

## PENGEMBANGAN INSTRUMEN ASESMEN *HIGHER ORDER THINKING SKILLS* (HOTS) PADA MATA PELAJARAN BAHASA INDONESIA SMA DAN SMK

### *The Development of Higher Order Thinking Skills (HOTS) Assessment Instruments in Indonesian Study of High School and Vocational High School*

Sutami<sup>1\*</sup>, Yusak Hudyono<sup>2</sup>, M. Ilyas<sup>3</sup>

<sup>1</sup> Magister Pendidikan Bahasa dan Sastra Indonesia, FKIP, Universitas Mulawarman

<sup>2,3</sup> Universitas Mulawarman

Pos-el Korespondensi: [tamisutami99@gmail.com](mailto:tamisutami99@gmail.com)

**Abstract:** *This research aims to produce a valid and reliable Indonesian language assessment instrument in form of HOTS test items and it describes the quality of HOTS test items to measure HOTS skill for the tenth grade of SMA and SMK students. This study was a research and development study adapted from Borg & Gall's development model, including the following steps: research and information collection, planning, early product development, limited try out, revising the early product, field try out, and revising the final product. The research's result shows that the HOTS assessment instrument in the form of HOTS test consists of 40 multiple choice items and 5 essay test items. Based on the judgment of the materials, construction, and language was valid and appropriate to be used. The reliability coefficients were 0.88 for the multiple-choice items, and 0.79 for essays. The multiple-choice items have the average difficulty 0.57 (average), the average of item discrimination 0.44 (good), and the distractors function well. The essay items have the average of item difficulty 0.60 (average) and the average of item discrimination 0.45 (good).*

**Keywords:** *development, assessment instrument, higher order thinking skills, Indonesian language in senior high school and vocational high school*

**Abstrak:** Penelitian ini bertujuan untuk menghasilkan instrumen asesmen bahasa Indonesia berupa soal tes *HOTS* yang valid dan reliabel dan mendeskripsikan kualitas soal tes *HOTS* untuk mengukur keterampilan berpikir tingkat tinggi pada peserta didik SMA dan SMK kelas X semester 1. Penelitian ini merupakan penelitian pengembangan yang diadaptasi dari model pengembangan Borg & Gall, yang meliputi langkah-langkah sebagai berikut: penelitian pendahuluan, perencanaan, pengembangan produk awal, revisi produk awal, uji coba terbatas, uji coba lapangan, dan revisi produk akhir. Hasil penelitian menunjukkan bahwa instrumen asesmen *HOTS* berupa soal tes *HOTS* yang terdiri dari 40 butir soal pilihan ganda dan 5 butir soal uraian dari aspek materi, konstruksi, dan bahasa dinyatakan valid dan layak digunakan. Instrumen tersebut mempunyai koefisien reliabilitas sebesar 0,88 kategori sangat tinggi untuk soal pilihan ganda dan sebesar 0,79 kategori tinggi untuk soal uraian. Soal pilihan ganda memiliki rata-rata tingkat kesukaran 0,57 (sedang), rata-rata daya pembeda 0,44 (baik), dan semua pengecoh berfungsi baik. Soal uraian memiliki rata-rata tingkat kesukaran 0,60 (sedang) dengan rata-rata daya pembeda 0,45 (baik).

**Kata kunci:** pengembangan, instrumen asesmen, *higher order thinking skills*, bahasa Indonesia SMA dan SMK

## A. PENDAHULUAN

Penilaian merupakan proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar peserta didik. Berdasarkan Permendikbud nomor 23 tahun 2016 tentang Standar Penilaian, penilaian hasil belajar peserta didik meliputi aspek sikap, pengetahuan, dan keterampilan. Penilaian hasil belajar oleh pendidik ini bertujuan untuk memantau dan mengevaluasi proses, kemajuan belajar, dan perbaikan hasil belajar peserta didik secara berkesinambungan.

Sebagai bagian dari evaluasi, Indonesia melakukan *benchmark* internasional dengan mengikuti *Trends in International Mathematics and Science Study (TIMSS)* dan *Programme for International Student Assessment (PISA)*. Hasil *TIMSS* tahun 2015 untuk kelas IV Sekolah Dasar, Indonesia mendapatkan rata-rata nilai 397 dan menempati peringkat 4 terbawah dari 43 negara yang mengikuti *TIMSS* (Sumber: *TIMSS* 2015 International Database). Sementara untuk *PISA* tahun 2015, Indonesia mendapatkan rata-rata nilai 397 untuk membaca (peringkat terakhir), 403 untuk sains (peringkat ketiga dari bawah), dan 386 untuk matematika (peringkat kedua dari bawah) dari 72 negara yang mengikuti (Sumber: *OECD, PISA* 2015 Database). Pada umumnya kemampuan peserta didik Indonesia sangat rendah dalam: (1) memahami informasi yang kompleks; (2) teori, analisis, dan pemecahan masalah; (3) pemakaian alat, prosedur dan pemecahan masalah; dan (4) melakukan investigasi (Kemdikbud, 2018).

Hasil pengukuran capaian siswa berdasar UN ternyata selaras dengan capaian *PISA* maupun *TIMSS*. Hasil UN tahun 2018 menunjukkan bahwa peserta didik masih lemah dalam keterampilan berpikir tingkat tinggi (*higher order thinking skill*) seperti menalar, menganalisis, dan mengevaluasi. Berdasarkan kenyataan-

kenyataan di atas, maka perlu adanya perubahan sistem dalam pembelajaran dan penilaian. Penilaian yang dikembangkan oleh guru diharapkan dapat mendorong peningkatan kemampuan berpikir tingkat tinggi, meningkatkan kreativitas, dan membangun kemandirian peserta didik untuk menyelesaikan masalah (Kemdikbud, 2018).

Penyempurnaan kurikulum 2013 antara lain pada standar isi diperkaya dengan kebutuhan peserta didik untuk berpikir kritis dan analitis sesuai dengan standar internasional, sedangkan pada standar penilaian memberi ruang pada pengembangan instrumen penilaian yang mengukur berpikir tingkat tinggi. Penilaian hasil belajar diharapkan dapat membantu peserta didik untuk meningkatkan kemampuan *higher order thinking skills (HOTS)*, karena berpikir tingkat tinggi dapat mendorong peserta didik untuk berpikir secara luas dan mendalam tentang materi pelajaran.

Pengertian kemampuan berpikir tingkat tinggi atau *higher order thinking skills* menurut Brookhart (2010, hlm. 5):

*“Higher order thinking conceived of as the top end of the Bloom’s cognitive taxonomy. The teaching goal behind any of the cognitive taxonomies is equipping students to be able to do transfer. “Being able to think” means students can apply the knowledge and skills they developed during their learning to new contexts. “New” here means applications that the student has not thought of before, not necessarily something universally new. Higher order thinking is conceived as students being able to relate their learning to other elements beyond those they were taught to associate with it.”*

Soal-soal *HOTS* merupakan instrumen pengukuran yang digunakan untuk mengukur kemampuan berpikir tingkat tinggi, yaitu kemampuan berpikir yang tidak sekadar mengingat (*recall*),

menyatakan kembali (*restate*), atau merujuk tanpa melakukan pengolahan (*recite*). Soal-soal *HOTS* pada konteks asesmen mengukur kemampuan: (1) transfer satu konsep ke konsep lainnya, (2) memproses dan menerapkan informasi, (3) mencari kaitan dari berbagai informasi yang berbeda-beda, (4) menggunakan informasi untuk menyelesaikan masalah, dan (5) menelaah ide dan informasi secara kritis. Dengan demikian soal-soal *HOTS* menguji keterampilan berpikir menganalisis, mengevaluasi, dan mencipta (Kemdikbud, 2018).

Penilaian berorientasi *HOTS* bukanlah sebuah bentuk penilaian yang baru bagi guru dalam melakukan penilaian. Tetapi penilaian berorientasi *HOTS* ini memaksimalkan keterampilan guru dalam melakukan penilaian. Guru dalam penilaian ini harus menekankan pada penilaian sikap, pengetahuan, dan keterampilan yang bisa meningkatkan keterampilan peserta didik dalam proses pembelajaran berorientasi *HOTS*. Kualitas instrumen penilaian hasil belajar berpengaruh langsung dalam keakuratan status pencapaian hasil belajar peserta didik. Oleh karena itu, kedudukan instrumen penilaian hasil belajar sangat strategis dalam pengambilan keputusan guru dan sekolah terkait pencapaian hasil belajar peserta didik yang di antaranya kemampuan berpikir tingkat tinggi.

Permasalahan yang terjadi di sekolah, soal-soal yang dibuat guru cenderung lebih banyak menguji aspek ingatan yang kurang melatih keterampilan berpikir tingkat tinggi peserta didik. Hasil wawancara dengan beberapa guru di SMA dan SMK kota Samarinda menggambarkan bahwa hanya 30% instrumen tes yang digunakan pada level menganalisis (C4), mengevaluasi (C5) dan mengkreasi (C6). Sementara itu, 70% instrumen tes yang digunakan sebagian besar berupa soal-soal pada level pengetahuan (C1), pemahaman (C2),

penerapan atau aplikasi (C3). Asesmen yang digunakan dalam pembelajaran kurang merangsang peserta didik untuk berpikir secara sistematis, kritis, logis, dan analitis. Permasalahan yang dihadapi oleh guru adalah kemampuan guru dalam mengembangkan instrumen asesmen *HOTS* masih kurang dan belum tersedianya instrumen asesmen yang didesain khusus untuk melatih *HOTS*.

Berdasarkan permasalahan di atas, perlu dikembangkan instrumen asesmen *HOTS* yang berupa soal tes *HOTS*. Instrumen asesmen *HOTS* yang dikembangkan bertujuan untuk menghasilkan instrumen yang valid dan reliabel untuk mengukur *HOTS* peserta didik. Penelitian ini mempunyai manfaat antara lain: instrumen asesmen yang sudah valid dan reliabel dapat digunakan oleh guru untuk mengukur *HOTS* peserta didik, sebagai acuan untuk mengembangkan instrumen asesmen *HOTS* pada Kompetensi Dasar (KD) yang lainnya, dan dapat digunakan oleh peserta didik sebagai bahan latihan soal dalam melatih *HOTS*.

Penilaian merupakan prosedur yang digunakan untuk mendapatkan informasi untuk mengukur taraf pengetahuan dan keterampilan subjek didik yang hasilnya akan digunakan untuk keperluan evaluasi. Informasi adalah data yang diperoleh melalui pengukuran dan nonpengukuran termasuk di dalamnya dengan melakukan observasi kelas, menggunakan tes yang standar atau buatan tes guru, proyek dan portofolio subjek belajar (Subali, 2010).

Dilihat dari tekniknyanya, Poerwanti (2001) mengatakan bahwa penilaian proses dan hasil belajar dibedakan menjadi dua macam, yaitu dengan teknik tes dan nontes. Namun, pada umumnya pengajar lebih banyak menggunakan tes sebagai alat ukur. Dalam penulisan soal, Sunarti dan Rahmawati (2014) mengatakan bahwa ada beberapa kaidah yang perlu diperhatikan yaitu ketepatan materi, konstruksi dan bahasa. Selain

memperhatikan kaidah dalam penulisan soal tes pilihan ganda, penulis soal juga harus memberikan skor dengan baik.

Analisis soal dilakukan untuk mengetahui berfungsi tidaknya sebuah soal. Analisis pada umumnya dilakukan melalui dua cara, yaitu analisis kualitatif (*qualitatif control*) dan analisis kuantitatif (*quantitatif control*). Analisis kualitatif sering pula dinamakan sebagai validitas logis (*logical validity*) yang dilakukan sebelum soal digunakan untuk melihat berfungsi tidaknya sebuah soal. Analisis soal secara kuantitatif sering pula dinamakan sebagai validitas empiris (*empirical validity*) yang dilakukan untuk melihat lebih berfungsi atau tidaknya sebuah soal, setelah soal itu diujicobakan kepada sampel yang representatif.

## **B. METODE**

Metode penelitian ini mengacu pada metode penelitian dan pengembangan atau *Research and Development*. Produk yang dikembangkan adalah instrumen asesmen *HOTS* berupa soal tes *HOTS* pilihan ganda dan soal tes *HOTS* uraian. Untuk mendapatkan produk awal pengembangan, pada penelitian ini dilakukan adaptasi dari model pengembangan Borg & Gall. Dari 10 langkah pengembangan model Borg & Gall diadaptasi menjadi delapan langkah pengembangan yaitu: (1) penelitian pendahuluan dan pengumpulan informasi, (2) perencanaan, (3) pengembangan produk awal, (4) validasi produk awal (5) revisi produk awal, (6) uji coba terbatas, (7) uji coba analisis lapangan, dan (8) revisi produk akhir.

Pada tahap analisis kebutuhan dilakukan observasi dan wawancara pada beberapa guru kelas X terkait penilaian yang biasanya dilakukan untuk mengetahui persentase penggunaan tingkat Taksonomi Bloom dalam soal. Pada tahap ini juga telah dilakukan studi pustaka, ini dilakukan untuk pengenalan sementara terhadap produk yang akan

dikembangkan. Validasi dilakukan pada tahap pengembangan produk awal oleh tiga ahli yaitu ahli materi, ahli konstruksi, dan ahli bahasa. Uji coba empiris butir soal tes *HOTS* dilakukan dengan menggunakan uji coba terbatas dan uji coba lapangan. Uji coba terbatas dilakukan pada 30 peserta didik SMK Negeri 15 Samarinda. Uji coba lapangan dilakukan pada 137 peserta didik yang berasal dari SMA Negeri 1 Samarinda, SMA Negeri 3 Samarinda, SMK Negeri 15 Samarinda, dan SMKS TI Airlangga Samarinda. Analisis data uji coba menggunakan parameter teori tes klasik untuk mengetahui kualitas soal tes *HOTS* secara empiris sebagai dasar untuk revisi dan perakitan soal tes *HOTS*.

Data dalam penelitian ini meliputi data kualitatif dan kuantitatif. Data-data ini bertujuan untuk memberi gambaran mengenai kualitas produk yang dikembangkan. Data kualitatif diperoleh dari hasil validasi ahli produk awal soal tes *HOTS*, sedangkan data kuantitatif diperoleh dari uji coba produk soal tes *HOTS*. Instrumen penelitian yang dikembangkan dalam penelitian ini diklasifikasikan menjadi dua macam yang masing-masing digunakan untuk memenuhi kriteria valid dan reliabel.

Data penilaian ahli terhadap soal dianalisis dengan teknik deskriptif dengan menggunakan persentase (Purwanto, 2010). Analisis data dilakukan dengan cara menghitung skor yang dicapai dari seluruh aspek yang dinilai. Instrumen untuk mengukur kevalidan digunakan lembar validasi (telaah soal tes) yang terdiri dari tiga aspek, yaitu materi, konstruksi, dan bahasa.

Soal tes valid atau layak digunakan berdasarkan penilaian dari validator. Kriteria validasi butir soal menggunakan empat pilihan yaitu nilai 4 untuk kategori sangat sesuai (sangat layak), nilai 3 untuk kategori sesuai (layak), nilai 2 untuk kategori kurang sesuai (perlu diperbaiki), dan nilai 1 untuk kategori tidak sesuai

(perlu diganti) untuk setiap nomor butir soal.

Instrumen untuk mengukur reliabilitas digunakan dua perangkat soal tes yaitu seperangkat soal pilihan ganda dan soal uraian. Soal tes *HOTS* diujikan secara individu dan hasilnya dianalisis secara kuantitatif untuk mengetahui estimasi koefisien reliabilitas instrumen asesmen yang dikembangkan.

Teknik pengumpulan data yang digunakan peneliti adalah sebagai berikut: (1) menyusun instrumen yang akan digunakan dalam penelitian, seperti soal tes *HOTS*, rubrik penskoran dan penilaian, (2) menentukan validitas isi instrumen dengan *expert judgment* atau meminta beberapa ahli untuk memvalidasi instrumen yang telah dibuat, (3) melakukan revisi instrumen sesuai dengan saran validator, (4) melakukan uji coba instrumen penelitian, (5) menentukan reliabilitas, tingkat kesukaran, daya pembeda butir soal dan fungsi pengecoh pilihan jawaban, (6) melakukan revisi instrumen berdasarkan analisis hasil uji coba.

Analisis kualitatif soal tes *HOTS* diperoleh dari hasil lembar validasi (telaah soal tes) yang dilakukan dengan cara deskriptif kualitatif dengan menggunakan penghitungan persentase. Analisis kuantitatif data yang diperoleh dari respons jawaban peserta didik dianalisis dengan menggunakan bantuan *software Anates 4.0* untuk analisis soal pilihan ganda dan esai. Analisis butir soal digunakan untuk mengetahui kelayakan butir soal yang meliputi reliabilitas, tingkat kesukaran, daya pembeda, dan penyebaran pilihan jawaban (pengecoh).

## C. PEMBAHASAN

### 1. Hasil Studi Pendahuluan

Penelitian pengembangan yang telah dilakukan dimulai dengan studi pendahuluan di sekolah. Studi pendahuluan ini dimaksudkan untuk mengetahui kebutuhan dan permasalahan

yang ada. Hasil studi pendahuluan ini digunakan sebagai dasar atau landasan untuk mengembangkan produk. Studi pendahuluan yang dilakukan adalah wawancara dengan beberapa guru mata pelajaran bahasa Indonesia dan penyebaran angket kebutuhan siswa.

Dari wawancara yang dilakukan diketahui bahwa Persentase level kognitif pada soal yang disusun oleh sebagian besar guru belum merujuk pada level kognitif *HOTS*. Rata-rata guru membuat soal dengan persentase level kognitif *LOTS* lebih besar dari pada level kognitif *HOTS* dengan perbandingan 70% dan 30%. Sebagian besar guru merasa masih kesulitan menyusun soal *HOTS*, hal ini dikarenakan guru belum memahami soal *HOTS* dan belum tersedianya referensi berupa kumpulan soal *HOTS*. Oleh karena itu, guru merasa sangat perlu adanya kumpulan soal *HOTS* sebagai bahan belajar dan sebagai referensi penulisan soal *HOTS*.

Kesimpulan dari angket kebutuhan peserta didik adalah 92% peserta didik memilih instrumen tes *Higher Order Thinking Skill* yang digunakan sebagai acuan untuk evaluasi hasil belajar. Sebanyak 89,8% peserta didik menginginkan adanya pengembangan soal *Higher Order Thinking Skill*. Secara umum alasan dari responden memilih soal *HOTS* karena dapat merangsang kemampuannya untuk berpikir tingkat tinggi sehingga peserta didik dapat melatih kemampuan berpikir kreatif dan kritis.

### 2. Hasil Desain Pengembangan Produk

Untuk menulis butir soal *HOTS*, terlebih dahulu penulis menentukan perilaku yang hendak diukur dan merumuskan materi yang akan dijadikan dasar pertanyaan dalam konteks tertentu sesuai dengan perilaku yang diharapkan. Adapun langkah yang dilakukan dalam penyusunan soal *HOTS* yakni (1)

menganalisis KD yang dapat dibuat soal-soal HOTS, (2) menyusun kisi-kisi soal, (3) merumuskan stimulus yang kontekstual, (4) menulis butir soal sesuai kisi-kisi, dan (5) membuat pedoman penskoran dan kunci jawaban.

Proses yang dilakukan dalam pengembangan ini meliputi penyusunan produk soal tes *HOTS*. Soal tes *HOTS* yang dirancang dilakukan penilaian oleh validator ahli, dilakukan revisi untuk kemudian diperoleh produk awal soal tes *HOTS* yang siap digunakan sebagai bahan uji coba terbatas. Hasil dari uji coba terbatas, sebagai bahan revisi untuk kemudian menjadi produk utama soal tes *HOTS* yang siap digunakan sebagai bahan uji coba lapangan. Setelah diperoleh estimasi koefisien reliabilitas, kriteria tingkat kesukaran, daya pembeda, dan alternatif pengecoh dari hasil uji coba lapangan, diperoleh produk akhir soal tes *HOTS* yang siap digunakan.

### 3. Hasil Validasi Produk

Validasi dilakukan untuk menguji kelayakan produk sebelum diujikan pada uji terbatas maupun uji lapangan. Rata-rata persentase keseluruhan aspek yang

dinilai pada soal pilihan ganda adalah 94,42% dengan kategori termasuk sangat baik. Aspek materi sebesar 98,28% dengan kriteria sangat baik, aspek konstruksi sebesar 90% dengan kriteria sangat baik, dan aspek bahasa sebesar 95% dengan kriteria sangat baik. Sementara itu rata-rata persentase keseluruhan aspek yang dinilai pada soal esai adalah 94,79% dengan kategori termasuk sangat baik. Aspek materi sebesar 99,37% dengan kriteria sangat baik, aspek konstruksi sebesar 95% dengan kriteria sangat baik, dan aspek bahasa sebesar 90% dengan kriteria sangat baik. Berdasarkan kriteria penilaian, baik soal pilihan ganda maupun soal esai dinyatakan layak untuk diujicobakan.

### 4. Hasil Uji Coba Produk

#### a. Hasil Uji Coba Skala Kecil (Uji Terbatas)

Reliabilitas atau keajekan pengukuran adalah konsistensi dari sebuah pengukuran. Hasil pengujian reliabilitas instrumen *HOTS* disajikan dalam Tabel 1.

**Tabel 1. Reliabilitas Produk Awal**

Nomor Butir Soal	Koefisien Reliabilitas	Kriteria
1 – 40 (Pilihan ganda)	0,94	Sangat tinggi
1 – 5 (Esai)	0,79	Tinggi

Berdasarkan Tabel 1, dapat diketahui nilai reliabilitas dari produk awal soal tes *HOTS* pilihan ganda berada pada rentang 0,80—1,00 yang dikategorikan sangat tinggi, soal tes *HOTS* esai berada pada rentang 0,60—0,79 yang dikategorikan tinggi artinya instrumen penilaian tersebut memiliki keajekan atau konsistensi yang

tinggi. Itu artinya produk asesmen berupa soal *HOTS* bentuk pilihan ganda dan esai yang dikembangkan layak digunakan. Tingkat kesukaran diperlukan untuk mengetahui seberapa sukar instrumen yang diujikan berdasarkan hasil tes yang dikerjakan oleh peserta didik.

**Tabel 2. Tingkat Kesukaran Produk Awal Soal Tes *HOTS* Pilihan Ganda**

Kategori	Nomor Butir Soal	Jumlah	Persentase
0,00—0,29 (Sukar)	3, 7, 8, 10, 15, 23, 38	7	17,5%
0,30—0,69 (Sedang)	1, 2, 4, 5, 6, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40	31	77,5%
0,70—1,00 (Mudah)	22, 29	2	5%

Berdasarkan Tabel 2 dapat diketahui bahwa tingkat kesukaran soal bentuk pilihan ganda berkisaran pada kategori sedang sebanyak 33 butir soal (77,5%).

Soal dengan kategori sukar sebanyak 7 butir soal (17,5%) dan soal dengan kategori mudah sebanyak 2 butir soal (5%).

**Tabel 3. Tingkat Kesukaran Produk Awal Soal Tes *HOTS* Esai**

Kategori	Nomor Butir Soal	Jumlah	Persentase
0,30—0,69 (Sedang)	1, 2, 3, 4, 5	5	100%

Berdasarkan Tabel 3 dapat diketahui bahwa tingkat kesukaran soal bentuk esai berada pada kategori sedang yaitu sebanyak 5 butir soal (100%).

Daya pembeda adalah kemampuan suatu butir item tes hasil belajar untuk dapat membedakan antara *testee* yang berkemampuan tinggi dengan *testee* yang kemampuannya rendah.

**Tabel 4. Daya Pembeda Produk Awal Soal Tes *HOTS* Pilihan Ganda**

Kategori	Nomor Butir Soal	Jumlah	Persentase
0,20—0,39 (Cukup)	5, 6, 10, 17, 20, 22, 24, 31, 32, 35	10	25%
0,40—0,69 (Baik)	1, 2, 3, 4, 7, 9, 11, 12, 13, 14, 18, 25, 26, 27, 28, 29, 30, 33, 36, 37, 38, 39, 40	23	57,5%
0,70—1,00 (Sangat baik)	8, 15, 16, 19, 21, 23, 34	7	17,5%

Berdasarkan Tabel 4 dapat diketahui bahwa daya pembeda produk awal soal tes *HOTS* pilihan ganda tidak terdapat daya pembeda kategori kurang. Soal dengan daya pembeda kategori cukup

sebanyak 10 butir soal (20%). Soal dengan daya pembeda kategori baik sebanyak 23 butir soal (57,5%) dan soal dengan daya pembeda kategori sangat baik sebanyak 7 butir soal (17,5%).

**Tabel 5. Daya Pembeda Produk Awal Soal Tes *HOTS* Esai**

Kategori	Nomor Butir Soal	Jumlah	Persentase
0,20—0,39 (Cukup)	3	1	20%
0,40—0,69 (Baik)	1, 2, 4, 5	4	80%

Berdasarkan Tabel 5 dapat diketahui bahwa daya pembeda produk awal soal tes *HOTS* esai terdapat 1 butir soal (20%) dengan kategori cukup dan terdapat 4 butir soal (80%) kategori baik. Sesuai dengan kriteria kualitas instrumen tes dapat diketahui bahwa 5 butir soal esai

memiliki daya pembeda yang memenuhi kriteria.

Penyebaran pilihan jawaban dijadikan dasar dalam penelaahan soal. Hal ini dimaksudkan untuk mengetahui berfungsi tidaknya pilihan jawaban yang tersedia.

**Tabel 6. Penyebaran Pilihan Jawaban Produk Awal Soal Tes *HOTS***

Kategori	Nomor Butir Soal	Jumlah	Persentase
Pilihan jawaban ≥ 5%	3, 4, 5, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 19, 20, 21, 23, 24, 25, 26, 27, 29, 31, 32, 33, 35, 36, 38	28	70%
Pilihan jawaban < 5%	1, 2, 6, 13, 17, 22, 28, 30, 34, 37, 39, 40	12	30%

Berdasarkan Tabel 6 dapat diketahui bahwa butir soal dengan penyebaran jawaban (pengecoh) yang berfungsi baik sebanyak 28 butir soal (70%) dan butir soal dengan penyebaran jawaban (pengecoh) yang berfungsi tidak baik sebanyak 12 butir soal (30%). Dengan demikian, 28 soal dikatakan memenuhi kriteria dan 12 soal perlu direvisi.

**b. Hasil Uji Coba Skala Luas (Uji Lapangan)**

Hasil pengujian reliabilitas instrumen *HOTS* pada uji coba skala luas atau uji coba lapangan disajikan dalam tabel berikut.

**Tabel 7. Reliabilitas Produk Utama Soal Tes *HOTS***

Nomor Butir Soal	Koefisien Reliabilitas	Kriteria
1—40 (Pilihan ganda)	0,88	Sangat tinggi
1—5 (Esai)	0,79	Tinggi

Tabel 7 menunjukkan nilai reliabilitas dari produk utama soal tes *HOTS* pilihan ganda berada pada rentang 0,80—1,00 yang dikategorikan sangat tinggi, dan soal esai berada pada rentang 0,60—0,79 yang dikategorikan tinggi artinya instrumen

penilaian tersebut memiliki keajekan atau konsistensi yang tinggi.

Hasil pengujian tingkat kesukaran terhadap instrumen penilaian *HOTS* disajikan dalam tabel berikut.



**Tabel 8. Tingkat Kesukaran Produk Utama Soal Tes *HOTS* Pilihan Ganda**

Kategori	Nomor Butir Soal	Jumlah	Persentase
0,00—0,29 (Sukar)	1, 10, 18, 38	4	10%
0,30—0,69 (Sedang)	2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 19, 21, 23, 24, 25, 26, 27, 31, 32, 33, 34, 35, 36, 37, 39, 40	30	75%
0,70—1,00 (Mudah)	13, 20, 22, 28, 29, 30	6	15%

Berdasarkan Tabel 8 dapat diketahui bahwa tingkat kesukaran soal bentuk pilihan ganda berkisaran pada kategori sedang sebanyak 30 butir soal (75%). Soal

dengan kategori sukar sebanyak 4 butir soal (10%) dan soal dengan kategori mudah sebanyak 6 butir soal (15%).

**Tabel 9. Tingkat Kesukaran Produk Utama Soal Tes *HOTS* Esai**

Kategori	Nomor Butir Soal	Jumlah	Persentase
0,30—0,69 (Sedang)	1, 2, 3, 4, 5	5	100%

Berdasarkan Tabel 9 dapat diketahui bahwa tingkat kesukaran soal bentuk esai semua berada pada kategori sedang sebanyak 5 butir soal (100%).

Secara umum daya pembeda butir soal pilihan ganda dan esai dapat dilihat pada tabel 10 dan 11 berikut.

**Tabel 10. Daya Pembeda Produk Utama Soal Tes *HOTS* Pilihan Ganda**

Kategori	Nomor Butir Soal	Jumlah	Persentase
0,20—0,39 (Cukup)	2, 7, 13, 14, 23, 36, 39	7	17,5%
0,40—0,69 (Baik)	1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 40	33	82,5%

Berdasarkan Tabel 10 dapat diketahui bahwa daya pembeda produk utama soal tes *HOTS* pilihan ganda mempunyai daya pembeda pada kategori cukup sebanyak 7 butir soal (17,5%), dan sebagian besar butir soal pada kategori baik yaitu sejumlah 33 butir soal (82,5%).

**Tabel 11. Daya Pembeda Produk Utama Soal Tes *HOTS* Esai**

Kategori	Nomor Butir Soal	Jumlah	Persentase
0,20—0,39 (Cukup)	5	1	20%
0,40—0,69 (Baik)	1, 2, 3, 4	4	80%

Berdasarkan Tabel 11 dapat diketahui bahwa daya pembeda produk utama soal tes *HOTS* esai pada kategori cukup sebanyak 1 butir soal (20%) dan

sebagian besar butir soal pada kategori baik yaitu sejumlah 4 butir soal (80%).

Penyebaran pilihan jawaban butir soal pilihan ganda dapat dilihat pada tabel berikut.

**Tabel 12. Penyebaran Jawaban Produk Utama Soal Tes *HOTS***

Kategori	Nomor Butir Soal	Jumlah	Persentase
Pilihan jawaban $\geq$ 5%	1—40	40	100%
Pilihan jawaban $<$ 5%	-	0	0%

Berdasarkan Tabel 12 dapat diketahui bahwa semua butir soal mempunyai penyebaran jawaban (pengecoh) yang berfungsi baik yakni sebanyak 40 butir soal (100%).

### 5. Revisi Produk

Revisi produk dilakukan untuk memperoleh produk akhir yang memenuhi kriteria valid dan reliabel. Revisi dilakukan berdasarkan hasil penilaian dan analisis instrumen asesmen pada setiap tahap uji coba produk. Revisi produk pada penelitian ini terdiri dari: revisi produk hasil validasi dan revisi produk hasil uji coba terbatas. Pertama, revisi produk dilakukan berdasarkan hasil validasi dan saran dari masing-masing ahli yaitu ahli materi, ahli konstruksi, dan ahli bahasa. Pada aspek bahasa terdapat beberapa perbaikan berdasarkan saran dari ahli bahasa antara lain kesalahan penulisan istilah asing, penggunaan kalimat tidak efektif, dan kesalahan penggunaan tanda baca. Kedua, revisi butir soal berdasarkan hasil uji coba terbatas produk awal soal tes *HOTS*. Butir soal kategori diterima dengan revisi, direvisi berdasarkan hasil analisis karakteristik butir soal. Beberapa dilakukan pada penyebaran pilihan jawaban (pengecoh) yang kurang berfungsi untuk soal tes *HOTS* pilihan ganda. Butir soal yang memiliki penyebaran jawaban (pengecoh) kurang atau tidak berfungsi maka direvisi dengan

mengubah posisi pilihan jawaban atau mengganti pilihan jawaban dari butir soal tersebut.

### 6. Pembahasan

Produk akhir penelitian pengembangan ini adalah instrumen asesmen *HOTS* bahasa Indonesia SMA dan SMK kelas X semester 1 yang berupa perangkat soal tes *HOTS*. Berdasarkan hasil validasi ahli, uji coba terbatas, uji coba lapangan, dan perbaikan-perbaikan, serta analisis data yang dilakukan dapat diketahui bahwa perangkat soal tes *HOTS* yang dikembangkan telah memenuhi kriteria valid dan reliabel, serta kualitas butir soal yang baik.

Pengembangan instrumen penilaian *HOTS* untuk mengukur keterampilan berpikir tingkat tinggi ini juga sejalan dengan hasil penelitian terdahulu yang relevan yaitu menghasilkan produk instrumen asesmen yang valid dan layak untuk digunakan sebagai alat evaluasi untuk mengukur keterampilan berpikir tingkat tinggi peserta didik. Meski demikian, ada perbedaan pada komposisi materi dan model pengembangan yang digunakan.

Kelebihan produk pengembangan instrumen asesmen *HOTS* pada mata pelajaran bahasa Indonesia ini adalah semua butir soal memenuhi kriteria valid dan layak berdasarkan uji teoretis (validasi materi, konstruksi, dan bahasa) dan uji empiris (uji skala terbatas dan skala luas).

Produk asesmen yang dihasilkan siap digunakan di sekolah untuk mengukur keterampilan berpikir tingkat tinggi peserta didik.

Keterbatasan instrumen asesmen HOTS pada penelitian ini hanya menggunakan dua bentuk soal yaitu pilihan ganda dan esai. Pemilihan bentuk soal itu disebabkan jumlah peserta tes cukup banyak, sedangkan penskoran dan analisis butir soal harus secepatnya dilakukan. Sehingga bentuk soal yang paling memungkinkan adalah soal bentuk pilihan ganda dan uraian.

#### D. SIMPULAN

Produk akhir dalam penelitian ini menghasilkan instrumen asesmen HOTS bahasa Indonesia untuk mengukur keterampilan berpikir tingkat tinggi peserta didik SMA dan SMK kelas X semester 1. Instrumen asesmen berupa perangkat soal tes HOTS yang terdiri dari 40 butir soal pilihan ganda dengan lima pilihan jawaban dan 5 butir soal uraian. Instrumen asesmen dikembangkan melalui delapan langkah pengembangan, yaitu: (a) penelitian dan pengumpulan informasi, (b) desain perencanaan, (c) pengembangan produk awal, (d) validasi produk, (e) revisi produk awal, (f) uji coba terbatas, (g) uji coba lapangan, (h) revisi produk akhir. Kevalidan instrumen asesmen dibuktikan dengan hasil penilaian ahli yang menunjukkan bahwa instrumen asesmen yang dikembangkan layak digunakan berdasarkan telaah aspek materi, konstruksi, dan bahasa. Berdasarkan hasil uji coba instrumen tersebut juga telah memenuhi kriteria layak dengan nilai koefisien reliabilitas kategori sangat tinggi untuk soal pilihan ganda dan kategori tinggi untuk soal uraian. Soal pilihan ganda memiliki tingkat kesukaran rata-rata sedang, daya pembeda rata-rata baik, dan semua pengecoh berfungsi baik. Soal uraian memiliki tingkat kesukaran rata-rata

kategori sedang dengan daya pembeda rata-rata kategori baik.

Saran pemanfaatan produk akhir instrumen asesmen HOTS adalah: (1) peserta didik dapat menggunakan produk akhir instrumen asesmen HOTS sebagai bahan latihan untuk melatih keterampilan berpikir tingkat tinggi; (2) guru mata pelajaran Bahasa Indonesia dapat menggunakan produk akhir instrumen asesmen HOTS untuk mengukur penguasaan pengetahuan dan keterampilan berpikir tingkat tinggi peserta didik; dan (3) produk akhir instrumen asesmen HOTS hasil pengembangan yang berupa soal tes HOTS dapat digunakan sebagai acuan dalam mengembangkan instrumen asesmen HOTS pada Kompetensi Dasar (KD) lainnya.

#### DAFTAR PUSTAKA

- Amiriono & Daryanto. (2016). *Evaluasi dan Penilaian Pembelajaran Kurikulum 2013*, Yogyakarta: Gava Media.
- Anderson, L.W. & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. A Bridged Edition*. New York: Addison Wesley Longman.
- Azwar, S. (2009). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Borg, W. R. & Gall, M. D. (1983). *Educational Research: An introduction*. 4<sup>th</sup> Ed. New York: Longman.
- Brookhart, S. M. (2010). *How to assess higher order thinking skills in your classroom*. Virginia: SCD Alexandria.
- Budiman, A., & Jailani, J. (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.
- Depdiknas. (2007). Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 20, Tahun 2007

- tentang Standar Penilaian Pendidikan.
- Dharmawati, R. S., & Mahanal, S. (2016). Pengembangan Instrumen Asesmen Berpikir Kritis Untuk Siswa SMP Kelas VII pada Materi Interaksi Makhluk Hidup dengan Lingkungannya. *Jurnal Pendidikan*, 1(8), 1598-1606.
- Gronlund, N. E. & Linn, R. L. (1995). *Measurement and Assessment in Teaching*. 7<sup>th</sup> Ed. New Jersey: Prentice-Hall.
- Kemdikbud. (2017). *Modul Penyusunan Soal Higher Order Thinking Skills*. Jakarta: Direktorat Pembinaan SMA Dirjen Pendidikan Dasar dan Menengah.
- Kemdikbud. (2018). *Buku Penilaian Berorientasi Higher Order Thinking Skills*. Jakarta: Direktorat Pembinaan SMA Dirjen Pendidikan Dasar dan Menengah.
- Koyan, W. (2012). *Konstruksi Tes*. Singaraja: Universitas Pendidikan Ganesha Press.
- Kusaeri & Suprananto. (2012). *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu.
- Nofiana, M., Sajidan, & Puguh. (2014). Pengembangan Instrumen Evaluasi *Two-Tier Multiple Choice Question* untuk Mengukur Keterampilan Berpikir 164 Tingkat Tinggi pada Materi Kingdom Plantae. *Jurnal Inkuiri*, 3(2), 60-74.
- Poerwanti, E. (2008). *Asesmen Pembelajaran*. Jakarta: Depdiknas.
- Purwanto. (2010). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.
- Sudjana, Nana. (2016). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Rosdikarya
- Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.
- Sumarna, S. (2007). *Panduan Penulisan Tes Tertulis*. Bandung: Remaja Rosdakarya.
- Taufiqurrahman, T., Heryandi, M. T., & Junaidi, J. (2018). Pengembangan Instrumen Penilaian *Higher Order Thinking Skill* pada Mata Pelajaran Pendidikan Agama Islam. *Jurnal Pendidikan Islam Indonesia*, 2(2), 199-206.
- Zainal, A. (2019). Pengembangan Instrumen Asesmen Berbasis Keterampilan Berpikir Tingkat Tinggi dengan Memanfaatkan Kumpulan Prosa “Filosofi Kopi”. *Tesis*. Program Pascasarjana UM.