

PENGEMBANGAN INSTRUMEN PENILAIAN BERBASIS HOTS (*HIGH ORDER THINKING SKILL*) MATA PELAJARAN SISTEM KOMPUTER KELAS X SMK NEGERI 2 BANGKALAN

Ni'mah Nadliroh¹, Sigit Dwi Saputro², Ariesta Kartika Sari³, Puji Rahayu Ningsih⁴,
Etistika Yuni Wijaya⁵

Universitas Trunojoyo Madura

*nikmahnadliroh@gmail.com*¹, *sigitd.saputra@trunojoyo.ac.id*², *ariestakartika@trunojoyo.ac.id*³,
*pujirahayuningasih86@gmail.com*⁴, *etistikaw@gmail.com*⁵

Abstrak

Penelitian ini bertujuan untuk menghasilkan instrumen penilaian berbasis HOTS (High Order Thinking Skill) berbentuk uraian yang layak berdasarkan kriteria kualitas instrumen penilaian, diantaranya yaitu validitas, reliabilitas, taraf kesukaran, dan daya pembeda serta memenuhi validasi dari segi konten, konstruksi, dan bahasa. Penelitian ini menggunakan jenis penelitian pengembangan dengan model penelitian formative research yang terdiri dari tahap preliminary, self evaluation, prototyping, dan field test. Instrumen dalam penelitian ini berupa angket validasi expert review, angket respon siswa, dan tes. Hasil analisis data dari tahap prototyping bagian expert review dan one-to-one memperoleh soal HOTS dengan kategori "sangat valid", namun dilakukan perubahan berdasarkan saran dan komentar dari keduanya dan menghasilkan prototype 2. Pada tahap uji coba one-to-one dilakukan revisi kembali pada penggunaan bahasa sehingga menghasilkan prototype 3. Kemudian berdasarkan hasil dari tahap field test, prototype 3 diujikan dan dianalisis validitas menghasilkan 3 soal tidak valid dan 7 soal valid dan memiliki nilai reliabilitas 0.83 dengan interpretasi sangat tinggi. Pada analisis taraf kesukaran menghasilkan 6 soal sedang dan 4 soal mudah, sedangkan pada analisis daya pembeda menghasilkan 5 soal baik dan 5 soal kurang baik.

Kata kunci : instrumen penilaian, high order thinking skill, formative research

Abstract

This study aimed to produce HOTS-based assessment instruments (High Order Thinking Skill) in the form of appropriate descriptions based on the quality criteria of assessment instruments, including validity, reliability, degree of difficulty, and differentiation as well as fulfilling validation in terms of content, constructs, and languages. This study used a type of development research with a formative research model consisting of the preliminary stage, self-evaluation, prototyping, and field tests. The instruments in this study were questionnaire validation expert reviews, student response questionnaires, and tests. Results from the prototyping stage, the expert review, and one-to-one section received HOTS questions with the category "very valid", but made changes based on suggestions and comments from both of them and produced a prototype 2. At the one-to-one trial phase revisions were made to the use of language so that it produces a prototype 3. Then based on the results of the field test stage, prototype 3 was tested and analyzed validity produces 3 invalid questions and 7 valid questions and has a reliability value of 0.83 with very high interpretation. On the analysis, the difficulty level produces 6 intermediate questions and 4 easy questions, while the distinguishing analysis produces 5 good questions and 5 poor questions.

Keyword: assessment instruments, high order thinking skills, formative research

PENDAHULUAN

Keberhasilan pelaksanaan pendidikan yang tujuan utamanya meningkatkan sumber daya manusia dipengaruhi oleh berbagai faktor. Salah satu faktor yang ikut mempengaruhi adalah kemampuan guru

dalam melakukan dan memanfaatkan penilaian, evaluasi proses, dan hasil belajar. Kemampuan tersebut sangat diperlukan untuk mengetahui tercapai tidaknya tujuan pembelajaran yang sudah ditetapkan dalam kurikulum. Selain itu, kemampuan tersebut

juga dapat digunakan untuk memperbaiki dan meningkatkan kualitas pembelajaran yang telah dilakukan guru. Berkaitan dengan hal tersebut pemerintah telah memberikan pedoman yaitu dengan mengeluarkan [1] tentang Standar Kualifikasi Akademik dan Kompetensi Guru mata pelajaran bahwa kompetensi guru mata pelajaran antara lain adalah mengembangkan instrumen penilaian dan evaluasi proses dan hasil belajar.

Evaluasi menurut [2] merupakan proses penilaian pertumbuhan siswa dalam proses belajar mengajar. Sesuai dengan [3] Permendiknas Nomor 20 Tahun 2007 tentang Standar Penilaian Pendidikan bahwa Penilaian pendidikan adalah proses pengumpulan dan pengolahan informasi untuk menentukan pencapaian hasil belajar peserta didik. Hasil penilaian juga dapat memberikan umpan balik yang konstruktif bagi guru dan dapat memberikan motivasi kepada peserta didik untuk berprestasi lebih baik. Bahkan penilaian dapat mempengaruhi perilaku belajar peserta didik karena mereka cenderung mengarahkan kegiatan belajarnya menuju penilaian yang dilakukan oleh guru.

Dalam melaksanakan evaluasi proses dan hasil belajar tentu tidak lepas dari kegiatan penilaian. Pada proses pelaksanaan penilaian guru memerlukan instrumen dalam bentuk soal-soal, baik untuk menguji kemampuan kognitif (pengetahuan), afektif (sikap), dan psikomotorik (keterampilan). Kualitas instrumen penilaian berpengaruh terhadap keakuratan pencapaian hasil belajar peserta didik. Oleh karena itu kedudukan instrumen penilaian hasil belajar sangat penting dalam pengambilan keputusan guru dan sekolah terkait pencapaian hasil belajar peserta didik.

Peraturan Menteri Pendidikan dan Kebudayaan Nomor 65 Tahun 2013 tentang Standar Proses menyatakan bahwa ranah kognitif (pengetahuan) dapat diperoleh melalui aktifitas mengingat, memahami, menerapkan, menganalisis, mengevaluasi, serta mencipta. Aktifitas yang tercantum dalam standar proses pendidikan tersebut merupakan enam tingkatan proses kognitif yang terdapat dalam taksonomi Bloom revisi. Berdasarkan taksonomi Bloom, kemampuan berpikir peserta didik diklasifikasikan menjadi tiga jenjang, yaitu berpikir tingkat

tinggi *Higher Order Thinking Skill* (HOTS), berpikir tingkat menengah atau *Middle Order Thinking Skill* (MOTS) dan berpikir tingkat rendah atau *Lower Order Thinking Skill* (LOTS). Kemampuan berpikir tingkat tinggi (HOTS) bukan sekadar kemampuan mengingat atau menghafal, melainkan kemampuan dalam menganalisis, berpikir kritis dan kreatif. Sesuai dengan pernyataan [3] bahwa HOTS merupakan cara berpikir yang lebih tinggi daripada menghafalkan fakta, mengemukakan fakta, atau menerapkan peraturan, rumus, dan prosedur.

Berkaitan dengan kemampuan berpikir siswa, Indonesia secara ilmiah dianggap masih rendah dibandingkan siswa negara lain. Hal ini ditunjukkan oleh salah satu studi internasional yaitu tes PISA (*Programme International for Student Assessment*) (2015) yang menunjukkan bahwa Indonesia memperoleh ranking 62 dari 70 negara [4]. Kondisi ini relevan dengan hasil penelitian dari TIMSS (*Trends In Mathematics and Science Study*) bahwa Indonesia menempati ranking 45 dari 48 negara dengan poin 397 pada domain sains dan ranking 45 dari 50 negara dengan skor 397 pada domain matematika. Dari hasil kedua penelitian tersebut dapat dikatakan bahwa Indonesia masih jauh tertinggal dari beberapa negara, dimana peserta didik Indonesia masih sangat kurang dalam menjawab soal yang mengukur kemampuan berpikir tingkat tinggi yang merupakan karakteristik dari soal-soal PISA dan TIMSS.

Rendahnya kemampuan berpikir tingkat tinggi peserta didik tersebut tentunya disebabkan oleh beberapa faktor. Salah satu faktor penyebabnya antara lain karena peserta didik di Indonesia kurang terlatih dalam menyelesaikan soal-soal yang berupa penalaran, argumentasi dan kreatifitas dalam menyelesaikannya. Hal tersebut sejalan dengan hasil wawancara yang dilakukan peneliti kepada salah satu guru mata pelajaran Sistem Komputer SMK Negeri 2 Bangkalan yakni Ibu Rindah S.T yang mengatakan bahwa soal untuk penilaian atau evaluasi hasil belajar rata-rata masih mengukur kemampuan berpikir pada tingkat rendah, yaitu level kognitif memahami dan mengingat. Soal-soal ujian diperoleh dari mengkombinasikan antara soal yang dibuat

sendiri dengan soal yang terdapat di dalam buku paket. Akibatnya peserta didik hanya menghafal materi yang ada di buku sehingga ketika peserta didik dihadapkan pada soal-soal dengan tingkat berpikir yang lebih tinggi mereka akan kesulitan dalam menyelesaikannya. Selain itu, peserta didik sering tidak dapat menggunakan pengetahuan yang mereka miliki dalam kehidupan sehari-hari, bahkan tidak dapat menggunakan keterampilan menyelesaikan soal apabila diberikan soal yang sedikit berbeda dari apa yang telah dipelajari. Dari hasil wawancara tersebut dapat dipahami bahwa di kelas X RPL pada mata pelajaran sistem komputer belum digunakan dan tidak ada instrumen penilaian berbasis HOTS yang dapat dijadikan sebagai acuan untuk menyusun soal berbasis HOTS.

Sementara itu, permasalahan di sekolah yang dihadapi oleh pendidik adalah kemampuan dalam mengembangkan soal HOTS yang masih kurang dan belum tersedianya instrumen penilaian yang didesain khusus untuk melatih HOTS. Sebagaimana hasil wawancara peneliti kepada salah satu staf kurikulum SMK Negeri 2 Bangkalan yaitu Ibu Ida yang mengatakan bahwa beberapa guru masih membuat soal pada tingkat *Low Order Thinking Skill* (LOTS) karena tidak semua guru paham dalam membuat soal HOTS. Oleh karena itu terdapat seorang guru SMK Negeri 2 Bangkalan yang dipanggil untuk mengikuti pelatihan pembelajaran dan membuat soal HOTS di tingkat Jawa Timur. Pelatihan tersebut dilakukan secara descending, dari lingkup besar ke lingkup kecil yakni kepada instruktur di tingkat nasional lalu ke tingkat provinsi.

Melihat ketidaksesuaian antara yang seharusnya dengan yang ada di lapangan berkaitan dengan penilaian hasil belajar, rendahnya kemampuan berpikir tingkat tinggi siswa dan instrumen penilaian yang digunakan, maka solusinya adalah dengan melatih peserta didik melalui soal HOTS agar terbiasa dan mampu mengembangkan kemampuan berpikir yang ada di dalam potensi dirinya masing-masing, supaya mereka tidak hanya menghafalkan rumus-rumus dalam menyelesaikan soal-soal. Sementara dari pihak pendidik sendiri perlu

kiranya mengembangkan instrumen penilaian dengan karakteristik model PISA dan TIMSS untuk melatih dan meningkatkan kemampuan berpikir tingkat tinggi sehingga memudahkan peserta didik dalam memecahkan masalah dan peserta didik dapat lebih terlatih dalam mengerjakan soal-soal berbasis HOTS. Oleh sebab itu peneliti memutuskan melakukan penelitian mengenai "Pengembangan Instrumen Penilaian Berbasis HOTS (*High Order Thinking Skill*) Mata Pelajaran Sistem Komputer Kelas X SMK Negeri 2 Bangkalan".

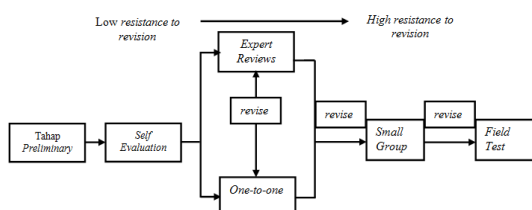
Tujuan penelitian ini adalah untuk menghasilkan instrumen penilaian berbasis HOTS pada mata pelajaran Sistem Komputer Kelas X SMK Negeri 2 Bangkalan yang layak digunakan serta memiliki validitas, reliabilitas, taraf kesukaran dan daya pembeda soal yang baik. Supaya dapat dimanfaatkan sebagai acuan dalam membuat instrument penilaian berbasis HOTS.

METODE

Metode Penelitian ini dilaksanakan di SMK Negeri 2 Bangkalan semester genap tahun pelajaran 2018/2019. Subjek penelitian adalah siswa kelas X RPL 1 SMK Negeri 2 Bangkalan yang berjumlah 23 orang terdiri atas 10 orang laki-laki dan 13 orang perempuan.

Penelitian ini menggunakan metode penelitian pengembangan atau *research and development* (R&D). Menurut Sugiyono [5] metode penelitian dan pengembangan atau dalam bahasa Inggrisnya *Research and Development* adalah metode penelitian yang digunakan untuk menghasilkan produk tertentu, dan menguji keefektifan produk tersebut. Penelitian ini merujuk pada model pengembangan tipe formative research Tessmer (1993). Peneliti memilih model ini karena sesuai dengan tujuan peneliti yaitu untuk menghasilkan produk instrumen penilaian berbasis HOTS yang layak digunakan serta untuk memecahkan masalah yang dialami guru dan siswa dalam pembelajaran. Selain itu peneliti dapat menilai kelebihan dan kekurangan produk dalam tahap-tahap pengembangan untuk tujuan merevisi agar dapat memperbaiki keefektifan produk. Berikut tahapan model

penelitian pengembangan menurut Tessmer dalam penelitian ini:



Gambar 1. Tahapan Model Formative Research[6]

Pada tahap *Preliminary* peneliti melakukan persiapan yakni menentukan tempat dan subjek penelitian serta pengkajian terhadap beberapa sumber referensi yang berkaitan dengan penelitian.

Dimana informasi ini didapat dengan cara melakukan wawancara kepada guru pengampu mata pelajaran Sistem Komputer di SMK Negeri 2 Bangkalan serta mengadakan persiapan-persiapan lain, seperti mengatur jadwal penelitian dan menggali informasi terhadap staf kurikulum di sekolah.

Tahap *Self Evaluation* dilakukan dengan tahapan: analisis dan desain. Pada tahap analisis peneliti melakukan analisis kurikulum, analisis siswa, dan analisis materi. Pada tahap desain peneliti mendesain kisi-kisi soal, soal berbasis HOTS, kunci jawaban dan pedoman penskoran sesuai dengan kebutuhan.

Pada tahap *Prototyping* produk yang telah dibuat dievaluasi. Tahap ini diuji cobakan kedalam tiga kelompok, yaitu:

Tahap *expert review* dilakukan untuk mengetahui penilaian pakar terhadap produk instrumen penilaian berbasis HOTS dari segi konten, konstruk, dan Bahasa. Para ahli diminta tanggapan dan saran pada lembar validasi dan menyatakan bahwa instrumen penilaian berbasis HOTS valid dari segi konten, konstruk, dan bahasa. Tahap ini biasa disebut uji validitas.

Tahap *one-to-one* dilakukan bersamaan dengan tahap *expert review*. Peneliti meminta 3 orang peserta didik non subjek penelitian sebagai *tester* untuk menjawab soal kemudian diminta untuk memberikan komentar terhadap soal berbasis HOTS. Tiga peserta didik terdiri atas siswa

dengan kemampuan tinggi, sedang, dan rendah. Komentar yang diperoleh digunakan untuk merevisi soal dan menghasilkan *prototype 2*.

Prototype 2 kemudian diujicobakan pada *small group*, terdiri atas 6 orang peserta didik dengan kemampuan tinggi, sedang, dan rendah masing-masing 2 orang. Peserta didik diminta menjawab soal dan memberikan komentar terkait soal berbasis HOTS. Dari komentar tersebut selanjutnya produk direvisi dan menghasilkan *prototype 3*.

Pada tahap *field test prototype 3* diujikan kepada subjek penelitian yakni siswa kelas X RPL 2 SMK Negeri 2 Bangkalan untuk mengetahui kelayakan dan kualitas instrumen penilaian yang telah disusun.

Instrumen pengumpulan data dalam penelitian ini meliputi: (1) angket yang terdiri atas angket praktisi, angket ahli materi, angket ahli bahasa dan angket respon siswa uji coba *small group* dan (2) tes.

Sedangkan teknik analisis data yang digunakan dalam penelitian ini adalah: (1) analisis deskriptif kualitatif untuk mengolah data berupa saran dan kritik dari tahap *expert review* dan uji coba *small group* yang terdapat pada lembar angket dan (2) analisis deskriptif kuantitatif meliputi analisis data angka pada lembar angket yang diberikan kepada *expert review* dan *small group* serta analisis setelah melalui tahap *field test* yang meliputi:

Uji validitas dalam penelitian ini menggunakan rumus korelasi *product-moment* [6]

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}}$$

Dari hasil perhitungan validitas selanjutnya dikonsultasikan pada tabel harga *r* product moment sebagaimana tabel 1.

Tabel 1. Tabel Kategori *r* hitung [7]

Nilai <i>r</i>	Kategori
0,80-1,00	Sangat Tinggi
0,60-0,79	Tinggi
0,40-0,39	Cukup
0,20-0,38	Rendah

0,00-0,19	Sangat Rendah
-----------	---------------

Jika harga r lebih kecil dari harga kritik dalam tabel maka korelasi tersebut tidak signifikan. Begitu juga arti sebaliknya (Arikunto, 2015: 89). Jadi, dapat dipahami bahwa jika $r_{xy} < r_{tabel}$ maka soal dinyatakan tidak valid.

Uji reliabilitas soal bentuk uraian menggunakan rumus menurut Arikunto (2006) dalam Ratnawulan dan Rusdiana, (2015: 175):

$$r_{11} = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right]$$

Selanjutnya nilai r yang diperoleh dari hasil perhitungan dikonsultasikan dengan harga r tabel dengan $\alpha = 0.05$ dan $dk = N-2$ ($N =$ banyaknya siswa). Apabila $r_{hit} > r_{tab}$ maka instrumen dinyatakan reliabel [7]. Untuk mengetahui tinggi rendahnya reliabilitas instrumen digunakan kategori berikut sebagaimana tabel 2 [7].

Tabel 2. Kategori Koefisien Reliabilitas (r_{11}) [7]

Nilai r_{11}	Kategori
0,800 – 1,00	Sangat Tinggi
0,600 – 0,799	Tinggi
0,400 – 0,599	Cukup
0,200 – 0,399	Rendah
0,000 – 0,199	Sangat Rendah

Rumus yang digunakan untuk menghitung tingkat kesukaran soal dalam penelitian ini adalah sebagai berikut [7]:

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

Kemudian untuk menafsirkan tingkat kesukaran soalnya dapat digunakan kriteria sebagai sebagaimana tabel 3.:

Tabel 3. Kategori Taraf Kesukaran [8]

Nilai Taraf Kesukaran	Kategori
0% - 27%	Mudah
28% - 72%	Sedang
73% - 100%	Sukar

Untuk menguji daya pembeda, digunakan rumus sebagai berikut [8]:

$$DP = \frac{XKA - XKB}{Skor Maks}$$

Kemudian membandingkan daya pembeda dengan kriteria sebagaimana tabel 4.

Tabel 4. Kategori Daya Pembeda[8]

Nilai D	Kategori
0,40 – 1,00	Sangat Baik
0,30 – 0,39	Baik
0,20 – 0,29	Cukup
0,19 – 0,00	Kurang Baik

HASIL

Prototype awal yang telah disusun diujicobakan kepada para ahli sesuai dengan tahapan-tahapan model penelitian. *Expert review* dalam penelitian ini terdiri atas 2 orang praktisi, 2 orang ahli, dan 1 orang ahli bahasa. Hasil analisis data dari lembar angket yang diberikan pada praktisi 1 adalah 90%, dan hasil dari praktisi adalah 96%. Sehingga dari kedua praktisi tersebut memperoleh nilai rata-rata 93%. Setelah dikonversikan pada tabel kategori validitas ahli diketahui bahwa instrumen penilaian berbasis HOTS menurut penilaian praktisi berada dalam kategori "sangat valid". Namun dalam lembar angket terdapat beberapa saran yang diberikan sehingga peneliti melakukan perbaikan berdasarkan saran tersebut.

Pada tahap validasi ahli materi, hasil yang diperoleh dari ahli materi 1 adalah 97.5%, sedangkan ahli materi 2 memperoleh nilai 92.5%. Sehingga didapat nilai rata-rata dari kedua ahli materi sebesar 95%. Kemudian hasil analisis data dari angket yang diberikan kepada ahli bahasa memperoleh nilai 96%. Hal ini menunjukkan bahwa instrumen penilaian berbasis HOTS "sangat valid" menurut penilaian ahli materi dan ahli bahasa.

Setelah melalui tahap validasi dari para ahli, bersamaan dengan ini dilakukan uji coba perorangan (*one-to-one*). Data yang diperoleh dari tahap ini berupa data kualitatif berbentuk saran dan kritik terhadap soal berbasis HOTS. Berikut ini disajikan tabel 5 yang menggambarkan saran dan kritik dari siswa uji coba *one-to-one* serta keputusan revisi yang dilakukan oleh peneliti.

Tabel 5. Keputusan Revisi Soal HOTS Uji Coba One-to-one

Sebelum revisi	Sesudah revisi
- Kalimat perintahnya sulit dimengerti di soal nomor 3	- Kalimat perintah di soal nomor 3 sudah direvisi
- Terlalu banyak kalimat terutama pada soal nomor 6	- Penggunaan kalimat pada soal nomor 6 sudah dikurangi
- Soal nomor 10 gambarnya kurang jelas dan sedikit membingungkan bagaimana membedakannya	- Gambar di soal nomor 10 sudah direvisi

Dari tabel 5 tersebut peneliti melakukan beberapa perbaikan dan menghasilkan *prototype 2* yang dapat dilanjutkan pada tahap uji coba small group.

Pada tahap small group, *prototype 2* diuji cobakan kepada 6 orang siswa. Siswa diminta mengerjakan soal dan memberikan saran dan kritik. Dari saran dan kritik tersebut peneliti melakukan perbaikan. Adapun saran dan kritik serta keputusan revisi yang dilakukan oleh peneliti dapat dilihat pada tabel 6.

Tabel 6. Keputusan Revisi Soal HOTS Uji Coba Small Group

Sebelum revisi	Sesudah revisi
- Soalnya tidak rumit, tetapi kalimat perintahnya saja yang rumit terutama di soal nomor 5	- Kalimat perintah soal nomor 5 sudah direvisi
- Sebaiknya kalimatnya jangan terlalu banyak dan rumit, khususnya soal nomor 6	- Kalimat atau bahasa soal nomor 6 sudah diperbaiki dan dikurangi
- Soalnya rumit karena bahasanya yang agak sulit, khususnya soal nomor 6	- Kalimat perintah soal nomor 10 dan soal yang sejenis yakni soal nomor 7 sudah direvisi
- Kalimat perintahnya agak rumit dan membingungkan terutama pada soal nomor 10	

Dari hasil perbaikan yang dilakukan oleh peneliti pada tahap ini selanjutnya menghasilkan *prototype 3*, dimana *prototype 3* merupakan *prototype final* dari penelitian ini yang kemudian dilanjutkan pada tahap uji lapangan (*field test*).

Tahap field test merupakan tahap akhir dari penelitian ini. Pada tahap ini *prototype 3* diujikan kepada subjek penelitian yakni siswa kelas X RPL 2 SMK Negeri 2 Bangkalan yang berjumlah 23 orang. Dari hasil jawaban siswa inilah kemudian dilakukan analisis kualitas soal dan butir soal meliputi analisis validitas, reliabilitas, taraf kesukaran, dan daya pembeda.

Hasil analisis validitas soal dapat dilihat pada tabel 7.

Tabel 7. Analisis Validitas Soal

Soal	r_{xy}	r_{tabel}	Kategori	Kategori Validitas
1	0.18	0.433	Tidak Valid	Sangat Rendah
2	0.26	0.433	Tidak Valid	Rendah
3	0.70	0.433	Valid	Tinggi
4	0.72	0.433	Valid	Tinggi
5	0.77	0.433	Valid	Tinggi
6	0.61	0.433	Valid	Tinggi
7	0.77	0.433	Valid	Tinggi
8	0.30	0.433	Tidak Valid	Rendah
9	0.80	0.433	Valid	Sangat Tinggi
10	0.50	0.433	Valid	Cukup

Berdasarkan tabel 7 dapat diketahui bahwa ada beberapa soal yang tidak valid, yaitu soal nomor 1, nomor 2, dan nomor 8. Dan soal yang valid ada 7 soal yakni soal nomor 3, 4, 5, 6, 7, 9, dan 10.

Setelah melakukan analisis validitas soal, langkah selanjutnya adalah menganalisis reliabilitas soal. Analisis reliabilitas soal hanya dilakukan pada soal yang sudah valid, hasil analisis reliabilitas soal dapat dilihat pada tabel 8 berikut.

Tabel 8. Analisis Reliabilitas Soal

Nomor Soal	σ_b^2	σ_t^2
3	0.824	31.28166352
4	1.558	31.28166352
5	1.040	31.28166352
6	1.758	31.28166352
7	1.474	31.28166352
9	1.497	31.28166352
10	0.802	31.28166352
Jumlah varians butir	8.953	-

Berdasarkan tabel 8 dapat diketahui bahwa jumlah varians butir soal adalah 8.953. Selanjutnya reliabilitas soal dapat diketahui dengan rumus:

$$r_{11} = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right]$$

Sehingga diperoleh:

$$r_{11} = \left[\frac{7}{7-1} \right] \left[1 - \frac{8.953}{31.28166352} \right] = 0.83$$

Dari hasil perhitungan reliabilitas soal diperoleh kesimpulan bahwa soal berbasis HOTS reliabel. Hal ini dibuktikan dengan hasil nilai koefisien reliabilitas yakni $r_{11} = 0.83 > r_{tabel} = 0.43$. Kemudian nilai koefisien reliabilitas tersebut dikonversikan dengan tabel 2 dan didapatkan reliabilitas soal berada dalam kategori "sangat tinggi". Sehingga hal ini menunjukkan tidak ada revisi instrumen penilaian menurut uji reliabilitas.

Analisis taraf kesukaran soal dilakukan untuk mengetahui soal yang mudah, sedang, dan sukar. Taraf kesukaran soal dapat diketahui dari skor siswa pada tiap butir soal. Hasil analisis taraf kesukaran soal instrumen penilaian berbasis HOTS disajikan pada tabel 9.

Tabel 9. Analisis Taraf Kesukaran Soal

Nomor Soal	Tingkat Kesukaran (%)	Kategori
1	22.7%	Mudah
2	4.5%	Mudah
3	31.8%	Sedang
4	36.4%	Sedang
5	31.8%	Sedang
6	22.7%	Mudah
7	54.5%	Sedang
8	50.0%	Sedang
9	22.7%	Mudah
10	63.6%	Sedang

Berdasarkan tabel 9 dapat diketahui bahwa soal nomor 1, 2, 6, dan 9 memiliki taraf kesukaran dengan kategori "mudah", artinya banyak siswa yang menjawab benar pada soal-soal tersebut. Sedangkan soal nomor 3, 4, 5, 7, 8, dan 10 memiliki taraf kesukaran "sedang". Hal ini menunjukkan bahwa siswa yang menjawab benar dan siswa yang menjawab salah seimbang.

Analisis daya pembeda soal diperoleh dari data hasil pekerjaan siswa pada uji coba lapangan (*field test*). Hasil analisis daya pembeda soal HOTS dijabarkan pada tabel 10.

Tabel 10. Analisis Daya Pembeda Soal

Nomor Soal	Daya Pembeda	Kategori
1	0.18	Kurang Baik
2	0.05	Kurang Baik
3	0.25	Cukup
4	0.31	Baik
5	0.25	Cukup
6	0.13	Kurang Baik
7	0.20	Cukup
8	0.04	Kurang Baik
9	0.27	Cukup
10	0.13	Kurang Baik

Dari tabel 10 diketahui bahwa soal nomor 4 memiliki daya pembeda dengan kategori "baik", artinya soal tersebut baik untuk membedakan siswa yang menguasai materi dan yang kurang menguasai materi. Soal nomor 3, 5, 7, dan 9 memiliki daya pembeda "cukup" yang berarti bahwa soal tersebut cukup baik untuk membedakan siswa yang menguasai materi dan yang kurang menguasai materi. Sedangkan soal nomor 1, 2, 6, 8, dan 10 memiliki daya pembeda dengan kategori "kurang baik" yang berarti bahwa soal tersebut kurang baik untuk membedakan siswa yang menguasai materi dan yang kurang menguasai materi sehingga soal tersebut tidak digunakan dalam membedakan kemampuan siswa yang menguasai dan yang kurang menguasai materi.

SIMPULAN

Hasil penelitian yang dilakukan menghasilkan produk instrumen penilaian berbasis HOTS berupa 10 butir soal berbentuk uraian yang telah melalui uji validasi ahli dari segi konten, konstruk, dan bahasa serta memenuhi kriteria kualitas soal dan butir soal yang baik meliputi validitas, reliabilitas, taraf kesukaran, dan daya pembeda. Hal ini menunjukkan bahwa instrumen penilaian berbasis HOTS layak digunakan, dibuktikan dengan perhitungan yang dilakukan pada tahap expert review.

Dari validator telaah soal HOTS 1 diperoleh hasil 90%, validator telaah soal HOTS 2 diperoleh hasil 96%, ahli materi 1 diperoleh hasil 97.5%, ahli materi 2 diperoleh 92.5%, dan ahli bahasa diperoleh hasil 96%. Seluruh hasil presentase tersebut mendapat kesimpulan bahwa validitas instrumen produk awal pengembangan berada dalam kategori "sangat valid", atau dapat digunakan tanpa revisi. Namun peneliti merevisi produk instrumen penilaian berdasarkan saran dan komentar pada lembar angket yang diberikan.

Perhitungan yang dilakukan pada tahap uji coba small group diperoleh hasil 74.6% yang diketahui bahwa validitas instrumen berada dalam kategori "valid", atau dapat digunakan sesuai dengan saran perbaikan. Kemudian tahap uji coba field test dilakukan analisis kualitas instrumen penilaian dan menghasilkan 7 soal dengan kategori valid dan 3 soal tidak valid, kemudian dari soal yang valid tersebut diperoleh nilai reliabilitas sebesar 0.83 dengan interpretasi "sangat tinggi". Pada analisis taraf kesukaran memperoleh 6 soal kategori "sedang" dan 4 soal kategori "mudah". Sedangkan pada analisis daya pembeda menghasilkan 5 soal kategori "kurang baik", 1 soal kategori "baik" dan 4 soal kategori "cukup".

Adapun saran dalam penelitian ini adalah: (a) dalam menyusun soal berbasis HOTS, sebaiknya gunakan kalimat yang mudah dipahami siswa dan bentuk soal lebih variatif supaya tidak monoton; (b) untuk pihak-pihak yang ingin mengembangkan produk lebih lanjut diharapkan dapat mengembangkan soal berbentuk objektif atau pilihan ganda; (c) soal HOTS umumnya memiliki tingkat kesukaran yang lebih tinggi daripada soal evaluasi biasa, sehingga disarankan dalam penyusunan soal berbasis HOTS hendaknya memuat stimulus yang memerlukan daya nalar dan analisa lebih tinggi; dan (d) untuk mengetahui lebih lanjut mengenai baik atau tidaknya instrumen penilaian yang telah dikembangkan, maka disarankan pada peneliti selanjutnya agar dapat menguji cobakan soal pada subjek uji coba yang lebih luas.

DAFTAR RUJUKAN

- [1] Depdiknas, "Permendiknas Nomor 16 Tahun 2007 tentang Standar Kualifikasi Akademik dan Kompetensi Guru Mata Pelajaran." Depdiknas, 2007.
- [2] Sukardi, *Evaluasi Pendidikan: Prinsip & Operasionalnya*. Jakarta: Bumi Aksara, 2009.
- [3] A. Nugroho, *Higher Order Thinking Skills (Kemampuan Berpikir Tingkat Tinggi: Konsep, Pembelajaran, Penilaian dan Soal-soal*. Jakarta: Gramedia, 2018.
- [4] OECD, "PISA (Programme International for Student Assesment): Result in Focus." 2018.
- [5] Sugiyono, *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta, 2017.
- [6] S. Arikunto, *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara, 2015.
- [7] E. Ratnawulan dan Rusdiana, *Evaluasi Pembelajaran*. Bandung: Pustaka Setia, 2015.
- [8] Z. Arifin, *Evaluasi Pembelajaran: Prinsip, Teknik, dan Prosedur*. Bandung: PT Remaja Rosdakarya, 2013.