
**EFEKTIVITAS INSTRUMEN UJI SOAL HOTS DAN BUDAYA
PENGAMBILAN NILAI ULANGAN HARIAN PADA MADRASAH**

***THE EFFECTIVENESS OF HOTS TESTING INSTRUMENTS
AND THE CULTURE OF DAILY REVIEWS AT MADRASAH***

IIN NURMAYANTI DAN ABDUL BASID

DOI: <https://doi.org/10.31330/penamas.v3i2.516>

Iin Nurmayanti

Guru MTs Al Hikmah
Tangkil Kulon Kedungwuni
Pekalongan Jawa Tengah
Indonesia
Email: iinnurmayanti01@gmail.com

Abdul Basid

Peneliti Balai Penelitian
dan Pengembangan Agama
Jakarta
Balai Penelitian dan
Pengembangan Agama
Jakarta
Jln. Rawa Kuning, Cakung,
Jakarta, Indonesia
Email: abd.basid19@gmail.com

Naskah diterima:
13 September 2021

Revisi: 25 September 2021

Disetujui: 27 Desember 2021

Abstract

This study was conducted to examine the design of an assessment instrument based on Higher Order Thinking Skills (HOTS) in the Al-Qur'an Hadith Subject at Madrasah Tsanawiyah (MTs) Class VIII. The implementation of HOTS-based learning in the 2013 Curriculum (K-13) has not been supported by the availability of HOTS-based practice questions in the K-13 Qur'an Hadith textbook. Researchers tested the HOTS-based assessment instrument independently to support the implementation of HOTS-based learning in MTs. The purpose of this study was to describe the HOTS assessment instrument for the subject of Al-Qur'an Hadith class VIII which was developed validly and reliably, to produce a HOTS-based description and assessment question design that was equipped with an assessment grid and rubric as well as to identify the culture or habits of educators in decision making, daily test scores. This study uses a quantitative approach. Data were collected from the HOTS questionnaire technique for the subjects of Al-Qur'an Hadith class VIII semester 1 at the Madrasah Tsanawiyah level regarding the material for "Kugapai Rezeki-Mu dengan Ikhtiarku" and small-scale and large-scale trial tests. The data were analyzed qualitatively in the expert test phase of the description of the question design and quantitatively by analyzing the small-scale and large-scale trial assessment instruments. The results of this study indicate that the design of the HOTS assessment instrument is in accordance with the guidelines for the preparation of the HOTS-based assessment instrument. The results of the content validity test of the panel of experts were declared valid, namely 0.95. The HOTS assessment instrument developed is valid based on the Pearson Correlation Output for each item because it is above 0.3 and reliable because the small scale reliability is 0.852 and the large scale is 0.886 which means it is greater than 0.7. Meanwhile, educators are not used to taking daily test scores with HOTS-based assessment instruments. Educators must be able to develop self-quality, MGMP must be able to facilitate the sharing of HOTS knowledge, and the Ministry of Religion must conduct guidance and improve the quality of educators as learning evaluators.

Keywords: Assessment Instrument, HOTS, Qur'an-Hadis Subjects

Abstrak

Penelitian ini dilakukan untuk menguji desain instrumen penilaian berbasis *Higher Order Thinking Skills* (HOTS) pada Mata Pelajaran Al-Qur'an-Hadis di Madrasah Tsanawiyah (MTs) Kelas VIII. Pemberlakuan pembelajaran berbasis HOTS pada Kurikulum 2013 (K-13) belum didukung oleh tersedianya latihan soal-soal berbasis HOTS dalam buku mata pelajaran Al-Qur'an-Hadis K-13. Peneliti melakukan uji coba instrumen penilaian berbasis HOTS secara mandiri untuk mendukung penerapan pembelajaran berbasis HOTS di MTs. Tujuan penelitian ini adalah mendeskripsikan instrumen penilaian HOTS untuk mata pelajaran Al-Qur'an-Hadis kelas VIII yang dikembangkan secara valid dan *reliabel*, menghasilkan desain soal uraian dan penilaian berbasis HOTS yang dilengkapi kisi-kisi dan rubrik penilaian serta mengidentifikasi budaya atau kebiasaan pendidik dalam pengambilan nilai ulangan harian. Penelitian ini menggunakan pendekatan kuantitatif. Data dikumpulkan dari teknik kuisioner HOTS mata pelajaran Al-Qur'an-Hadis kelas VIII semester 1 tingkat Madrasah Tsanawiyah tentang materi Kugapai Rezeki-Mu dengan Ikhtiarku dan tes uji coba skala kecil dan skala besar. Data dianalisis secara kualitatif dalam tahap uji ahli desain soal uraian dan secara kuantitatif dengan menganalisis instrumen penilaian uji coba skala kecil dan skala besar. Hasil penelitian ini menunjukkan bahwa desain instrumen penilaian HOTS sesuai rambu-rambu penyusunan instrumen penilaian berbasis HOTS. Hasil uji validitas isi panel para ahli dinyatakan valid, yaitu 0,95. Instrumen penilaian HOTS yang dikembangkan valid berdasarkan *Output Pearson Corelation* tiap butir soal karena diatas 0,3 dan reliabel karena reliabilitas skala kecil 0,852 dan skala besar 0,886 artinya lebih besar dari 0,7. Sedangkan para pendidik belum terbiasa mengambil nilai ulangan harian dengan instrumen penilaian berbasis HOTS. Para pendidik harus mampu mengembangkan kualitas diri, MGMP harus bisa memfasilitasi *sharing* pengetahuan HOTS, dan Kementerian Agama harus melakukan pembinaan dan peningkatan kualitas pendidik sebagai evaluator pembelajaran.

Kata Kunci : Instrumen Penilaian, HOTS, Mata Pelajaran Qur'an-Hadis

PENDAHULUAN

Pembelajaran abad 21 dapat diartikan sebagai pembelajaran yang memberikan kecakapan abad 21 kepada peserta didik, yaitu 4C yang *Communication, Collaboration, Critical Thinking and problem solving*, dan *Creative and Innovative*. Kurikulum 2013 ingin mengantarkan pada peserta didik mencapai kecakapan di abad 21, salah satunya dengan menerapkan pembelajaran dan penilaian berbasis HOTS.

Menurut Bloom, Kemampuan dibagi menjadi dua bagian. *Pertama, Lower Order Thinking Skills* (LOTS) atau kemampuan berpikir tingkat rendah. Kemampuan ini lebih menekankan pada aspek mengingat (*remembering*), memahami (*understanding*), dan menerapkan (*applying*) dalam proses pembelajaran. *Kedua, Higher Order Thinking Skills* (HOTS) atau kemampuan berpikir tingkat tinggi, merupakan kemampuan menganalisis (*analysing*), mengevaluasi (*evaluating*), dan mengkreasi (*creating*) (Setiawati, dkk., 2018:5).

Kemampuan berpikir tingkat tinggi dalam bahasa umum dikenal sebagai *Higher Order Thinking Skills* (HOTS). Definisi kemampuan berpikir tingkat tinggi yang salah satunya dari Resnick yang dikutip oleh Yoki Ariana dkk. adalah proses berpikir kompleks yang menguraikan materi kemudian membuat kesimpulan, selanjutnya disampaikan dalam presentasi dan dianalisis, sehingga dapat tercipta hubungan yang melibatkan aktivitas mental yang paling mendasar. Kemampuan ini juga digunakan untuk menggarisbawahi berbagai proses tingkat tinggi menurut jenjang *taksonomi Bloom*.

Instrumen penilaian yang berorientasi HOTS bukanlah sebuah instrumen penilaian yang baru bagi guru dalam melakukan penilaian. Instrumen penilaian HOTS sudah digunakan pada mata pelajaran ujian nasional, tetapi untuk mata pelajaran yang diujikan pada ujian akhir madrasah berstandar nasional (UAMBN) masih jarang dilakukan, sehingga harapannya guru dapat maksimal dalam melakukan kegiatan penilaian dengan instrumen penilaian soal yang berorientasi HOTS. Penilaian hasil belajar peserta didik diharapkan dapat membantu untuk meningkatkan kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skills*), karena berpikir tingkat tinggi dapat mendorong peserta didik untuk berpikir secara luas dan mendalam tentang materi pelajaran yang dipelajarinya.

Efektivitas instrumen penilaian HOTS disini adalah tercapainya tujuan yang telah ditentukan sebelumnya yang merupakan sebuah pengukuran dimana suatu target telah tercapai sesuai dengan apa yang telah direncanakan. Pendefinisian uraian efektivitas berbeda-beda sesuai dengan dasar ilmu yang dimiliki walaupun tujuan akhir dari efektivitas adalah selalu sama yaitu pencapaian tujuan (Anggraini, t.t:14). Suatu usaha dikatakan efektif kalau usaha itu mencapai tujuan (Hartini, 2009:3). Dalam penelitian efektivitas instrumen penilaian HOTS ini dikatakan efektif apabila yang diteliti valid dan reliabel.

Instrumen penilaian HOTS di sini mengarah pada pembuatan instrumen penilaian dalam bentuk soal uraian pada mata pelajaran Al-Qur'an-Hadis kelas VIII, alasannya agar peserta didik terbiasa berlatih soal-soal yang levelnya HOTS, sehingga soal-soal UAMBN yang akan

mereka hadapi terjawab dengan mudah. Pembuatan instrumen soal difokuskan pada mata pelajaran agama Islam di lingkungan Madrasah Tsanawiyah (MTs), yaitu Al-Qur'an-Hadis. Mata pelajaran al-Quran Hadits adalah merupakan unsur mata pelajaran pendidikan agama Islam (PAI) pada Madrasah Tsanawiyah yang memberikan pendidikan kepada peserta didik supaya dapat memahami Al-Qur'an dan hadits yang merupakan sumber hukum Islam dan dapat menerapkannya dalam kehidupan sehari-hari. Oleh karena itu, tujuan mempelajari mata pelajaran Al-Qur'an hadits membantu memahami ilmu secara teoretis, sehingga dapat membentuk sikap, kepribadian juga dapat mengamalkan isi kandungan Al-Qur'an dan hadits sebagai petunjuk atau pedoman dalam kehidupan sehari-hari (Umam, 2008:1), selain memang mata pelajaran Al-Qur'an hadits termasuk mata pelajaran yang diujikan dalam Ujian Madrasah Berstandar Nasional atau UAMBN.

Adanya penerapan HOTS dalam pengembangan kurikulum 2013 standar isi tentang pengembangan kebutuhan peserta didik untuk berpikir kritis dan analisis, belum diimbangi dengan soal yang tersedia dalam buku paket mata pelajaran Al-Qur'an-Hadis kurtilas. Fenomena tersebut menjadikan guru sangat membutuhkan pengembangan intrumen penilaian berupa soal HOTS dalam mata pelajaran Al-Qur'an-Hadis Kurtilas.

Sejalan dengan hal diatas, pengawas madrasah juga menyarankan pada guru-guru MTs untuk mata pelajaran agama perlu mencoba menerapkan instrumen penilaian berbasis HOTS agar peserta didik terbiasa berpikir kritis dengan mengerjakan soal-soal dan pembelajaran yang berorientasi pada

kemampuan berpikir tinggi. Harapannya agar hasil UAMBN menjadi lebih baik lagi (Aminudin, 2018), mengingat hasil pengukuran capaian peserta didik dalam Ujian Akhir Madrasah Berstandar Nasional (UAMBN) tahun 2019 menunjukkan bahwa peserta didik masih lemah dalam kemampuan berpikir tingkat tinggi.

Model-model soal dari *Trends in Internasional Mathematics and Science Study* (TIMMS) dan *Programme for International Student Assessment* (PISA) ternyata memakai desain soal yang mengandung kriteria *Higher Order Thinking Skills*/(HOTS) yang bangsa Indonesia belum membiasakan di dalam kegiatan belajar di sekolah tingkat dasar wajib belajar (Wajar) 12 tahun. Maka di dalam kurikulum 13 pemerintah ingin penerapan *Higher Order Thinking Skills* (HOTS) dalam pelaksanaan pendidikan dan penilaiannya, tujuannya agar ketika ada soal-soal yang diujikan itu mengandung HOTS baik itu tingkat internasional maupun nasional maka peserta didik sudah terbiasa dan mudah untuk menjawabnya.

Sedangkan hasil pengukuran capaian peserta didik berdasar Ujian Akhir Nasional (UAN) dan Ujian Akhir Madrasah Berstandar Nasional (UAMBN) ternyata selaras dengan capaian *Programme for International Student Assessment* (PISA) maupun *Trends in Internasional Mathematics and Science Study* (TIMMS) yaitu hasil UAN tahun 2019 menunjukkan bahwa peserta didik masih lemah dalam kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skills*) (Kebudayaan, 2018:5) yaitu proses kognitif C4 dalam kemampuan menganalisis (*analysing*), proses kognitif C5 dalam kemampuan mengevaluasi/ menilai (*evaluating*), dan proses kognitif

C6 dalam kemampuan mengkreasi/mencipta (*creating*). Rerata Ujian Madrasah Berstandar Nasional (UAMBN) tahun 2019 rumpun Pendidikan Agama Islam (PAI) yaitu mata pelajaran Al-Qur'an-Hadis, fikih, dan sejarah kebudayaan Islam pada MTs se-Kabupaten Pekalongan juga masih rendah. Berdasarkan hal tersebut maka diperlukan adanya pengembangan instrumen penilaian yang disesuaikan dengan materi yang mengandung penalaran, argumentasi dan kreativitas dalam menyelesaikannya, sesuai soal-soal seperti dalam *Trends in Internasional Mathematics and Science Study* (TIMM) dan *Programme for International Student Assessment* (PISA) di tingkat internasional.

Kebijakan pemerintah dalam pengembangan kurikulum 2013 terdapat penyempurnaan yang meliputi: a) Standar isi, yaitu mengurangi materi yang tidak relevan serta pendalaman dan perluasan materi yang relevan bagi peserta didik serta diperkaya dengan kebutuhan peserta didik untuk berpikir kritis dan analisis sesuai dengan standar internasional; b) Standar penilaian, dengan mengadaptasi secara bertahap model-model penilaian standar internasional (Setiawati, dkk., 2018:5). Standar internasional disini dapat kita contohkan yaitu soal-soal seperti dalam *Trends in Internasional Mathematics and Science Study* (TIMM) dan *Programme for International Student Assessment* (PISA) di tingkat internasional.

Berdasarkan latar belakang masalah diatas maka diperlukan adanya pengembangan instrumen penilaian yang disesuaikan dengan materi yang mengandung penalaran, argumentasi dan kreativitas dalam menyelesaikannya. Fokus

obyek yang diteliti pada saat ini adalah efektivitas instrumen penilaian kemampuan berpikir tingkat tinggi (HOTS) dalam mata pelajaran Al-Qur'an-Hadis di Madrasah Tsanawiyah (MTs) kabupaten Pekalongan dengan mengembangkan instrumen penilaian yang berbentuk soal uraian. Lebih spesifik kami akan mengambil mata pelajaran Al-Qur'an hadits kelas VIII untuk dilakukan *treathment* penilaian dengan instrumen penilaian/soal yang mengandung kemampuan berpikir tingkat tinggi (HOTS) yang disesuaikan dengan silabus yang terperinci dalam kompetensi inti dan kompetensi dasar.

Penelitian ini menggunakan kelas VIII MTs Al-Hikmah Proto, kecamatan Kedungwuni, kabupaten Pekalongan sebanyak 4 kelas, yaitu kelas A, B, C dan D sebagai sampel skala besar dan kelas A sebagai sampel skala kecil. Peneliti hanya menggunakan satu pokok bahasan mata pelajaran Al-Qur'an-Hadis kelas VIII yang disesuaikan dengan silabus yang terperinci dalam kompetensi inti dan kompetensi dasar, selanjutnya diharapkan para pendidik mampu mengembangkan pada kelas, materi, situasi dan kondisi yang berbeda. Pemilihan sampel di tingkat SMP/MTs karena anak tingkat SMP/MTs yang berumur mulai 12 tahun ke atas, mereka sudah mempunyai kemampuan menguji hipotesis, mengembangkan kemampuan nalar secara ilmiah, mampu menyadari kemampuan kognitif/pengetahuannya, dan mampu mengembangkan pengetahuannya menjadi efisien atau tidak.

Peneliti memilih MTs al-Hikmah Proto karena nilai ujian Al-Qur'an-Hadis kemarin kurang begitu bagus. Hal tersebut dapat dibuktikan pada tabel berikut ini.

Tabel 1. Hasil Rata-rata Nilai UAMBN MTs al-Hikmah Proto

Tahun Pelajaran/TP	Jumlah Peserta Didik	Nilai Rata-rata	Ket. Pelaksanaan Ujian
2015 – 2016	143 Peserta Didik	74	Kertas dan kesepakatan nilai se-Kabupaten
2016 – 2017	132 Peserta Didik	73	Kertas dan kesepakatan nilai se-Kabupaten
2017 – 2018	162 Peserta Didik	61	Komputer dan nilai murni
2018 – 2019	127 Peserta Didik	62	Komputer dan nilai murni

Sumber: Data dari Istitlah, S. Ag (Guru Al-Qur'an-Hadis) dan Isma Fuaida, M. Si (Wakil Kepala Bagian Kurikulum)

Rumusan masalah penelitian ini dituangkan dalam dua pertanyaan penelitian, yaitu: (1) Bagaimana pendidik mengambil nilai ulangan harian peserta didik?; (2) bagaimana desain instrumen tes HOTS untuk mengukur kemampuan berpikir tingkat tinggi dalam mata pelajaran Al-Qur'an Hadits kelas VIII?; (2) bagaimana uji keterpakaian instrumen tes HOTS Mata Pelajaran Al-Qur'an-Hadis Kelas VIII?; dan (3) bagaimana validitas dan reliabilitas instrumen tes HOTS Mata Pelajaran Al-Qur'an-Hadis Kelas VIII?

METODE PENELITIAN

Peneliti menggunakan research and development (R&D) dalam pengembangan instrumen ini. Pendekatan ini dapat diartikan sebagai cara ilmiah untuk meneliti, merancang, memproduksi dan menguji validitas produk yang telah dihasilkan. Kegiatan penelitian dan pengembangan dapat disingkat menjadi 4P, yaitu penelitian, perancangan, produksi dan pengujian (Sugiyono, 2017:30). Produk riset ini adalah konstruksi soal berbasis HOTS yang siap diujikan untuk mengukur kognitif peserta didik (KI3).

Peneliti melakukan kajian kebutuhan instrumen penilaian berbasis HOTS,

merancang soal tes uraian berbasis HOTS, uji keterbacaan, uji skala kecil dan uji skala besar. Soal tes uraian disusun sesuai materi yang dijabarkan dalam kompetensi dasar (KD). Uji keterbacaan instrumen penilaian menggunakan 30 responden. Penelitian ini menggunakan Kelas VIII MTs Al-Hikmah Tangkil Kulon Pekalongan (4 kelas), 1 kelas untuk uji coba sampel instrumen penilaian skala kecil dan 3 kelas untuk uji coba sampel instrumen penilaian skala besar.

Pengumpulan data menggunakan teknik kuesioner dari uji ahli tentang desain instrumen soal HOTS mata pelajaran Al-Qur'an-Hadis dan tes untuk uji skala kecil dengan tujuan mencari kevalidan dan realibilitas desain instrumen soal HOTS mata pelajaran Al-Qur'an-Hadis dan uji coba soal untuk uji skala besar dengan tujuan membuat penilaian HOTS.

Teknik analisis data secara singkat menggunakan uji ahli dengan teknik *Walk through*. *Walk through* adalah suatu cara untuk mengevaluasi atau memvalidasi suatu rancangan yang dilakukan oleh ahli di bidangnya secara langsung dipilih karena memungkinkan peneliti berinteraksi secara langsung dengan ahli sebagai validator. *Walk through* ini para pakar menganalisis secara deskriptif dengan mempertimbangkan komentar atau saran yang ada, (Safari, 2017:341) analisis terhadap hasil uji coba skala kecil dan skala besar menggunakan statistik parametrik (Santosa, t.t:7).

HASIL PENELITIAN DAN PEMBAHASAN

Hasil penelitian ini menggunakan alur research and development yang disingkat dengan 4P, yaitu penelitian, perancangan (desain), produksi (penyusunan soal) dan

pengujian. Penelitian dalam alur ini adalah penelitian tentang budaya pendidik dalam mengambil nilai ulangan harian berbasis HOTS dengan menggunakan langkah *Taksonomi Bloom*.

Desain Instrumen Tes HOTS

Langkah dalam penelitian dan pengembangan tes hasil belajar, pembuatan soal ulangan harian isian non obyektif HOTS menggunakan sembilan langkah, yaitu: 1. Menyusun spesifikasi tes; 2. Menulis soal tes; 3. Menelaah soal tes; 4. Melakukan uji coba tes; 5. Menganalisis butir soal; 6. Memperbaiki tes; 7. Merakit tes; 8. Melaksanakan tes; dan 9. Menafsirkan hasil tes (Mardapi, 2018:88).

Langkah-langkah pengembangan penyusunan soal HOTS yang dilakukan peneliti, yaitu: 1. Menyusun spesifikasi tes (menggunakan bentuk tes non obyektif) 2. Menulis soal tes (soal uraian HOTS dilengkapi dengan pedoman peskoran dan kunci jawaban) 3. Menelaah soal tes (soal ditelaah oleh ahli dan akademisi) 4. Melakukan uji coba tes (uji coba tes digunakan untuk memperoleh data tentang reliabilitas, validitas, daya beda, tingkat keterbacaan, dan tingkat kesukaran) 5. Menganalisis butir soal (analisis data tentang reliabilitas, validitas, daya beda, tingkat keterbacaan, dan tingkat kesukaran menghasilkan kualitas soal yang diharapkan) 6. Memperbaiki tes (dengan adanya kemungkinan soal yang dibuang, diperbaiki maupun soal yang sudah tidak perlu diperbaiki) 7. Merakit tes (mengurutkan nomer urut soal dan pengelompokkan bentuk soal) 8. Melaksanakan tes (dilaksanakan sesuai dengan waktu yang telah ditentukan dan diawasi oleh pengawas atau guru

untuk mendapatkan hasil tes yang jujur) 9. Menafsirkan hasil tes (data kuantitatif skor kemudian ditafsirkan menjadi nilai).

Adapun secara spesifik dalam menyusun spesifikasi tes/soal HOTS yang peneliti lakukan adalah: 1. Menganalisis KD yang dapat dibuat soal-soal HOTS; 2. Menyusun kisi-kisi soal; 3. Merumuskan stimulus yang menarik dan kontekstual; 4. Menulis butir pertanyaan sesuai dengan kisi-kisi soal; dan 5. Membuat pedoman penskoran (rubrik) atau kunci jawaban. Desain instrumen penilaian atau butir soal selain sesuai dengan pedoman pembuatan soal HOTS dan rambu-rambu dari beberapa uji ahli yang berupa saran dan pendapat juga menjadikan *Taksonomi Bloom* revisi sebagai acuan dalam menyusun soal mulai dari jenjang mengingat (C1) sampai jenjang mencipta/mengkreasi (C6). Adapun perbedaan antara *taksonomi Bloom* Versi lama dan Versi Revisi dapat dilihat dalam tabel di bawah ini.

Tabel 2. Taksonomi Bloom Versi Lama dan Versi Revisi

Taksonomi Bloom Versi Lama	Taksonomi Bloom Versi Baru
Evaluation/Evaluasi	Creating/Mencipta
Synthesis/Sintesis	Evaluating/Mengevaluasi
Application/Aplikasi	Analysing/Menganalisis
Comperhension/Pemahaman	Understanding/Memahami
Knowledge/Pengetahuan	Remembering/Mengingat

Sumber: Ismet Basuki, Hariyanto, *Assesmen Pembelajaran*, hlm. 15.

Tabel 3. Proses Kognitif Sesuai dengan Level Kognitif Bloom

PROSES KOGNITIF		DEFINISI
C4	Menganalisis	Memecahkan materi ke dalam bagian-bagiannya dan menentukan bagaimana bagian-bagian itu terhubung antara bagian dan ke struktur atau tujuan keseluruhan
C5	Menilai/ Mengevaluasi	Membuat pertimbangan berdasarkan kriteria atau standar
	Mengkreasi/ Mencipta	Menempatkan unsur-unsur secara bersama-sama untuk membentuk keseluruhan secara koheren atau fungsional; menyusun kembali unsur-unsur ke dalam pola atau struktur baru

Sumber: Wiwik Setiawati dkk pada *Buku Penilaian Berorientasi pada Kemampuan Berpikir Tingkat Tinggi (Program*

Peningkatan Kompetensi Pembelajaran Berbasis Zonasi) hlm. 5.

Tabel 4. Kata Kerja Operasional dalam Perumusan Indikator

Dimensi Proses Kognitif dan Kategori	Kata Kerja Operasional untuk Perumusan Indikator/Tujuan
Menganalisis (C4)	Pengertian: Memecah-mecah materi jadi bagian-bagian penyusunnya dan menentukan hubungan hubungan antarbagian itu dan hubungan antara bagian-bagian tersebut dan keseluruhan struktur atau tujuan
4.1. Membedakan	membedakan, menganalisis perbedaan, mengorganisasikan, membuat diagram, menunjukkan bukti, menghubungkan, menganalisis kesalahan, menganalisis kelebihan, menunjukkan sudut pandang
4.2. Mengorganisasi	mengorganisasikan, membuat diagram, menunjukkan bukti, menghubungkan
4.3. Mengatribusikan	menganalisis kesalahan, menganalisis kelebihan, menunjukkan sudut pandang
Mengevaluasi (C5)	Pengertian: Mengambil keputusan berdasarkan kriteria dan atau standar
5.1. Memeriksa	memeriksa, menunjukkan kelebihan, menunjukkan kekurangan, membandingkan, menilai, mengkritik
5.2. Mengkritik	menilai, mengkritik
Mencipta (C6)	Pengertian: Memadukan bagian-bagian untuk membentuk sesuatu yang baru dan koheren atau untuk membuat suatu produk yang orisinal
6.1. Merumuskan	Merumuskan, merencanakan, merancang, mendisain, memproduksi, membuat
6.2. Merencanakan	merencanakan, merancang, mendisain
6.3. Memproduksi	memproduksi, membuat

Indikator untuk menyusun butir soal HOTS disesuaikan dengan kata kerja operasional. Kata kerja butir soal disesuaikan dengan kedalaman soal yang diinginkan, diantaranya dapat menggunakan kata kerja operasional pada *taksonomi Bloom* yang telah direvisi. Kata kerja disini dipakai sebagai cara berpikir dalam menjawab pertanyaan, bukan harus dipakai dalam kalimat pertanyaan. Adapun beberapa kata kerja operasional yang terangkum dapat kita lihat apada tabel 4 di bawah ini:

Tabel 5. Kata Kerja Operasional Ranah Kognitif (Ariyana, 2018:10)

KEMAMPUAN YANG DIUKUR	KATA KERJA OPERASIONAL		
Kemampuan Menganalisis (C4)	Menganalisis	Mengkorelasikan	Memaksimalkan
	Mengaudit	Merasionalalkan	Memerintahkan
	Memecahkan	Menguji	Mengedit
	Mengagaskan	Mencerahkan	Mengaitkan
	Mendeteksi	Menjelajah	Memilih
	Mendiagnosis	Membagikan	Mengukur
Kemampuan Mengevaluasi (C5)	Menyeleksi	Menyimpulkan	Melatih
	Memerinci	Menemukan	Menransfer
	Menominasikan	Menelaah	Mendiagramkan
	Membandingkan	Membuktikan	Menugaskan
	Menyimpulkan	Memutuskan	Menvaliditasi
	Menilai	Memisahkan	Mengedit
Kemampuan Kreasi / Mencipta (C6)	Mengarahkan	Mempridiksi	Mendukung
	Mengkritik	Memperjelas	Memilih
	Menimbang	Mengukur	Memproyeksikan
	Menafsirkan	Merangkum	Memerinci
	Mempertahankan	Mempertahankan	
	Mengabstraksi	Menanggulangi	Menfasilitasi
Kemampuan Kreasi / Mencipta (C6)	Mengatur	Menghubungkan	Membentuk
	Menganimasi	Menciptakan	Merumuskan
	Mengumpulkan	Mengkreasikan	Menggeneralisasi
	Mengkategorikan	Mengoreksi	Menggabungkan
	Mengkode	Merancang	Memadukan
	Mengkombinasikan	Merencanakan	Membatas
	Menyusun	Mendikte	Mereparasi
	Mengarang	Meningkatkan	Menampilkan
	Membangun	Memperjelas	Menyiapkan
	Merekonstruksi	Merangkum	Memproduksi
	Membuat		

Sumber: Yoko Ariyana Dkk, Buku Pegangan Pembelajaran HOTS, hlm. 10

Peneliti dalam menulis soal uraian menggunakan teori karakteristik soal-soal HOTS, antara lain: mengukur ketrampilan berpikir tingkat tinggi, berbasis permasalahan kontekstual, menarik, dan mengusung kebaruan.

Indikator soal dan butir soal disusun berdasarkan teori kemampuan berpikir tingkat tinggi yang menggunakan pembagian level kognitif HOTS dari Bloom serta kata kerja operasional soal HOTS sebagai acuan dalam kerangka berpikir. Indikator soal dikembangkan dengan mengakomodir kemampuan menganalisis, menilai atau mengevaluasi, dan mengkreasi atau mencipta. Satu indikator pencapaian dijadikan menjadi 1 sampai 2 butir soal kemampuan berpikir tingkat tinggi.

Penilaian tes uraian untuk mengukur ketrampilan berpikir tingkat tinggi dilakukan di dalam uji coba soal skala besar. Penilaian menggunakan pedoman penskoran. Kurikulum 2013 mata pelajaran Al-Qur'an-

Hadis kelas VIII menjadi dasar dalam menentukan materi tes, sehingga materi sesuai dengan kurikulum yang berlaku pada saat ini.

Menurut Rusilowati (Rusilowati, 2017:54) yang dikutip oleh Arta mahindra Diputera bahwa spesifikasi instrumen disusun dari 3 faktor, yaitu faktor kontekstual luar, atribut internal dan spesifikasi tes di masa yang akan datang. Faktor kontekstual luar yang dimaksud peneliti disini terkait dengan karakteristik peserta tes, pengalokasian waktu, pengadministrasian tes dan standarisasi tes. Karakteristik peserta tes adalah peserta didik kelas VIII tingkat MTs. Alokasi waktu 2 jam pelajaran atau 80 menit dengan jumlah butir soal 14 soal pada skala kecil dan pada skala besar setelah adanya revisi menjadi 10 soal. Pengadministrasian tes guru cukup menggunakan lembar jawab peserta tes sebagai bahan analisis soal dan mendapatkan nilai tes. Standarisasi tes yang dimaksud disini adalah adanya petunjuk soal dalam mengerjakan, pemberitahuan waktu yang disediakan, dan materi tes yang akan diujikan. Atribut internal yang dimaksud peneliti adalah adanya kisi-kisi soal sebagai acuan atau gambaran peserta tes tentang materi apa yang akan keluar dalam ujian ulangan harian nanti. Spesifikasi tes yang dimaksud peneliti adalah tes uraian yang lebih sesuai bagi guru untuk melihat pengukuran kemampuan belajar peserta didik dalam ketrampilan berpikir tingkat tinggi.

Adanya pedoman penskoran memudahkan guru untuk membuat penilaian. Kategori penilaian dibuat menjadi 5 kategori dengan penambahan 1 kategori untuk mengakomodir jawaban peserta didik yang tidak sesuai dengan 5 kategori. Guru

dengan menggunakan pedoman penskoran tersebut diharapkan tidak perlu untuk memberikan nilai bonus atau nilai tambahan yang bersifat subyektif kepada peserta tes ulangan harian.

Dalam pembuatan pedoman penskoran masing-masing guru berbeda tergantung jumlah soal dan rentan nilai yang akan dibuat. Tabel yang ada di bawah ini adalah pedoman penskoran soal uji skala kecil yang dikembangkan oleh peneliti, hasil dari penilaian nantinya akan dianalisis untuk mencari validitas dan reliabilitas instrumen soal.

Tabel 6. Pedoman Penskoran Nilai yang Dikembangkan

PEDOMAN PENSKORAN		
Satuan Pendidikan	: MTs	Jumlah Soal : 14
Mata Pelajaran	: Al-Qur'an-Hadis	Bentuk Soal : Uraian
Kelas/Semester	: VIII/1	Alokasi Waktu : 2 JP (80 menit)
Kurikulum	: 2013	
Skor	Interpretasi	Keterangan
5	Jawaban benar Penjelasan Penuh/Lengkap	Peserta didik menjawab soal dengan jelas dan benar, sesuai dengan kategori.
3	Jawaban Benar tapi penjelasan tidak lengkap	Peserta didik menjawab dengan jelas dan benar, tetapi tidak lengkap. Maksudnya: Peserta didik hanya menjawab 1 atau lebih komponen dalam satu kategori tetapi tidak lengkap.
0	Tidak menjawab sesuai kategori	Peserta didik tidak menjawab soal sesuai dengan kategori kunci jawaban.
Kategori Penilaian " jawaban tidak mencakup kategori di atas".		
2	Jawaban masih berhubungan dengan pertanyaan	Peserta didik menjawab soal. Namun, tidak mencakup 2 kategori sebelumnya dan jawaban masih ada hubungan dengan pertanyaan.
1	Jawaban tidak ada hubungan dengan pertanyaan	Peserta didik menjawab soal. Namun, tidak mencakup 2 kategori sebelumnya dan jawaban tidak berhubungan dengan pertanyaan.

Produksi Soal

Hasil penelitian ini adalah instrumen penilaian atau soal uraian berbasis HOTS untuk mata pelajaran Al-Qur'an-Hadis Kelas VIII semester 1 dengan materi Kugapai Rezeki-Mu dengan Ikhtiarku. Berikut adalah desain soal, kunci jawaban dan pedoman penskorannya:

Tabel 7. Desain Soal, Kunci Jawaban dan Penskoran

No	Desain Soal	Kunci Jawaban	Skor
1	Allah Swt telah menjamin rezeki manusia dan seluruh makhluk hidup. Simpulkan pengertian rezeki menurut pendapatmu?	Pengertian rezeki dari Allah adalah segala sesuatu dari Allah SWT yang berdaya guna, serta dapat dimanfaatkan oleh manusia sebagai sumber kehidupan.	
	Sesuatu yang berdaya guna		0-5
	Sesuatu yang dapat dimanfaatkan		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		10
2	Jawablah dengan bahasamu sendiri, bagaimana cara Allah mendatangkan rezeki kepada makhluknya ?	Cara Allah mendatangkan rezeki kepada umatNya adalah dengan cara yang tidak terduga-duga oleh manusia dan juga atas usaha manusia baik dengan ikhtiar sedekah, bersyukur dan lain-lain.	
	Dengan cara tidak terduga-duga		0-5
	Dengan usaha manusia		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		10
3	Seluruh makhluk yang ada di muka bumi telah dijamin rezekinya oleh Allah Swt. Buatlah jawaban dengan bahasamu sendiri, bagaimana cara akal pikiran manusia dapat memperoleh rezeki yang dijamin oleh Allah Swt tersebut?	Cara manusia memperoleh rezeki dari Allah yaitu dengan berusaha, bekerja dengan sungguh-sungguh dan berdoa	
	Dengan berusaha dan bekerja sungguh-sungguh		0-5
	Dengan berdoa		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		10
4	Ajaran di dalam Al-Qur'an-Hadis mendorong manusia mencari rezeki yang halal dan bagus agar kebutuhan hidup manusia terpenuhi. Simpulkan apa yang dimaksud halal dan bagus pada kalimat di atas!	Halal dan bagus artinya mencari rezeki dengan cara yang baik dengan landasan takwa kepada Allah SWT.	
	Dengan cara yang baik		0-5
	Dengan landasan takwa kepada Allah SWT		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		10
5	<p>الَّذِي أَطْعَمَهُمْ مِنْ جُوعٍ وَآمَنَهُمْ مِنْ خَوْفٍ</p> <p>Pada QS. Al-Quraisy ayat 4 diatas, rezeki yang diberikan oleh Allah kepada orang-orang Quraisy adalah makanan dan mengamankan (keamanan). Pilihlah kata yang menunjukkan arti memberi makan kepada mereka dan mengamankan mereka!</p>	Memberi makan kepada mereka: اطعمهم, mengamankan mereka: آمنهم	
	اطعمهم		0-5
	آامنهم		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		10
6	<p>أَلَمْ نَفْرَحْ لَكَ صَدْرَكَ</p> <p>Pada QS. Al-Insyirah ayat 1, disebutkan bahwa Kami (Allah) telah melapangkan dadamu (Muhammad Saw). Pilihlah kata yang menunjukkan arti melapangkan!</p>	Arti melapangkan: نشرح	
	نشره		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		5
7	<p>إِنلَفَهِم رِحْلَةَ الِشَّيْءِ وَالصَّيْفِ</p> <p>2. (yaitu) kebiasaan mereka bepergian pada musim dingin dan musim panas. Jelaskan maksud dari kebiasaan bepergian pada musim dingin dan musim panas orang-orang Quraisy pada ayat di atas!</p>	Perjalanan dagang suku Quraisy pada musim dingin ke yaman, dan pada musim panas ke Syam dalam setiap tahunnya	
	Perjalanan dagang suku Quraisy pada musim dingin ke yaman		0-5
	Perjalanan dagang suku Quraisy ke Syam pada musim panas		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		10
8	<p>أَلَمْ نَفْرَحْ لَكَ صَدْرَكَ</p> <p>1. Bukankah Kami telah melapangkan untukmu dadamu? Dari ayat dan terjemahannya diatas, jelaskan maksud "menghilangkan daripadamu bebanmu"?</p>	Menghilangkan beban/ ujian ketika nabi Muhammad berdakwah	
	Menghilangkan beban/ujian		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		5
9	<p>Jelaskan yang termasuk rezeki dalam kandungan QS. Al-Quraisy!</p>	Rezeki kepada kaum Quraisy berupa makanan dan rasa aman	
	Makanan		0-5
	Rasa aman		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		10
10	<p>Jelaskan yang dimaksud rezeki yang ada dalam kandungan QS. Al-Insyirah!</p>	Rezeki dalam surah al-Insyirah berupa peninggian sebutan untuk Nabi Muhammad SAW, dengan ditetapkan bersanding dengan nama Allah SWT.	
	Meninggikan nama Nabi Muhammad		0-5
	disandingkan dengan nama Allah SWT		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		10
11	<p>Allah mengingatkan kepada orang-orang Quraisy bahwa setelah Allah Swt memberi kenikmatan kepada mereka maka mereka diminta untuk mengikuti perintahNya. Jelaskan apa perintah Allah yang sesuai dengan QS. Al-Quraisy ayat 3!</p>	Perintah untuk bersyukur dan beribadah kepada Allah SWT	
	Bersyukur		0-5
	Beribadah kepada Allah SWT		0-5
	Jawaban tidak mencakup kategori di atas		1-2
	Skor maksimal		10

12	Allah memberikan penghargaan kepada Nabi Muhammad Saw, yaitu meninggikan sebutan untuk Nabi yang sesuai dengan QS. Al-Insyirah ayat 4. Jelaskan bukti peristiwa tersebut dalam lingkungan kehidupan kita sehari-hari!	Bukti dalam lingkungan kehidupan kita sehari-hari : ketika kita mengucapkan kalimat syahadat, adzan, serta iqamat
	Ketika mengucapkan kalimat syahada	0-5
	Saat adzan dan iqamat	0-5
	Jawaban tidak mencakup kategori di atas	1-2
Skor maksimal		10
13	Buatlah kesimpulan kandungan QS. Al-Quraisy tentang cara Allah memberikan rezeki kepada orang-orang Quraisy!	Dengan cara memberi makanan ketika mereka lapar dan memberi mereka rasa aman ketika mereka merasa takut
	Memberi makanan ketika mereka lapar	0-5
	Memberi mereka rasa aman ketika mereka merasa takut	0-5
	Jawaban tidak mencakup kategori di atas	1-2
Skor maksimal		10
14	Kandungan QS. Al-Insyirah menjelaskan tentang "pemberian rejeki Allah Swt sebagai pelipur lara kepada nabi Muhammad Saw". Telaah dan buat kesimpulan maksud kalimat tersebut!	Rezeki pelipur lara kepada nabi Muhammad SAW disini diartikan sebagai sesuatu yang bermanfaat bagi beliau karena namanya selalu disandingkan dengan nama Allah SWT disetiap penyebutannya, tiada rasul yang disebut secara bersamaan dengan Allah SWT kecuali beliau nabi Muhammad SAW
	Disebut bersamaan dengan nama Allah SWT	0-5
	Nabi dan rasul terakhir yang paling dekat dengan Allah SWT	0-5
	Jawaban tidak mencakup kategori di atas	1-2
Skor maksimal		10

Desain soal di atas digunakan untuk uji coba test skala kecil yang akan menghasilkan kevalidan desain instrumen HOTS.

Pengujian

Sebelum instrumen penilaian (soal) ini digunakan untuk mengevaluasi kemampuan berpikir tingkat tinggi peserta didik harus dilakukan pengujian terlebih dahulu. Pengujian instrumen dilakukan dengan uji ahli, uji validitas isi dan uji butir soal (IRT), uji daya beda dan tingkat kesukaran, uji keterbacaan, uji reliabilitas, dan uji kemampuan peserta tes. Tahapan pengujian tersebut dikelompokkan dalam 3 uji, yaitu uji ahli, uji skala kecil dan uji skala besar. Rincian tahapan uji penilaian item soal sebagai berikut:

1. Uji Ahli

Hasil instrumen penilaian/soal untuk mengukur *higher order thinking skills* mata pelajaran Al-Qur'an-Hadis kelas VIII dilakukan uji validitas isi oleh beberapa ahli. Uji validitas isi dilakukan oleh 5 ahli untuk melihat kesesuaian materi, kontruksi dan bahasa. Lima ahli yang dimaksud yaitu 2 dosen ahli dari IAIN Pekalongan berpendidikan doktoral, dan 3 praktisi berpendidikan Magister Pendidikan Islam yang terdiri dari 1 orang pengawas dan 2 guru pns bersertifikasi mata pelajaran Al-Qur'an-Hadis Kelas VIII Madrasah Tsanawiyah.

Uji ahli menggunakan rumus formula Aiken's V yang didasarkan pada hasil panel ahli dan praktisi sebanyak 5 orang terhadap suatu butir mengenai sejauh mana butir tersebut mewakili konstruk yang diukur. Rumus Aiken' V seperti di bawah ini.

$$[V = \frac{\sum s \pi n}{c - 1}]$$

Keterangan:

- Lo : Angka penilaian validasi yang terendah (dalam hal ini 1)
- c : Angka penilaian validasi yang tertinggi (dalam hal ini 5)
- r : Angka yang diberikan oleh seorang penilai
- s : r - lo

Tabel 8. Koefisien Kesepakatan Ahli

No. Soal	Indeks Aiken's V	Simpulan
1	1	valid
2	0,95	valid
3	0,95	valid
4	0,95	valid
5	1	valid
6	1	valid
7	1	valid
8	1	valid
9	1	valid
10	1	valid
11	1	valid
12	1	valid
13	1	valid
14	1	valid

Kategori Kesepakatan Ahli untuk nilai A, dikatakan valid bila nilai A sama dengan $0,70 \leq A \leq 1$, dan bila nilai A kurang dari 0,70 ($A < 0,70$) maka tidak valid. Hasil validitas 14 butir soal dari ahli dengan rumus Aiken V diperoleh bahwa nilai V ke-14 soal tersebut antara 0,95 – 1. Artinya, ke-14 soal tersebut valid dan memiliki nilai tertinggi serta dapat digunakan untuk pengambilan nilai ulangan harian dengan materi tersebut.

Para penguji ahli memberikan saran perbaikan, antara lain:

- a. Memperbaiki tata bahasa dalam kata-kata tertentu agar mudah dipahami peserta didik atau tidak membingungkan;
- b. Memperbaiki indikator soal pada kisi-kisi;
- c. Memperbaiki susunan kalimat pertanyaan soal uraian agar sesuai dengan Pedoman Umum Ejaan Bahasa Indonesia (PUEBI).

Saran perbaikan dari para uji ahli digunakan sebagai dasar pertimbangan pembuatan soal tes yang akan diujicobakan kepada peserta didik. Uji coba instrumen kepada peserta didik dibagi menjadi dua, uji skala kecil dan besar.

Uji skala kecil dilakukan pada satu kelas dengan 30 peserta didik dan uji skala besar dilakukan pada 3 kelas dengan 117 peserta didik di MTs Al-Hikmah Proto, Pekalongan. Uji pada peserta didik meliputi uji validitas isi, respon butir atau item response theory (IRT), uji beda dan uji kesukaran.

2. Uji Skala Kecil

Hasil Uji Validitas Isi dan Hasil Validitas Perhitungan Teori Respon Butir atau *Item Response Theory* (IRT). Adapun hasil Validitas Perhitungan Teori Respon Butir atau *Item Response Theory* (IRT).

Saat validasi IRT skala kecil ada 2 item soal dengan t-hitung < t-tabel (1,734), yaitu soal nomor 4 (1,32) dan 8 (1,38). Artinya kedua soal tersebut invalid dan harus dibuang.

Tabel 9. Uji Validitas, Uji Beda dan Kesukaran Skala Kecil

Item	Uji Ahli	Skala Kecil		
		Validitas	Uji Beda	Kesukaran
1	valid	valid	Soal dibuang	Mudah
2	valid	valid	Soal diterima perlu diperbaiki	Sedang
3	valid	valid	Soal diterima perlu diperbaiki	Sedang
4	valid	invalid	Soal dibuang	Mudah
5	valid	valid	Soal diterima	Mudah
6	valid	valid	Soal dibuang	Mudah
7	valid	valid	Soal diterima perlu diperbaiki	Mudah
8	valid	invalid	Soal dibuang	Mudah
9	valid	valid	Soal diterima perlu diperbaiki	Sedang
10	valid	valid	Soal diterima perlu diperbaiki	Sedang
11	valid	valid	Soal diterima perlu diperbaiki	Sedang
12	valid	valid	Soal diterima	Sedang
13	valid	valid	Soal diterima perlu diperbaiki	Sedang
14	valid	valid	Soal diterima perlu diperbaiki	Sedang

Dari hasil analisis uji coba skala kecil instrumen penilaian/soal HOTS uraian ulangan harian mata pelajaran Al-Qur'an-Hadis kelas VIII menunjukkan bahwa dari 14 soal yang dipakai uji coba terdapat 2 soal diterima, 8 soal yang diterima perlu diperbaiki, dan 4 soal dibuang. Peneliti membuang soal nomer

1, 4, 6, dan 8. Tabelnya dapat kita lihat di bawah ini.

Dari hasil pengujian skala kecil setelah dianalisis maka yang layak dilanjutkan untuk diuji cobakan di skala besar ada 10 soal. 4 soal yang dibuang berdasarkan hasil analisis tingkat kesukaran, daya beda, dan dalam kategori mudah.

3. Uji Skala Besar

Hasil Analisis Daya Beda dan Tingkat Kesukaran Perhitungan Teori Respon Butir atau *Item Response Theory* (IRT)

Nilai validitas berdasar *output* Pearson Correlation tiap butir. Butir dikatakan valid apabila r hitung ≥ 0.3 (Diputera, 2018:53). Ada juga pendapat lain, menurut Sugiono bahwa kriteria pengujian validitas adalah dengan membandingkan antara t -tabel (1,6716) dengan t -hitung. Kriterianya: “Jika t -hitung $>$ dari t -tabel, maka instrumen valid, sebaliknya jika t -hitung $<$ t -tabel maka instrumen tidak valid.

Dari uji validitas skala besar, diperoleh t -hitung $>$ t -tabel dengan kisaran angka antara 8,4789 - 16,15. Artinya, butir-butir soal yang disusun tersebut layak digunakan untuk melakukan penilaian kompetensi siswa dengan berbasis HOTS.

Kategori Daya Beda (D), bila $D \leq 0,30$ (soal dibuang), $0,30 < D < 0,40$ (Soal diterima perlu diperbaiki) dan $D \geq 0,40$ (Soal diterima). Peneliti melakukan analisis tingkat kesukaran dari hasil ulangan harian soal HOTS pada uji coba skala kecil yang berjumlah 14 butir soal. Hasil analisis menunjukkan bahwa dari 14 soal yang di buat memiliki tingkat

kategori sedang sebanyak 8 soal dan kategori mudah sebanyak 6 buah.

Dari uji coba skala besar maka analisis tingkat kesukaran menunjukkan bahwa 4 soal memiliki tingkat kesukaran pada kategori sedang, 6 soal memiliki tingkat kesukaran pada kategori mudah, dan tidak memiliki soal tingkat kesukaran pada kategori sukar.

Kategori Tingkat Kesukaran (Sumarna, 2005:21), $p \leq 0,3$ (sukar), $0,3 \leq p \leq 0,7$ (Sedang) dan $P > 0,7$ (mudah)

4. Hasil Angket Keterbacaan

Uji keterbacaan dilakukan kepada peserta uji skala kecil dengan jumlah responden 30. Uji coba keterbacaan peneliti membagikan angket ke peserta didik kelas VIII satu kelas yang jumlahnya dibatasi 30 peserta didik. Uji keterbacaan digunakan untuk mengetahui keterbacaan yang digunakan untuk soal uraian ulangan harian *higher order thinking skills* mata pelajaran Al-Qur'an-Hadis Madrasah Tsanawiyah.

Tabel 10. Hasil Angket Keterbacaan Yang Dikembangkan

No.	Pertanyaan	Jawaban	
		Ada	Tidak
1.	Adakah huruf yang terlalu kecil?	6	24
2.	Adakah huruf yang terlalu besar?	0	30
3.	Adakah petunjuk pengerjaan yang tidak jelas?	1	29
4.	Adakah pertanyaan yang membingungkan	16	14
5.	Adakah kata/istilah yang tidak kamu mengerti?	5	25
6.	Adakah pertanyaan yang terlalu pendek sehingga sulit dipahami?	4	26
7.	Adakah pertanyaan yang terlalu panjang sehingga sulit dipahami?	4	26
8.	Adakah pertanyaan yang matetinya belum pernah kalian pelajari?	0	30
9.	Adakah soal yang tidak sesuai dengan materi?	0	30
10.	Adakah petunjuk di dalam soal yang tidak jelas?	0	30
11.	Adakah pertanyaan yang membingungkan harus menjawab beberapa point?	7	23

12.	Adakah pertanyaan yang jawabannya terlalu luas?	18	12
13.	Adakah pertanyaan yang jawabannya sama?	0	30
JUMLAH		61	329

Hasil data tabel diatas menunjukkan bahwa tingkat keterbacaan desain instrumen penilaian/butir soal mata pelajaran Al-Qur'an-Hadis kelas VIII sebanyak 329, maka dapatlah dikatakan dalam kategori baik. Angket keterbacaan yang merupakan saran peserta didik adalah perbaikan dalam hal terdapat pertanyaan yang membingungkan dan jawaban beberapa butir masih sangat luas.

5. Hasil Reliabilitas Perhitungan Teori Respon Butir atau *Item Response Theory* (IRT)

Instrumen penilaian soal tes ulangan harian bentuk uraian yang mengandung *higher order thinking skills* mata pelajaran Al-Qur'an-Hadis Madrasah Tsanawiyah kelas VIII dilakukan uji reliabilitas skala kecil dan uji reliabilitas skala besar. Uji reliabilitas menggunakan aplikasi SPSS. Adapun hasilnya dapat dilihat pada tabel 19 di bawah ini.

Tabel 11. *Reliability Statistics*

Nama Uji Coba	Cronbach's Alpha	N of items
Skala Kecil	0,852	14
Skala Besar	0,886	10

Menurut Mardapi (2007:99), instrumen dikatakan reliabel apabila $\alpha \geq 0,7$. Pada Tabel 11, angka alpha (α) untuk skala kecil menunjukkan skala 0,852 dan 0,886 untuk skala besarnya. Artinya instrumen tersebut reliabel sebab kedua angka skala tersebut lebih besar dari 0,7.

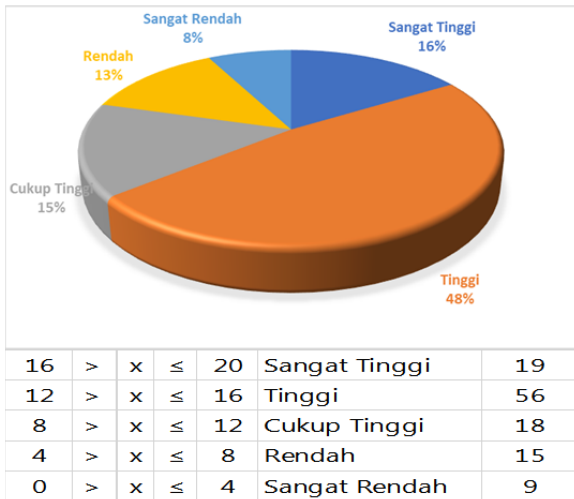
6. Hasil Uji HOTS pada Peserta Didik

Peneliti menganalisis hasil HOTS menggunakan skala besar, selama 80 menit setara 2 jam pelajaran, dan diawasi oleh pengawas 1 orang. Butir soal yang digunakan sebanyak 10 soal dari yang semula 14 soal. Soal dibuang karena ada 4 soal yang tidak bisa membedakan kemampuan peserta didik, serta dari 4 soal tersebut ada 2 soal yang tingkat kesukarannya diatas 1 sehingga dianggap terlalu mudah. Soal yang jumlahnya lebih sedikit menjadikan peserta didik mempunyai waktu lebih banyak untuk mengerjakan tes daripada pada waktu uji skala kecil. Uji coba skala besar dilakukan terhadap 117 peserta didik di MTs al-Hikmah Proto, Kecamatan Kedungwuni, Kabupaten Pekalongan.

Pada uji coba skala besar analisis *higher order thinking skills* ini mampu memberikan informasi peserta didik tentang kemampuan mereka dalam tingkatan kategori berpikir menganalisis (C4), menilai/mengevaluasi (C5), dan mengkreasi/mencipta(C6).

a. Hasil Uji HOTS Menganalisa (C4)

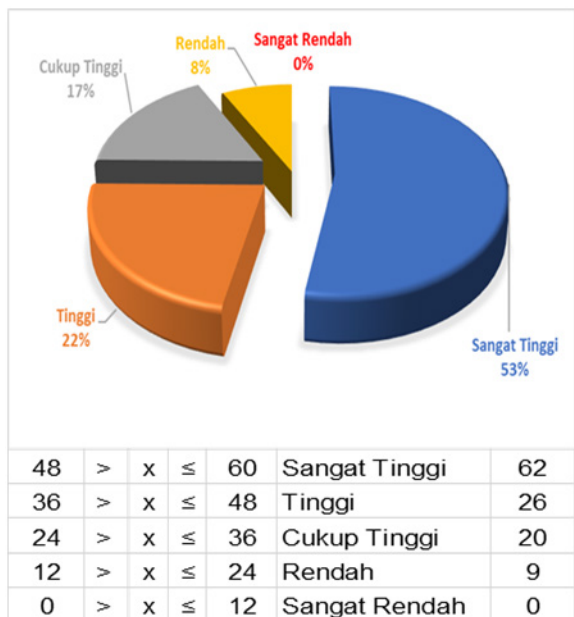
Hasil uji coba skala besar mendapat hasil rincian, yaitu kemampuan peserta didik pada tingkat menganalisis sebesar 16% peserta didik pada kategori sangat tinggi, 48% peserta didik pada kategori tinggi, 15% peserta didik pada kategori cukup tinggi, 13% peserta didik pada kategori rendah, dan 8% peserta didik pada kategori sangat rendah, dapat dilihat pada gambar 1 di bawah ini:



Gambar 1. Analisis HOTS Menganalisa (Hasil Olah Data Uji Soal C4)

b. Hasil Uji HOTS Menilai (C5)

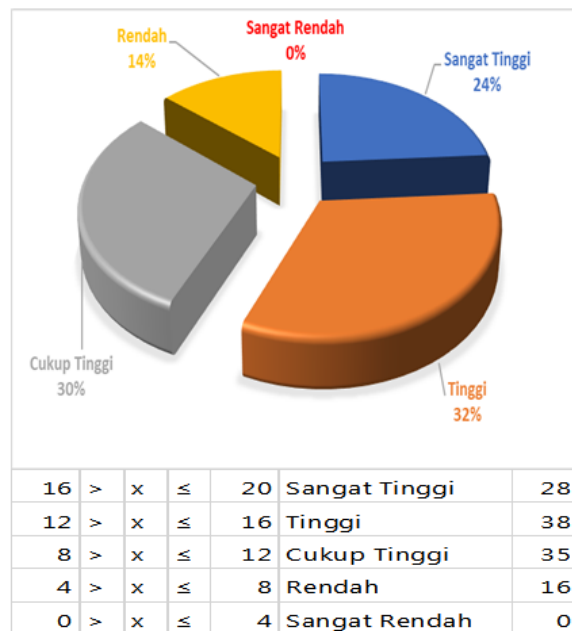
Pada Gambar 2, kita dapat melihat skala besar analisis HOTS untuk kemampuan menilai atau mengevaluasi sangat tinggi mencapai 53%. Sedangkan yang kategori tinggi kemampuan menilai atau mengevaluasi dengan prosentasi 22%, kemudian turun lagi di kategori cukup tinggi yaitu 17% kemudian turun terus di kategori rendah 8% dan sangat rendah 0%.



Gambar 2. Analisis HOTS Mengevaluasi (Hasil Olah Data Uji Soal C5)

c. Hasil Uji HOTS Mengkreasi/ Mencipta (C6)

Pada Gambar 3, kita dapat melihat hasil kemampuan peserta didik pada tingkat mengkreasi atau mencipta sebesar 24% peserta didik pada kategori sangat tinggi, 32% peserta didik pada kategori tinggi, 30% peserta didik pada kategori cukup tinggi, 14% peserta pada kategori rendah dan 0% peserta didik pada kategori sangat rendah.



Gambar 3. Analisis HOTS Mengkreasi (Hasil Olah Data Uji Soal C6)

Budaya Pendidik dalam Pengambilan Nilai Ulangan Harian

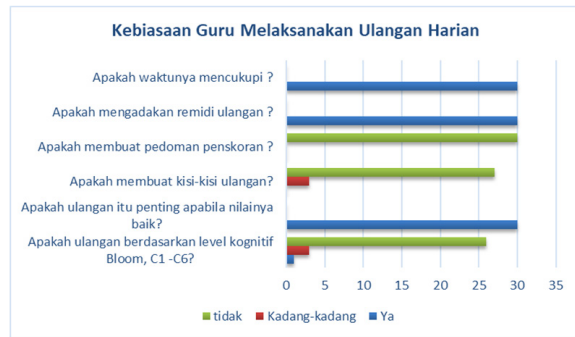
Pengertian budaya disini sebagai suatu kebiasaan yang sudah sukar diubah (kbbi. *online*), yaitu kebiasaan para pendidik dalam melaksanakan penilaian dalam bentuk ulangan harian. Untuk menyusun sebuah soal penilaian kognitif, minimal ada acuan yang digunakan, mengkaji kompetensi dasar dan indikator, menyiapkan materi yang akan diuji, membuat kisi-kisi soal, menuliskan butir soal, menyiapkan butir pendukung seperti pedoman penskoran

serta menganalisisnya untuk mengetahui mutu soal.

Menurut H. Aminudin, Pengawas Madrasah, soal-soal Ujian Nasional sudah berbasis HOTS namun belum digunakan untuk Ujian Akhir Madrasah Berbasis Nasional (UAMBN). Mata Pelajaran yang di-Ujian Nasional-kan sementara pendidik madrasah belum banyak memahami tentang prosedur penyusunan soal-soal berbasis HOTS.

Selama ini guru belum terbiasa membuat soal ulangan harian memakai level *Taksonomi Bloom* secara lengkap, yaitu mengingat, memahami, menerapkan, menganalisis, mengevaluasi, dan mencipta (Setiawati, dkk., 2018:5). Guru selama ini masih menggunakan level kognitif tingkat mengingat, memahami dan menerapkan, karena masih ada paradigma bahwa akan sangat baik apabila nilai anak tinggi, walaupun level soal yang diujikan tidak mengandung ketrampilan berpikir tingkat tinggi atau *higher order thinking skills* di level menganalisis, mengevaluasi, dan mencipta. Dalam membuat soal ulangan harian, guru tidak membuat kisi-kisi terlebih dahulu untuk disampaikan ke anak, tidak membuat pedoman penskoran, hingga nilai bonus tidaklah memiliki patokan yang jelas.

Gambar 4 merupakan hasil pengolahan data wawancara terhadap 30 guru MTs di Pekalongan tentang kebiasaan pelaksanaan ulangan harian. Semua pendidik bersepakat ulangan harian itu penting dilaksanakan (100%), namun mayoritas pendidik tidak memperhatikan kaidah *Taksonomi Bloom* dalam menyusun soal-soal harian (C1-C6), mereka juga tidak membuat kisi-kisi dan pedoman penskoran. Secara detail hasil rekap hasil wawancara tersebut dapat dilihat pada Gambar 4 di bawah ini:



Gambar 4. Kebiasaan Guru dalam Melaksanakan Ulangan Harian (Hasil rekapitulasi wawancara terhadap 30 Guru MTs di Pekalongan)

Menurut Iin Nurmayanti, peneliti sekaligus pendidik MTs menyatakan bahwa rerata para pendidik dalam melakukan penilaian ulangan harian berbentuk uraian menggunakan soal-soal yang ada di buku atau lembar kerja siswa (LKS). Kelompok kerja guru seperti MGMP (Musyawarah Guru Mata Pelajaran) Al-Qur'an-Hadis pun belum melakukan *sharing* tentang tata cara pembuatan soal-soal harian, apalagi soal harian berbasis HOTS.

PENUTUP

Desain instrumen penilaian HOTS untuk mengukur kemampuan berpikir tingkat tinggi yang meliputi menilai (C4), menganalisis (C5), dan mengkreasi (C6) pada Mata Pelajaran Al-Qur'an-Hadis Kelas VIII, materi “Kugapai Rezeki-Mu dengan Ikhtiarku” sudah dapat digunakan untuk melakukan evaluasi pembelajaran materi tersebut.

Langkah-langkah yang dilakukan sudah sesuai panduan soal HOTS dan hasil uji instrumen tersebut valid dan reliabel. Hasil dari uji validitas isi panel para ahli dinyatakan valid karena nilai 0,95 lebih dari 0,70. Validitas isi dari uji ahli memberikan gambaran bahwa desain instrumen penilaian atau soal yang digunakan untuk uji skala

kecil valid. Dari rekap uji keterbacaan juga valid. Sedangkan instrumen penilaian soal tes ulangan harian bentuk uraian HOTS valid berdasarkan *output Pearson Corelation* tiap butir, karena diatas 0,3 dan juga *reliabel* karena diatas 0,7 dari penghitungan teori respon butir (*Item Response Theory/IRT*) melalui SPSS. Menurut hasil analisa tersebut, soal-soal yang telah diujikan dianggap dapat digunakan untuk evaluasi penilaian ulangan harian pada materi “Kugapai Rezeki-Mu dengan Ikhtiarku” kelas VIII MTs.

Meskipun tahapan-tahapan dalam menyusun instrumen penilaian berbasis HOTS sudah benar, namun akan sangat sulit untuk dilaksanakan apabila para pendidik masih membiasakan mengambil soal-soal dari Lembar Kerja Siswa (LKS) atau soal-soal buku ajar untuk dijadikan instrumen penilaian ulangan harian tanpa menggunakan analisa apakah soal tersebut mengandung HOTS atau tidak.

DAFTAR PUSTAKA

- Addison Wesley Longman, 2001, *A Taxonomy For Learning, Teaching, and Assesing: A Revision of Bloom's Taxonomy of Educational Objectives. A Bridged Edition*, Edisi terjemahan alih bahasa oleh Agung Prihantoro, 2010, Kerangka landasan Untuk Pembelajaran, Pengajaran dan Asesmen Revisi Taksonomi Pendidikan Blomm, Yogyakarta : Pustaka Pelajar
- Anggraini, Lysa, & Yusliati, T.T., *Efektivitas Rehabilitasi Pencandu Narkotika Serta Pengaruhnya Terhadap Tingkat Kejahatan di Indosnesia*. Ponorogo: Uwais Inspirasi Indonesia.
- Ari Syahidul Shidiq, dkk, 2015, *Analisis Higher Order Thinking Skills (HOTS) Menggunakan Instrumen Two-Tier Multiple Choice pada ateri Kelarutan dan Hasil Kali Kelarutan untuk Siswa Kelas XI SMA N Surakarta*, Magister Pendidikan Sains dan Doktor Pendidikan IPA FKIP UNS.
- Diputera, Arta Mahindra, 2018, “Desain Instrumen Tes essai Standar Untuk Mengukur Kemampuan Berpikir Tngkat Tinggi Mata Pelajaran IPS SMP Kelas 8,” *Tesis Magister Penelitian dan Evaluasi Pendidikan*, Semarang : Universitas Negeri Semarang.

Untuk itu, penting kiranya para pendidik melatih diri dengan membiasakan membuat soal-soal berbasis HOTS dengan melihat kaidah atau tata cara penyusunan soal berbasis HOTS. Kementerian Agama Kabupaten Pekalongan harus memberikan arahan dan pembinaan kepada para pendidik, dan para guru dapat menggunakan forum MGMP untuk sharing pengetahuan tentang penyusunan soal-soal berbasis HOTS.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terimakasih kepada Kepala MTs al-Hikmah Proto, Kecamatan Kedungwuni, Kabupaten Pekalongan yang telah membantu dalam proses pengumpulan data. Terima kasih juga disampaikan kepada para dosen uji ahli, pengawas dan guru yang turut membantu menvalidasi desain instrumen penilain HOTS.

- Hartini, CH, 2009, *Penelitian Evaluatif tentang Efektivitas Pelaksanaan Pembinaan Kesejahteraan Sosial Anak di PSBR "Taruna Yudha" Sukoharjo*, Departemen Sosial RI, Badan Penelitian dan Pengembangan Pelayanan Kesehatan Sosial.
- Ismet, Basuki, Hariyanto, *Asesmen Pembelajaran*, (Bandung: PT. Remaja Rosdakarya, 2015).
- Mardapi, Djemari, 2018, *Teknik Penyusunan Instrumen Tes dan Nontes*, Yogyakarta: Parama Publising.
- Rusilowati, A., 2017. *Pengembangan Instrumen Penilaian*, Semarang: Unnes Press.
- Safari, Indra., dkk., 2017. *Prosiding Seminar Nasional, 2017, Membangun Generasi Emas 2045 yang Berkarakter dan Melek IT dan Pelatihan Suprarasional*, Universitas Pendidikan Indonesia Kampus Sumedang, Sumedang: UPI Sumedang Press.
- Sasongko, Purwo, "Perbandingan Keefektifan Bentuk Tes Uraian dan Testlet dengan Penerapan Graded Response Model GRM", *Jurnal Penelitian dan Evaluasi Pendidikan*, Tahun 14 Nomor 2, Jurusan matematika FKIP UPS Tegal.
- Setiawati, Wiwik. dkk., 2018, *Buku Penilaian Berorientasi Higher Order Thinking Skills (Program Peningkatan Kompetensi Pembelajaran Berbasis Zonasi)*, Jakarta: Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan.
- Singgih Santosa, t.t., *Statistik Parametrik Konsep dan Aplikasi dengan SPSS*, Jakarta : PT Elex Media Komputindo.
- Sugiyono, 2017, *Metode Penelitian & Pengembangan Research and Development untuk Bidang Pendidikan, Manajemen, Sosial Teknik*, Bandung: Alfabeta.
- Sumarna Surapranata, 2005, *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*, Bandung : PT. Remaja Rosdakarya.
- Umam, Khoirul, 2018, Peranan Mata Pelajaran Al-Qur'an-Hadis dalam Melaksanakan Pendidikan Agama Islam di MTs Negeri Malang I, *Tesis Magister Evaluasi dan Penelitian*, Malang : Universitas Islam Negeri Maulan Malik Ibrahim.
- Umi Pratiwi, & Eka farida Fasha, 2015, Pengembangan Instrumen Penilaian HOTS Berbasis Kurikulum 2013 Terhadap Sikap Disiplin, *Jurnal Penelitian dan Pengembangan IPA, JPPI*, Vol 1, No. 1, Pendidikan Fisika, FKIP Universitas Muhammadiyah Purworejo.
- Yoko Ariyana Dkk, 2018, *Buku Pegangan Pembelajaran Berorientasi pada Kemampuan Berpikir Tingkat Tinggi (Program Peningkatan Kompetensi Pembelajaran Berbasis Zonasi)*, Jakarta: Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan.

