan revisi, untuk diujicobakan kembali pada uji coba kelas.

Tabel 7. Analisis Validitas Butir Bentuk Uraian Pada Uji Kelas

No.	Kriteria	No. Soal	Jml	%
1	Tidak Valid	-	20	20
2	Valid	1, 2, 3, 4, 5	80	80
Jumlah			5	100

Berdasarkan Tabel 5, hasil analisis uji validitas soal uraian pada uji kelas, 100% soal dinyatakan valid. Berdasarkan hasil analisis validitas soal pilihan ganda dan uraian pada uji kelas, soal dinyatakan valid. Hasil penelitian ini sejalan dengan pendapat (Sudijono, 2013: 184), yaitu soal dinyatakan valid apabila korelasi skor butir soal memiliki range validitas positif.

2) Reliabilitas

Berdasarkan hasil penelitian, pada uji coba kelompok soal bentuk PG memiliki reliabilitas sangat tinggi, yaitu 0,96. Sedangkan, soal bentuk uraian memiliki reliabilitas tinggi, yaitu 0,81. Berdasarkan hasil penelitian, pada uji kelas soal bentuk PG memiliki reliabilitas sangat tinggi, yaitu 0,91. Sedangkan, soal bentuk uraian memiliki reliabilitas tinggi, yaitu 0,88. Oleh karena itu, instrumen penilaian yang dibuat oleh peneliti memiliki reliabilitas tinggi.

Penelitian ini sejalan dengan pendapat (Arifin, 2009: 69) suatu alat ukur dapat dikatakan reliabel atau andal jika ia mempunyai hasil yang taat asas. Misalnya, suatu alat ukur diberikan kepada sekelompok peserta didik saat ini, kemudian diberikan lagi kepada kelompok peserta didik yang sama pada saat yang akan datang, dan ternyata hasilnya sama atau mendekati sama, maka dapat dikatakan alat ukur tersebut mempunyai tingkat reliabilitas tinggi.

3) Tingkat kesukaran

Tabel 8. Analisis Tingkat Kesukaran Bentuk PG Pada Uji Kelompok

No.	Kriteria	Nomor Soal	Jml	%
1.	Sangat Mudah	34	1	3
2.	Mudah	2, 3, 4, 5, 16, 35	6	17
3.	Sedang	1, 6, 7, 11, 12, 13, 14, 15, 17, 18, 19, 21, 22, 23, 26, 28, 29, 30, 32, 33	20	57
4.	Sukar	8, 9, 10, 20, 24, 25, 27, 31	8	23
Jumlah			35	100

Tabel 9. Analisis Tingkat Kesukaran Bentuk Uraian Pada Uji Kelompok

No.	Kriteria	No. Soal	Jml	%
1	Mudah	1	1	20
2	Sedang	2, 3, 4	3	60
3	Sukar	5	5	20
Jumlah			5	100

Berdasarkan hasil penelitian pada uji coba kelompok, instrumen penilaian memiliki tingkat kesukaran yang proporsional antara sulit, sedang, dan mudah. Menurut (Arifin, 2009: 69) salah satu karakteristik instrumen yang baik harus proporsional, artinya suatu alat ukur harus memiliki tingkat kesulitan yang proporsional antara sulit, sedang dan mudah. Soal yang diterima adalah soal yang memiliki tingkat kesukaran sedang, sedangkan soal kriteria sangat mudah, mudah, dan sukar harus direvisi, untuk diujicobakan kembali pada uji kelas.

Tabel 10. Analisis Tingkat Kesukaran Bentuk PG Pada Uji Kelas

No.	Kriteria	No. Soal	Jml	%
1.	Mudah	11, 14, 19	3	8,6
2.	Sedang	1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 15, 16, 17, 18, 20, 21, 22, 23, 24, 26, 28, 29, 30, 32, 33, 34, 35	28	80
3.	Sukar	10, 25, 27, 31	4	11,4
Jumlah			35	100

Tabel 11. Analisis Tingkat Kesukaran Bentuk Uraian Pada Uji Kelas

No.	Kriteria	No. Soal	Jml	%
1	Mudah	-	0	0
2	Sedang	1, 2, 3, 4, 5	5	100
3	Sukar	-	0	0
Jumlah			5	100

Berdasarkan hasil penelitian pada uji coba kelompok, instrumen penilaian rata-rata memiliki tingkat kesukaran sedang. Hal ini sejalan dengan penelitian (Wardany, et al., 2015: 539-542) 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.

4) Daya Beda

Tabel 12. Analisis Daya Pembeda Bentuk PG Pada Uji Kelompok

No.	Kriteria	Nomor Soal	Jml	%
1	Jelek	-	0	0
2	Cukup	-	0	0
3	Baik	2, 16, 34	3	85,7
4	Sangat Baik	1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 35.	32	91,3
Jumlah			35	100

Berdasarkan Tabel 12, 91,3% soal bentuk PG pada uji coba kelompok memiliki daya pembeda yang sangat baik.

Tabel 13. Analisis Daya Pembeda Bentuk PG Pada Uji Kelompok

No.	Kriteria	No. Soal	Jml	%
1	Jelek	-	0	0
2	Cukup	-	0	0
3	Baik	1	1	20
4	Sangat Baik	2, 3, 4, 5	4	80
Jumlah			5	100

Berdasarkan Tabel 13, 80% soal bentuk uraian pada uji coba kelompok memiliki daya pembeda yang sangat baik.

Tabel 14. Analisis Daya Pembeda Bentuk PG Pada Uji Kelas

No.	Kriteria	Nomor Soal	Jml	%
1	Jelek	-	0	0
2	Cukup	-	0	0
3	Baik	27, 32	2	6
4	Sangat Baik	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34, 35.	33	94
Jumlah			35	100

Berdasarkan Tabel 14, 94% soal bentuk PG pada uji coba kelas memiliki daya pembeda yang sangat baik.

Tabel 15. Analisis Daya Pembeda Bentuk Uraian Pada Uji Kelas

No.	Kriteria	Nomor Soal	Jml	Persentase
1	Jelek	-	0	0%
2	Cukup	-	0	0%
3	Baik	1, 4	2	40%
4	Baik Sekali	2, 3, 5	3	60%
Jumlah			5	100%

Berdasarkan Tabel 12, 60% soal bentuk uraian pada uji coba kelas memiliki daya pembeda yang sangat baik. Secara keseluruhan soal yang dibuat oleh peneliti sudah baik, dan butir soal tersebut mampu membedakan mana peserta didik yang pandai dan mana peserta didik yang kurang pandai.

Hasil daya beda ini sejalan dengan pernyataan (Sudijono, 2013: 386) mengetahui daya pembeda item itu penting sekali, sebab salah satu dasar yang dipegang untuk menyusun butir soal tes hasil belajar adalah adanya anggapan, bahwa kemampuan antara siswa yang satu dengan siswa yang lain itu berbeda-beda, dan bahwa butir-butir tes hasil belajar itu haruslah mampu memberikan hasil tes yang mencerminkan adanya perbedaan-perbedaan kemampuan siswa tersebut.

5) Distraktor

Berdasarkan hasil penelitian, secara keseluruhan distraktor dalam instrumen penilaian ini sudah berfungsi dengan baik dan tidak ada omit. Menurut (Arikunto, 2012: 220) pola jawaban soal adalah distribusi testee dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Pola jawaban soal diperoleh dengan menghitung banyaknya testee yang memilih pilihan jawaban a, b, c, atau d atau yang tidak memilih pilihan manapun. Dalam istilah evaluasi disebut omit. Dari pola jawaban soal dapat ditentukan apakah pengecoh berfungsi sebagai pengecoh dengan baik atau tidak.

2. Karakteristik Instrumen Penilaian *Higher Order Thinking Skills*

Karakteristik instrumen penilaian *Higher Order Thinking Skills* menurut ahli terdapat pada Tabel 3.

Tabel 3. Hasil Analisis Karakteristik Instrumen Penilaian Berbasis HOTS

No	Validasi	Bentuk Stimulus	Berpikir Kritis	Berpikir Kreatif	Rata-rata Persentase
1	V-I	100%	90%	67,5%	85%
2	V-III	100%	87,5%	77,5%	88,3%
3	V-III	100%	92,5%	80%	90,8%
Rata-rata Total		100%	90%	75%	88%
Kriteria		Terpenuhi	Terpenuhi	Terpenuhi	Terpenuhi

Berdasarkan hasil penelitian, instrumen penilaian telah memenuhi karakteristik *Higher Order Thinking Skills*, meliputi: aspek stimulus, kemampuan berpikir kritis, dan kemampuan berpikir kreatif. Hal ini sejalan dengan penelitian (Rochmah & Asih, 2015: 37) diperoleh hasil karakteristik soal tipe *Higher Order Thinking Skills*, yaitu stimulus terdapat pada soal. Sedangkan untuk karakteristik kemampuan berpikir kritis dan kreatif belum ditemui dalam item soal karena pada pelaksanaan penelitian hanya dilakukan analisis terhadap butir soal, tidak beserta penyelesaian soal oleh siswa.

Penelitian ini memiliki perbedaan, sehingga pada penelitian ini bukan hanya aspek stimulus saja, melainkan aspek kemampuan berpikir kritis dan aspek kemampuan berpikir kreatif dapat diukur berdasarkan indikator yang sudah ditetapkan. Selain itu, penelitian ini terdapat penyelesaian soal oleh peserta didik. Sehingga instrumen penilaian yang peneliti kembangkan sudah memenuhi karakteristik *Higher Order Thinking Skills*.

Penelitian ini didukung oleh (Wardany, et.al., 2015: 542) berdasarkan hasil penelitiannya dapat ditarik kesimpulan bahwa 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.

Berdasarkan hasil penelitian uji coba kelas, maka dapat disimpulkan bahwa instrumen penilaian yang dikembangkan sudah memenuhi kelayakan secara empiris dan dapat digunakan sebagai masukan untuk guru kelas III dalam menyusun instrumen penilaian hasil belajar. Hal ini sejalan dengan penelitian yang dilakukan oleh (Budiman & Jailani, 2014) bahwa instrumen yang layak memiliki tingkat kesukaran sedang, daya pembeda baik, semua pengecoh berfungsi baik untuk soal PG. Sedangkan untuk soal uraian, memiliki tingkat kesukaran sedang dengan daya pembeda baik.

3. Efektivitas Instrumen Penilaian

Berdasarkan hasil perhitungan uji hipotesis, maka diperoleh jawaban dari rumusan masalah dan tujuan penelitian, yaitu perhitungan uji perbedaan rata-rata antara kelompok tinggi, sedang, dan rendah diperoleh $F_{hitung} = 238,417 > F_{tabel} = 3,115366$ pada taraf signifikansi 0,05 dilambangkan $F_{hitung} > F_{tabel}$ atau sig. 0,000 ($<0,05$). Sehingga untuk jawaban dari rumusan masalah bagaimanakah efektivitas instrumen penilaian berbasis *Higher Order Thinking Skills* adalah terdapat perbedaan skor rata-rata hasil belajar antara peserta didik yang berkemampuan tinggi, sedang, dan rendah.

Perbedaan hasil belajar masing-masing kelompok ini disebabkan karena alat ukur hasil belajar yang digunakan untuk menilai terbukti efektif, karena instrumen penilaian tersebut dapat membedakan antara peserta didik yang berkemampuan tinggi, rendah, dan sedang. Berdasarkan penjabaran tersebut maka

instrumen penilaian digunakan untuk mencapai tujuan yaitu untuk mengukur hasil belajar peserta didik. Efektivitas berkaitan dengan pencapaian hasil, oleh karena itu instrumen penilaian dikatakan efektif apabila mampu membedakan hasil belajar antara peserta didik yang berkemampuan tinggi, sedang dan rendah pada suatu kelas tertentu.

Menurut (Sudijono, 2013: 259) dengan menggunakan rumus formula C. Hoyt, dalam menentukan reliabilitas tes hendaknya kita menganggap bahwa data yang berupa skor-skor hasil tes itu kita anggap sebagai data hasil eksperimen, dimana faktor pertama adalah subyek, sedangkan faktor kedua adalah item. Masing-masing sel di sini terdiri atas 1 subyek. Selanjutnya di cari interaksi antara subyek dengan item. Teknik analisis seperti inilah yang terkenal dengan nama teknik analisis varian (=ANOVA), dengan menentukan reliabilitas tes adalah merupakan perkembangan baru dalam dunia evaluasi pendidikan.

Peneliti memodifikasi formula tersebut, bahwa data berupa skor tersebut adalah eksperimen, faktor subjek adalah subjek yang berasal dari kelompok tinggi, subjek yang berasal dari kelompok sedang, dan subjek yang berasal dari kelompok rendah. Faktor kedua adalah item. Item disini adalah rata-rata skor hasil belajar. Hasilnya adalah interaksi interaksi subyek dengan item, yaitu skor rata-rata hasil belajar dari kelompok tinggi, sedang, dan rendah menunjukkan perbedaan yang signifikan, pada uji lapangan di tiga sekolah dasar.

Berdasarkan hal tersebut, menunjukkan bahwa instrumen penilaian dinyatakan efektif karena mampu membedakan peserta didik yang pandai dan kurang pandai.

Artinya, ketika instrumen penilaian itu diujicobakan dimanapun hasilnya tetap sama.

Hal ini juga berhubungan dengan reliabilitas instrumen. (Hopkins & Antes, 1990: 5) menyatakan reliabilitas sebagai konsistensi pengamatan yang diperoleh dari pencatatan berulang baik pada satu subjek maupun sejumlah subjek.

SIMPULAN DAN SARAN

Berdasarkan hasil analisis data dan pembahasan dalam penelitian ini, maka dapat ditarik simpulan bahwa: 1) instrumen penilaian layak digunakan, baik layak secara teoritis maupun empiris. Layak secara teoritis yaitu telah memenuhi aspek konstruksi, bahasa, dan materi yang dinilai oleh para ahli dan praktisi. Layak secara empiris yaitu melalui uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan distraktor; 2) instrumen penilaian telah memenuhi karakteristik *Higher Order Thinking Skills*; dan 3) instrumen penilaian efektif digunakan sebagai alat evaluasi hasil belajar karena instrumen penilaian dapat membedakan siswa yang berkemampuan tinggi, sedang, dan rendah.

Saran bagi peserta didik diharapkan dapat termotivasi dan tertantang untuk mengerjakan setiap soal berpikir tingkat tinggi yang dikerjakan. Sehingga diharapkan peserta didik memiliki kemampuan berpikir kritis dan kreatif dalam memecahkan masalah, bukan hanya penyelesaian soal, melainkan memecahkan persoalan tantangan hidup.

Saran bagi guru, guru harus lebih kreatif dalam menciptakan instrumen penilaian berbasis *Higher Order Thinking Skills*. Berinovasi

untuk mengembangkan instrumen penilaian yang berkualitas dan efektif, sehingga guru dapat mengetahui peserta didik yang memiliki kemampuan tinggi, sedang, dan rendah.

Saran untuk peneliti selanjutnya, yaitu penelitian ini dapat dijadikan penelitian lanjutan untuk dilakukan diseminasi produk di sekolah-sekolah, sehingga produk instrumen penelitian ini dapat dikembangkan menjadi soal yang baku yang dimanfaatkan bagi banyak sekolah. Adapun produk yang dikembangkan sebaiknya tidak hanya bentuk soal pilihan ganda dan uraian.

DAFTAR PUSTAKA

- Abdullah, Ridwan Sani. 2016. *Penilaian Autentik*. Jakarta: Bumi Aksara.
- Abosalem, Yousef. 2016. Assessment Techniques and Students Higher-Order Thinking Skills. *International Journal of Secondary Education*, 4 (1): 1-11.
<http://www.sciencepublishinggroup.com/j/ijsedu>. (Diakses pada tanggal 1 Agustus 2016).
- Amnie, Erlida. 2016. Summative Assessment Design through the PjBL to Improve Students' Higher-Order Thinking Skills. *Makalah disajikan dalam International Conference in Education Research and Evaluation*, UNY.
[https://pps.uny.ac.id/sites/pps.uny.ac.id/files/Prosiding%20I-CERE%202016-ilovepdf-compressed%20\(1\).pdf](https://pps.uny.ac.id/sites/pps.uny.ac.id/files/Prosiding%20I-CERE%202016-ilovepdf-compressed%20(1).pdf). Diakses pada 1 September 2016.
- Anderson, L.W., & Krathwohl, D.R. 2001. *A Taxonomy for Learning, Teaching, and*

- Assesing: A Revision of Bloom's Taxonomy of Educatioanl Objectives*. New York: Addison Wesley Longman, Inc.
- Arifin, Zainal. 2009. *Evaluasi Pembelajaran*. PT Remaja Rosdakarya, Bandung.
- Arikunto, Suharsimi. 2012. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Budiman, Agus & Jailani. 2014. Pengembangan Instrumen Asesmen Higher Order Thinking Skill (HOTS) Pada Mata Pelajaran Matematika SMP Kelas VIII Semester 1. *Jurnal Riset Pendidikan Matematika*, 1 (2): 139-151. <http://www.sciencepublishinggroup.com/j/ijsedu>. (Diakses pada tanggal 1 Agustus 2016).
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skill in Your Classroom*, Virginia: ASCD
- Costa, A.L. 1985. *Developing Minds. A resource Book for Teaching Thinking Association for Supervision and Curriculum Development*. Alexandria, Virginia.
- Filsaime, K. D. 2008. *Menguak Rahasia Berpikir Kritis dan Kreatif*. Prestasi Pustaka, Jakarta.
- Gronlund, N.E. & Linn, R.L. 1990. *Measurement and evaluation in teaching*. Macmillan Publishing, NewYork.
- Gunawan, Imam & Anggarini Retno Palupi. 2008. Taksonomi Bloom – Revisi Ranah Kognitif: Kerangka Landasan Untuk Pembelajaran, Pengajaran, dan Penilaian. *Jurnal Pendidikan Dasar*, 16-40. <https://akhmadsudrajat.files.wordpress.com/2008/01/revisi-taksonomi-bloom.pdf>. Diakses pada 1 Januari 2016.
- Heong, Y. M.,Othman, W.D.,Md Yunos, J., Kiong, T.T., Hassan, R., & Mohamad,M.M. 2011. The Level of Marzano Higher Order Thinking Skills Among Technical Education Students. *International Journal of Social and humanity*,Vol. 1,No. 2, July 2011, 121-125.
- Hopkins, Charles D & Antes. 1990. *Classroom Meseasurement and Evaluation*. I tasca, Illinois: F.E. Peacock Publisher. Inc.
- Kemendikbud. 2013. *Peraturan Pemerintah No.32 Tahun 2013 tentang Standar Nasional Pendidikan*. Jakarta: Kementrian Pendidikan dan Kebudayaan.
- _____. 2013. *Permendikbud No. 81 Standar Penilaian Pendidikan*. Jakarta: Kementrian Pendidikan dan Kebudayaan.
- King, F. J., Goodson, L., Rohani, F. 2010. Higher Order Thinking Skills, Definition, Teaching Strategis, Assesment. *A Publication of The Educational Services Program*. www.Cala.fsu.ed. (Diakses pada tanggal 1 November 2015).
- Krulik, S & Rudnick. 1999. Innovative Taks to Improve Critical and Creative Thinking Skills. *Develovin g Mathematical Raesoning in Grades K-12*. pp.138-145.
- Mangiante, Elaine Silva. 2013. Planning Science Instruction for Critical Thinking: Two Urban Elementary Teachers'

- Responses to a State Science Assessment. *Journal Education Science*, Vol 3: 222-258. www.mdpi.com/journal/education. (Diakses pada tanggal 12 November 2015).
- Mohammed, Gulistan Saido., et.al. 2015. Higher Order Thinking Skills Among Secondary School Students in Science Learning. *The Malaysian Online Journal of Educational Science*, Vol 3 (3): 13-20. <http://www.moj-es.net/frontend/articles/pdf/v03i03/v03-i03-02.pdf>. (Diakses pada tanggal 1 Agustus 2016).
- Peter, Ebiendele Ebosele. 2012. Critical thinking: Essence for teaching mathematics and mathematics problem solving skills. *African Journal of Mathematics and Computer Science Research*, 5(3) : 39-43.
- Rochmah, Nur Lailly & Asih Widi Widiastuti. 2015. Analisis Soal Tipe Higher Order Thinking Skill (HOTS) Dalam Soal UN Kimia SMA Rayon B tahun 2012/2013. *Kaunia*. Vol 11. No. 1: 27-39
- Rofiah, Emi., Nonoh Siti Aminah dan Elvin Yusliana Ekawati. 2013. Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika Pada Siswa SMP. *Jurnal Pendidikan Fisika*, Vol 1 (2): 17-22. (Diakses pada tanggal 24 Maret 2016).
- Rubin, Jim & Manikya Rajakaruna. 2015. Teaching and Assessing Higher Order Thinking in the Mathematics Classroom with Clickers. *International Society of Educational Research*, 10 (1): 37-51. [indir/](http://iejme.com/makale_indir/) 88. (Diakses pada tanggal 1 September 2016).
- Sastrawati, Eka. et, al,. 2011. Problem Based Learning, Strategi Metakognisi, dan Kemampuan Berfikir Tingkat Tinggi Siswa. *Teno-Pedagogi*. Vol. 1 No. 2: 1-14.
- Sudijono, Anas. 2013. *Pengantar Evaluasi Pendidikan*. Jakarta: Rajawali Pers.
- Sugiyono. 2014. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: ALFABETA.
- Wardany, K., Sajidan, & Murni R. (2015). Penyusunan Instrumen Tes Higher Order Thinking Skill Pada Materi Ekosistem SMA Kelas X. *Seminar Nasional XII Pendidikan Biologi FKIP UNS 2015*, Hal 538-543.
- Wening, Sri. 2015. Asesmen Berbasis Keterampilan Berpikir Kritis dan Kreatif Untuk Menghasilkan Guru Vokasi Profesional di Masa Datang. *Jurnal Aptekindo*, 1046-1055. <http://jurnal.upi.edu/pene-litianpendidikan/view/3128/asesmen-berbasis-keterampilan-berpikir-kritis-dan-kreatifuntuk-menghasilkan-guru-vokasi-profesional-di-masa-datang.html>. Diakses pada tanggal 3 November 2015.