

**PENGEMBANGAN INSTRUMEN PENILAIAN BENTUK *MULTIPLE CHOICE*
BERBASIS HOTS PADA MATA PELAJARAN PENDIDIKAN AGAMA ISLAM
DI SMPN 1 UJUNG BATU ROKAN HULU**

Afini Nurul Hidayat¹, Miftahurizqa², Risnawati³

^{1,2,3} Pendidikan Agama Islam, UIN Suska Riau

¹22390125351@students.uin-suska.ac.id, ²miftahir.rizqa@uin-suska.ac.id,

³risnawati@uin-suska.ac.id

ABSTRACT

This research aims to develop an assessment instrument in the form of multiple choice based on Higher Order Thinking Skills (HOTS) in Islamic Religious Education (PAI) subjects at SMPN 1 Ujung Batu, Rokan Hulu. The assessment instruments developed are expected to measure students' high-level thinking skills, which include analysis, evaluation, and creation skills. The research method used is Research and Development (R&D) with the Borg and Gall development model which consists of ten steps, but in this study, it is simplified into seven main steps: (1) initial research and information collection, (2) planning, (3) initial product development, (4) initial field trials, (5) product revision, (6) main field trials, and (7) final product refinement. The subject of this study is grading VII A students at SMPN 1 Ujung Batu which totals 30 students. Data was collected through questionnaires and tests that were analyzed in a quantitative and qualitative descriptive manner. The results of the study show that the HOTS-based multiple choice assessment instrument developed is valid and reliable to be used in assessing students' high-level thinking skills in PAI subjects. This instrument has been proven to be able to improve students' critical thinking skills, as well as provide challenges that are appropriate to their level of cognitive development. Feedback from teachers and students also shows that this instrument can motivate students to learn more deeply and critically in understanding PAI material.

Keywords: development of assessment instruments, HOTS, PAI

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan instrumen penilaian berbentuk *multiple choice* berbasis *Higher Order Thinking Skills* (HOTS) pada mata pelajaran Pendidikan Agama Islam (PAI) di SMPN 1 Ujung Batu, Rokan Hulu. Instrumen penilaian yang dikembangkan diharapkan dapat mengukur kemampuan berpikir tingkat tinggi siswa, yang meliputi kemampuan analisis, evaluasi, dan kreasi. Metode penelitian yang digunakan adalah *Research and Development* (R&D) dengan model pengembangan Borg and Gall yang terdiri dari sepuluh langkah, namun dalam penelitian ini disederhanakan menjadi tujuh langkah utama: (1) penelitian dan pengumpulan informasi awal, (2) perencanaan, (3) pengembangan produk awal, (4) uji coba lapangan awal, (5) revisi produk, (6) uji coba lapangan utama, dan (7) penyempurnaan produk akhir. Subjek penelitian ini adalah siswa kelas VII A di SMPN 1 Ujung Batu yang berjumlah 30 siswa. Data dikumpulkan melalui kuesioner dan tes yang dianalisis secara deskriptif kuantitatif dan kualitatif.

Hasil penelitian menunjukkan bahwa instrumen penilaian berbentuk *multiple choice* berbasis HOTS yang dikembangkan valid dan reliabel untuk digunakan dalam menilai kemampuan berpikir tingkat tinggi siswa pada mata pelajaran PAI. Instrumen ini terbukti mampu meningkatkan keterampilan berpikir kritis siswa, serta memberikan tantangan yang sesuai dengan tingkat perkembangan kognitif mereka. Umpan balik dari guru dan siswa juga menunjukkan bahwa instrumen ini dapat memotivasi siswa untuk belajar lebih mendalam dan kritis dalam memahami materi PAI.

Kata kunci: Pengembangan instrumen penilaian, HOTS, PAI

A. Pendahuluan

Pendidikan di Indonesia terus mengalami perkembangan guna meningkatkan kualitas dan daya saing sumber daya manusia. Salah satu komponen penting dalam pendidikan adalah penilaian, yang berfungsi tidak hanya untuk mengukur hasil belajar siswa tetapi juga untuk mendorong pembelajaran yang lebih mendalam dan bermakna. Dalam konteks ini, penguatan kemampuan berpikir tingkat tinggi atau Higher Order Thinking Skills (HOTS) menjadi sangat penting. HOTS mencakup keterampilan analisis, evaluasi, dan kreasi, yang sangat dibutuhkan dalam menghadapi kompleksitas dan tantangan abad ke-21. (Sari, 2021)

Pembelajaran pada abad 21 ini didesain dengan dilengkapi keterampilan-keterampilan. Keterampilan berpikir kritis dan pemecahan masalah, kreativitas, inovasi, komunikasi, dan kolaborasi sebagai upaya dari tantangan yang ada. Keterampilan hidup dan karir meliputi kemampuan dalam hal fleksibilitas dan adaptif, berinisiatif dan mandiri, keterampilan sosial dan budaya, kepemimpinan dan tanggung jawab, serta produktif dalam rangka upaya pemenuhan kebutuhan masa depan. Adapun keterampilan informasi, media, dan teknologi

sehingga peserta didik peka terhadap informasi yang berkembang. (Diah Rusmala, 2019)

Dengan keterampilan pembelajaran di atas, maka kurikulum yang diterapkan haruslah dapat membentuk generasi yang mampu berpikir kritis, analitis, kreatif, inovatif, mampu memecahkan persoalan, mampu beradaptasi dengan cepat dalam lingkungan dan teknologi, serta memiliki karakter yang kuat dan positif. Oleh karena itu, perlu disusun model penilaian yang mampu memberikan solusi atas tuntutan yang ada. (Marasabessy, 2021)

Pendidikan Agama Islam (PAI) merupakan salah satu mata pelajaran yang tidak hanya mengajarkan pengetahuan agama tetapi juga berperan dalam pembentukan karakter dan moral siswa. Namun, berdasarkan pengamatan di SMPN 1 Ujung Batu, Rokan Hulu, instrumen penilaian yang digunakan saat ini masih banyak yang bersifat konvensional dan belum sepenuhnya mampu mengukur kemampuan berpikir tingkat tinggi siswa. Instrumen penilaian yang ada cenderung fokus pada kemampuan mengingat dan memahami, sehingga belum memberikan tantangan yang cukup bagi siswa untuk mengembangkan keterampilan analisis, evaluasi, dan

kreasi. (Anderson, 2001)

Kurikulum Merdeka yang diterapkan di Indonesia menekankan pentingnya pengembangan HOTS dalam proses pembelajaran. Hal ini mendorong perlunya pengembangan instrumen penilaian yang mampu mengukur HOTS secara efektif. Salah satu bentuk penilaian yang dianggap efektif dalam mengukur HOTS adalah soal *multiple choice* yang dirancang dengan baik. Soal *multiple choice* berbasis HOTS dapat mengukur berbagai aspek kognitif siswa, termasuk kemampuan untuk menganalisis, mensintesis, dan mengevaluasi informasi. (Rahmawati, 2020)

Penelitian ini bertujuan untuk mengembangkan instrumen penilaian berbentuk *multiple choice* berbasis HOTS pada mata pelajaran PAI di SMPN 1 Ujung Batu, Rokan Hulu. Instrumen ini diharapkan dapat memberikan alat ukur yang valid dan reliabel bagi guru dalam menilai kemampuan berpikir tingkat tinggi siswa. Selain itu, diharapkan instrumen ini dapat memotivasi siswa untuk belajar dengan pendekatan yang lebih mendalam dan kritis, sesuai dengan tuntutan Kurikulum Merdeka. (Kemendikbud, 2013)

Fokus penelitian ini meliputi: (1) bagaimana mengembangkan instrumen penilaian *multiple choice* berbasis HOTS yang sesuai dengan kurikulum PAI di SMPN 1 Ujung Batu, (2) bagaimana validitas dan reliabilitas instrumen yang dikembangkan, dan (3) bagaimana efektivitas instrumen tersebut dalam meningkatkan kemampuan berpikir tingkat tinggi siswa. Metode penelitian yang digunakan adalah Research and Development (R&D) dengan model pengembangan Borg and Gall yang

disederhanakan menjadi beberapa langkah utama yang sistematis. (Borg & Gall, 1982)

Hasil penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan instrumen penilaian yang lebih baik dan relevan, tidak hanya untuk mata pelajaran PAI, tetapi juga untuk mata pelajaran lainnya. Selain itu, penelitian ini juga diharapkan dapat menjadi acuan bagi para pendidik dalam merancang penilaian yang mampu mendorong pengembangan kemampuan berpikir tingkat tinggi siswa, sesuai dengan kebutuhan dan tantangan pendidikan modern. (Zohar, (2003).

B. Metode Penelitian

Penelitian ini menggunakan metode Research and Development (R&D) dengan tujuan utama mengembangkan instrumen penilaian berbentuk *multiple choice* berbasis *Higher Order Thinking Skills* (HOTS) pada mata pelajaran Pendidikan Agama Islam (PAI) di SMPN 1 Ujung Batu, Rokan Hulu. Metode R&D dipilih karena mampu menghasilkan produk yang valid, reliabel, dan praktis untuk diterapkan dalam konteks pendidikan. Model pengembangan yang digunakan dalam penelitian ini adalah model Borg and Gall yang telah disederhanakan menjadi beberapa langkah utama yang sistematis. (Sugiyono, 2016). R&D adalah metode penelitian yang digunakan untuk menghasilkan produk tertentu, dan menguji keefektifan produk tersebut. Selain itu, Borg & Gall (Sumitro dkk, 2017) menyatakan bahwa pengembangan bertujuan

menghasilkan produk berdasarkan temuan dari serangkaian uji coba, melalui perorangan, kelompok kecil, kelompok sedang, uji coba lapangan, direvisi untuk mendapatkan hasil atau produk yang memadai atau layak dipakai.

Dua jenis data yang diterapkan pada penelitian ini adalah data kualitatif dan kuantitatif. Data kualitatif didapat dari *output* lembar validasi dari tiga validator pakar yakni validator guru pendidikan agama islam, materi dan evaluasi dengan disertai kritik dan saran. Sedangkan data kuantitatif diperoleh dari analisa tiap soal dengan memakai aplikasi *Iteman* yaitu mulai dari uji validitas, reliabilitas, tingkat kesulitan soal, kemampuan pembeda soal, tipu daya soal, dan analisis keterampilan berpikir tingkat tinggi.

Data kualitatif didapat dari *output* penilaian validator yang dianalisis dengan menggunakan perhitungan skala *Guttman* yakni dari hasil “Ya” dapat nilai 1 dan dari hasil “Tidak” dapat nilai 0 (Riduwan, 2015). Adapun hasil perhitungan nilai validasi para ahli dikatakan di jenis “Sangat Lemah” saat memperoleh hasil diantara 0%-20%, guna jenis “Lemah” berada pada hasil 21%-40%, pada kriteria “Cukup” berada pada nilai 41%-60%, guna syarat “Kuat” ada di hasil 61%-80%, tapi guna syarat “Sangat Kuat” ada di hasil 81%-100%. Penetapan minimal kelayakan instrumen penilaian oleh ahli adalah pada kriteria “Cukup” (Kinanti & Subagio, 2020, p. 542).

Sedangkan data kuantitatif didapatkan dari hasil analisis 30 butir soal yang telah dikembangkan. Adapun analisis butir soal yang

pertama yakni uji validitas, butir soal dinyatakan memenuhi aspek valid apabila *r tabel* (0,3291) lebih kecil dari *r hitung* yang terdapat pada tabel *item statistic* (Emy, 2019, p. 3; Maulidia & Pahlevi, 2020, p. 139; Najihah et al., 2018, p. 3).

Tahap Penelitian

1. Penelitian dan Pengumpulan Informasi
 - Deskripsi: Mengidentifikasi kebutuhan dan masalah yang ada di lapangan.
 - Instrumen: Wawancara, observasi, dan kuesioner.
 - Rumus yang digunakan: Statistik deskriptif untuk menganalisis data awal (misalnya, persentase, frekuensi).
2. Perencanaan
 - Deskripsi: Merencanakan pengembangan instrumen penilaian.
 - Instrumen: Dokumentasi dan review literatur.
 - Rumus yang digunakan: Tidak ada rumus khusus, lebih pada pengumpulan dan analisis kualitatif.
3. Pengembangan Produk Awal
 - Deskripsi: Membuat draf awal instrumen penilaian berbentuk multiple choice berbasis HOTS.
 - Instrumen: Soal-soal multiple choice.
 - Rumus yang digunakan: Validasi isi melalui expert judgment menggunakan skala Likert.
4. Uji Coba Lapangan Awal

- Deskripsi: Menguji cobakan produk awal pada sampel kecil.
- Instrumen: Tes soal-soal multiple choice.
- Rumus yang digunakan
 - Validitas butir soal menggunakan Korelasi Point-Biserial (r_{pb}):

$$r_{pb} = \frac{M_1 - M_0}{S} \sqrt{\frac{PQ}{n}}$$

di mana:

- M_1 = rata-rata skor total siswa yang menjawab benar
- M_0 = rata-rata skor total siswa yang menjawab salah
- S = standar deviasi skor total
- P = proporsi siswa yang menjawab benar
- $Q = 1 - P$
- n = jumlah siswa
- Reliabilitas menggunakan koefisien Cronbach's Alpha (α):

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_t^2} \right)$$

Rumus α digunakan untuk mengukur reliabilitas atau konsistensi internal dari instrumen penilaian, khususnya soal-soal yang bersifat multiple choice. Reliabilitas mengacu pada seberapa konsisten hasil pengukuran jika dilakukan pengukuran berulang dengan instrumen yang sama.

- k : Jumlah butir soal.

- s_i^2 : Varians dari setiap butir soal ke-i.
- s_t^2 : Varians total dari seluruh butir soal.

5. Revisi Produk

- Deskripsi: Melakukan revisi terhadap instrumen berdasarkan hasil uji coba awal.
- Instrumen: Hasil analisis dari uji coba awal.
- Rumus yang digunakan: Tidak ada rumus khusus, lebih pada penyesuaian berdasarkan hasil analisis.

6. Uji Coba Lapangan Utama

- Deskripsi: Menguji cobakan instrumen yang telah direvisi pada sampel yang lebih besar.
- Instrumen: Tes soal-soal multiple choice.
- Rumus yang digunakan:

- Validitas kontruksi menggunakan Analisis Faktor Eksploratori (EFA)

$$\lambda_i = \frac{Cov(X_i, F)}{Var(F)}$$

di mana λ_i adalah beban faktor untuk butir ke-i, X_i adalah variabel butir soal, dan F adalah faktor yang diekstraksi.

- Reliabilitas menggunakan koefisien Cronbach's Alpha (α)

7. Penyempurnaan Produk Akhir

- Deskripsi: Melakukan penyempurnaan akhir berdasarkan hasil uji coba lapangan utama.
- Instrumen: Data hasil uji coba lapangan utama.

- Rumus yang digunakan: Tidak ada rumus khusus, lebih pada penyesuaian final.

Adapun teknik Pengumpulan Data

- Wawancara: Menggali informasi mendalam dari guru dan siswa terkait kebutuhan dan masalah dalam penilaian PAI.
- Observasi: Mengamati proses pembelajaran dan penilaian di kelas.
- Kuesioner: Mengumpulkan data dari siswa mengenai pengalaman dan tanggapan mereka terhadap instrumen penilaian yang dikembangkan.
- Tes: Mengukur kemampuan berpikir tingkat tinggi siswa dengan menggunakan instrumen yang telah dikembangkan.

Teknik Analisis Data

- Statistik Deskriptif: Menggambarkan karakteristik data awal dan hasil uji coba.
- Validitas Butir Soal: Menggunakan Korelasi Point-Biserial.
- Reliabilitas: Menggunakan koefisien Cronbach's Alpha.
- Validitas Konstruksi: Menggunakan Analisis Faktor Eksploratori (EFA).

C. Hasil Penelitian dan Pembahasan

1. Proses pengembangan Instrumen Penilaian Berbasis HOTS

Adapun hasil pengembangan instrumen penilaian berbasis HOTS yang telah diterapkan dengan menggunakan model

pengembangan Sugiyono (2015) dengan sepuluh langkah tahap pengembangan akan tetapi dibatasi hanya tujuh langkah pengembangan yakni:

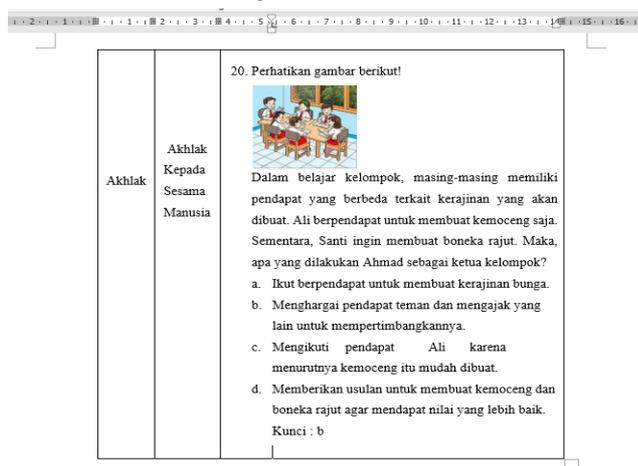
- a. Tahap pertama yakni potensi dan masalah, peneliti menentukan lokasi penelitian dan menganalisis potensi dan masalah di SMPN 1 Ujung Batu, Rokan Hulu. Berdasarkan hasil wawancara yang dilakukan oleh peneliti pada mata pelajaran PAI belum menerapkan instrumen penilaian berbasis HOTS. Kemudian siswa belum terbiasa berargumen dan memecahkan permasalahan faktual, bentuk soal yang diberikan guru sebatas pada kemampuan memahami dan mengingat.
- b. Tahap kedua yakni pengumpulan data, peneliti melakukan wawancara tidak terstruktur dengan guru mata pelajaran PAI. Hasil wawancara menunjukkan bahwa instrumen tes yang biasa

digunakan guru berpacu pada teori dan butir soal yang ada pada modul, LKS, dan kumpulan soal latihan sehingga siswa belum terbiasa memecahkan bentuk soal faktual, serta kegiatan penilaian masih dilakukan secara konvensional (paper-based test).

- c. Tahap ketiga yakni desain produk, peneliti menyusun soal HOTS dengan berbasis pilihan ganda (*multiple choice*), sejalan dengan pendapat Widiyawati et al., (2019) instrumen penilaian berbentuk pilihan ganda sering digunakan karena lebih objektif dan mudah dalam pengolahan data. Instrumen penilaian yang peneliti kembangkan berjumlah 30 soal dan diambil 25 soal terbaik.
- d. Tahap keempat yakni validasi desain, validasi instrumen penilaian dilakukan oleh tiga validator ahli yaitu validator guru PAI yang merupakan guru

agama sekaligus sebagai pimpinan di SMPN 1 Ujung Batu, validator materi yang merupakan guru PAI yang merupakan guru agama dan Validator evaluasi yang merupakan dosen ahli assessment di UIN SUSKA RIAU dengan hasil penilaian sebesar 97% dengan kriteria “Sangat Kuat”.

- e. Tahap kelima yakni revisi desain, instrumen penilaian HOTS dilakukan perbaikan dari kritik dan saran yang diberikan guru sehingga menjadi produk yang layak untuk uji coba.



Gambar 1. Instrumen Penilaian berbasis HOTS

- f. Tahap Keenam yakni uji coba produk, instrumen penilaian yang telah diperbaiki kemudian

diujicobakan kepada siswa kelas VII A SMPN 1 Ujung Batu sebanyak 30 siswa.

- g. Langkah terakhir yakni perbaikan produk 1, berdasarkan hasil analisis data secara kuantitatif, didapatkan 25 butir soal valid dan 5 soal tidak valid. Sehingga dari hasil tersebut, diambil 25 soal terbaik sebagai produk akhir instrumen penilaian yang dikembangkan peneliti. Adapun soal yang dikatakan layak terdapat pada nomor: 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13,14, 15, 18, 19, 20, 21, 23, 24, 25, 26, 27, 28, 29, 30, sedangkan 5 soal tidak layak dan harus dibuang terdapat pada nomor 3, 8, 16, 17, dan 22. Sesuai dengan pendapat Arifin & Retnawati (2017); Budiman & Jailani (2014); Desilva et al., (2020); Najihah et al., (2018); Widiyawati et al., (2019) yang juga mengembangkan instrumen berbasis HOTS tipe multiple choice test dan menetapkan soal terbaik

setelah dikatakan valid dan layak.

2. Kelayakan Instrumen HOTS

Analisis kelayakan produk diperoleh dari hasil perhitungan data secara kualitatif dan kuantitatif. Untuk uji kualitatif didapatkan dari validasi oleh tiga validator ahli yaitu validator PAI, materi dan konstruksi atau evaluasi. Hal tersebut sesuai dengan penelitian yang dikembangkan oleh Desilva et al., (2020, p. 48); Maulidia & Pahlevi (2020, p. 141); Najihah et al., (2018, p. 24); Sutami et al., (2020, p. 105); Wardany et al., (2017, p. 3) menyatakan bahwa aspek validasi instrumen penilaian HOTS terdiri dari aspek bahasa, materi dan konstruksi.

Adapun hasil penilaian validator disajikan pada tabel berikut:

Tabel 1. Hasil Penilaian Validator

Validator	Penilaian Validator	Kriteria
Validator Bahasa	87%	Sangat Kuat
Validator Materi	97%	Sangat Kuat
Validator Evaluasi	99%	Sangat Kuat
Rata-Rata	94%	Sangat Kuat

Sumber (Data Diolah Peneliti)

Sesuai dengan hasil pada tabel tersebut, diperoleh rata-rata penilaian

dari ketiga ahli tersebut sebesar 94% dengan kategori “Sangat Kuat”. Dikarenakan minimal kelayakan produk pada penelitian ini adalah pada kategori “Cukup” sehingga hasil validasi ahli pada instrumen yang dikembangkan dapat dikatakan layak. Sedangkan untuk uji kuantitatif diperoleh berdasarkan hasil analisis butir soal yang meliputi validitas, reliabilitas, tingkat kesukaran soal, daya pembeda, dan distraktor. Validitas pada soal jenis pilihan ganda memiliki tingkat valid yang tinggi (Budiman & Jailani, 2014, p. 89). Soal dikatakan valid jika r_{hitung} lebih besar dari r_{tabel} (0,3291), tingkat validitas soal dapat diketahui dengan melihat hasil analisis pada tabel *point biser*. Setelah dilakukan uji produk, didapatkan 25 soal dengan nilai lebih dari 0,3291 dan 5 soal tidak valid yaitu terdapat pada soal nomor 3 dengan nilai $0,221 < 0,3291$, soal nomor 8 dengan nilai $0,294 < 0,3291$, soal nomor 16 dengan nilai $0,158 < 0,3291$, soal nomor 17 dengan nilai $0,043 < 0,3291$ dan soal nomor 22 dengan nilai $0,067 < 0,3291$.

Reliabilitas soal dapat diketahui pada tabel *alpha*. Soal dapat dikatakan reliabel apabila skor yang didapatkan di atas 0,60 (Akhsan et al.,

2020, p. 38). Dari hasil uji coba, didapatkan nilai reliabilitas sebesar 0,77. Nilai tersebut menunjukkan $0,77 > 0,60$, sehingga soal dikatakan reliabel dengan kategori “Tinggi”. Selanjutnya adalah analisis tingkat kesukaran soal. Soal dikatakan semakin sulit jika indeks tingkat kesukaran soal kecil, dan begitu pun sebaliknya (Wantoro et al., 2019). Tingkat kesukaran soal dapat diketahui pada tabel *Prop Correct*. Adapun hasil uji coba diperoleh rata-rata soal dengan kriteria “Sedang” ditunjukkan dengan 1 soal pada nomor 18 dengan nilai sebesar 0,26 pada kriteria soal “Sulit”, 4 soal pada nomor 9, 14, 21 dan 22 dengan nilai antara 0,71- 0,89 pada kriteria soal “Mudah” dan 25 soal dengan nilai antara 0,31- 0,70 pada kriteria soal “Sedang”.

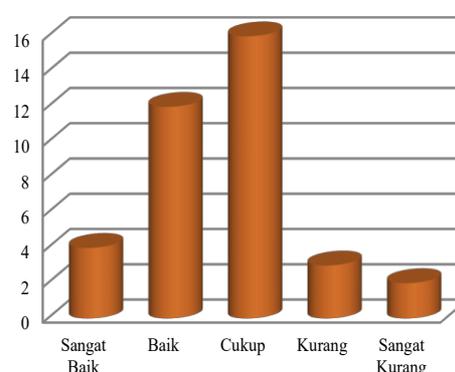
Selanjutnya yakni analisis daya pembeda soal. Tujuan analisis butir soal dengan daya pembeda berguna untuk menganalisis antara siswa *upper group* dan *lower group* (Wardany et al., 2017). Daya pembeda soal dapat diketahui pada tabel *Biser*. Adapun hasil uji coba menunjukkan tidak semua butir soal memiliki daya pembeda yang baik, sehingga butir-butir soal perlu direvisi,

atau dibuang. Terdapat 25 soal dengan kriteria “Sangat Baik”, 1 soal yakni pada soal nomor 8 dengan nilai sebesar 0,389 dengan kriteria “Baik”, 2 soal pada nomor 3 sebesar 0,280 dan soal nomor 16 sebesar 0,203 dengan kriteria “Cukup”, dan 2 soal pada nomor 17 sebesar 0,054 dan nomor 22 sebesar 0,093 dengan kriteria “Kurang”. Sehingga 2 soal dengan kriteria “Kurang” yang akan dibuang, soal tersebut juga termasuk soal tidak valid.

Pengecoh soal (*distractor*), pengecoh soal dapat diketahui pada tabel *prop endorsing*. Sesuai dengan pendapat Arifin & Retnawati (2017) pengecoh soal dikatakan efektif jika dipilih 5% dari jumlah siswa. Dari hasil uji coba pengecoh soal dapat dikatakan sudah berfungsi dengan baik karena secara keseluruhan pilihan (*option*) jawaban melebihi 5%, namun terdapat 2 soal dengan pengecoh tidak berfungsi yaitu terdapat pada nomor 8 yaitu pada *option* A dan D tidak dipilih 5% dari keseluruhan siswa dan soal nomor 22 yaitu pada *option* B dan E tidak dipilih 5% dari keseluruhan siswa sehingga 2 soal harus dibuang, karena soal tersebut juga merupakan soal tidak valid

3. Analisis Kemampuan Berpikir Tingkat Tinggi Siswa Dilihat dari Hasil Pengerjaan Soal Berbasis HOTS

Kemampuan berpikir tinggi siswa dapat diukur setelah siswa selesai mengerjakan soal. Adapun hasil penilaian kemampuan berpikir tinggi



siswa terdapat pada diagram berikut:

Gambar 2. Kemampuan Berpikir Tingkat Tinggi Siswa (Sumber: Data Diolah Peneliti)

Berdasarkan data hasil jawaban siswa dapat diketahui hasil kemampuan berpikir tingkat tinggi siswa pada tabel berikut:

Tabel 2. Perolehan Nilai Siswa

Banyak Siswa	Interpretasi Skor Siswa	Kriteria
4	81,7 – 88,4	Sangat Baik
12	64,6 – 71,5	Baik
13	44,2 – 54,4	Cukup
4	23,8 – 34	Kurang
1	20,4	Sangat Kurang

(Sumber: Data Diolah Peneliti)

Berdasarkan hasil pengujian siswa dengan kategori “kurang” dan “sangat kurang” disebabkan oleh beberapa faktor :1) keterbatasan pemahaman konseptual, 2) keterbatasan pemikiran, 3) keterbatasan penalaran, dan 4) keterbatasan keterampilan memecahkan masalah. Faktor-faktor tersebut juga dijelaskan oleh Budiman & Jailani (2014); Widiyawati et al., (2019). Faktor yang menyebabkan siswa belum berpikir kritis dikarenakan peserta didik juga kurang terlatih mengerjakan soal dengan level tinggi, sehingga diperlukan kemampuan guru untuk memberikan model pembelajaran serta menyusun pertanyaan-pertanyaan berbasis HOTS supaya siswa terlatih untuk berpikir dengan level tinggi mereka (Arifin & Retnawati, 2017).

Pembahasan Hasil Penelitian

Hasil pengembangan produk pada penelitian ini yakni pada 30 butir soal HOTS menunjukkan beberapa soal masih memerlukan perbaikan bahkan tidak layak digunakan. Soal dikatakan tidak layak tersebut merupakan soal yang dilihat dari segi kualitas soal tidak berfungsi dengan baik, yaitu dapat dilihat dari hasil analisis soal yang membuktikan beberapa soal tidak valid, tingkat pembeda pada pilihan jawaban masih kurang yakni kecenderungan siswa memilih jawaban yang sama antara satu siswa

dengan yang lain, pengecoh pilihan jawaban yang tidak berfungsi yakni ada beberapa *option* jawaban yang dipilih kurang dari 5% total keseluruhan siswa, sehingga ada beberapa soal yang dibuang dan tidak layak digunakan. Sehingga dari hasil analisis diambil beberapa soal yang terbaik dan layak sebagai produk akhir dari pengembangan instrumen pada penelitian ini. Adapun butir soal terbaik sejumlah 25 soal dan 5 soal tidak layak digunakan. Hal tersebut sejalan dengan hasil pengembangan instrumen HOTS oleh Najihah et al., (2018); Sutami et al., (2020) yang juga melakukan uji kualitatif oleh validator ahli untuk menetapkan kelayakan soal, dan uji kuantitatif dari analisis butir soal untuk merevisi dan mengambil soal-soal terbaik.

Sedangkan perolehan tingkat berpikir tingkat tinggi siswa didapatkan setelah siswa selesai mengerjakan soal. Adapun rata-rata tingkat berpikir tinggi siswa diperoleh dengan kriteria “Cukup”. Perolehan kriteria berpikir tingkat tinggi pada kriteria “Cukup” pada siswa SMPN 1 Ujung Batu tersebut dilihat dari segi jawaban siswa yang sifatnya masih homogen artinya sebagian besar siswa memiliki jawaban yang sama

pada setiap butir soal baik untuk soal dengan jawaban benar maupun yang salah. Selain itu siswa kebanyakan menjawab benar pada butir soal dengan kriteria *option* jawaban singkat. Rata-rata hasil jawaban siswa menunjukkan 40% pada kategori “baik” dan “sangat baik” sedangkan 60% pada kategori “cukup”, “kurang” dan “sangat kurang”. Adapun langkah yang dapat diambil untuk meningkatkan kemampuan berpikir tinggi siswa adalah dengan membiasakan siswa untuk berlatih mengerjakan soal level tinggi. Selain itu pendidik perlu memberikan stimulus yang inovatif yaitu dengan mengaitkan dengan permasalahan sekitar tidak hanya berpaku pada soal yang ada pada teks atau modul (Purbaningrum, 2017). Hasil uji coba pada penelitian ini juga sejalan dengan penelitian yang dilakukan oleh Purbaningrum (2017); R. R. Putri et al., (2018); Rodiana & Pahlevi (2020); Wardany et al., (2017) yang juga mengembangkan instrumen penilaian HOTS. Hasil penelitian oleh beberapa peneliti tersebut menunjukkan kemampuan berpikir tingkat tinggi siswa yang beragam artinya tidak semua siswa memiliki kemampuan berpikir pada kategori sangat baik,

ditunjukkan dengan hasil presentase dari masing-masing kategori hasil belajar yang masih diduduki oleh beberapa siswa. Penyusunan instrumen HOTS pada penelitian ini dirancang dengan berbantuan Penilaian siswa menunjukkan respon yang “Sangat Positif” sebesar 89,5%. Analisis respon siswa dalam penelitian ini juga sejalan dengan penelitian respon siswa oleh (Amri & Shobri 2020; Dewi et al., 2020; Noor 2020; Purba 2019).

D. Kesimpulan

Sesuai Hasil Pembahasan di atas dapat ditarik simpulan berikut ini: 1) Instrumen penilaian dikembangkan dan diterapkan dengan bentuk soal pilihan ganda pada ranah C4, C5 dan C 6 sejumlah 30 soal dengan diambil 25 soal terbaik dan layak, 2) Uji kelayakan Instrumen HOTS dianalisis secara kualitatif yaitu dengan penilaian tiga validator ahli dan diperoleh rata-rata penilaian 94% dengan kriteria “Sangat Kuat” dan uji data secara kuantitatif yaitu analisa pertanyaan mulai dari validitas, reliabilitas, kemampuan pembeda, level kesulitan dan *distractor* dengan menggunakan software Iteaman. 3) Kemampuan berpikir tingkat tinggi

siswa kelas VII A SMPN 1 Ujung Batu memiliki rata-rata “Cukup”

.21831/pg.v12i1.14058.

DAFTAR PUSTAKA

Akhsan, H., Wiyono, K., Novianti, R., & Dkk. (2020). Pengembangan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Materi Fluida dan Getaran Harmonis. *Jurnal Inovasi Pendidikan*, 9(2), 33–40. <http://sij-inovpend.ejournal.unsri.ac.id/index.php/sij-inovpend/index>.

Amiriono, & Daryanto. (2016). *Evaluasi & Penilaian Pembelajaran Kurikulum 2013*. Gava Media.

Amri, M., & Shobri, Y. A. (2020). Persepsi Mahasiswa Terhadap Penggunaan Quizizz Dalam Pembelajaran Akuntansi Konsolidasi Bank Syariah Di IAIN Ponorogo. *Jurnal Teknologi Informasi Dan Pendidikan*, 13(1), 128–138. <https://doi.org/https://doi.org/10.24036/tip.v13i1>.

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.

Arifin, & Retnawati, H. (2017). Pengembangan Instrumen Pengukur Higher Order Thinking Skills Matematika Siswa SMA Kelas X. *Jurnal Pendidikan Matematika*, 12(1), 98–108. <https://doi.org/http://dx.doi.org/10>

Borg, W. R., & Gall, M. D. (1983). *Educational Research: An Introduction* (4th ed.). Longman.

Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. ASCD.

Budiman, A., & Jailani. (2014). Pengembangan Instrumen Asesmen Higher Order Thinking

Desilva, D., Sakti, I., & Medriati, R. (2020). Pengembangan Instrumen Penilaian Hasil Belajar Fisika Berorientasi HOTS(Higher Order Thinking Skills) Pada Materi Elastisitas Dan Hukum Hooke. *Jurnal Kumparan Fisika* 3(1), 41–50. https://ejournal.unib.ac.id/index.php/kumparan_fisika.

Kemendikbud. (2013). *Kurikulum 2013: Kompetensi Dasar Sekolah Menengah Pertama/Madrasah Tsanawiyah (SMP/MTs)*. Kementerian Pendidikan dan Kebudayaan Republik Indonesia.

Maulidia, F., & Pahlevi, T. (2020). Pengembangan Instrumen Penilaian Tes Soal Pilihan Ganda Berbasis HOTS Pada Mata Pelajaran Administrasi Umum Jurusan OTKP SMK Negeri 1 Lamongan. *Jurnal Pendidikan Administrasi Perkantoran*, 8(1), 136–145. <https://journal.unesa.ac.id/index.php/jpap/article/view/8114>.

Najihah, A. R., Serevina, V., & Et al.

- (2018). The Development of High Order Thinking Skills (HOTS) Assessment Instrument for Temperature and Heat Learning. *Jurnal Penelitian Dan Pengembangan Pendidikan Fisika*, 4(1), 19–26.
<https://doi.org/doi.org/10.21009/1.04103>.
- Purbaningrum, K. A. (2017). Kemampuan Berpikir Tingkat Tinggi Siswa Smp Dalam Pemecahan Masalah Matematika Ditinjau Dari Gaya Belajar. *Jurnal Penelitian Dan Pembelajaran Matematika*, 10(2), 40–49.
<http://jurnal.untirta.ac.id/index.php/JPPM/article/viewFile/2029/1571>.
- Rahmawati, Y., & Ridwan, A. (2020). Pengembangan Instrumen Penilaian Berbasis HOTS pada Pembelajaran PAI. *Jurnal Pendidikan Islam*, 11(1), 21-35.
- Sari, D. P., & Susilo, H. (2021). Implementasi HOTS dalam Penilaian Pembelajaran PAI di Era Digital. *Jurnal Teknologi Pendidikan*, 23(3), 211-225.
- Skill (HOTS) Pada Mata Pelajaran Matematika SMP Kelas VIII Semester 1. *Jurnal Riset Pendidikan Matematika*, 1(2), 139–151.
<https://journal.uny.ac.id/index.php/jrpm/article/view/2671>.
- Sugiyono. (2015). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung:Alfabeta.
- Sutami, Hudiyono, Y., & Ilyas, M. (2020). Pengembangan Instrumen Asesmen Higher Order Thinking Skills (HOTS) Pada Mata Pelajaran Bahasa Indonesia SMA Dan SMK. *Jurnal Kajian Bahasa, Sastra, Dan Pengajarannya*, 3(1), 102–113.
<http://diglosiaunmul.com/index.php/diglosia/article/view/24>.
- Wantoro, J., Utama, Zuhriah, S., & Hafida, S. H. N. (2019). Pengembangan Instrumen Penilaian Pendidikan Profesi Guru Sekolah Dasar Berbasis HOTS. *Jurnal Profesi Pendidikan Dasar*, 6(1), 11–20.
<https://doi.org/10.23917/ppd.v1i1.8453>.
- Wardany, K., Sajidan, & Ramly, M. (2017). Pengembangan Penilaian Untuk Mengukur Higher Order Thinking Skills Siswa. *Jurnal Inkuiri*, 6(2), 1–16.
<http://jurnal.uns.ac.id/inkuiri>.
- Widiyawati, Y., Nurwahidah, I., & Sari, D. S. (2019). Pengembangan Instrumen Integrated Science Test Tipe Pilihan Ganda Beralasan Untuk Mengukur HOTS Peserta Didik. *Jurnal Saintifika*, 21(2), 1–14.
<https://jurnal.unej.ac.id/index.php/STF/article/view/13394>
- Zohar, A., & Dori, Y. J. (2003). Higher Order Thinking Skills and Low-Achieving Students: Are They Mutually Exclusive? *The Journal of the Learning Sciences*, 12(2), 145-181.