

PENGEMBANGAN INSTRUMEN ASESMEN *HIGHER ORDER THINKING SKILL* (HOTS) PADA MATA PELAJARAN MATEMATIKA SMP KELAS VIII SEMESTER 1

Agus Budiman¹⁾, Jailani²⁾
SMP Negeri 2 Mandiraja¹⁾, Universitas Negeri Yogyakarta²⁾
agusbudimath@yahoo.com¹⁾, jailani@uny.ac.id²⁾

Abstrak

Penelitian ini bertujuan untuk menghasilkan instrumen asesmen matematika berupa soal tes *HOTS* yang valid dan reliabel, dan mendeskripsikan kualitas soal tes *HOTS* untuk mengukur keterampilan berpikir tingkat tinggi pada peserta didik SMP kelas VIII. Penelitian ini merupakan penelitian pengembangan yang diadaptasi dari model pengembangan Borg & Gall, yang meliputi langkah-langkah sebagai berikut: penelitian dan pengumpulan informasi, perencanaan, pengembangan produk awal, uji coba terbatas, revisi produk awal, uji coba lapangan, dan revisi produk akhir. Hasil penelitian menunjukkan bahwa instrumen asesmen *HOTS* berupa soal tes *HOTS* yang terdiri dari 24 butir soal pilihan ganda dan 19 butir soal uraian dari aspek materi, konstruksi, dan bahasa dinyatakan valid dan layak digunakan. Instrumen tersebut mempunyai koefisien reliabilitas sebesar 0,713 (soal pilihan ganda) dan sebesar 0,920 (soal uraian). Soal pilihan ganda memiliki rata-rata tingkat kesukaran 0,406 (sedang), rata-rata daya pembeda 0,330 (baik), dan semua pengecoh berfungsi baik. Soal uraian memiliki rata-rata tingkat kesukaran 0,373 (sedang) dengan rata-rata daya pembeda 0,508 (baik).

Kata kunci: pengembangan, instrumen asesmen, *Higher Order Thinking Skills* (HOTS), matematika SMP

DEVELOPING AN ASSESSMENT INSTRUMENT OF HIGHER ORDER THINKING SKILL (HOTS) IN MATHEMATICS FOR JUNIOR HIGH SCHOOL GRADE VIII SEMESTER 1

Abstract

This research aims to produce a valid and reliable mathematics assessment instrument in the form of HOTS test items, and describe the quality of HOTS test items to measure the high order thinking skill of grade VIII SMP students. This study was a research and development study adapting Borg & Gall's development model, including the following steps: research and information collection, planning, the 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 items consists of 24 multiple – choice test items and 19 essay test items, based on the judgement of the materials, construction, and language is valid and appropriate to be used. The reliability coefficients of the instrument are 0.713 for the multiple choice items, and 0.920 for essays. The multiple choice items has the average of item difficulty 0.406 (average), the average of item discrimination 0.330 (good), and the distractors function well. The essay test items has the average of item difficulty 0.373 (average) and the average of item discrimination 0.508 (good).

Keywords: development, assessment instrument, *Higher Order Thinking Skills* (HOTS), mathematics in the junior high school

PENDAHULUAN

Keberhasilan pendidikan yang tujuan utamanya meningkatkan sumber daya manusia, dipengaruhi oleh berbagai faktor. Salah satu faktor yang ikut mempengaruhi keberhasilan ini 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 atau meningkatkan proses pembelajaran yang telah dilakukan guru. Berkaitan dengan hal tersebut pemerintah telah memberikan pedoman yaitu dengan mengeluarkan Permendiknas Nomor 16 Tahun 2007 tentang Standar Kualifikasi Akademik dan Kompetensi Guru mata pelajaran (termasuk guru matematika SMP/MTs) dinyatakan bahwa kompetensi guru mata pelajaran antara lain adalah mengembangkan instrumen penilaian.

Penilaian Pendidikan menurut Permendiknas Nomor 20 Tahun 2007 tentang Standar Penilaian Pendidikan adalah proses pengumpulan dan pengolahan informasi untuk menentukan pencapaian hasil belajar peserta didik. Prinsip dan standar penilaian menekankan dua ide pokok yaitu penilaian harus meningkatkan belajar peserta didik dan penilaian merupakan sebuah alat yang berharga untuk membuat keputusan pengajaran (Van de Walle, 2007, p.78). Penilaian tidak sekedar pengumpulan data peserta didik, tetapi juga pengolahannya untuk memperoleh gambaran proses dan hasil belajar peserta didik. Penilaian tidak sekedar memberi soal peserta didik kemudian selesai, tetapi guru harus menindaklanjutinya untuk kepentingan pembelajaran. Untuk melaksanakan penilaian, guru memerlukan instrumen penilaian dalam bentuk soal-soal baik untuk menguji kemampuan kognitif, afektif, maupun psikomotor.

Penilaian merupakan kegiatan sangat penting dalam pembelajaran matematika. Penilaian dapat memberikan umpan balik yang konstruktif bagi guru maupun peserta didik. Hasil penilaian juga dapat memberikan motivasi kepada peserta didik untuk berprestasi lebih baik. Bahkan penilaian dapat mempengaruhi perilaku belajar karena peserta didik cenderung mengarahkan kegiatan belajarnya menuju muara penilaian yang dilakukan guru. 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 diantaranya kemampuan berpikir tingkat tinggi.

Rendahnya kemampuan berpikir tingkat tinggi di kalangan peserta didik telah menarik para pendidik dan peneliti pendidikan matematika seperti tersirat dalam pernyataan Henningsen & Stein (1997, p.524) "*much discussion and concern have been focused on limitations in students' conceptual understanding as well as on their thinking, reasoning, and problem-solving skills in mathematics*", maknanya banyak diskusi dan perhatian telah difokuskan pada keterbatasan dalam pemahaman konseptual peserta didik, serta pada pemikiran, penalaran, dan keterampilan pemecahan masalah dalam matematika. Termasuk di Indonesia rendahnya pengetahuan matematika peserta didik senantiasa menjadi topik pembicaraan yang hangat dalam masyarakat. Peserta didik sering tidak dapat menggunakan pengetahuan matematika yang mereka miliki dalam kehidupan sehari-hari, bahkan tidak dapat menggunakan keterampilan menyelesaikan soal apabila diberikan soal yang sedikit berbeda dari apa yang dipelajari. Hasil survei mengenai prestasi peserta didik yang dilaksanakan secara internasional nilai peserta didik Indonesia masih jauh di bawah rata-rata. Kemampuan anak Indonesia secara ilmiah tetap dianggap masih dalam kategori rendah. Hal ini sempat dimunculkan lewat penelitian *Trends in International Mathematics and Science Study (TIMSS)* setiap empat tahun sekali yang mengukur kemampuan peserta didik kelas VIII SMP.

Mullis, *et al.* (2012, p.56) menyatakan hasil prestasi *TIMSS* tahun 2007 dan 2011 menunjukkan skor pencapaian prestasi belajar peserta didik kelas VIII SMP (*eight grade*) berturut-turut 397 dan 386 (skala 0 sampai 800) dengan skor rata-rata 500. Keadaan kemampuan peserta didik kelas VIII SMP Indonesia berada di bawah rata-rata. Hasil tidak menunjukkan banyak perubahan pada setiap keikutsertaannya. Hasil prestasi *TIMSS* yang rendah tersebut tentunya disebabkan oleh beberapa faktor. Salah satu faktor penyebabnya antara lain karena peserta didik di Indonesia kurang terlatih dalam menyelesaikan soal-soal kontekstual, menuntut penalaran, argumentasi dan kreativitas dalam menyelesaikannya, dimana soal-soal tersebut merupakan karakteristik soal-soal *TIMSS*. Hal

ini sesuai dengan Kemdikbud (2013, p.2) yang menyatakan bahwa rendahnya prestasi siswa Indonesia tersebut disebabkan oleh banyaknya materi uji di TIMSS yang tidak terdapat dalam kurikulum Indonesia.

Mullis, *et al.* (2012, p.30) menyatakan domain penilaian pada TIMSS 2011 pada peserta didik kelas VIII SMP meliputi domain konten (*content domains*) dan domain kognitif (*cognitive domains*) yang masing-masing terdiri dari beberapa domain. Domain konten sejalan dengan materi (konten) pada standar isi mata pelajaran Matematika SMP, yaitu: bilangan (*number*), aljabar (*algebra*), geometri (*geometry*), data dan peluang (*data and chance*). Domain kognitif terdiri dari pengetahuan (*knowing*), penerapan (*applying*), dan penalaran (*reasoning*). Soal-soal matematika yang dikembangkan oleh TIMSS menuntut peserta didik untuk berpikir tingkat rendah sampai tingkat tinggi. Soal-soal yang memuat tuntutan berpikir tingkat tinggi berkaitan dengan ranah kognitif penalaran yang antara lain mencakup kemampuan menemukan konjektur, analisis, generalisasi, koneksi, sintesis, pemecahan masalah tidak rutin, dan justifikasi atau pembuktian.

Karakteristik HOTS yang diungkapkan Resnick (1987, p.3) diantaranya adalah *non algoritmik*, bersifat kompleks, *multiple solutions* (banyak solusi), melibatkan variasi pengambilan keputusan dan interpretasi, penerapan *multiple criteria* (banyak kriteria), dan bersifat *effortful* (membutuhkan banyak usaha). Conklin (2012, p.14) menyatakan karakteristik HOTS sebagai berikut: "*characteristics of higher-order thinking skills: higher-order thinking skills encompass both critical thinking and creative thinking*" artinya, karakteristik keterampilan berpikir tingkat tinggi mencakup berpikir kritis dan berpikir kreatif. Berpikir kritis dan kreatif merupakan dua kemampuan manusia yang sangat mendasar karena keduanya dapat mendorong seseorang untuk senantiasa memandang setiap permasalahan yang dihadapi secara kritis serta mencoba mencari jawabannya secara kreatif sehingga diperoleh suatu hal baru yang lebih baik dan bermanfaat bagi kehidupannya.

Instrumen penilaian yang digunakan guru untuk menilai hasil belajar peserta didik pada aspek kognitif biasanya diambil dari berbagai buku paket atau kumpulan soal. Soal dapat berupa uraian atau pilihan ganda. Jenis pertanyaan yang diajukan atau tugas yang diberikan oleh guru sangat berpengaruh terhadap perkembangan keterampilan berpikir peserta didik. Perta-

nyaan atau tugas tersebut bukan hanya untuk memfokuskan peserta didik pada kegiatan, tetapi juga untuk menggali potensi belajar mereka. Pertanyaan atau tugas yang memicu peserta didik untuk berpikir analitis, evaluatif, dan kreatif dapat melatih peserta didik dalam keterampilan berpikir tingkat tinggi. Terkait dengan aspek kognitif ini, (*National Council of Teacher Mathematics (NCTM)*, 2000, p.7) mengemukakan "*the next five Standards address the processes of problem solving, reasoning and proof, connections, communication, and representation*". Maksudnya bahwa terdapat lima keterampilan proses yang harus dimiliki peserta didik melalui pelajaran matematika yang tercakup dalam standar proses, yaitu pemecahan masalah, penalaran dan pembuktian, komunikasi, koneksi, dan representasi. Keterampilan-keterampilan tersebut termasuk pada berpikir matematika tingkat tinggi (*high order mathematical thinking*). Kenyataan yang terjadi disekolah, soal-soal cenderung lebih banyak menguji aspek ingatan yang kurang melatih HOTS atau keterampilan berpikir tingkat tinggi peserta didik, pada hal beberapa Standar Kompetensi (SK) dan Kompetensi Dasar (KD) pada mata pelajaran matematika tersebut dapat dikembangkan soal HOTS.

Peningkatan keterampilan berpikir tingkat tinggi telah menjadi salah satu prioritas dalam pelajaran matematika sekolah. Peserta didik tingkat SMP/MTs harus mulai dilatih berpikir tingkat tinggi sesuai dengan usianya, hal ini sesuai dengan BSNP (2006, p.139) yang menyatakan bahwa mata pelajaran matematika diberikan kepada semua peserta didik untuk membekali mereka dengan kemampuan berpikir logis, analitis, sistematis, kritis, dan kreatif, serta kemampuan bekerjasama. Sama halnya dengan Permendikbud Nomor 81a Tahun 2013 tentang Implementasi Kurikulum dalam lampiran I Pedoman Penyusunan dan Pengelolaan Kurikulum Tingkat Satuan Pendidikan yang menyatakan bahwa kemampuan peserta didik yang diperlukan yaitu antara lain kemampuan berkomunikasi, berpikir kritis dan kreatif. Selain itu, hasil Konvensi Ujian Nasional (UN) Tahun 2013 yang diselenggarakan oleh KEMDIKBUD memutuskan bahwa pada penentuan kelulusan untuk meningkatkan kredibilitas dan reliabilitas UN maka ke depan UN mengukur ranah kognitif yang lebih tinggi (*higher order thinking*). Melatih peserta didik untuk terampil ini dapat dilakukan guru dengan cara melatih soal-soal yang berkarakteristik HOTS. Untuk menunjang itu

guru tidak mungkin asal memindah materi dalam buku paket tetapi harus mencari rujukan lain yang lebih berbobot. Masalah yang dihadapi oleh guru adalah kemampuan guru dalam mengembangkan instrumen asesmen *HOTS* masih kurang, selain itu belum tersedianya instrumen asesmen yang didesain khusus untuk melatih *HOTS* atau keterampilan berpikir tingkat tinggi peserta didik. Hal ini sesuai dengan hasil penelitian Thompson (2008, p.96) yang menyatakan bahwa interpretasi guru matematika dari 32 orang mengalami kesulitan menafsirkan keterampilan berpikir dalam Taksonomi Bloom dan membuat item tes untuk berpikir tingkat tinggi.

Permasalahannya, 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 *HOTS*, dan masalah 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*, sehingga perlu dikembangkan instrumen asesmen *HOTS*. Pengembangan keterampilan berpikir tingkat tinggi peserta didik akan menghasilkan: kemahiran peserta didik dalam strategi pemecahan masalah menjadi baik, tingkat keyakinan peserta didik dalam matematika meningkat, dan prestasi belajar peserta didik pada masalah non-rutin yang menuntut keterampilan berpikir tingkat tinggi meningkat (Butkowski, *et al.*, 1994).

Bentuk instrumen asesmen terdiri dari instrumen tes dan non tes. Bentuk instrumen asesmen yang dikembangkan dalam penelitian ini menggunakan bentuk instrumen tes pilihan ganda dan uraian. Tes berbentuk pilihan ganda dan uraian dapat digunakan untuk mengukur *HOTS* atau keterampilan berpikir tingkat tinggi, hal ini sesuai dengan pendapatnya Brookhart (2010, p.33), Nitko & Brookhart (2011, p.223), Kubiszyn & Borich (2013, p.143), dan Sumarna Surapranata (2007, p.137). Pendekatan yang disarankan untuk mengukur berpikir tingkat tinggi yaitu dengan menggunakan *context-dependent item sets* atau seperangkat butir soal yang terdiri dari pengantar dan diikuti oleh pilihan jawaban dan *Context-dependent item sets* atau latihan menafsirkan. Materi pengantar

untuk membuat butir soal tes *HOTS* diantaranya menggunakan gambar, grafik, tabel dan sebagainya yang menuntut peserta didik pada tingkat penerapan taksonomi tujuan pendidikan dan melibatkan proses kognitif tingkat yang lebih tinggi.

Berdasarkan permasalahan di atas, perlu dikembangkan instrumen asesmen *HOTS* berupa soal tes *HOTS* berbentuk pilihan ganda dan uraian pada mata pelajaran matematika SMP Kelas VIII Semester 1. 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 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*.

METODE

Penelitian ini merupakan penelitian pengembangan (*developmental research*). Produk yang dikembangkan adalah instrumen asesmen *HOTS* berupa soal tes *HOTS* pilihan ganda dan soal tes *HOTS* uraian. Untuk mendapatkan prototipe pengembangan, pada penelitian ini dilakukan adaptasi dari model pengembangan Borg & Gall. Dari 10 langkah pengembangan model Borg & Gall diadaptasi menjadi tujuh langkah pengembangan yaitu: (1) penelitian dan pengumpulan informasi, (2) perencanaan, (3) pengembangan produk awal, (4) uji coba terbatas, (5) revisi produk awal, (6) uji coba lapangan, dan (7) revisi produk akhir. Penelitian dan pengumpulan informasi dilakukan untuk telaah konsep berdasarkan kajian teori yang relevan. Validasi instrumen asesmen dilakukan untuk mengevaluasi kevalidan instrumen asesmen yang berupa butir soal tes *HOTS*. Validasi dilakukan pada tahap pengembangan produk awal oleh tiga ahli pendidikan matematika. Uji coba empiris butir soal tes *HOTS* dilakukan dengan menggunakan uji coba terbatas dan uji coba lapangan. Uji coba terbatas dilakukan pada 31 peserta didik SMP Negeri 2 Banjarnegara. Uji coba lapangan dilakukan pada 178 peserta didik yang berasal dari SMP Negeri 1 Banjarnegara, SMP Negeri 2 Banjarnegara, dan SMP Negeri 2 Mandiraja. Analisis data uji coba menggunakan parameter teori tes klasik untuk mengetahui kualitas soal tes *HOTS* secara em-

piris sebagai dasar untuk revisi dan perakitan soal tes *HOTS*.

Data, Instrumen, dan Teknik Pengumpulan Data

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.

Instrumen untuk mengukur kevalidan digunakan lembar validasi (telaah soal tes) yang dianalisis secara kualitatif. Validasi ditinjau dari tiga aspek, yaitu materi, konstruksi, dan bahasa. Soal tes valid atau layak digunakan berdasarkan penilaian dari validator. Kriteria validasi butir soal menggunakan tiga pilihan yaitu layak digunakan (nilai 3), perlu diperbaiki (nilai 2), dan perlu diganti (nilai 1) 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.

Soal tes *HOTS* disusun berdasarkan indikator *HOTS* dan indikator KD. Indikator *HOTS* disintesis dari indikator berpikir kritis dan kreatif menurut Nitko & Brookhart (2011, p.232), Arends & Kilcher (2010, pp.214-233), Presseisen (1985, p.45), Szetela (1993, p.143), Krulik & Rudnick (1999, p.139), O'Daffer & Thornquist (1993, p.40), Maite & Laura (2011, p.609), dan Perkins (1985, p.58). Indikator yang dimaksud antara lain: (1) mengidentifikasi dan mengaitkan informasi yang relevan dari suatu situasi/masalah, (2) membuat simpulan yang tepat berdasarkan informasi dari suatu situasi/masalah, (3) menemukan konsistensi/ inkonsistensi dalam suatu operasi/produk, (4) menilai suatu operasi/produk yang relevan berdasarkan kriteria/standar, (5) memadukan ide/strategi untuk menyelesaikan suatu masalah, (6) menggunakan ide/strategi yang tepat untuk menyelesaikan suatu masalah, (7) mengembangkan atau membuat alternatif baru dalam menyelesaikan suatu masalah.

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 pendidikan matematika untuk memvalidasi instrumen yang telah dibuat, (3) melakukan revisi instrumen sesuai dengan saran validator, (4) melakukan uji coba instrumen penelitian, (5) menentukan reliabilitas, tingkat kesukatan, dan daya pembeda butir soal, (6) melakukan revisi instrumen berdasarkan analisis hasil uji coba.

Teknik Analisis Data

Analisis Kualitatif Soal Tes HOTS

Analisis kualitatif soal tes *HOTS* diperoleh dari hasil lembar validasi (telaah soal tes) yang dilakukan dengan cara deskriptif kualitatif. Data berupa nilai tiap butir soal hasil penilaian ahli dianalisis dengan menggunakan formula Aiken's V untuk menghitung *content validity coefficient*. Rentang angka V yang dapat diperoleh antara 0 sampai dengan 1,00.

Analisis Kuantitatif Soal Tes HOTS

Data yang diperoleh dari respon jawaban peserta didik dianalisis dengan menggunakan bantuan *software MicroCAT ITEMAN 3.00* untuk analisis soal pilihan ganda, sedangkan bantuan program *Microsoft Excel* untuk analisis soal uraian. Analisis butir soal digunakan untuk mengetahui karakteristik butir soal yang meliputi tingkat kesukaran, daya pembeda, dan penyebaran pilihan jawaban/*options* (pengecoh) untuk soal pilihan ganda, sedangkan statistik soal akan diperoleh karakteristik perangkat soal yaitu rata-rata, standar deviasi, tingkat kesukaran, daya pembeda, koefisien reliabilitas, dan *SEM*.

HASIL DAN PEMBAHASAN

Hasil Pengembangan

Hasil pengembangan dalam penelitian ini adalah instrumen asesmen *HOTS* yang berupa soal tes pilihan ganda dan uraian mata pelajaran matematika SMP kelas VIII semester 1 yang valid dan reliabel. Instrumen asesmen yang dikembangkan telah melewati dua tahap penilaian. Penilaian tahap pertama dilakukan untuk menilai kevalidan instrumen asesmen yang dilakukan oleh ahli pendidikan matematika. Penilaian tahap kedua dilakukan uji coba lapangan

yang melibatkan 178 peserta didik dari tiga sekolah, penilaian difokuskan pada karakteristik butir soal tes *HOTS*.

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.

Hasil Uji Coba Produk

Validasi oleh ahli dilakukan untuk melihat isi produk awal. Validasi ini bertujuan untuk mendapatkan masukan, saran perbaikan, dan sekaligus penilaian terhadap produk awal sebelum dilakukan uji coba terbatas. Kegiatan validasi dilakukan dengan cara memberikan naskah produk awal yaitu berupa kisi-kisi soal dan soal tes *HOTS* serta lembar validasi kepada tiga validator ahli. Selanjutnya dilakukan analisis penilaian butir soal tes *HOTS* sesuai dengan penilaian validator dengan menggunakan formula Aiken's V untuk menghitung *content validity coefficient*. Data hasil analisis validasi ahli dapat dilihat pada Tabel 1 dan Tabel 2 berikut.

Tabel 1. Hasil Analisis Validasi Ahli Soal Tes *HOTS* Pilihan Ganda

Nomor Butir Soal	Koefisien Aiken's V	Kriteria
1 – 30	0,67 – 1,00	Layak digunakan

Tabel 2. Hasil Analisis Validasi Ahli Soal Tes *HOTS* Uraian

Nomor Butir Soal	Koefisien Aiken's V	Kriteria
1 – 5	0,67 – 1,00	Layak digunakan

Berdasarkan hasil analisis menggunakan formula Aiken's V soal tes *HOTS* yang terdiri atas 30 butir soal pilihan ganda dan 5 butir soal uraian semuanya dinyatakan layak digunakan. Namun demikian, ada beberapa soal yang diperbaiki sesuai masukan dan saran dari ketiga validator yaitu mengenai perbaikan pada stem

diantaranya rumusan kalimat, kelengkapan informasi bahan pengantar pada stem, dan indikator yang kurang sesuai dengan butir soal.

Hasil uji coba terbatas diperoleh informasi waktu yang diperlukan untuk menyelesaikan soal tes *HOTS*, untuk paket soal pilihan ganda dan uraian waktu yang diperlukan masing-masing kurang lebih 120 menit. Selain itu melalui interpretasi analisis butir soal dapat diketahui kualitas butir soal berdasarkan karakteristik butir soal yang meliputi tingkat kesukaran, daya pembeda, dan juga penyebaran pilihan jawaban/*options* (pengecoh) untuk soal pilihan ganda serta dapat juga diketahui statistik soalnya.

Karakteristik Butir Soal Tes *HOTS* Pilihan Ganda Hasil Uji Coba Terbatas

Tingkat kesukaran butir soal pilihan ganda dapat dilihat pada Tabel 3 berikut.

Tabel 3. Tingkat Kesukaran Produk Awal Soal Tes *HOTS* Pilihan Ganda

Kategori	Nomor Butir Soal	Jumlah	%
TK < 0,25 (Sukar)	4, 11, 15, 22, 23, 28, 29	7	23,33
$0,25 \leq TK \leq 0,80$ (Sedang)	1, 2, 3,5, 6, 7, 8, 9, 10, 12, 13, 14, 16, 17, 18, 19, 20, 21, 24, 25, 26, 27, 30	23	76,67
TK > 0,80 (Mudah)	-	0	0

Berdasarkan Tabel 3 dapat diketahui bahwa tingkat kesukaran berkisaran pada kategori sedang sebanyak 23 butir soal (76,67%)

Daya pembeda butir soal diketahui dengan melihat koefisien korelasi point biser (r_{pbis}). Secara umum daya pembeda butir soal pilihan ganda dapat dilihat pada Tabel 4 berikut.

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

Kategori	Nomor Butir Soal	Jumlah	%
DP $\geq 0,40$ (Baik)	-	0	0
$0,30 \leq DP \leq 0,39$ (Diterima tanpa revisi)	2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29	25	83,33
$0,20 \leq DP \leq 0,29$ (Diterima dengan revisi)	1, 5, 6, 16, 30	5	16,67
DP $\leq 0,19$ (Diganti/benar-benar direvisi)	-	0	0

Berdasarkan Tabel 4 dapat diketahui bahwa daya pembeda berkisaran pada kategori diterima tanpa revisi sebanyak 25 butir soal (83,33%).

Penyebaran pilihan jawaban/*options* (pengecoh) butir soal pilihan ganda dapat dilihat pada Tabel 5 berikut.

Tabel 5. Efektifitas Pengecoh Produk Awal Soal Tes *HOTS* Pilihan Ganda

Kategori	Nomor Butir Soal	Jumlah	%
r_{pbis} kunci jawaban positif, Respons $\geq 5\%$, dan r_{pbis} pengecoh negatif	2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29	25	83,3
r_{pbis} kunci jawaban negatif, Respons $< 5\%$, dan r_{pbis} pengecoh positif	1, 5, 6, 16, 30	5	16,67

Berdasarkan Tabel 5 dapat diketahui bahwa butir soal dengan penyebaran pilihan jawaban/*options* (pengecoh) yang berfungsi baik sebanyak 25 butir soal (83,33%).

Hasil analisis karakteristik butir soal di atas, berkesimpulan bahwa jumlah butir soal yang baik dan diterima tanpa revisi, diterima dengan revisi, dan diganti/benar-benar direvisi dapat terlihat pada Tabel 6 berikut.

Tabel 6. Hasil Analisis Karakteristik Produk Awal Soal Tes *HOTS* Pilihan Ganda

Kategori	Nomor Butir Soal	Jumlah	%
Baik dan diterima tanpa revisi	2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29	25	83,33
Diterima dengan revisi	1, 5, 6, 16, 30	5	16,67
Diganti/benar-benar direvisi	-	0	0

Berdasarkan Tabel 6 butir soal kategori baik dan diterima tanpa revisi sebanyak 25 butir soal (83,33%). Butir soal kategori baik dan diterima tanpa revisi langsung digunakan dalam produk utama. Butir soal kategori diterima dengan revisi sebanyak 5 butir soal (16,67%), direvisi terlebih dahulu sebelum digunakan dalam produk utama. Butir soal kategori baik dan diterima tanpa revisi dan yang telah direvisi dirakit kembali menjadi produk utama soal tes

HOTS pilihan ganda yang akan diujikan dalam uji coba lapangan.

Statistik Butir Soal Tes *HOTS* Pilihan Ganda Hasil Uji Coba Terbatas

Statistik soal produk awal soal tes *HOTS* pilihan ganda berdasarkan uji coba terbatas dapat dilihat pada Tabel 7 berikut.

Tabel 7. Statistik Hasil Analisis Produk Awal Soal Tes *HOTS* Pilihan Ganda

Skala Statistik	
Mean	11,161
Standar Deviasi	4,378
Median	10,000
Koefisien Reliabilitas	0,702
SEM	2,390
Rata-rata tingkat kesukaran	0,372
Rata-rata daya pembeda	0,327

Berdasarkan Tabel 7 Koefisien reliabilitas tes ini 0,702 dan *SEM* untuk tes ini 2,390. Mean P/rata-rata tingkat kesukaran soal tes ini 0,372, berarti soal-soal pada tes ini adalah sedang. Mean Item-Tot./rata-rata daya pembeda dengan menghitung nilai rata-rata point biserial soal tes ini 0,327, berarti soal-soal pada tes ini sudah baik (diterima) menunjukkan bahwa soal tes *HOTS* pilihan ganda mampu membedakan peserta didik kelompok atas dan kelompok bawah.

Karakteristik Butir Soal Tes *HOTS* Uraian Hasil Uji Coba Terbatas

Tingkat kesukaran butir soal uraian dapat dilihat pada Tabel 8 berikut.

Tabel 8. Tingkat Kesukaran Produk Awal Soal Tes *HOTS* Uraian

Kategori	Nomor Butir Soal	Jumlah	%
TK $< 0,25$ (Sukar)	1e, 2b, 2c, 3d, 3e, 4a, 4b, 4c	8	42,10
$0,25 \leq TK \leq 0,80$ (Sedang)	1a, 1b, 1c, 1d, 2a, 3a, 3b, 3c, 5c	9	47,37
TK $> 0,80$ (Mudah)	5a, 5b	2	10,53

Berdasarkan Tabel 8 dapat diketahui bahwa tingkat kesukaran berkisar pada kategori sedang sebanyak 9 butir soal (47,37%) dan kategori sukar sebanyak 8 butir soal (42,10%).

Daya pembeda butir soal uraian dapat dilihat pada Tabel 9 berikut.

Tabel 9. Daya Pembeda Produk Awal Soal Tes *HOTS* Uraian

Kategori	Nomor Butir Soal	Jumlah	%
$DP \geq 0,40$ (Baik)	1a, 1b, 1c, 1d, 3a, 3c, 4a, 4b, 5b, 5c	10	52,63
$0,30 \leq DP \leq 0,39$ (Diterima tanpa revisi)	2a, 3b, 5a	3	15,79
$0,20 \leq DP \leq 0,29$ (Diterima dengan revisi)	1e, 2b, 2c, 3d, 3e, 4c	6	31,58
$DP \leq 0,19$ (Diganti/benar-benar direvisi)	-	0	0

Berdasarkan Tabel 9 dapat diketahui bahwa daya pembeda berkisar pada kategori baik dan diterima tanpa revisi sebanyak 13 butir soal (68,42%) dan kategori diterima dengan revisi sebanyak 6 butir soal (31,58%).

Hasil analisis karakteristik butir soal di atas, berkesimpulan bahwa jumlah butir soal yang baik dan diterima tanpa revisi, diterima dengan revisi, dan diganti/benar-benar direvisi dapat terlihat pada Tabel 10 berikut.

Tabel 10. Hasil Analisis Karakteristik Produk Awal Soal Tes *HOTS* Uraian

Kategori	Nomor Butir Soal	Jumlah	%
Baik dan diterima tanpa revisi	1a, 1b, 1c, 1d, 2a, 3a, 3b, 3c, 4a, 4b, 5a, 5b, 5c	13	68,42
Diterima dengan revisi	1e, 2b, 2c, 3d, 3e, 4c	6	31,58
Diganti/benar-benar direvisi	-	0	0

Berdasarkan Tabel 10 butir soal kategori baik dan diterima tanpa revisi sebanyak 13 butir soal (68,42%). Butir soal kategori baik dan diterima tanpa revisi langsung digunakan dalam produk utama. Butir soal kategori diterima dengan revisi sebanyak 6 butir soal (31,58%), direvisi terlebih dahulu sebelum digunakan dalam produk utama. Butir soal baik dan diterima tanpa revisi dan yang telah direvisi dirakit kembali menjadi produk utama soal tes *HOTS* uraian yang akan diujikan dalam uji coba lapangan.

Statistik Butir Soal Tes *HOTS* Uraian Hasil Uji Coba Terbatas

Statistik soal produk awal soal tes *HOTS* uraian berdasarkan uji coba terbatas dapat dilihat pada Tabel 11 berikut.

Tabel 11. Statistik Hasil Analisis Produk Awal Soal Tes *HOTS* Uraian

Skala Statistik	
Mean	33,935
Standar Deviasi	19,477
Median	30,000
Koefisien reliabilitas	0,910
SEM	5,838
Rata-rata tingkat kesukaran	0,378
Rata-rata daya pembeda	0,493

Berdasarkan Tabel 11 koefisien reliabilitas tes ini 0,910 dan *SEM* untuk tes ini 5,838. Rata-rata tingkat kesukaran soal 0,378, berarti soal-soal pada tes ini adalah sedang dan rata-rata daya pembeda soal 0,493 berarti soal-soal pada tes ini sudah baik (diterima). Daya pembeda sudah baik (diterima) menunjukkan bahwa soal tes *HOTS* uraian mampu membedakan peserta didik kelompok atas dan kelompok bawah.

Uji coba lapangan dilaksanakan untuk mengetahui kualitas soal tes *HOTS* berdasarkan karakteristik butir soal dan statistik soal produk utama soal tes *HOTS* pilihan ganda yang dikembangkan hasil dari uji coba terbatas.

Karakteristik Butir Soal Tes *HOTS* Pilihan Ganda Hasil Uji Coba Lapangan

Tingkat kesukaran butir soal pilihan ganda dapat dilihat pada Tabel 12 berikut.

Tabel 12. Tingkat Kesukaran Produk Utama Soal Tes *HOTS* Pilihan Ganda

Kategori	Nomor Butir Soal	Jumlah	%
$TK < 0,25$ (Sukar)	-	0	0
$0,25 \leq TK \leq 0,80$ (Sedang)	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	30	100
$TK > 0,80$ (Mudah)	-	0	0

Berdasarkan Tabel 12 dapat diketahui bahwa tingkat kesukaran soal tes *HOTS* pilihan ganda pada kategori sedang.

Daya pembeda butir soal diketahui dengan melihat koefisien korelasi point biser (r_{pbis}). Secara umum daya pembeda butir soal

pilihan ganda dapat dilihat pada Tabel 13 berikut.

Tabel 13. Daya Pembeda Produk Utama Soal Tes *HOTS* Pilihan Ganda

Kategori	Nomor Butir Soal	Jumlah	%
$DP \geq 0,40$ (Baik)	8, 15, 21, 26, 30	5	16,67
$0,30 \leq DP \leq 0,39$ (Diterima tanpa revisi)	2, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 17, 19, 22, 23, 24, 27, 28, 29	19	53,33
$0,20 \leq DP \leq 0,29$ (Diterima dengan revisi)	1, 12	2	6,67
$DP \leq 0,19$ (Diganti/benar-benar direvisi)	16, 18, 20, 25	4	13,33

Berdasarkan Tabel 13 dapat diketahui bahwa daya pembeda pada kategori baik dan diterima tanpa revisi sebanyak 24 butir soal (80%) dan kategori diterima dengan revisi dan diganti sebanyak 6 butir soal (20%).

Penyebaran pilihan jawaban/*options* (pengecoh) butir soal pilihan ganda dapat dilihat pada Tabel 14 berikut.

Tabel 14. Efektifitas Pengecoh Produk Utama Soal Tes *HOTS* Pilihan Ganda

Kategori	Nomor Butir Soal	Jumlah	%
r_{pbis} kunci jawaban positif, Respons $\geq 5\%$, dan r_{pbis} pengecoh negatif	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 17, 19, 21, 22, 23, 24, 26, 27, 28, 29, 30	25	83,33
r_{pbis} kunci jawaban negatif, Respons $< 5\%$, dan r_{pbis} pengecoh positif	1, 16, 18, 20, 25	5	16,67

Berdasarkan Tabel 14 dapat diketahui bahwa butir soal dengan penyebaran pilihan jawaban/*options* (pengecoh) yang berfungsi baik sebanyak 25 butir soal (83,33%).

Hasil analisis karakteristik butir soal di atas, berkesimpulan bahwa jumlah butir soal yang baik dan diterima tanpa revisi, diterima dengan revisi, dan diganti/benar-benar direvisi dapat terlihat pada Tabel 15 berikut.

Tabel 15. Hasil Analisis Karakteristik Produk Utama Soal Tes *HOTS* Pilihan Ganda

Kategori	Nomor Butir Soal	Jumlah	%
Baik dan diterima tanpa revisi	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 17, 19, 21, 22, 23, 24, 26, 27, 28, 29, 30	24	80
Diterima dengan revisi	1, 12	2	6,67
Diganti/benar-benar direvisi	16, 18, 20, 25	4	13,33

Berdasarkan Tabel 15 butir soal kategori baik dan diterima tanpa revisi sebanyak 24 butir soal (80%). Butir soal kategori baik dan diterima tanpa revisi digunakan langsung sebagai produk akhir. Butir soal kategori diterima dengan revisi dan diganti sebanyak 6 butir soal (20%) tidak digunakan. Butir soal kategori baik dan diterima tanpa revisi dirakit kembali menjadi produk akhir soal tes *HOTS* pilihan ganda yang siap digunakan.

Statistik Butir Soal Tes *HOTS* Pilihan Ganda Hasil Uji Coba Lapangan

Statistik soal produk utama soal tes *HOTS* berdasarkan uji coba lapangan dapat dilihat pada Tabel 16 berikut.

Tabel 16. Statistik Hasil Analisis Produk Utama Soal Tes *HOTS* Pilihan Ganda

Skala Statistik	
Mean	12,185
Standar Deviasi	4,627
Median	11,000
Koefisien reliabilitas	0,713
SEM	2,480
Rata-rata tingkat kesukaran	0,406
Rata-rata daya pembeda	0,330

Berdasarkan Tabel 16 Mean Item-Tot./rata-rata daya pembeda dengan menghitung nilai rata-rata point biserial soal tes ini 0,330, berarti soal-soal pada tes ini sudah baik (diterima) menunjukkan bahwa soal tes *HOTS* pilihan ganda mampu membedakan peserta didik kelompok atas dan bawah. Mean P/rata-rata tingkat kesukaran perangkat soal adalah 0,406, berarti tingkat kesukaran produk utama soal tes *HOTS* dalam kategori sedang. Koefisien reliabilitas perangkat soal ini besarnya 0,713 dan SEM untuk tes ini 2,480.

Karakteristik Butir Soal Tes HOTS Uraian Hasil Uji Coba Lapangan

Tingkat kesukaran butir soal uraian dapat dilihat pada Tabel 17 berikut.

Tabel 17. Tingkat Kesukaran Produk Utama Soal Tes *HOTS* Uraian

Kategori	Nomor Butir Soal	Jumlah	%
TK < 0,25 (Sukar)	1d, 1e, 2b, 2c, 3d, 3e, 4a, 4b, 4c	9	47,37
0,25 ≤ TK ≤ 0,80 (Sedang)	1a, 1b, 1c, 2a, 3a, 3b, 3c, 5b,5c	9	47,37
TK > 0,80 (Mudah)	5a	1	5,26

Berdasarkan Tabel 17 dapat diketahui bahwa tingkat kesukaran berkisar pada kategori sukar sebanyak 9 butir soal (47,37%) dan kategori sedang sebanyak 9 butir soal (47,37%).

Daya pembeda butir soal uraian dapat dilihat pada Tabel 18 berikut.

Tabel 18. Daya Pembeda Produk Utama Soal Tes *HOTS* Uraian

Kategori	Nomor Butir Soal	Jumlah	%
DP ≥ 0,40 (Baik)	1b, 1c, 1d, 2a, 3b, 3c, 4a, 5b, 5c	9	47,37
0,30 ≤ DP ≤ 0,39 (Diterima tanpa revisi)	1a, 1e, 2b, 2c, 3a, 3d, 3e, 4b, 4c, 5a	10	52,63
0,20 ≤ DP ≤ 0,29 (Diterima dengan revisi)	-	0	0
DP ≤ 0,19 (Diganti/benar-benar direvisi)	-	0	0

Berdasarkan Tabel 18 dapat diketahui bahwa daya pembeda butir soal pada kategori baik dan diterima tanpa revisi, hal ini menunjukkan bahwa soal tes *HOTS* uraian mampu membedakan peserta didik kelompok atas dan kelompok bawah.

Hasil analisis karakteristik butir soal di atas, berkesimpulan bahwa jumlah butir soal yang baik dan diterima tanpa revisi, diterima dengan revisi, dan diganti/benar-benar direvisi dapat terlihat pada Tabel 19 berikut.

Tabel 19. Hasil Analisis Karakteristik Butir Soal Produk Utama Soal Tes *HOTS* Uraian

Kategori	Nomor Butir Soal	Jumlah	%
Baik dan diterima tanpa revisi	1a, 1b, 1c, 1d, 1e, 2a, 2b, 2c, 3a, 3b, 3c, 3d, 3e, 4a, 4b, 4c, 5a, 5b, 5c	19	100
Diterima dengan revisi	-	0	0
Diganti/benar-benar direvisi	-	0	0

Berdasarkan Tabel 19 semua butir soal produk utama soal tes *HOTS* uraian kategori baik dan diterima tanpa revisi, berarti produk akhir soal tes *HOTS* uraian sudah siap digunakan.

Statistik Butir Soal Tes *HOTS* Uraian Hasil Uji Coba Lapangan

Statistik soal produk utama soal tes *HOTS* uraian berdasarkan uji coba lapangan dapat dilihat pada Tabel 20 berikut.

Tabel 20. Statistik Hasil Analisis Produk Utama Soal Tes *HOTS* Uraian

Skala Statistik	
Mean	31,657
Standar Deviasi	20,926
Median	23
Koefisien reliabilitas	0,920
SEM	5,927
Rata-rata tingkat kesukaran	0,373
Rata-rata daya pembeda	0,508

Berdasarkan Tabel 20 Koefisien reliabilitas tes ini 0,920 dan *SEM* untuk tes ini 5,927. Rata-rata tingkat kesukaran soal 0,373, berarti soal-soal pada tes ini adalah sedang dan rata-rata daya pembeda soal 0,508 berarti soal-soal pada tes ini sudah baik (diterima). Daya pembeda sudah baik (diterima) menunjukkan bahwa soal tes *HOTS* uraian mampu membedakan peserta didik kelompok atas dan kelompok bawah.

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, revisi produk hasil uji coba terbatas, dan revisi produk hasil uji coba lapangan.

Berdasarkan hasil evaluasi setelah penilaian ahli, uji coba terbatas, dan uji coba lapangan, instrumen asesmen yang dikembangkan mengalami beberapa revisi. Pertama, revisi butir soal didasarkan masukan dan saran validator. Secara umum masukan dan saran tersebut mengenai perbaikan pada stem diantaranya rumusan kalimat, kelengkapan informasi bahan pengantar pada soal, dan indikator yang kurang sesuai dengan butir soal. 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. Sebagian besar dilakukan pada penyebaran pilihan jawaban/*options* (pegecoh) yang kurang berfungsi untuk soal tes *HOTS* pilihan ganda dan perbaikan pada rumusan kalimat dan kelengkapan informasi bahan pengantar pada soal untuk soal tes *HOTS* uraian. Ketiga, revisi butir soal berdasarkan hasil uji coba terbatas produk utama soal tes *HOTS*. Butir soal kategori diterima dengan revisi dan diganti tidak digunakan (dibuang). Butir soal yang baik dan diterima tanpa revisi diverifikasi kembali dengan indikator *HOTS* untuk mengetahui semua indikator terwakili. Hasil verifikasi butir soal dirakit kembali menjadi produk akhir soal tes *HOTS* yang siap digunakan.

Kajian Produk Akhir

Produk akhir penelitian pengembangan ini adalah instrumen asesmen *HOTS* matematika SMP kelas VIII 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.

Kevalidan instrumen asesmen yang berupa soal tes *HOTS* berdasarkan pada kriteria validasi produk hasil pengembangan yang telah ditetapkan. Validasi dilakukan oleh tiga orang ahli yang berasal dari akademisi dosen Pascasarjana UNY. Validasi oleh ahli pada produk soal tes *HOTS* sudah memenuhi validitas logis. Ketiga validator menyatakan bahwa produk soal tes *HOTS* pilihan ganda dan uraian yang dikembangkan memenuhi kriteria layak digunakan.

Reliabilitas instrumen asesmen yang berupa soal tes *HOTS* berdasarkan pada hasil analisis butir soal produk utama soal tes *HOTS*. Koefisien reliabilitas yang diperoleh dari hasil analisis

soal tes *HOTS* pilihan ganda adalah 0,713 dengan *SEM* 2,480, sedangkan koefisien reliabilitas soal tes *HOTS* uraian adalah 0,920 dengan *SEM* 5,927.

Kualitas instrumen asesmen yang berupa soal tes *HOTS* berdasarkan pada hasil analisis butir soal produk utama soal tes *HOTS* yaitu menganalisis semua butir soal berdasarkan data empirik. Untuk soal tes *HOTS* pilihan ganda memiliki rata-rata tingkat kesukaran 0,406 (sedang) dan rata-rata daya pembeda 0,330 (baik), dan semua pengecoh berfungsi baik. Sedangkan untuk soal tes *HOTS* uraian memiliki rata-rata tingkat kesukaran 0,373 (sedang) dan rata-rata daya pembeda 0,508 (baik).

SIMPULAN DAN SARAN

Simpulan

Berdasarkan hasil penelitian dan pembahasan diperoleh simpulan sebagai berikut: (1) produk akhir dalam penelitian ini menghasilkan instrumen asesmen *HOTS* untuk mengukur keterampilan berpikir tingkat tinggi peserta didik SMP kelas VIII. Instrumen asesmen berupa perangkat soal tes *HOTS* yang terdiri dari 24 butir soal pilihan ganda dengan empat pilihan jawaban dan 19 butir soal uraian. Instrumen asesmen dikembangkan melalui tujuh langkah pengembangan yaitu: (a) penelitian dan pengumpulan informasi, (b) perencanaan, (c) pengembangan produk awal, (d) uji coba terbatas, (e) revisi produk awal, (f) uji coba lapangan, (g) revisi produk akhir. Kevalidan instrumen dibuktikan dengan hasil penilaian ahli yang menunjukkan bahwa instrumen layak digunakan berdasarkan telaah aspek materi, konstruksi, dan bahasa. Instrumen tersebut juga telah memenuhi kriteria reliabel. (2) Soal pilihan ganda memiliki tingkat kesukaran sedang, daya pembeda baik, semua pengecoh berfungsi baik, dan soal uraian memiliki tingkat kesukaran sedang dengan daya pembeda baik.

Saran

Berdasarkan hasil penelitian dan simpulan di atas, ada beberapa saran pemanfaatan produk akhir instrumen asesmen *HOTS* yaitu sebagai berikut: (1) peserta didik dapat menggunakan produk akhir instrumen asesmen *HOTS* sebagai bahan latihan untuk melatih keterampilan berpikir tingkat tinggi, (2) guru matematika SMP dapat menggunakan produk akhir instrumen asesmen *HOTS* untuk mengukur penguasaan pengetahuan dan keterampilan berpikir tingkat

tinggi peserta didik, (3) produk akhir instrumen asesmen *HOTS* hasil pengembangan yang berupa soal tes *HOTS* dapat digunakan sebagai acuan dalam mengembangkan instrumen asesmen *HOTS* pada SK dan KD lainnya.

DAFTAR PUSTAKA

- Arends, R. I., & Kilcher, A. (2010). *Teaching for student learning becoming an accomplished teacher*. New York and London: Routledge Taylor and Francis Group.
- Badan Standar Nasional Pendidikan (BSNP). (2006). *Standar isi untuk satuan pendidikan dasar dan menengah. Standar kompetensi dan kompetensi dasar*. Jakarta: BSNP.
- Borg, W. R. & Gall, M.D. (1983). *Educational researcher: An introduction, (4th ed.)*. New York: Longman.
- Brookhart, S. M. (2010). *How to assess higher order thinking skills in your classroom*. Virginia USA : SCD Alexandria.
- Butkowski, J., Corrigan, C., Nemeth, T., & Spencer, L. (1994). Improving student higher order thinking skills in mathematics. *Theses, Mathematics Education Research*. Saint Xavier University-IRI, Field-Based Master's Program.
- Conklin, W. (2012). *Higher-order thinking skills to develop 21st century learners*. Huntington Beach: Shell Educational Publishing, Inc.
- Depdiknas. (2007). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 16, Tahun 2007, tentang Standar Kualifikasi Akademik dan Kompetensi Guru*.
- Depdiknas. (2007). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 20, Tahun 2007, tentang Standar Penilaian Pendidikan*.
- Henningsen, M., & Stein, M.K. (1997). Mathematical task and student cognition: classroom based factors that support and inhibit level mathematical thinking and reasoning. *Journal for research in mathematics education, Vol. 28 No. 5. (Nov., 1997), pp.524-549*.
- Kemdikbud. (2013). *Hasil Konvensi Ujian Nasional, Tahun 2013*. Diambil pada tanggal 23 Oktober 2013, dari <http://www.kemdiknas.go.id/kemdikbud/sites/default/files/Hasil%20Konvensi%20OUN.pdf>.
- Kemdikbud. (2013). *Peraturan Menteri Pendidikan dan Kebudayaan RI Nomor 68, Tahun 2013, tentang Kerangka Dasar dan Struktur Kurikulum SMP/MTs*.
- Kemdikbud. (2013). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 81a, Tahun 2013, tentang Implementasi Kurikulum: Lampiran I Pedoman Penyusunan dan Pengelolaan Kurikulum Tingkat Satuan Pendidikan*.
- Krulik, S., & Rudnick, J. A. (1999). Innovative tasks to improve critical and creative thinking skills. Dalam Lee V. Stiff & Frances R. Curcio (Editor), *Developing mathematical reasoning in grades K-12, 1999 yearbook*. Reston, VA: The National Council of Teachers of Mathematics, Inc.
- Kubiszyn, T. & Borich, G. D. (2013). *Educational testing & measurement. Classroom application and practice, (10th ed.)*. New York: John Wiley & Sons, Inc.
- Maite, G & Laura, B. (2011). Effect of a play program on creative thinking of preschool children. *Journal of Psychology, vol. 14, num. 2, 2011, pp. 608-618*, Universidad Complutense de Madrid. Espana, Diambil pada tanggal 08 Oktober 2013, dari <http://www.redalyc.org/articulo.oa?id=17220620009>
- Mullis, I. V. S., Martin M. O., Foy P., & Arora A. (2012). *TIMSS 2011 international results in mathematics*. Boston: TIMSS & PIRLS International Study Center.
- Nasional Council of Teacher of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: The National Council of Teachers of Mathematics, Inc.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of student, (6th ed.)*. Boston: Pearson Education.

- O'Daffer, P. G., & Thornquist, B. A. (1993). Critical thinking, mathematical reasoning, and proof. Dalam P. S. Wilson (Editor), *In research ideas for the classroom: High school mathematics* (pp. 39-56). New York: Maxwell Macmillan International.
- Perkins, D. N. (1985). What Creative Thinking Is. Dalam Arthur L. Costa (Edited), *Developing minds: A resource book for teaching thinking* (pp.43-48). Alexandria, Virginia: ASCD.
- Preseisen, B. Z. (1985). Thinking skills: meanings and models. Dalam Arthur L. Costa (Edited), *Developing minds: A resource book for teaching thinking* (pp. 43-48). Alexandria, Virginia: ASCD.
- Resnick, L. B. (1987). *Education and learning to think*. Washington, D.C: National Academy Press.
- Sumarna Surapranata. (2007). *Panduan penulisan tes tertulis. Implementasi kurikulum 2004*. Bandung: PT Remaja Rosdakarya.
- Szetela, W. (1993). Facilitating communication for assessing critical thinking in problem solving. Dalam Webb, N. L. & Coxford, A. (Editor), *Assessment in the mathematics classroom, 1993 yearbook*. Reston, VA: The National Council of Teachers of Mathematics, Inc.
- Thompson, T. (2008). Mathematics teachers' interpretation of higher-order thinking in bloom's taxonomy. *International Electronic Journal of Mathematics Education: Diambil pada tanggal 25 April 2013, dari <http://www.doaj.org>*.
- Van de Walle, J. A. (2007). *Elementary and middle school mathematics: teaching developmentally, (6th ed.)*. United States of America: Pearson Education, Inc.