Loghat Arabi: ***Jurnal Bahasa Arab dan Pendidikan Bahasa Arab***
Vol. 7, No. 2, Juli 2026 | DOI: https://doi.org/10.36915/la.v7i2.697

LOGHAT ARABI
Jurnal Bahasa Arab dan Pendidikan Bahasa Arab
https://journal.iaiddipolman.ac.id/index.php/loghat/index

Pengembangan Instrumen Evaluasi Berbasis *Higher Order Thinking Skills (HOTS)* Dalam Pembelajaran Bahasa Arab Pada Kurikulum Merdeka Belajar di MA DDI Takkalasi/ Development of Evaluation Instrument Based on Higher Order Thinking Skills in Arabic Language Learning in the Independent Learning Curriculum at MA DDI Takkalasi

Najmalia Fitra[1]*, Herdah[2], Kaharuddin[3], Saepudin[4], Darmawati[5]
[1,2,3,4,5,6]Pascasarjana Institut Agama Islam Negeri Parepare, Sulawesi Selatan, Indonesia

Article Information:
Received : 02 Januari 2026
Revised : 28 Fabruari 2026
Accepted : 25 Maret 2026

Keywords:
Instrument Development;
Higher Order Thinking Skills Evaluation;
Independent Learning Curriculum;
Arabic Language Learning

***Correspondence Address:**
Najmaliafitra0@gmail.com

Abstract: This study is motivated by the need for an evaluation instrument that aligns with the learning objectives of the Independent Learning Curriculum and the competency levels of grade XI students at MA DDI Takkalasi. This study aimed to develop a Higher Order Thinking Skills (HOTS)-based Arabic language evaluation instrument and examine its quality through validity, reliability, discriminating power, difficulty level, and distractor effectiveness tests. This research employed a Research and Development (R&D) approach by adapting the Borg and Gall model, consisting of needs analysis, planning, product development, testing, and revision stages. The results indicate that: (1) a HOTS-based Arabic evaluation instrument is highly needed because previous assessments were predominantly focused on Lower Order Thinking Skills (LOTS); (2) the developed prototype based on cognitive indicators C4, C5, and C6 with contextual stimuli was highly valid, with most I-CVI values ≥ 0.78 and S-CVI values ≥ 0.90; (3) field and final trials showed that multiple-choice and essay items were valid and reliable, with adequate to very good discriminating power, moderate difficulty levels, and effective distractors; and (4) the instrument demonstrated effectiveness through varied student scores and 58.3% learning completion, with high practicality from teachers (93%) and good student readability. Therefore, the developed HOTS-based evaluation instrument is feasible as a valid, reliable, effective, and practical assessment tool to support the implementation of the Merdeka Curriculum.

Abstrak: Penelitian ini dilatarbelakangi oleh kebutuhan terhadap instrumen evaluasi yang selaras dengan tujuan pembelajaran Kurikulum Merdeka dan tingkat kompetensi peserta didik kelas XI MA DDI Takkalasi. Penelitian ini bertujuan untuk mengembangkan instrumen evaluasi bahasa Arab berbasis *Higher Order Thinking Skills (HOTS)* serta menguji kualitasnya melalui uji validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Penelitian ini menggunakan pendekatan Research and Development (R&D) dengan mengadaptasi model Borg and Gall yang meliputi tahap analisis kebutuhan, perencanaan, pengembangan produk, pengujian, dan revisi. Hasil penelitian menunjukkan bahwa: (1) instrumen evaluasi bahasa Arab berbasis HOTS sangat dibutuhkan karena instrumen sebelumnya masih didominasi oleh kemampuan berpikir tingkat rendah (LOTS); (2) prototipe instrumen HOTS yang dikembangkan berdasarkan indikator kognitif C4, C5, dan C6 dengan stimulus kontekstual dinyatakan sangat valid, dengan sebagian besar nilai I-CVI ≥ 0,78 dan nilai S-CVI seluruh aspek ≥ 0,90; (3) hasil uji coba lapangan dan uji akhir menunjukkan bahwa seluruh butir pilihan ganda dan esai valid dan reliabel, memiliki daya pembeda yang cukup hingga sangat baik, tingkat kesukaran dominan sedang, serta fungsi pengecoh yang efektif; dan (4) instrumen terbukti efektif melalui variasi skor peserta didik dan tingkat ketuntasan belajar sebesar 58,3%, serta sangat praktis dengan tingkat kepraktisan guru sebesar 93% dan keterbacaan peserta didik dalam kategori baik hingga sangat baik. Dengan demikian, instrumen evaluasi berbasis HOTS yang dikembangkan layak digunakan sebagai alat ukur yang valid, reliabel, efektif, dan praktis untuk mendukung implementasi asesmen dalam Kurikulum Merdeka.

Loghat Arabi: Jurnal Bahasa Arab dan Pendidikan Bahasa Arab
Vol. 7, No. 2, Juli 2026 | DOI: https://doi.org/10.36915/la.v7i2.697
This is an open access journal, licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
https://journal.iaiddipolman.ac.id/index.php/loghat/index

Pendahuluan

Bahasa Arab sebagai salah satu mata pelajaran utama di Madrasah Aliyah, turut mengalami transformasi dalam pendekatan pembelajarannya. Sejak tahun 2020, pembelajaran Bahasa Arab yang berorientasi pada *Higher Order Thinking Skill (HOTS)* mulai diperkenalkan dan terus berkembang hingga saat ini. Tuntutan integrasi HOTS dalam pembelajaran Bahasa Arab telah menarik perhatian banyak kalangan, termasuk para akademisi. Nailur Rahmati dalam artikelnya yang berjudul *Pembelajaran Bahasa Arab: Menuju Higher Order Thinking Skill (HOTS)* menyebutkan bahwa pendekatan ini mampu meningkatkan antusiasme belajar peserta didik, menumbuhkan motivasi, serta mendorong mereka untuk tidak mudah menyerah dalam menghadapi kesulitan.[1]

Pentingnya penguasaan HOTS dalam pembelajaran Bahasa Arab tidak hanya terletak pada aspek umum berpikir kritis, tetapi juga memiliki dampak langsung terhadap kemampuan berbahasa yang lebih mendalam dan fungsional. Dengan demikian, HOTS tidak hanya meningkatkan kualitas berpikir peserta didik, tetapi juga memperkuat keterampilan bahasa Arab mereka secara menyeluruh.[2] Berbagai metode pembelajaran aktif seperti diskusi, studi kasus, *project-based learning, problem-based learning, problem solving, discovery/inquiri*, banyak digunakan oleh guru untuk menerapkan kegiatan pembelajaran pada level HOTS untuk mendorong peserta didik berpikir di luar batas pemahaman literal. Selain metode pembelajaran, instrumen evaluasi pembelajaran juga telah disusun dan dikembangkan berbasis HOTS, dimana alat penilaian ini dirancang untuk mengukur kemampuan peserta didik dalam berpikir tingkat tinggi yaitu analisis, evaluasi, dan kreasi.[3]

Kehadiran instrumen evaluasi berbasis HOTS menjadi penting untuk mendukung implementasi kurikulum yang berorientasi pada kompetensi, seperti kurikulum merdeka. Instrumen evaluasi dengan HOTS dapat membantu guru memastikan bahwa

[1] Nailur Rahmawati, 'Pembelajaran Bahasa Arab: Menuju Higher Order Thinking Skills (HOTS)', *Prosiding Konferensi Nasional Bahasa Arab*, 4.4 (2018), 149–54.
[2] Nailur Rahmawati, 'Pembelajaran Bahasa Arab: Menuju Higher Order Thinking Skills (HOTS)', *Prosiding Konferensi Nasional Bahasa Arab*, 4.4 (2018), 149–54.
[3] Yuniar Handayani, Eni Asia, and Saleh Hidayat, 'Peningkatan Kemampuan High Order Thinking Skills (HOTS) Melalui Project-Based Learning (PjBL) Dalam Implementasi Kurikulum Merdeka', *PTK: Jurnal Tindakan Kelas*, 4.1 (2023), 48–60.

pembelajaran tidak hanya berpusat pada penguasaan materi, tetapi juga pada pengembangan kemampuan berpikir kritis dan kreatif dari peserta didik.[4]

Pengembangan instrumen evaluasi berbasis HOTS dalam bidang pendidikan selalu terkait dengan penelitian dan pengembangan atau *research and development.* Menurut Sugiyono penelitian dan pengembangan adalah metode penelitian yang digunakan untuk menghasilkan produk tertentu (seperti media, model pembelajaran, perangkat ajar, modul, instrumen evaluasi) dan menguji keefektifan produk tersebut.[5] Mc Kenney dan Reeves melalui model *Educational Design Research* mendefinisikan pengembangan sebagai proses yang sistematis dan berulang untuk merancang, menguji, dan menyempurnakan suatu intervensi pendidikan, seperti kurukulum, media pembelajaran, instrumen, dan sejenisnya, sambil diujicobakan langsung dalam lingkungan yang nyata.[6]

Dalam pembahasan HOTS tentu tidak terlepas dari Taksonomi Bloom yang diungkapkan oleh Bemanjamin S. Bloom yang mengenalkan teori beberapa tingkatan berpikir yaitu pengetahuan, pemahaman, penerapan, analisis, sintesis, dan evaluasi. Akan tetapi teori ini sudah direvisi oleh murid dari Bloom sendiri yaitu Krathwohl dan Anderson menjadi mengingat (*remembering*), memahami (*understnding*), menerapkan (*applying*), merupakan LOTS, sedangkan menganalisis (*analyzing),* mengevaluasi (*evaluating*) dan mencipta (*creating*) adalah HOTS.[7]

Tabel 1. Revisi Taksonomi Bloom

Taksonomi Bloom			
Benjamin S. Bloom (1956)		Revisi Anderson & Krathwohl (2001)	kategori
C1	Pengetahuan	Mengingat	LOTS
C2	Pemahaman	Memahami	
C3	Aplikasi	Mengaplikasikan	
C4	Analisis	Menganalisis	HOTS
C5	Sintesis	Mengevaluasi	
C6	Evaluasi	Mencipta	

[4] Ahmad Ashfia and Ubaid Ridlo, 'Optimalisasi Higher Order Thinking Skill (HOTS) Dalam Kurikulum Merdeka: Strategi Dan Konsep Penyusunan Soal Bahasa Arab Di MTs Pembangunan Jakarta', *Mauriduna: Journal of Islamic Studies*, 5.1 (2024), 330–42.

[5] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D* (Bandung: Alfabeta, 2020).

[6] S Putrawangsa, *Design Research Sebagai Framework Desain Pembelajaran* (Mataram: Sinabil Publishing, 2019).

[7] Yayuk Susilowati and Sumaji Sumaji, 'Interseksi Berpikir Kritis Dengan High Order Thinking Skill (Hots) Berdasarkan Taksonomi Bloom', *JURNAL SILOGISME: Kajian Ilmu Matematika Dan Pembelajarannya*, 5.2 (2021), 62–71.

Brookhart menjelaskan jenis HOTS didasarkan pada tujuan pembelajaran di kelas, yaitu terdiri dari 3 kategori, yaitu: 1) HOTS sebagai *transfer* mencakup *analyzing, evaluating, creating*, berpikir kreatif, berpikir logis dengan dirangkum menjadi menganalisa, mengevaluasi, dan mencipta. 2) HOTS sebagai *critical thingking* didefinisikan sebagai keterampilan memberikan keputusan (*judgment*) menggunakan alasan yang logis dan ilmiah, ini mencakup berpikir kritis dan metakognitif. 3) HOTS sebagai *problem solving* didefinisikan sebagai keterampilan mengidentifikasi masalah dan menyelesaikannya dengan menggunakan strategi yang nonautomatic.[8]

Klasifikasi Brookhart menunjukkan bahwa HOTS bukan hanya sekedar level kognitif tertinggi dalam hierarki berpikir tetapi sebagai tujuan fungsional pembelajaran, yang berarti HOTS bukan hanya tentang tingkat kesulitan soal melainkan kualitas pemprosesan pengetahuan dan bagaimana pengetahuan digunakan dalam konteks yang autentik. Sehingga HOTS tidak lagi dinilai sebagai ranah taksonomi melainkan sebagai fungsi berpikir dalam ranah pedagogis dan asesmen.

Dalam menyusun butir soal HOTS, penulis soal dituntut untuk dapat menentukan perilaku yang hendak diukur dan merumuskan materi yang akan dijadikan dasar penetapan pertanyaann (stimulus) dalam konteks tertentu sesuai dengan perilaku yang diharapkan.[9] Selain itu soal yang akan menjadi pertanyaan (yang menuntut penalaran tinggi) tidak selalu tersedia dalam buku pelajaran, oleh karena itu dalam penulisan soal HOTS dibutuhkan penguasaan materi ajar, keterampilan dalam menulis soal (konstruksi soal), dan kreatifitas guru dalam memilih stimulus soal.

Penyusunan soal HOTS bukan sekedar menaikkan tingkat kesulitan soal, melainkan sebuah proses yang sistematis. Secara teoritis hal ini sejalan dengan pendekatan asesmen yang dikemukakan oleh Brookhart bahwa HOTS harus dirancang secara sengaja, bukan muncul secara kebetulan dari soal yang sulit. Sehingga kualitas soal HOTS ditentukan oleh kejelasan konstruk kognitif yang diukur, kesesuaian antara stimulus dan soal, serta ketetapan indikator dan kompetensi dasar yang hendak diukur.

Adapun langkah-langkah penyusunan soal HOTS yang dapat digunakan dalam mengembangkan instrumen evaluasi berbasis HOTS diantaranya: 1) menganalisis

[8] Nailur Rahmawati, 'Pembelajaran Bahasa Arab: Menuju Higher Order Thinking Skills (HOTS)', *Prosiding Konferensi Nasional Bahasa Arab*, 4.4 (2018), 149–54.

[9] Indri Sari and others, 'Langkah Penyusunan Dan Analisis Butir Soal Hots (Higher Order Thinking Skills) Pada Mata Pelajaran Pendidikan Agama Islam (PAI) Dan Budi Pekerti', *Jurnal Insan Pendidikan Dan Sosial Humaniora*, 1.4 (2023), 56–73.

kompetensi dasar yang dapat dibuat soal-soal HOTS, 2) menyususn kisi-kisi, 3) memilih stimulus yang menarik dan konstektual, 4) menulis butir pertanyaan sesuai dengan kisi-kisi soal, dan 5) membuat pedoman penskoran (rubrik) dan kunci jawaban.[10] Hal ini sejalan dengan teori evaluasi perpektif Norman E. Gronlund yang menekankan bahwa instrumen evaluasi yang baik harus selaras dengan tujuan pembelajaran, disusun berdasarkan kisi-kisi yang sistematis, dan memiliki kejelasan indikator yang hendah diukur. Gronlund juga menjelaskan bahwa kegagalan utama dalam penyusunan instrumen evaluasi adalah tidak jelasnya hubungan antara tujuan pembelajaran dan soal yang dibuat.

Dalam menyusun instrumen evaluasi khususnya yang berbasis HOTS harus berdasarkan dengan beberapa karakteristik, agar instrumen tersebut dapat mengukur hasil belajar secara tepat, akurat, dan efektif. Berikut adalah karakteristik utama instrumen evaluasi yaitu, valid, reliabel, objektivitas yang tinggi, praktis, relevan, berkeadilan, sensitivitas, tingkat keterbacaan, efisien, dan memliki daya tarik. [11]

Selanjutnya, teknik analisis butir soal merupakan suatu prosedur yang digunakan sebagai alat untuk menganalisis item tes. Analisis item tes ini dapat dilakukan dengan dua cara, yaitu analisis kualitatif (*qualitative control*) dan analisis kuantitatif (*quantitative control*). Analisis kualitatif atau yang disebut validitas logis yaitu analisis yang digunakan sebelum soal dipakai, apakah berfungsi atau tidak. Sedangkan analisis soal secara kuantitatif atau validitas empiris adalah menganalisis butir soal dengan diuji cobakan. kepada populasi atau sampel untuk mengetahui berfungsi tidaknya sebuah butir soal.[12]

Adapun aspek yang dianalisis secara kualitatif dalam tes bahasa [13], yaitu: 1) Aspek materi; kriteria analisis butir soal aspek materi adalah soal sesuai dengan indikator, materi sesuai dengan komptensi, pilihan jawaban homogen dan logis, hanya terdapat satu kunci jawaban. 2) Aspek konstruksi soal; diantara kriteria yang dianalisis dalam aspek kontruksi soal adalah soal harus dirumuskan secara singkat, jelas, dan tegas,

[10] Sihabuddin Sihabuddin, 'Prosedur Penyusunan Tes Berbasis HOTS Pada Empat Keterampilan Berbahasa Arab', *LEARNING: Jurnal Inovasi Penelitian Pendidikan Dan Pembelajaran*, 3.1 (2023), 40–47.

[11] Abdul Munip, *Penilaian Pembelajaran Bahasa Arab* (Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta, 2019).

[12] Najmalia Fitra, Herdah Herdah, and Amira Ezzat Mahrous, 'Language Test Item Analysis Techniques', 11.1 (2025), 158–79 <https://doi.org/10.14421/almahara.2025.>.

[13] Muhammad Atanda Musa, Rara Mutiah, and Rahmani, 'Analisis Butir Soal Bahasa Arab Di Mtsn Kota Parepare', *Sao Jurnal IAIN Parepare*, 1.2 (2022), 11–23.

soal tidak memberikan petunjuk ke arah jawaban yang benar, soal tidak mengandung pertanyaan yang bersifat negatif,gambar, garifk, tabel, diagram dan sejenisnya harus jelas dan berfungsi, panjang pilihan jawaban relatif sama, pilihan jawaban tidak menggunakan pertanyaan "semua jawaban benar/salah", dan rumusan soal dan pilihan jawaban merupakan pernyataan yang diperlukan saja.

Selanjutnya, 3) Aspek budaya/bahasa; kriteria analisis butir soal aspek bahasa adalah menggunakan bahasa yang sesuai dengan kaidah, menggunakan bahasa yang komunikatif, tidak menggunakan bahasa setempat/bahasa daerah, pilihan jawaban tidak mengulang kata yang sama. 4) Kesesuaian materi terhadap kemampuan peserta didik; Analisis materi bertujuan untuk menelaah substansi keilmuan yang ditanyakan dalam soal dengan memperhatikan tingkat kemampuan peserta didik. Dengan mempertimbangkan materi terdapat dalam kompetensi dasar dan telah diajarkan kepada peserta didik, guru telah mengajarkan secara maksimal materi sesuai dengan tuntutan kompetensi yang harus dikuasai peserta didik, dan guru memiliki kreativitas dalam menyajikan materi pembelajaran.

Analisis kuantitatif dilakukan dengan cara mengujicobakan instrumen yang telah dianalisis secara kualitatif kepada sejumlah peserta didik yang memiliki karakteristik sama dengan peserta didik yang akan diuji dengan instrument tersebut. Analisis soal secara kuantitatif menekankan pada analisis karakteristik internal tes melalui data yang diperoleh secara empiris. Karakteristik internal secara kuantitatif meliputi validitas, reliabilitas, daya pembeda, tingkat kesukaran serta efektifitas fungsi pengecoh (distraktor).[14]

Dalam pengembangan dan pengimplementasian instrumen evaluasi yang berbasis HOTS terdapat beberapa permasalahan dalam pengembangan dan penerapannya, diantaranya penelitian yang dilakukan oleh Fakturmen pada tahun 2022, diketahui bahwa terdapat permasalahan atau kendala yang dialami oleh guru dalam menyusun soal HOTS pada penilaian akhir sekolah di MAN 3 Sleman yaitu terletak pada: 1) kendala menyusun indikator KD yang mengarah ke C4, C5, dan C6, 2) kendala menyusun indikator KD yang mengarah ke dimensi pengetahuan metakognitif, 3) kendala menyusun level kognitif, 4) kendala menyusun indikator soal lebih mengarah ke komponen Behavior

[14] khaerudin, 'Kualitas Instrumen Tes Hasil Belajar', *Jurnal Madaniyah*, 2 (2015), 212–35.

berbasis HOTS, 5) kendala menyusun stimulus yang menarik, 6) kendala menyusun stimulus yang kontekstual.[15]

Research gap penelitian ini terletak pada belum adanya kajian yang secara komprehensif mengembangkan dan menguji instrumen evaluasi Bahasa Arab berbasis HOTS yang selaras dengan Kurikulum Merdeka. Secara *knowledge gap*, penelitian sebelumnya belum menghasilkan model desain instrumen HOTS yang sistematis dan kontekstual. Dari sisi *empirical dan evidence gap*, penelitian terdahulu, seperti Fakturmen, masih bersifat deskriptif dan belum menyajikan data empiris serta bukti statistik mengenai validitas, reliabilitas, efektivitas, dan kepraktisan instrumen HOTS. Secara *methodology gap*, penelitian sebelumnya menggunakan pendekatan *mixed methods* untuk memotret kemampuan guru, sementara penelitian ini menggunakan metode pengembangan (R&D) yang lebih aplikatif dan menghasilkan produk instrumen yang siap digunakan. Selain itu, terdapat *population gap*, karena penelitian terdahulu hanya melibatkan guru, sedangkan penelitian ini berfokus pada peserta didik sebagai pengguna utama instrumen evaluasi, sehingga mampu memberikan gambaran yang lebih utuh tentang kelayakan dan keberfungsian instrumen HOTS dalam pembelajaran Bahasa Arab.

Selanjutnya penelitian Heni Verawati pada tahun 2022, berfokus pada analisis keterampilan berpikir tingkat tinggi (HOTS) berdasarkan taksonomi Bloom revisi (Anderson & Krathwohl) dalam instruksi tugas buku ajar Bahasa Arab kelas X Madrasah Aliyah yang diterbitkan oleh Kementerian Agama. Hasil analisis menunjukkan bahwa buku teks memuat sebanyak 78 instrusksi tugas, hanya 32 instruksi (41%) yang merepresentasikan HOTS. Sedangkan LOTS mendominasi isi buku dengan 46 instruksi (59%) dari total keseluruhan.[16] Namun, penelitian tersebut bersifat deskriptif dan belum mengembangkan perangkat evaluasi yang dapat mengukur kemampuan HOTS secara langsung. Berdasarkan hal tersebut, penelitian ini hadir untuk mengisi berbagai kesenjangan penelitian (*research gaps*) yang meliputi:

Penelitian sebelumnya, seperti Verawati, masih terbatas pada analisis konten buku ajar dengan pendekatan deskriptif kuantitatif dan belum menghasilkan produk

[15] Fakturmen, 'Kemampuan Guru Bahasa Arab Dalam Menyusun Soal Higher Order Thingking Skills (Hots) Pada Penilaian Akhir Sekolah (Pas) Di Man 3 Sleman', 2022.
[16] Heni Verawati and others, 'HOTS Analysis of Task Instructions in Bahasa Arab Madrasah Aliyah Textbook Published by The Ministry of Religious Affairs', *Edukatif: Jurnal Ilmu Pendidikan*, 4.1 (2022), 944–51.

evaluasi yang dapat digunakan langsung oleh guru, sehingga menyisakan *knowledge gap* dan *methodology gap*. Selain itu, penelitian terdahulu bersifat literer-dokumentatif tanpa melibatkan praktik pembelajaran nyata, sehingga belum menyediakan data empiris maupun bukti validitas dan reliabilitas instrumen (*empirical gap* dan *evidence gap*). Perbedaan mendasar juga terlihat pada subjek penelitian, yang sebelumnya berfokus pada teks, sedangkan penelitian ini melibatkan peserta didik sebagai pengguna utama instrumen evaluasi (*population gap*). Oleh karena itu, penelitian ini mengisi kesenjangan tersebut melalui penelitian pengembangan (R&D) yang menghasilkan instrumen evaluasi HOTS yang valid, reliabel, empiris, dan aplikatif sesuai konteks Kurikulum Merdeka.

Meskipun beberapa penelitian telah mengkaji keberadaan unsur HOTS dalam buku teks bahasa Arab untuk Madrasah Aliyah, namun belum ditemukan penelitian yang secara khusus mengambangkan instrumen evaluasi berbasis HOTS yang sesuai dengan pendekatan kurikulum merdeka, khususnya untuk kelas XI Madrasah Aliyah. Selain itu, masih minimnya pengembangan instrumen evaluasi yang mengintegrasikan capaian pembelajaran dan kompetensi berpikir kritis sebagai bagian dari profil belajar pancasila.

Berdasarkan beberapa kajian literatur tersebut maka dapat dipahami bahwa dalam pembelajaran bahasa Arab ada kecenderungan guru bahasa Arab membuat soal dengan kategori LOTS karena lebih mudah dan cepat, namun di sisi lain terlihat belum maksimalnya pemahaman guru terhadap konsep soal HOTS dan cara mengimplementasikannya ke dalam soal, juga belum maksimalnya penggunaan soal HOST dalam buku ajar bahasa Arab, serta peserta didik kurang terbiasa dengan pola pikir kritis dan analitis sehingga ketika menemukan soal HOTS peserta didik mengalami kesulitan saat menjawab soal.

Hal yang sama juga terjadi di MA DDI Takkalasi, berdasarkan hasil observasi di kelas XI MA diketahui bahwa dalam proses pembelajaran, pendekatan dan strategi yang digunakan oleh guru belum sepenuhnya mengakomodasi pengembangan berpikir tingkat tinggi, serta masih dominan pada mengingat dan memahami yang merupakan keterampilan berpikir tingkat rendah, penggunaan soal berbasis HOTS oleh guru belum sepenuhnya optimal, kemampuan guru dalam menyusun instrumen berbasis HOTS masih terbatas, kebiasaan peserta didik dalam mengerjakan soal lebih dominan pada tahap mengingat dan memahami, serta belum tersedianya soal yang khusus mengukur HOTS di sekolah.

Metode Penelitian

Penelitian ini merupakan penelitian dan pengembangan (*Research and Development/R&D*) yang mengadaptasi model Borg dan Gall. Prosedur pengembangan dilaksanakan melalui sepuluh tahapan, yaitu riset dan pengumpulan informasi awal, perencanaan, pengembangan produk awal, uji coba produk awal, revisi produk awal, uji coba lapangan, revisi produk hasil uji coba lapangan, uji lapangan pada skala yang lebih luas (uji coba final), revisi produk final, dan diseminasi.[17] Subjek penelitian meliputi peserta didik kelas XI MA DDI Takkalasi, Guru mata pelajaran Bahasa Arab, Kepala Madrasah, serta pakar/ahli dalam bidang validasi materi, media, dan soal. Teknik pengumpulan data dilakukan melalui observasi, wawancara, angket, tes dan dokumentasi.

Observasi, wawancara dan angket dilakukan pada tahap pengumpulan data awal terkait instrumen yang dikembangkan. Observasi dilakukan untuk mengetahui sejauh mana pembelajaran dan pelaksanaan evaluasi yang diberikan kepada peserta didik di kelas. Wawancara dilakukan bersama kepala Madrasah, guru mata pelajaran, dan peserta didik untuk memperoleh data mengenai kurikulum, proses pembelajaran dan evalasi bahasa Arab, kemampuan awal peserta ddidik, dan menganalisis kebutuhan instrumen evaluasi berbasis HOTS, kemudian data ini menjadi landasan bagi pengembangan instrumen evaluasi berbasis HOTS pada Kurikulum Merdeka dalam pembelajaran bahasa Arab.

Selanjutnya angket diberikan kepada pakar/ahli digunakan untuk menilai kelayakan dan kepraktisan penggunaan instrumen yang dikembangkan. Penelitian ini menggunakan tiga jenis angket yaitu: angket validasi ahli materi, angket validasi ahli media dan angket kepraktisan instrumen. Angket validasi materi digunakan untuk menilai aspek konten, konstruk, bahasa dan penyajian, serta teknis instrumen. Angket validasi media digunakan untuk menilai aspek tampilan desain, susunan isi dan keterbacaan. Angket kepraktisan instrumen digunakan untuk menilai kejelasan, tingkat kesulitan dan kemudahan penggunaan instrumen setelah diuji cobakan. Analisis data menggunakan pendekatan kualitatif dan kuantitatif dengan menggunakan teknik analisis validitas isi menggunakan *Item Content Validity Index* (I-CVI) dan *Scale Content Validity Index* (S-CVI).

[17] Marinu Waruwu, 'Metode Penelitian Dan Pengembangan (R&D): Konsep, Jenis, Tahapan Dan Kelebihan', *Jurnal Ilmiah Profesi Pendidikan*, 9.2 (2024).

Tes dan dokumentasi dilakukan pada tahap uji coba final pengembangan instrumen. Subjek penelitian pada tahap ini adalah peserta didik kelas XI BI MA DDI Takkalasi yang berjumlah 20 orang. Pada tahap ini digunakan desain eksperimen kuasi (*quasi experimental design*) dengan model *one-shot case study* dimana suatu kelompok diberi treartmen/perlakuan kemudian diobservasi hasilnya.[18] Tes yang diberikan kepada peserta didik merupakan instrumen evaluasi berbasis HOTS yang telah dikembangkan, selanjutnya hasil tes tersebut akan dianalisis untuk mengetahui validitas tes, reliabilitas tes, tingkat kesukaran, daya pembeda, dan efektivitas fungsi distraktor. Dokumentasi juga dilakukan untuk mengumpulkan data berupa foto kegiatan evaluasi yang berlangsung, bertujuan untuk memperkuat hasil observasi, wawancara, angket, dan tes yang telah dilakukan sebelumnya.

Hasil dan Pembahasan

Analisis kebutuhan instrumen evaluas berbasis HOTS dalam pembelajaran bahasa Arab pada kurikulum merdeka di kelas XI MA DDI Takkalasi

Hasil analisis kebutuhan instrumen evaluasi berbasis HOTS dalam pembelajaran bahasa Arab di kelas XI MA DDI Takkalasi, ditemukan bahwa instrumen evaluasi yang selama ini digunakan, khususnya pada penilaian sumatif akhir semmester, masih didominasi oleh soal yang berorientasi Lower Order Thingking Skills (LOTS). Bentuk soal berupa pilihan ganda sebanyak 20 butir dan uraian sebanyak 5 butir yang sebagian besar menuntut kemampuan mengingat, menghafal, menyusun kosakata, serta menerjemahkan mufradat yang bersumber dari buku paket. Kondisi ini menunjukkan bahwa proses evaluasi pembelajaran bahasa Arab belum sepenuhnya selaras dengan tuntutan kurikulum merdeka yang menekankan pengembangan kemampuan berpikir tingkat tinggi peserta didik.

Kondisi ini menunjukkan adanya kesenjangan antara praktik evaluasi yang dilakukan di kelas dengan tuntutan kurikulum merdeka yang menekankan kemampuan berpikir tingkat tinggi. Soal-soal yang hanya menekankan pada kemampuan mengingat dan memahami tentu belum mampu mengukur kamampuan analisis dan evaluasi peserta didik secara komprehensif. Sedangkan Ralph W. Tyelr menyatakan bahwa evaluasi pembelajaran harus selaras dengan tujuan pembelajaran yang ingin dicapai sesuai dengan tuntutan kurikulum yang digunakan. Jika tujuan pembelajaran menuntut pengembangan

[18] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D* (Bandung: Alfabeta, 2020.

kemampuan berpikir tingkat tinggi, maka instrumen evaluasi yang digunakan juga harus dirancang untuk mengukur kemampuan tersebut.[19]

Hasil wawancara juga menunjukkan bahwa peserta didik memiliki minat yang tinggi terhadap penerapan instrumen evaluasi HOTS dalam pembelajaran Bahasa Arab. peserta didik memandang evaluasi HOTS sebagai bentuk penilaian yang adil dan mampu mempresentasikan kemampuan berpikir mereka. Temuan ini menunjukkan bahwa peserta didik tidak hanya siap menerima bentuk evaluasi yang lebih menantang, tetapi juga memandang evaluasi sebagai sarana untuk menunjukkan kemampuan berpikir kritis, melakukan analisis, serta memecahkan masalah secara lebih autentik.

Disamping itu, Madrasah juga telah menunjukkan komitmen yang positif terhadap implementasi evaluasi berbasis HOTS. Kepala Madrasah dan jajaran guru telah memberikan dukungan melalui penyediaan pelatihan, sarana dan prasarana yang menunjang pengembangan instrumen evaluasi HOTS. Dukungan institusional ini mencerminkan keselarasan visi Madrasah dengan arah kebijakan pendidikan nasional yang mendorong asesmen autentik dan pengembangan keterampilan abad ke 21. Namun, dalam praktiknya guru bahasa Arab masih menghadapi sejumlah kendalam dalam menyusun instrumen evaluasi berbasis HOTS, seperti kesulitan menemukan ide soal, menyusun stimulus yang konstektual dan menarik, merancang pertanyaan pemantiik berpikir tingkat tinggi, serta keterbatasan waktu dalam pengembangan soal. Hambatan ini menunjukkan bahwa meskipun pemahaman konseptual guru tentang HOTS telah ada, namun guru masih membutuhkan panduan praktis dan instrumen siap pakai yang sistematis.

Temuan tersebut sejalan dengan penelitian yang dilakukan oleh Fakturmen bahwa implementasi pembelajaran bahasa Arab yang berorientasi pada HOTS masih menghadapi berbagai tantangan, seperti keterbatasan pengalaman, kurangnya pelatihan teknis, serta keterbatasan referensi menjadi faktor utama yang menghambat pengembangan instrumen evaluasi HOTS.[20] Selain itu, penelitian yang dilakukan oleh Heni Verawati juga menemukan bahwa sebagian besar aktivitas pembelajaran pada buku

[19] Stephen Kelvin Sata, 'A Critical Analysis of Ralph Tyler ' s Principles : Their Role in Shaping Historical and Foundational Curriculum Theories and Contemporary Educational Practices', 1.3 (2024), 1–8.
[20] Fakturmen, 'Kemampuan Guru Bahasa Arab Dalam Menyusun Soal Higher Order Thingking Skills (Hots) Pada Penilaian Akhir Sekolah (Pas) Di Man 3 Sleman', 2022.

paket masih berada pada level kognitif rendah sehingga belum sepenuhnya mendorong peserta didik untuk berpikir kritis dan analitis.[21]

Berdasarkan hasil analisis tersebut, maka penelitian ini penting terhadap pengembangan evaluasi pembelajaran bahasa Arab. Pertama, diperlukan pengembangan instrumen evaluasi yang secara sistematis dirancang untuk mengukur kemampuan berpikir tingkat tinggi. Kedua, guru membutuhkan panduan praktis dan contoh instrumen yang dapat membantu mereka dalam merancang soal HOTS secara lebih efektif. Ketiga, pengembangan instrumen evaluasi HOTS juga dapat menjadi sarana untuk meningkatkan kualitas pembelajaran bahasa Arab secara keseluruhan, karena evaluasi yang baik akan mendorong proses pembelajaran yang lebih bermakna.

Dengan demikian temuan penelitian ini telah menjawab rumusan masalah, bahwa madrasah sangat membutuhkan instrumen evaluasi berbasis HOTS dalam pembelajaran bahasa Arab di kelas XI MA DDI Takkalasi. Kebutuhan tersebut muncul karena adanya kesenjangan antara tuntutan kurikulum merdeka yang menekankan kemampuan berpikir tingkat tinggi dengan praktik evaluasi yang masih didominasi oleh soal-soal LOTS. Oleh karena itu, pengembangan instrumen evaluasi berbasis HOTS dalam penelitian ini diharapkan dapat menjadi solusi untuk menjembatani kesenjangan tersebut.

Prototipe instrumen evaluasi berbasis HOTS dalam pembelajaran bahasa Arab pada Kurikulum Merdeka di Kelas XI MA DDI Takkalasi

Proses pengembangan prototipe instrumen evaluasi berbasis HOTS dalam pembelajaran bahasa Arab dilakukan secara sistematis melalui tahapan penyusunan kisi-kisi isntrumen, pengembangan draf butir soal, penyusunan kunci jawaban, penyusunan rubrik penilaian dan pedoman penskoran, serta uji produk awal melalui proses validasi ahli. Prototipe instrumen yang dikembangkan terdiri dari 20 butir soal pilihan ganda dan 5 butir soal uraian yang dirancang untuk mengukur kemampuan berpikir tingkat tinggi peserta didik pada level analisis (C4), evaluasi (C5), dan Kreasi (C6) sebagaimana terdapat dalam revisi taksonomi Bloom oleh W. Anderson yang mengembangkan klasifikasi kognitif dari teori awal Benjamin S. Bloom.[22]

[21] Verawati and others, 'HOTS Analysis of Task Instructions in Bahasa Arab Madrasah Aliyah Textbook Published by The Ministry of Religious Affairs', *Edukatif: Jurnal Ilmu Pendidikan*, 4.1 (2022), 944-51.
[22] Rifda Haniefa, 'Implementasi Model Penilaian HOTS Pada Penilaian Keterampilan Berbahasa Arab', *Journal of Arabic Education & Arabic Studies*, 1.1 (2022), 49–71.

Proses pengembangan instrumen yang dilakukan dalam penelitian ini sejalan dengan teori evaluasi pendidikan yang dikemukakan oleh Ralph W. Tyler yang menegaskan bahwa evaluasi harus didasarkan pada tujuan pembelajaran yang jelas serta indikator yang terukur.[23] Dalam penelitian ini, tujuan pembelajaran dikembangkan ke dalam indikator HOTS kemudian disusun dalam bentuk kisi-kisi instrumen dan penyusunan butir soal. Kisi-kisi merupakan komponen penting karena berfungsi sebagai peta konseptual yang menjadi panduan antara kompetensi yang diukur dengan bentuk soal yang dikembangkan. Hal ini menunjukkan bahwa proses pengembangan instrumen telah memperhatikan prinsip validitas isi pada tahap awal pengembangan instrumen.

Pada tahapan uji coba produk awal, hasil validasi dari para ahli menunjukkan bahwa:

a. Ahli Media

Hasil validasi ahli menunjukkan bahwa prototipe instrumen yang dikembangkan memiliki tingkat validitas yang sangat tinggi. Pada validitas media, seluruh item memperoleh nilai I-CVI sebesar 1.00 dan nilai S-CVI/Ave sebesar 1.00 yang mengindikasikan adanya kesepakatan antar validator mengenai relevansi desain dan tampilan instrumen. Temuan ini mengindikasikan bahwa aspek tampilan, susunan isi, serta keterbacaam instrumen telah memenuhi standar kualitas media evaluasi pembelajaran yang baik. Meskipun demikian, validator tetap memberikan beberapa saran perbaikan terkait estetika visual, kejelasan petunjuk penggunaan, serta tata letak soal dan pilihan jawaban agar instrumen lebih mudah digunakan oleh peserta didik.

b. Ahli Materi

Hasil validasi ahli materi menunjukkan bahwa sebagian besar butir soal memiliki nilai I-CVI yang berada pada kategori sangat valid dengan nilai rata-rata S-CVI/Ave sebesar 0.97, hal ini menunjukkan bahwa secara keseluruhan instrumen telah memenuhi kriteria validitas isi yang sangat baik pada aspek konten, konstruk, bahasa, serta teknis penyusunan soal. Temuan ini juga menegaskan bahwa sebagian besar butir soal telah sesuai dengan indikator HOTS. Tapi terdapat beberapa butir soal yang masih perlu direvisi, terutama terkait ketepatan tata bahasa Arab, kesesuaian level kognitif soal, serta kualitas distraktor pada soal pilihan ganda.

[23] Stephen Kelvin Sata, 'A Critical Analysis of Ralph Tyler ' s Principles : Their Role in Shaping Historical and Foundational Curriculum Theories and Contemporary Educational Practices', 1.3 (2024) 1-8.

Proses revisi ini menunjukkan bahwa pengembangan instrumen evaluasi berbasis HOTS tidak hanya menuntut kompleksitas kognitif, tetapi juga ketelitian dalam aspek kebahasaan dan logika penyusunan soal.

Temuan penelitian ini sejalan dengan hasil penelitian Nailur Rahmawati yang menegaskan bahwa pembelajaran bahasa Arab yang mengarah pada HOTS memerlukan instrumen evaluasi yang mampu mengukur kemampuan HOTS peserta didik secara autentik.[24] Selain itu, hasil penelitian ini juga mendukung temuan penelitian Heni Verawati yang menunjukkan sebagian besar instruksi tugas dalam buku teks Bahasa Arab Madrasah Aliyah belum sepenuhnya mengembangkan keterampilan berpikir tingkat tinggi peserta didik.[25] Sehingga pengembangan instrumen evaluasi berbasis HOTS dalam penelitian ini dapat dipandang sebagai upaya untuk mengisi kesenjangan antara tuntutan kurikulum dan praktik evaluasi yang selama ini masih didominasi oleh soal LOTS.

Implikasi dari temuan penelitian ini menunjukkan bahwa pengembangan instrumen evaluasi berbasis HOTS dapat menjadi solusi dalam meningkatkan kualitas evaluasi pembelajaran bahasa Arab di Madrasah. Instrumen yang dirancang secara sistematis tidak hanya mampu mengukur kemampuan berpikir tingkat tinggi peserta didik, tetapi juga dapat membantu guru dalam merancang evaluasi yang lebih bermakna dan sesuai dengan tuntutan kurikulum merdeka serta pengembangan keterampilan abad ke-21.[26] Berdasarkan hasil analisis tersebut, dapat disimpulkan bahwa prototipe instrumen evaluasi berbasis HOTS yang dikembangkan dalam penelitian ini telah memenuhi kriteria validitas yang tinggi sehingga layak digunakan sebagai alat evaluasi pembelajaran bahasa Arab di kelas XI MA DDI Takkalasi.

Validitas dan reliabilitas instrumen evaluasi berbasis HOTS pada Pembelajaran Bahasa Arab yang dikembangkan untuk kelas XI MA DDI Takkalasi

a. Uji Validitas

Hasil uji coba lapangan menunjukkan bahwa sebagian besar butir soal pilihan ganda memiliki nilai signifikansi (Sig. 2-tailed) ≤ 0.05, sehingga dinyatakan valid. Hal ini menunjukkan bahwa butir soal memiliki korelasi yang signifikan dengan skor

[24] Rahmawati, Pembelajaran Bahasa Arab: Menuju Higher Order Thinking Skills (HOTS)', *Prosiding Konferensi Nasional Bahasa Arab*, 4.4 (2018), 149–54.
[25] Verawati and others, 'HOTS Analysis of Task Instructions in Bahasa Arab Madrasah Aliyah Textbook Published by The Ministry of Religious Affairs', *Edukatif: Jurnal Ilmu Pendidikan*, 4.1 (2022), 944-51.

total dan mampu mengukur konstruk yang sama, yaitu kemampuan berpikir tingkat tinggi peserta didik. Namun, terdapat beberapa butir soal yang tidak memnuhi kriteria validitas, yaitu soal nomor 1, 4, 6, 7, 11, dan 14 karena memiliki nilai signifikansi > 0.05, butir-butir tersebut kemudian direvisi sebelum dilakukan uji coba tahap akhir. Sementara itu, hasil analisis validitas butir soal uraian menunjukkan bahwa seluruh butir memiliki nilai sig. 2-tailed ≤ 0.05 dengan korelasi yang signifikan terhadap skor total dan soal esai yang dikembangkan memiliki validitas yang sangat baik.

Hasil uji coba final menunjukkan bahwa seluruh butir soal pilihan ganda maupun uraian memiliki nilai signifikansi ≤ 0.05 dengan koefisien korelasi diatas r tabel. Hal ini menunjukkan bahwa setelah dilakukan revisi, seluruh butir soal telah memenuhi kriteria validitas dan memiliki daya ukur yang baik dalam menilai kemampuan berpikir tingkat tinggi peserta didik. Temuan ini sejalan dengan teori validitas instrumen yang dikemukakan oleh Anastasi yang menyatakan bahwa suatu tes dikatakan valid apabila mampu mengukur apa yang seharusnya diukur.[27]

b. Uji Reliabilitas

Hasil analisis reliabilitas pada tahap uji coba lapangan menunjukkan bahwa instrumen pilihan ganda memperoleh nilai Alpha Cronbach sebesar 0.863, sedangkan instrumen uraian memperoleh nilai sebesar 0.741. nilai tersebut berada di atas batas minimal 0.70 sehingga instrumen dapat dikategorikan reliabel. Hasil ini menunjukkan bahwa butir-butir soal dalam instrumen memiliki konsistensi internal yang baik dan menghasilkan pengukuran yang relatif stabil.

Pada tahap uji coba final, nilai reliabilitas instrumen bahkan mengalami peningkatan, yaitu sebesar 0.868 untuk soal pilihan ganda dan 0.876 untuk soal uraian. Nilai ini berada pada kategori reliabilitas sangat tinggi yang menunjukkan bahwa instrumen memiliki tingkat konsistensi yang sangat baik dalam mengukur kemampuan berpikir tingkat tinggi peserta didik.

Temuan ini sesuai dengan teori reliabilitas tes yang dikemukakan oleh Cronbach yang menyatakan bahwa reliabilitas merupakan indikator konsistensi suatu

[26] Rina Mahmudati, 'The Characteristics And Developing An Assessment Instrument Of High Order Thinking Skill (HOTS) Items With 4-D Models', *Jurnal Pendidikan Matematika Dan IPA*, 14.1 (2023), 107–17.

[27] Indah Rahmi Nur Fauziah, Syihabudin Syihabudin, and Asep Sopian, 'Analisis Kualitas Tes Bahasa Arab Berbasis Higher Order Thinking Skill (Hots)', لسـانـنـا *(LISANUNA): Jurnal Ilmu Bahasa Arab Dan Pembelajarannya*, 10.1 (2020), 45–54.

instrumen dalam mengukur konstruk tertentu. Instrumen dengan nilai Alpha Cronbach yang tinggi menunjukkan bahwa seluruh butir soal memiliki keterkaitan yang kuat dalam mengukur konstruk yang sama.[28]

Hasil uji validitas dan reliabilitas dari penelitian ini menunjukkan bahwa instrumen evaluasi HOTS yang dikembangkan telah memenuhi kriteria instrumen evaluasi yang baik dari aspek validitas dan reabilitas. Dengan demikian, instrumen tersebut layak digunakan sebagai alat ukur dalam mengevaluasi kemampuan berpikir tingkat tinggi peserta didik dalam pembelajaran bahasa Arab pada Kurikulum Merdeka.

Keefektifan dan kepraktisan instrumen evaluasi berbasis HOTS dalam Pembelajaran Bahasa Arab yang dikembangkan untuk kelas XI MA DDI Takkalasi.

a. Keefektifan instrumen

1) Daya Pembeda, Tingkat Kesukaran, dan Analisis Distraktor

Hasil analisis daya pembeda pada tahap uji coba lapangan menunjukkan bahwa sebagian besar butir soal memiliki nilai corrected item–total correlation pada rentang 0.30–0.79 yang termasuk kategori cukup hingga sangat baik, sehingga mampu membedakan kemampuan peserta didik secara efektif. Namun beberapa butir soal, yaitu nomor 1, 4, 6, 7, 11, dan 14, memiliki daya pembeda rendah bahkan negatif sehingga perlu direvisi. Setelah dilakukan perbaikan dan uji coba final, seluruh butir soal menunjukkan daya pembeda yang berada pada kategori cukup hingga sangat baik, dengan rentang nilai 0.30–0.77 pada soal pilihan ganda dan 0.60–0.82 pada soal uraian.

Hasil analisis tingkat kesukaran menunjukkan bahwa sebagian besar soal pilihan ganda berada pada kategori sedang dengan indeks kesukaran 0,30–0,70. Beberapa soal berada pada kategori mudah dan sukar tapi tetap dipertahankan karena memiliki daya pembeda yang baik. Sementara itu, seluruh soal uraian berada pada kategori sedang hingga mendekati mudah, yang menunjukkan bahwa soal mampu menuntut peserta didik melakukan proses berpikir tingkat tinggi seperti analisis dan kreasi. Hal ini sejalan dengan prinsip penyusunan tes yang dijelaskan oleh Gronlund yang menyatakan bahwa instrumen evaluasi yang baik

[28] Keith S Taber, 'The Use of Cronbach ' s Alpha When Developing and Reporting Research Instruments in Science Education', 2018, 1273–96.

sebaiknya memiliki distribusi tingkat kesukaran yang seimbang sehingga mampu mengukur kemampuan peserta didik secara optimal.[29]

Analisis distraktor pada tahap uji coba lapangan menunjukkan bahwa sebagian besar opsi pengecoh telah berfungsi dengan baik karena dipilih oleh lebih dari 5% peserta didik. Namun beberapa distraktor pada soal nomor 1, 7, 11, dan 18 belum berfungsi optimal karena dipilih kurang dari 5% peserta didik. Setelah dilakukan revisi, hasil uji coba final menunjukkan bahwa seluruh distraktor telah berfungsi dengan baik dengan distribusi pemilihan jawaban yang relatif merata.

2) Hasil tes peserta didik

Tabel 2. Hasil Tes Peserta Didik

Descriptive Statistics

	N	Min.	Max.	Mean	Std. Deviation
Nilai_100	24	13.33	93.33	64.72	28.79817
Valid N (listwise)	24				

Hasil analisis menunjukkan bahwa nilai peserta didik berada pada rentang yang cukup lebar, dengan skor minimum sebesar 13.33 dan skor maksimum sebesar 93.33. Rentang nilai yang besar ini menunjukkan adanya variasi kemampuan berpikir tingkat tinggi peserta didik dalam menyelesaikan soal evaluasi berbasis HOTS. Nilai rata-rata sebesar 64.72 berada pada kategori cukup yang mengindikasikan bahwa secara umum peserta didik telah mampu mengerjakan soal-soal yang menuntut keterampilan berpikir tingkat tinggi khususnya pada level analisis dan mencipta, meskipun pencapaiannya belum merata pada seluruh peserta didik. Selain itu, nilai standar deviasi sebesar 28.79 menunjukkan tingkat penyebaran skor yang relatif tinggi, yang menunjukkan adanya kesenjangan kemampuan HOTS antar peserta didik.

Implikasi dari temuan ini menunjukkan bahwa penggunaan instrumen evaluasi berbasis HOTS dapat membantu guru memperoleh gambaran yang lebih akurat mengenai kemampuan berpikir peserta didik. Rentang nilai yang cukup luas menunjukkan bahwa instrumen mampu membedakan tingkat kemampuan

[29] Aulia Rahman and Heni Rusnayati, 'Analysis of the Characteristics of Higher Order Thinking Skills (HOTS) Test on Momentum and Impulse for Senior High School Student Using Item Response Theory', *Jurnal Penelitian Fisika Dan Aplikasinya (JPFA)*, 11.2 (2021), 127–37.

peserta didik secara efektif, sehingga guru dapat mengidentifikasi peserta didik yang telah memiliki kemampuan berpikir tingkat tinggi maupun yang masih memerlukan bimbingan lebih lanjut. Selain itu, hasil ini juga memberikan implikasi bahwa pengembangan evaluasi pembelajaran yang berorientasi HOTS perlu diiringi dengan strategi pembelajaran yang mendukung pengembangan kemampuan berpikir analisis, evaluasi, dan kreas peserta didik.

3) Ketuntasan Hasil Belajar Peserta Didik

Tabel 3. Ketuntasan Hasil Belajar Peserta Didik

Tuntas

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	10	41.7	41.7	41.7
	1.00	14	58.3	58.3	100.0
	Total	24	100.0	100.0	

Hasil analisis ketuntasan belajar menunjukkan bahwa dari 24 peserta didik, sebanyak 14 peserta didik (58.3%) telah mencapai nilai di atas Kriteria Ketuntasan Minimal (KKM), sedangkan 10 peserta didik (41.7%) belum mencapai ketuntasan. Persentase ketuntasan sebesar 58.3% menunjukkan bahwa tingkat keberhasilan peserta didik dalam menyelesaikan soal evaluasi berbasis HOTS berada pada kategori cukup. Hal ini mengindikasikan bahwa sebagian besar peserta didik telah mampu menyelesaikan soal-soal yang menuntut keterampilan berpikir tingkat tinggi, meskipun masih terdapat sejumlah peserta didik yang mengalami kesulitan.

Persentase ketuntasan yang belum mencapai tingkat tinggi dapat dipahami karena karakteristik soal HOTS menuntut kemampuan analisis, evaluasi, dan kreasi yang lebih kompleks dibandingkan soal pada tingkat kognitif rendah. Dalam perspektif teori kognitif kemampuan berpikir tingkat tinggi memerlukan proses penalaran, pengolahan informasi, serta integrasi pengetahuan yang lebih mendalam. Oleh karena itu, tidak semua peserta didik dapat dengan mudah mencapai ketuntasan ketika dihadapkan pada soal-soal yang menuntut keterampilan tersebut.[30] Dengan demikian, rumusan masalah penelitian mengenai efektivitas instrumen evaluasi HOTS dalam pembelajaran bahasa Arab dapat

dijawab bahwa instrumen yang dikembangkan mampu mengukur kemampuan berpikir tingkat tinggi peserta didik secara cukup efektif.

4) Ketercapaian Indikator

Hasil analisis ketercapaian indikator menunjukkan bahwa kemampuan peserta didik dalam menjawab soal berbasis HOTS berada pada kategori cukup tercapai hingga tercapai. Dari 23 indikator yang dianalisis, sebagian besar indikator berada pada rentang 60–75%, yang mengindikasikan bahwa peserta didik telah mampu memenuhi tuntutan kognitif tingkat tinggi, meskipun belum sepenuhnya optimal pada semua indikator. Indikator dengan level kognitif C4 (menganalisis) cenderung menunjukkan ketercapaian yang bervariasi, beberapa indikator C5 (mengevaluasi) berada pada kategori belum tercapai (≤ 60%), khususnya pada indikator yang evaluasi kritis. Temuan ini mengindikasikan bahwa peserta didik masih mengalami kesulitan dalam mengintegrasikan informasi secara mendalam dan melakukan penilaian terhadap stimulus yang disajikan.

Sebaliknya, indikator pada level C6 (mencipta), khususnya pada soal uraian, menunjukkan persentase ketercapaian yang relatif lebih tinggi, yaitu berada pada kategori cukup tercapai hingga tercapai (≥ 65%). Hal ini menunjukkan bahwa peserta didik lebih mampu mengekspresikan ide, menyusun jawaban secara terbuka, dan mengonstruksi respons berdasarkan pemahaman mereka sendiri ketika diberikan ruang berpikir yang lebih luas melalui soal uraian.

b. Kepraktisan instrumen

Hasil analisis angket kepraktisan menunjukkan bahwa instrumen evaluasi berbasis HOTS memperoleh nilai persentase sebesar 93%, yang berada pada kategori sangat praktis. Temuan ini menunjukkan bahwa instrumen mudah digunakan oleh guru, baik dari segi kejelasan petunjuk, kemudahan penerapan, maupun efisiensi waktu dalam pelaksanaan evaluasi. Tingginya tingkat kepraktisan tersebut mengindikasikan bahwa instrumen telah dirancang secara sistematis dan aplikatif sehingga dapat mendukung guru dalam melaksanakan evaluasi HOTS sesuai dengan tuntutan kurikulum merdeka.

[30] Ihwan Mahmudi, Neni Naqiyah, and Alif Cahya Setiadi, 'The Influence Of High Order Thinking Skill (Hots) Based Questions On Arabic Language Learning Outcomes Of Madrasah Tsanawiyah Student',

Selain itu, hasil analisis angket keterbacaan yang diberikan kepada 16 peserta didik menunjukkan bahwa persentase tiap butir berada pada rentang 70%–92% dengan kategori baik hingga sangat baik. Nilai rata-rata sebagian besar item berada di atas skor 4.00, yang menunjukkan bahwa instrumen memiliki kejelasan bahasa, struktur kalimat yang mudah dipahami, serta mampu menantang kemampuan berpikir peserta didik.

Hasil uji reliabilitas angket keterbacaan juga menunjukkan nilai Cronbach's Alpha sebesar 0.736 yang berada pada kategori reliabel. Nilai ini menunjukkan bahwa instrumen memiliki konsistensi internal yang baik. Dengan demikian, dapat disimpulkan bahwa instrumen evaluasi HOTS yang dikembangkan tidak hanya valid dan reliabel, tetapi juga praktis dan memiliki tingkat keterbacaan yang baik sehingga layak digunakan dalam evaluasi pembelajaran Bahasa Arab yang berorientasi pada pengembangan keterampilan berpikir tingkat tinggi.

Simpulan

Instrumen evaluasi Bahasa Arab berbasis HOTS di kelas XI MA DDI Takkalasi terbukti sangat dibutuhkan untuk memenuhi tuntutan Kurikulum Merdeka Belajar, mengingat evaluasi sebelumnya masih didominasi LOTS dan belum optimal mengukur kemampuan berpikir tingkat tinggi. Penelitian ini menunjukkan peserta didik memiliki minat dan kesiapan tinggi terhadap soal HOTS, terutama pada level C4 dan C6, meskipun guru masih menghadapi kendala teknis dalam penyusunannya. Prototipe instrumen yang dikembangkan secara sistematis dinyatakan sangat valid berdasarkan hasil validasi ahli materi (I-CVI ≥ 0,78 dan S-CVI/Ave = 0,90) serta ahli media (I-CVI dan S-CVI/Ave = 1,00), sehingga layak digunakan dengan perbaikan minor.

Hasil uji coba menunjukkan instrumen efektif dan sangat praktis, ditandai dengan daya pembeda yang baik, tingkat kesukaran sedang, variasi skor yang luas, serta respons guru sebesar 93%, meskipun ketuntasan belajar baru mencapai 58,3% akibat kesiapan kognitif peserta didik dan dominasi pembelajaran LOTS sebelumnya. Namun, generalisasi hasil masih terbatas karena karakteristik sampel yang homogen berbasis pesantren. Oleh karena itu, penelitian selanjutnya disarankan memperluas cakupan materi, jumlah responden, dan pendekatan analisis, termasuk penggunaan item response

theory dan mixed methods, guna memperoleh gambaran yang lebih komprehensif tentang kualitas instrumen dan proses berpikir peserta didik.

Daftar Rujukan

Ashfia, Ahmad, and Ubaid Ridlo, 'Optimalisasi Higher Order Thinking Skill (HOTS) Dalam Kurikulum Merdeka: Strategi Dan Konsep Penyusunan Soal Bahasa Arab Di MTs Pembangunan Jakarta', *Mauriduna: Journal of Islamic Studies*, 5.1 (2024), 330–42

Fakturmen, 'Kemampuan Guru Bahasa Arab Dalam Menyusun Soal Higher Order Thingking Skills (Hots) Pada Penilaian Akhir Sekolah (Pas) Di Man 3 Sleman', 2022

Fauziah, Indah Rahmi Nur, Syihabudin Syihabudin, and Asep Sopian, 'Analisis Kualitas Tes Bahasa Arab Berbasis Higher Order Thinking Skill (Hots)', *لسانـنـا (LISANUNA): Jurnal Ilmu Bahasa Arab Dan Pembelajarannya*, 10.1 (2020), 45–54

Fitra, Najmalia, Herdah Herdah, and Amira Ezzat Mahrous, 'Language Test Item Analysis Techniques', 11.1 (2025), 158–79 <https://doi.org/10.14421/almahara.2025.>

Handayani, Yuniar, Eni Asia, and Saleh Hidayat, 'Peningkatan Kemampuan High Order Thinking Skills (HOTS) Melalui Project-Based Learning (PjBL) Dalam Implementasi Kurikulum Merdeka', *PTK: Jurnal Tindakan Kelas*, 4.1 (2023), 48–60 <https://doi.org/10.53624/ptk.v4i1.236>

Khaerudin, 'Kualitas Instrumen Tes Hasil Belajar', *Jurnal Madaniyah*, 2 (2015), 212–35

Mahmudati, Rina, 'The Characteristics And Developing An Assessment Instrument Of High Order Thinking Skill (HOTS) Items With 4-D Models', *Jurnal Pendidikan Matematika Dan IPA*, 14.1 (2023), 107–17

Mahmudi, Ihwan, Neni Naqiyah, and Alif Cahya Setiadi, 'The Influence Of High Order Thinking Skill (Hots) Based Questions On Arabic Language Learning Outcomes Of Madrasah Tsanawiyah Student', *Jurnal At-Ta'dib*, 18.2 (2023)

Munip, Abdul, *Penilaian Pembelajaran Bahasa Arab* (Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta, 2019)

Musa, Muhammad Atanda, Rara Mutiah, and Rahmani, 'Analisis Butir Soal Bahasa Arab Di Mtsn Kota Parepare', *Sao Jurnal IAIN Parepare*, 1.2 (2022), 11–23

Putrawangsa, S, *Design Research Sebagai Framework Desain Pembelajaran* (Mataram: Sinabil Publishing, 2019)

Rahman, Aulia, and Heni Rusnayati, 'Analysis of the Characteristics of Higher Order Thinking Skills (HOTS) Test on Momentum and Impulse for Senior High School Student Using Item Response Theory', *Jurnal Penelitian Fisika Dan Aplikasinya (JPFA)*, 11.2 (2021), 127–37 <https://doi.org/10.26740/jpfa.v11n2.p127-137>

Rahmawati, Nailur, 'Pembelajaran Bahasa Arab: Menuju Higher Order Thinking Skills (HOTS)', *Prosiding Konferensi Nasional Bahasa Arab*, 4.4 (2018), 149–54

Rifda Haniefa, 'Implementasi Model Penilaian HOTS Pada Penilaian Keterampilan Berbahasa Arab', *Journal of Arabic Education & Arabic Studies*, 1.1 (2022), 49–71

Sari, Indri, Danil Usama, Dwi Noviani, and Firdaus Basuni, 'Langkah Penyusunan Dan Analisis Butir Soal Hots (Higher Order Thinking Skills) Pada Mata Pelajaran Pendidikan Agama Islam (PAI) Dan Budi Pekerti', *Jurnal Insan Pendidikan Dan Sosial Humaniora*, 1.4 (2023), 56–73

Sata, Stephen Kelvin, 'A Critical Analysis of Ralph Tyler ' s Principles : Their Role in

Shaping Historical and Foundational Curriculum Theories and Contemporary Educational Practices', 1.3 (2024), 1–8

Sihabuddin, Sihabuddin, 'Prosedur Penyusunan Tes Berbasis HOTS Pada Empat Keterampilan Berbahasa Arab', *LEARNING: Jurnal Inovasi Penelitian Pendidikan Dan Pembelajaran*, 3.1 (2023), 40–47

Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D* (Bandung: Alfabeta, 2020)

Susilowati, Yayuk, and Sumaji Sumaji, 'Interseksi Berpikir Kritis Dengan High Order Thinking Skill (Hots) Berdasarkan Taksonomi Bloom', *JURNAL SILOGISME: Kajian Ilmu Matematika Dan Pembelajarannya*, 5.2 (2021), 62–71

Taber, Keith S, 'The Use of Cronbach ' s Alpha When Developing and Reporting Research Instruments in Science Education', 2018, 1273–96 <https://doi.org/10.1007/s11165-016-9602-2>

Verawati, Heni, Evi Febriani, Intan Muflihah, Uswatun Hasanah, Agus Susanti, and Fitriani Fitriani, 'HOTS Analysis of Task Instructions in Bahasa Arab Madrasah Aliyah Textbook Published by The Ministry of Religious Affairs', *Edukatif: Jurnal Ilmu Pendidikan*, 4.1 (2022), 944–51

Waruwu, Marinu, 'Metode Penelitian Dan Pengembangan (R&D): Konsep, Jenis, Tahapan Dan Kelebihan', *Jurnal Ilmiah Profesi Pendidikan*, 9.2 (2024), 1220–30 <https://doi.org/10.29303/jipp.v9i2.2141>