EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

ANALISIS PENGEMBANGAN INSTRUMEN PENILAIAN TES BERORIENTASI HOTS PADA MATERI PERKEMBANGAN ISLAM: TINJAUAN LITERATUR SISTEMATIS

Ida Nur Azizah, Wasino
Penelitian dan Evaluasi Pendidikan, Sekolah Pascasarjana, Universitas Negeri Semarang
e-mail: idanurazizah219@gmail.com, wasino@mail.unnes.ac.id

Diterima: *16/05/2026*; Direvisi: *23/05/2026*; Diterbitkan: *18/06/2026*

ABSTRAK

Perkembangan ilmu pengetahuan dan teknologi yang pesat harus diimbangi dengan kemampuan peserta didik dalam menyaring dan memanfaatkan informasi secara tepat dan bertanggung jawab. Hal ini sejalan dengan kompetensi yang dibutuhkan pada abad 21 yang menekankan kemampuan berpikir kritis, kreatif, komunikatif, dan kolaboratif. Oleh karena itu, diperlukan instrumen asesmen yang tepat untuk mengukur kompetensi dan keterampilan peserta didik tersebut. Tujuan penelitian ini ialah mengkaji secara sistematis berbagai studi tentang pengembangan instrumen penilaian tes berorientasi *Higher-Order Thinking Skills* (HOTS) pada materi perkembangan islam. Pendekatan yang digunakan adalah *Systematic Literature Review* (SLR) dengan menganalisis 7 artikel jurnal nasional, 3 artikel jurnal prosiding, dan 21 artikel jurnal internasional terindeks Scopus yang terbit antara tahun 2014 hingga 2025. Proses seleksi dilakukan melalui tahapan identifikasi, penyaringan, kelayakan, dan sintesis menggunakan kriteria PRISMA. Hasil penelitian menunjukkan bahwa pengembangan instrumen tes HOTS umumnya dilakukan melalui pendekatan penelitian dan pengembangan (*Research and Development*) dengan model-model Borg & Gall, ADDIE, 4D, dan Tessmer. Bentuk instrumen yang banyak digunakan berupa pilihan ganda, uraian, dan two-tier dengan menekankan kemampuan menganalisis, mengevaluasi, dan mencipta sesuai taksonomi Bloom revisi. Sementara itu, sebagian besar instrumen yang dikembangkan memiliki tingkat validitas dan reliabilitas yang baik sehingga layak digunakan dalam proses evaluasi pembelajaran. Kajian ini diharapkan dapat menjadi referensi bagi peneliti dan pendidik dalam pengembangan instrumen HOTS yang berkualitas sesuai tuntutan keterampilan abad-21.
Kata Kunci: *Systematic Literature Review, PRISMA, Pengembangan Instrumen Penilaian, HOTS*

ABSTRACT

Rapid advances in science and technology must be balanced by students' ability to filter and utilize information appropriately and responsibly. This aligns with the competencies required in the 21st century, which emphasize critical, creative, communicative, and collaborative thinking skills. Therefore, appropriate assessment instruments are needed to measure these competencies and skills in students. The purpose of this study is to systematically review various studies on the development of assessment instruments for Higher-Order Thinking Skills (HOTS)-oriented tests on Islamic development materials. The approach used was a Systematic Literature Review (SLR), analyzing 7 national journal articles, 3 conference proceedings articles, and 21 Scopus-indexed international journal articles published between 2014 and 2025. The selection process involved the stages of identification, screening, eligibility, and synthesis using the PRISMA criteria. The research results indicate that the development of HOTS test instruments is generally carried out through a Research and Development (R&D) approach

https://doi.org/10.51878/educational.v6i3.11239

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

using the Borg & Gall, ADDIE, 4D, and Tessmer models. The most commonly used instrument formats are multiple-choice, essay, and two-tier questions, emphasizing the abilities to analyze, evaluate, and create in accordance with the revised Bloom's Taxonomy. Meanwhile, most of the instruments developed demonstrate good levels of validity and reliability, making them suitable for use in the learning evaluation process. This study is expected to serve as a reference for researchers and educators in developing high-quality HOTS instruments that meet the demands of 21st-century skills.

Keywords: *Systematic Literature Review, PRISMA, Assessment Instrument Development, HOTS*

PENDAHULUAN

Kemampuan berpikir tingkat tinggi menjadi kebutuhan fundamental dalam membentuk generasi yang reflektif dan berdaya saing. Merujuk pada *Framework for 21st Century Learning,* terdapat tiga kompetensi yang diharapkan dapat dikuasai peserta didik dalam dunia nyata dan dunia kerja. Keterampilan abad ke-21 tersebut terdiri atas keterampilan belajar dan berinovasi, keterampilan hidup dan karier, serta keterampilan informasi, media, dan teknologi (*Partnership for 21st Century Learning*, 2009). Integrasi keterampilan tersebut merupakan proses kognitif tingkat tinggi melalui dorongan berpikir kritis, pemecahan masalah kompleks, serta implementasi pengetahuan konteks terkini. Keterampilan berpikir tingkat tinggi menjadi kemampuan esensial yang akan membentuk pola paradigma berpikir global dan menjawab tantangan abad ke-21.

Namun demikian, dalam paparan pembelajaran mendalam (Kemendikdasmen, 2025) dinyatakan bahwa keterampilan berpikir tingkat tinggi peserta didik di Indonesia masih sangat rendah. Hasil *Programme for International Student Assessment* (PISA) tahun 2022 menunjukkan bahwa > 99 % murid di Indonesia hanya bisa menjawab soal level dasar 1-3 atau *Lower Order Thinking Skills* (LOTS). Sedangkan < 1% yang mampu menjawab soal level tinggi 4-6 atau *Higher Order Thinking Skills* (HOTS).

Fenomena serupa juga terjadi di SMK Takhassus Al-Qur'an, di mana pada tahun 2025 kompetensi memahami, menggunakan, merefleksi, serta mengevaluasi teks informasi menurun 0,49 % dari tahun sebelumnya, yaitu 89,25. Rapor pendidikan juga menunjukkan bahwa kompetensi menemukan, mengidentifikasi serta mendeskripsikan suatu ide atau informasi secara eksplisit mengalami penurunan 1,86% dari tahun sebelumnya. Keadaan ini mengindikasikan bahwa terjadi penurunan kemampuan berpikir tingkat tinggi pada peserta didik pada satuan pendidikan. Oleh karena itu, peran guru menjadi sangat krusial dalam peningkatan kualitas pembelajaran, salah satunya melalui evaluasi.

Menurut Shalikhah & Nugorho (2023), kemampuan berpikir tingkat tinggi dapat ditingkatkan melalui penguatan model, media, dan asesmen secara terpadu. Sistem penilaian yang berkualitas dapat mendorong peserta didik untuk berpikir kritis, kreatif, dan reflektif dalam penyelesaian suatu masalah. Namun demikian, studi pendahuluan yang dilakukan di SMK Takhassus menunjukkan bahwa instrumen penilaian yang digunakan masih bersifat sederhana tanpa pemenuhan substansi, konstruksi, bahasa, bahkan validitas empirik. Sebagian besar guru masih menggunakan penilaian konvensional dalam proses evaluasi pembelajaran. Penilaian konvensional menganalisis hasil belajar peserta didik dengan menggunakan tes tertulis yang hanya mampu mengukur aspek kognitif dan keterampilan sederhana. Asesmen tradisional hanya berfokus pada kognitif saja, sering kali mengabaikan aspek afektif dan psikomotorik (Nurhaibi, 2025).

https://doi.org/10.51878/educational.v6i3.11239

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

Penelitian pengembangan instrumen HOTS menjadi sangat perlu untuk dilakukan sebagai sarana pengukuran keterampilan berpikir tingkat tinggi secara komprehensif. Penelitian ini pernah dilakukan oleh Ndoen *et al.* tentang pengembangan instrumen HOTS berbasis sejarah lokal NTT menggunakan model *Formative Research Tessmer* (validasi ahli, uji coba individu, uji coba kelompok kecil, dan uji coba lapangan). Penelitian ini menghasilkan instrumen penilaian yang memenuhi tingkat kelayakan. Integrasi sejarah lokal terbukti mampu mendorong peserta didik untuk melakukan analisis, evaluasi, dan penalaran sejarah tingkat tinggi. Uji coba instrumen dilakukan pada lingkup peserta didik yang terbatas dan belum dilakukan diseminasi secara luas (Ndoen *et al*., 2025). Penelitian sejenis juga dilakukan oleh Radiyansah et al. yang berjudul Development of evaluation tools HOTS-based project learning model to improve critical thinking *ability*. Analisis deskriptif dilakukan untuk gambaran skor awal dan akhir. Sementara itu, pendekatan kuantitatif dilakukan untuk uji statistik skor tes, uji homogenitas, N-Gain, dan uji t. Instrumen terbukti meningkatkan kemampuan kritis menggunakan analisis N-gain 0.75 (Radiansyah *et al.*, 2024).

Meskipun penelitian mengenai pengembangan instrumen tes HOTS telah banyak dilakukan, perlu dilakukan kajian mendalam yang merangkum penelitian secara sistematis. Sebagian besar penelitian memiliki variasi bentuk instrumen, model pengembangan, mata pelajaran, materi, serta teknik analisis yang beragam. Di sisi lain, kebutuhan akan gambaran menyeluruh sangat diperlukan untuk mengetahui arah perkembangan penelitian yang relevan. Oleh karena itu, diperlukan suatu tinjauan literatur sistematis untuk mengidentifikasi, menganalisis, dan menyintesis hasil-hasil penelitian pengembangan secara terstruktur.

Berdasarkan elaborasi di atas, penelitian dilakukan dengan tujuan untuk menganalisis hasil penelitian pengembangan instrumen tes HOTS melalui *Systematic Literature Review* sehingga dapat diketahui tren, metode, dan pola pengembangan. Fokus penelitian ini ialah untuk mengetahui bentuk instrumen yang dikembangkan, model dan pendekatan metodologis, kualitas psikometrik instrumen, serta kesenjangan yang masih ditemukan dalam pengembangan instrumen penilaian tes berorientasi HOTS.

METODE PENELITIAN

Penelitian ini menggunakan metode Systematic Literature Review (SLR) untuk mengidentifikasi, mengevaluasi, dan mensintesis hasil penelitian terkait pengembangan instrumen penilaian tes berorientasi Higher Order Thinking Skills (HOTS). Proses kajian dilakukan mengacu pada pedoman PRISMA yang meliputi identifikasi, penyaringan, penilaian kelayakan, dan sintesis artikel. Sumber data diperoleh dari database Google Scholar dan Scopus melalui bantuan perangkat lunak Publish or Perish. Penelusuran dilakukan menggunakan kata kunci yang berkaitan dengan *HOTS*, *assessment instrument*, *test development*, dan *history learning*. Artikel yang dipilih merupakan publikasi tahun 2014–2025 yang membahas pengembangan instrumen penilaian berorientasi HOTS.

Kriteria inklusi meliputi artikel yang memuat proses pengembangan instrumen, laporan validitas dan reliabilitas, serta tersedia dalam teks lengkap. Artikel yang tidak relevan, duplikat, atau tidak memenuhi kriteria kualitas dikeluarkan dari analisis. Setelah proses seleksi, diperoleh 31 artikel yang memenuhi syarat untuk dikaji. Data diekstraksi berdasarkan aspek penulis, tahun publikasi, model pengembangan, bentuk instrumen, serta hasil validitas dan reliabilitas. Selanjutnya data dianalisis secara deskriptif untuk mengidentifikasi tren, pola, dan kesenjangan penelitian pengembangan instrumen HOTS.

https://doi.org/10.51878/educational.v6i3.11239

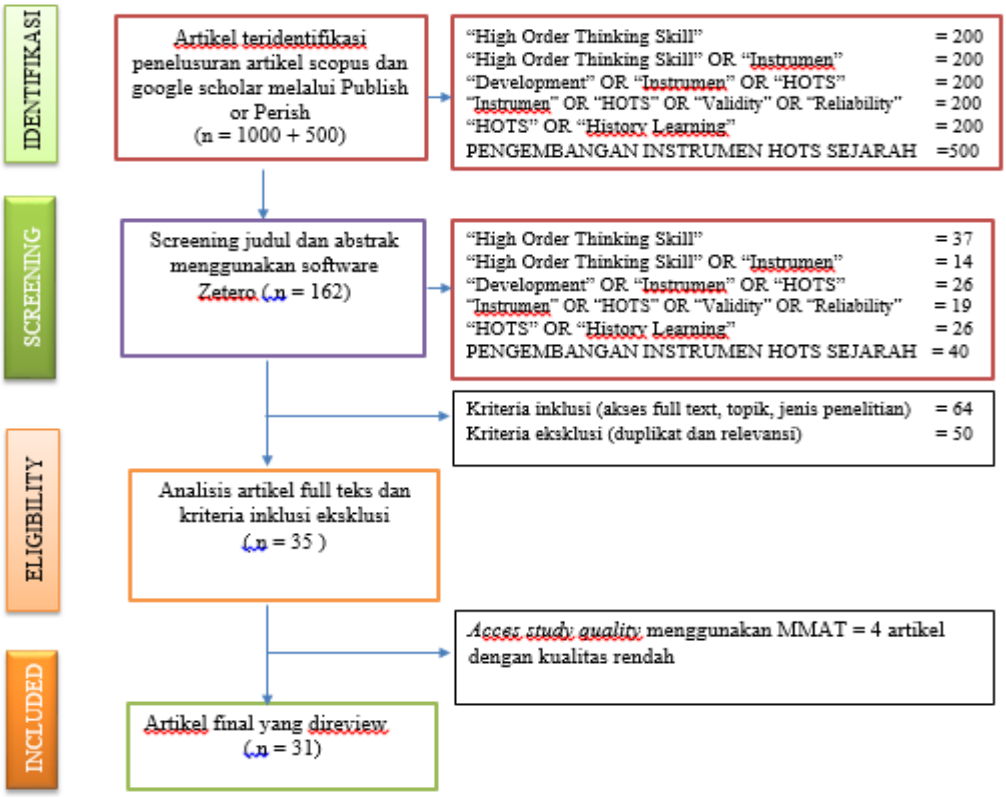
EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

HASIL DAN PEMBAHASAN

Hasil

Hasil penelitian yang dilakukan secara sistematis melalui beberapa tahapan, mulai dari *planning, collect paper, select eligible, assess study quality, extract data, analysis and synthesis,* hingga penyusunan manuskrip. Melalui tahapan tersebut, diperoleh 31 artikel ilmiah dari jurnal nasional dan internasional yang relevan. Artikel tersebut kemudian dianalisis secara mendalam dan disintesis untuk memperoleh gambaran secara komprehensif terkait topik yang dikaji dan mengidentifikasi kecenderungan hasil penelitian sehingga peneliti dapat menemukan kesenjangan dan mengetahui penelitian yang masih perlu dikembangkan.

Adapun proses seleksi artikel atau sumber data yang digunakan sebagai pendekatan dari penelitian ini disajikan pada Gambar 1. Diagram Proses Seleksi PRISMA. Visualisasi tersebut diharapkan mampu menggambarkan proses mengidentifikasi sumber literatur, menyaring artikel yang relevan, menilai kelayakan artikel, serta menentukan artikel akhir yang digunakan dalam analisis.



Gambar 1. Diagram Proses Seleksi PRISMA

Visualisasi pada Gambar 1 menunjukkan bahwa penentuan artikel akhir berasal dari proses identifikasi 1000 artikel jurnal internasional dan 500 artikel jurnal nasional melalui software Publish or Perish. Sejumlah 1500 artikel kemudian dilakukan *screening* pada software Zetero sehingga menghasilkan 162 artikel terpilih. Peninjauan kriteria inklusi dengan screening judul, metode penelitian, jenis penelitian, dan topik penelitian menghasilkan 64 artikel jurnal relevan. Sementara, hasil kriteria eksklusi dari duplikat dan relevansi menghasilkan 50 artikel jurnal. Pada tahap pembacaan artikel, didapatkan 35 karena masih terdapat konteks yang kurang relevan dengan tujuan penelitian. Meskipun demikian, peneliti memasukkan 2 penelitian

https://doi.org/10.51878/educational.v6i3.11239

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

kualitatif (Palgunadi et al., 2023 dan Khalid Rahman et al., 2024) karena konteks HOTS yang relevan dengan tujuan penelitian. Sementara itu, pada tahap evaluasi kualitas metodologi, ditemukan 4 dengan skor 3,3,3,dan 4 dengan kriteria rendah. Hal ini dipengaruhi oleh desain produk, produk pengembangan, integrasi, analisis, dan revisi berbasis data yang tidak tersedia. Keempat artikel tersebut hanya menyediakan analisis data, implementasi, dan evaluasi, sehingga artikel final yang akan dijadikan objek kajian berjumlah 31 artikel. Artikel-artikel tersebut disajikan pada Tabel 1. Daftar Ringkasan Artikel Hasil Seleksi SLR.

Tabel 1. Daftar Ringkasan Artikel Hasil Seleksi SLR

No	Penulis (Tahun)	Metode & Model	Bentuk Instrumen	Hasil Penelitian
1	Nasution, *et al.,* (2021)	Research and Development (R&D), Model Borg & Gall yang diadopsi sesuai kebutuhan.	Pilihan ganda (multiple choice) dan uraian (essay)	Kelayakan isi instrumen mempunyai kategori sangat valid. Instrumnen juga membuktikan efektif meningkatkan keterampilan berpikir tingkat tinggi.
2	Budiastuti, *et al.,* (2023)	R&D menggunakan Model 4D oleh Thiagarajan (*Define, Design, Develop, Disseminate*).	Instrumen penilaian diri dalam praktikum pembuatan busana.	Instrumen dinyatakan valid oleh ahli desain busana dengan kategori "Hampir Sempurna". Penilaian diri mampu menumbuhkan karakter kemandirian dan refleksi kritis mahasiswa selama praktikum.
3	Bunari & Yuliantoro, (2020)	R&D dengan pengumpulan data melalui wawancara dan studi literatur.	Pilihan ganda (*multiple choice*	Menghasilkan draf instrumen evaluasi Sejarah Riau yang valid berdasarkan penilaian para ahli materi dan terbukti efektif mengukur kemampuan analisis historis mahasiswa.
4	Bunari, *et al.,* (2022)	R&D dengan fokus pada langkah pengujian validitas oleh pakar materi dan reliabilitas item tes.	Pilihan ganda (*multiple choice*	Menunjukkan bahwa instrumen tes yang dikembangkan memenuhi syarat validitas dan reliabilitas. Tes ini layak digunakan untuk melatih kemampuan berpikir kritis mengenai sejarah maritim lokal.
5	Darmana , *et al.*, (2020)	R&D dengan fokus pada validasi ahli materi kimia dan ahli integrasi agama Islam.	Instrumen tes pilihan ganda (*multiple choice*)	Instrumen dinyatakan valid dan reliabel untuk mengukur kemampuan berpikir tingkat tinggi sekaligus menginternalisasi nilai-nilai keagamaan (tauhid) dalam pembelajaran kimia.
6	Dewi, *et al.,* (2020)	R&D dengan dengan model penelitian formatif.	Lembar evaluasi	Instrumen terbukti sangat layak digunakan berdasarkan hasil validasi ahli dan mampu memetakan tingkat berpikir tingkat tinggi.
7	Fadlila & Sagala, (2021)	R&D dengan pendekatan *Realistic Mathematics Education* (RME).	Instrumen tes matematika	Instrumen tes dinyatakan valid, reliabel, dan praktis. Hasil implementasi menunjukkan penggunaan instrumen ini secara efektif merangsang penalaran dan meningkatkan skor HOTS siswa.
8	Ndoen, *et al.,* (2025)	R&D dengan model procedural.	Instrumen tes berbasis HOTS	Instrumen valid, praktis, dan memiliki keefektifan yang baik dalam mengevaluasi kemampuan kognitif tingkat tinggi siswa.

https://doi.org/10.51878/educational.v6i3.11239

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

9	Handayani, *et al.,* (2019)	R&D dengan fokus pada tahap Analisis Kebutuhan (*Need Analysis*).	Instrumen penilaian proyek (*study project assessment*)	Guru sangat membutuhkan instrumen penilaian proyek berorientasi HOTS berbasis Android pada materi kalor.
10	Setiawan, *et al.,* (2021)	R&D dengan model Borg & Gall.	Instrumen tes pilihan ganda (*multiple choice*)	Instrumen sangat valid dan butir soal memiliki reliabilitas tinggi, tingkat kesukaran seimbang, daya beda kokoh, dan fungsi pengecoh yang berjalan efektif.
11	Rahman, *et al.,* (2024)	*Library Research* dengan pendekatan analisis konten dan deskriptif-analitis berbasis perspektif neurosains.	Model kerangka instrumen pembelajaran.	Penerapan strategi berbasis neurosains mampu mengoptimalkan fungsi kognitif, afektif, dan psikomotorik, serta membantu siswa mengaitkan peristiwa sejarah.
12	Kim How, *et al.,* (2023)	*Design and Development Research* (DDR) dengan 7 tahapan	Instrumen tes subjektif terstruktur (17 butir soal terstruktur)	Instrumen HOTS-QE valid secara konten, bahasa, serta memiliki akurasi domain HOTS yang tinggi untuk mengukur kemampuan berpikir tingkat tinggi.
13	Maxnun, *et al.,* (2024)	R&D model ADDIE	Instrumen penilaian kognitif	Instrumen valid, reliabel, dan efektif secara signifikan dalam mengukur serta melatih HOTS
14	Nursalam, *et al.,* (2018)	R&D model tessmer (self evaluation, expert judgment, one-to-one, small group, dan field test)	Instrumen tes tertulis	Uji validitas isi menunjukkan instrumen berada dalam kategori sangat valid. Uji coba lapangan membuktikan produk memiliki reliabilitas tinggi serta daya beda yang baik.
15	Sagala & Andriani, (2019)	Research and Development (R&D) model tessmer	Butir soal tes evaluasi	Instrumen valid secara substansi dan konstruksi, serta efektif melatih penalaran abstrak level menganalisis dan mengevaluasi.
16	Rintayati, *et al.,* (2020)	R&D dengan fokus pada tahap implementasi dan pengujian kelayakan.	Instrumen tes pilihan ganda dua tingkat (Two-Tier Multiple-Choice Test / TTMCT)	Instrumen sangat layak, valid, dan andal untuk mengukur kemampuan HOTS. Format dua tingkat efektif mengurangi faktor menebak dan mampu mengidentifikasi argumen.
17	Radiansyah, *et al.,* (2024)	R&D dengan mengintegrasikan evaluasi berbasis HOTS	Alat evaluasi pembelajaran proyek	Instrumen valid dan praktis, serta menunjukkan peningkatan signifikan pada kemampuan berpikir kritis siswa.
18	Ramadhan, *et al.,* (2019)	R&D modifikasi Model Oriendo	Instrumen tes diagnostik	Instrumen yang dikembangkan memenuhi validitas isi dan reliabilitas (rasch).
19	Rosidin, *et al.,* (2019)	R&D model Borg & Gall 7 langkah.	Model perangkat pembelajaran terintegrasi STEM	Model kombinasi secara signifikan terbukti mempercepat lompatan kemampuan siswa dalam menganalisis, mengevaluasi, dan berkreasi.
20	Serevina, *et al.,* (2019)	R&D dengan model Borg&Gall	Instrumen penilaian/soal tes berbasis HOTS	Instrumen yang valid, reliabel, dan memiliki tingkat kesukaran proporsional. Soal-soal terbukti mampu memicu keterampilan analisis fisis dan memecahkan fenomena fluida di kehidupan.

https://doi.org/10.51878/educational.v6i3.11239

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

21	Widyaningsih, *et al.,* (2021)	Research and Development (R&D) dengan model ADDIE	Tes fisika berbasis HOTS melalui LMS Moodle	Instrumen efektif digunakan secara online untuk mengukur kemampuan HOTS siswa.
22	Yunita, *et al.,* (2018)	R&D dengan model Tessmer	Instrumen tes matematika berbasis HOTS	Instrumen yang dikembangkan valid dan reliabel untuk mengukur HOTS siswa kelas VIII.
23	Zaki, *et al.,* (2020)	R&D dengan model Tessmer	Instrumen tes HOTS pada materi eksponen	Soal-soal yang dihasilkan memenuhi kriteria HOTS dan layak digunakan
24	Zhou, *et al.,* (2023)	Exploratory Factor Analysis & Confirmatory Factor Analysis	Skala asesmen kuantitatif HOTS	Skala asesmen valid dan reliabel untuk mengukur kemampuan HOTS calon guru.
25	Mildasari & Aisiah (2022)	R&D dengan model ADDIE.	Pilihan ganda	Instrumen dinyatakan valid dan layak serta praktis setelah melalui uji kelayakan.
26	Jallapaksi, J., (2024)	R&D dengan model ADDIE yang dibatasi hanya sampai 3 tahap.	Asesmen interaktif digital bentuk pilihan ganda	Produk dinilai sangat layak dengan validasi materi dan uji coba kelompok kecil berkategori sangat baik
27	Aryanti, *et al.,* (2023)	R&D model tessmer	Instrumen evaluasi kognitif	Penelitian menghasilkan instrumen yang valid dan reliabel. Butir soal mampu mengukur kemampuan berpikir tingkat tinggi siswa secara efektif.
28	Kartikaningtyas. D.Y., *et al.,* (2025)	Research and Development (R&D) menggunakan Model ADDIE.	Instrumen tes diagnostik digital (pilihan ganda dua tingkat)	Instrumen valid dan reliabel. Tingkat kesukaran dan daya beda terdistribusi ideal (Rasch) dan efektif untuk memetakan profil miskonsepsi prasyarat siswa.
29	Melati, M. N., *et al.,* (2022)	R&D dengan model Borg & Gall modifikasi 7 langkah	Instrumen tes tertulis berbentuk soal uraian/esai	Instrumen menunjukkan kategori sangat baik (skala kecil) dan Baik (uji lapangan)
30	Palgunadi, K., *et al.,* (2023)	Classroom Action Research yang dilaksanakan dalam dua siklus tindakan.	Media pembelajaran interaktif berupa pameran virtual	Meningkatkan HOTS siswa kelas 8A SMP Negeri 10 Denpasar.
31	Sari, R.W.A & Purwaningsih, M., (2024)	R&D model ADDIE secara lengkap.	Soal evaluasi formatif digital	Produk terbukti sangat valid dan tingkat kepraktisan dengan kategori sangat baik.

Berdasarkan Tabel 1, 7 artikel jurnal nasional, 21 artikel jurnal internasional terindeks scopus dan 3 artikel prosiding didapatkan data bahwa kerangka konseptual paling dominan yang digunakan dalam pengembangan instrumen terdiri atas kerangka atau model 4D (*Define, Design, Develop, dan Disseminate*), model ADDIE (*Analyze, Design, Develop, Implementation, Evaluation*), *Formative Research* (Tessmer), model Borg & Gall, model oriondo modifikasi. Terdapat pula beberapa peneliti yang melakukan modifikasi langkah seperti halnya Kim How, *et al.,* yang modifikasi model Borg & Gall menjadi 7 tahap.

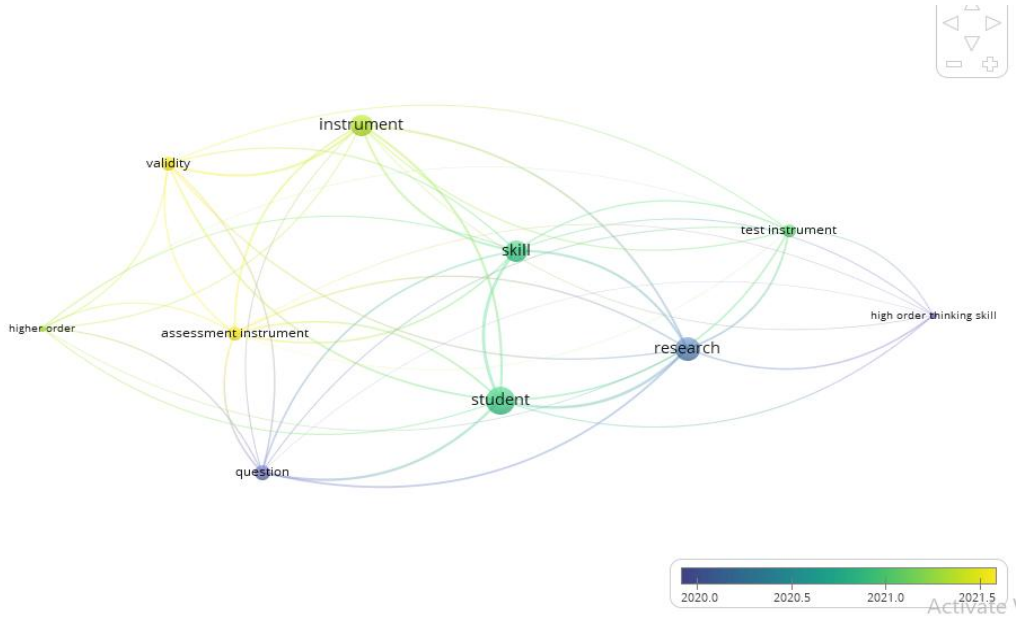
https://doi.org/10.51878/educational.v6i3.11239

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

Sementara itu, bentuk instrumen yang dikembangkan berupa instrumen tes dan instrumen non tes. Instrumen tersebut menunjukkan keberagaman seperti soal subjektif terstruktur, instrumen *self assesment* menggunakan skala likert, pilihan ganda, *Two-Tier Multiple Choice*, dan tes esai/uraian. Variasi bentuk soal yang dikembangkan dipengaruhi oleh tujuan penelitian karena setiap bentuk soal memiliki karakteristik, kelebihan, dan fungsi pengukuran yang berbeda. Kartikaningtyas, *et.al* (2025) mengembangkan instrumen diagnostik pilihan ganda untuk mengukur materi prasyarat. Bentuk tes pilihan ganda diuji cobakan melalui *Google Form*.

Kualitas psikometrik instrumen yang dikembangkan mengacu pada sejauh mana instrumen mempunyai tingkat kelayakan yang tinggi dan dapat mengukur suatu konstruk. Instrumen dikatakan baik secara ilmiah jika memenuhi tingkat validitas (isi, konstruk, empiris) dan reliabilitas yang tinggi. Berdasarkan studi literatur menunjukkan bahwa semua instrumen yang dikembangkan valid dan reliabel serta memenuhi tingkat kepraktisan. Penelitian Kim How *et al*., (2022) menghasilkan temuan yaitu 17 butir soal yang dikembangkan mempunyai tingkat Reliabilitas 0.79, I-CVI > 0.70, S-CVI 0.98, sehingga instrumen layak digunakan. Begitu juga Radiyansah (2024) yang mengembangkan butir soal berjumlah 10 soal pilihan ganda dan 5 esai menghasilkan tingkat validitas 88% dan kepraktisan 91%, serta efektif mengukur keterampilan berpikir tingkat tinggi. Ramadhan *et al*. (2019), menggunakan teori klasik melalui *iteman* dan teori modern *rasch* melalui *software winsteps*. Hasil analisis menunjukkan tingkat reliabilitas sebesar 0.81. Penelitian yang dilakukan Dwi Yuni *et al.,* (2025) menunjukkan bahwa nilai Aiken's V sebesar 0,90 dan *Cronbach's Alpha* 0,83. Sementara itu, analisis tingkat kesukaran dan analisis daya beda menunjukkan distribusi yang ideal.

Pada tahap analisis dan sintesis, peneliti menggunakan *software VOS viewer* untuk menemukan pola, tren, gap penelitian, dan rekomendasi. Hal ini dilakukan untuk memperoleh gambaran secara menyeluruh sehingga peneliti dapat mengidentifikasi tema-tema yang paling banyak diteliti, hubungan antarkonsep, arah perkembangan penelitian, serta topik yang masih jarang dikaji sehingga dapat menjadi peluang penelitian selanjutnya. Hasil analisis tersebut tersaji pada Gambar 2. Hasil Analisis *VOS viewer*.



Gambar 2. Hasil Analisis *VOS viewer*

https://doi.org/10.51878/educational.v6i3.11239

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

Berdasarkan Gambar 2 terakit visualisasi yang terdapat pada *software VOS viewer*, terlihat penelitian membentuk empat klaster. Klaster pertama berupa instrumen dan validitas yang menunjukkan bahwa hampir semua penelitian pengembangan instrumen penilaian menekankan uji validitas dan reliabilitas. Teknik validasi didominasi oleh ahli materi dan ahli penilaian menggunakan Aiken's V atau CVI disertai dengan reliabilitas internal melalui *Cronbach's Alpha*. Namun, terdapat pula penelitian Zhou. Y. *et al*., (2023) yang khusus menganalisis validitas konstruk melalui EFA (*Exploratory Factor Analysis*) dan CFA (*Confirmatory Factor Analysis*). Klaster kedua menunjukkan orientasi pengukuran menggunakan Taksonomi Bloom revisi untuk menentukan indikator soal. Klaster ketiga ialah subjek penelitian didominasi oleh peserta didik pada mata pelajaran tertentu. Klaster keempat berupa model pengembangan sebagai pedoman metodologi penelitian.

Sementara itu, terlihat bahwa tren penelitian dari hasil *overlay visualization dan density visualization menunjukkan adanya pergeseran tren, dari penyusunan soal dan instrumen dasar menjadi pengukuran keterampilan seperti* berpikir tingkat tinggi (HOTS) dengan aspek karakter, spiritual, dan literasi sains. Terdapat tren digitalisasi melalui LMS, Android dan *Computer Based Test*. Tren masih berfokus pada aspek teknis instrumen dan keterampilan HOTS siswa, belum banyak mengeksplorasi konteks atau integrasi lokal, meskipun ditemukan integrasi tauhid dan sejarah lokal, namun masih terbatas. Sementara itu, pendalaman aspek validitas lanjutan belum banyak, seperti validitas konstruk dan analisis Rasch. Ini terlihat pada penelitian Widyaningsih (2021).

Berdasarkan analisis tersebut ditemukanlah gap penelitian antara lain validitas instrumen masih bersifat dasar, termasuk analisis DIF untuk melihat apakah butir bias gender, wilayah, atau sosial ekonomi, penggunaan rasch hanya terdapat dibeerapa penelitian seperti Dwi Yuni Karti Kaningtyas *et al*. (2025) dan Syahrul Ramadhan *et al*. (2019), pengukuran HOTS lebih pada objek yang diukur, bukan proses penalaran, argumentasi, dan representasi mathematics, subjek penelitian peserta didik dari Pendidikan Anak Usia Dini (PAUD) masih terbatas, minimnya integrasi konteks lokal, serta belum ada data stabilitas instrumen jika digunakan dalam waktu yang lama dan berulang kali.

Pembahasan

Hasil penelitian *Systematic Literature Review* menunjukkan bahwa model pengembangan Borg & Gall menjadi model paling dominan dalam pengembangan instrumen. Hal ini karena langkah-langkah yang sistematis, rinci, dan sesuai dengan karakteristik penelitian pengembangan**.** Adanya tahap validasi dan revisi berulang memungkinkan produk yang dikembangkan memiliki kualitas dan tingkat kelayakan yang tinggi (Husnayayin *et al.,* 2024). Model tersebut telah teruji secara ilmiah dan diakui secara akademik. Model ini memungkinkan peneliti melakukan proses secara bertahap mulai dari analisis kebutuhan, pengembangan produk awal, desain produk, validasi desain, perbaikan desain, uji coba produk, revisi produk, uji coba pemakaian, revisi produk, dan pembuatan produk massal (Mesra, 2023). Tahap tersebut membantu peneliti untuk menghasilkan instrumen yang valid sehingga dapat mengukur kemampuan peserta didik dengan baik. Sementara itu, model Borg & Gall juga dinilai fleksibel sehingga dapat disesuaikan dengan kebutuhan pengembangan pada mata pelajaran dan jenjang tertentu.

Selain model Borg & Gall, model ADDIE juga digunakan dalam beberapa penelitian pengembangan karena mudah diadaptasi dengan memberikan alur yang sistematis untuk pengembangan instrumen penilaian. Model pengembangan ADDIE efektif dalam merancang dan mengembangkan pengalaman belajar yang berkualitas. Model ini telah menjadi salah satu

https://doi.org/10.51878/educational.v6i3.11239

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

kerangka kerja yang paling banyak digunakan dalam pengembangan instruksional (Zamsiswaya, 2024). Kedua model ini seringkali digunakan untuk mengembangkan media, bahan ajar, serta instrumen penilaian.

Pada penelitian pengembangan, bentuk instrumen yang dikembangkan mempertimbangkan variabel yang akan diteliti, misalnya tes esai digunakan karena dianggap mampu mengorganisasikan gagasan secara mendalam menggunakan narasi sendiri dan mampu mengukur kemampuan berpikir tingkat tinggi. Tes yang disajikan dalam bentuk esai cenderung memiliki nilai fungsi informasi item yang lebih tinggi dibandingkan dengan tes pilihan ganda. Pengembangan butir pilihan ganda mempertimbangkan jumlah responden yang besar dan dapat dianalisis menggunakan teori tes klasik atau teori tes modern (Rasch). Tes pilihan ganda dapat mengukur kemampuan secara objektif, jawaban dapat dikoreksi dengan mudah dan cepat dengan kunci jawaban, memiliki reliabilitas yang tinggi. Tes pilihan ganda dikembangkan karena memungkinkan pengukuran kemampuan peserta didik secara objektif dan efisien serta mendukung analisis kualitas butir soal secara psikometrik.

Hasil kajian juga menunjukkan bahwa sebagian penelitian masih berfokus pada validitas dasar, terutama validitas isi dan validitas konstruk sederhana melalui penilaian ahli. Validitas ini cenderung relatif sederhana dan membutuhkan validitas lanjutan. Sementara itu, pengujian validitas lanjut seperti CFA atau pendekatan teori respons butir masih belum banyak digunakan. Hal ini disebabkan oleh karena CFA memerlukan jumlah subjek penelitian yang lebih besar, penguasaan analisis statistik, serta perangkat lunak yang khusus. Oleh karena itu, kualitas pengembangan instrumen masih perlu ditingkatkan agar tidak hanya valid secara isi, tetapi juga kuat secara empiris dan secara konstruk.

Instrumen yang valid dan reliabel mampu mengukur keterampilan peserta didik secara tepat, salah satunya berpikir tingkat tinggi (HOTS). Sekarang ini, pengembangan instrumen HOTS sangat dibutuhkan untuk menjawab tantangan kurikulum dan tuntutan abad 21. Kurikulum 2013 tidak hanya menekankan kemampuan mengingat dan memahami, tetapi juga menganalisis, mengevaluasi, dan mencipta. Keterampilan berpikir tingkat tinggi merupakan aktivitas berpikir kompleks yang menuntut peserta didik untuk dapat menganalisis informasi secara mendalam, menilai keabsahan argumen berdasarkan kriteria/bukti, menciptakan solusi, gagasan, atau interpretasi baru (Bloom Revisi Anderson & Krathwohl, 2010). Oleh karena itu, perlu dirancang instrumen penilaian untuk mengukur keterampilan berpikir tingkat tinggi peserta didik.

Pengembangan instrumen HOTS menjadi sangat penting dan direkomendasikan untuk mengintegrasikan konteks lokal, misalnya etnomatematika, budaya lokal, sejarah lokal, atau permasalahan kontekstual daerah, sehingga instrumen lebih bermakna. Sementara itu, peneliti selanjutnya direkomendasikan untuk melakukan analisis validitas lanjutan. Teori modern (analisis Rasch) dapat digunakan untuk memperkuat validitas konstruk dan keandalan instrumen secara empirik. Namun demikian, jika akan menggunakan model teori klasik, peneliti disarankan untuk melakukan analisis validitas konstruk melalui *Exploratory Factor Analysis* (EFA) atau *Confirmatory Factor Analysis* (CFA). Peneliti selanjutnya juga diharapkan dapat melakukan analisis perbandingan metode tes konvensional dengan metode berbasis komputer serta kombinasi model pengembangan instrumen dari Borg & Gall, ADDIE dengan 4D atau *Formative Research Tessmer* dan lainnya. Hasil penelitian ini diharapkan dapat menjadi rujukan bagi peneliti, guru, maupun pengembang instrumen dalam menyusun alat evaluasi pembelajaran yang valid dan reliabel sehingga dapat mengukur keterampilan peserta didik sesuai perkembangan zaman.

https://doi.org/10.51878/educational.v6i3.11239

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

KESIMPULAN

Penelitian pengembangan instrumen penilaian merupakan kajian yang terus berkembang dalam dunia pendidikan. Selain mampu meningkatkan kualitas instrumen evaluasi, juga mampu mengukur kompetensi peserta didik. Berdasarkan tinjauan sistematis terhadap 31 artikel jurnal nasional, artikel jurnal internasional terindeks Scopus dan artikel prosiding, dapat disimpulkan bahwa umumnya instrumen dikembangkan untuk mengukur kemampuan kognitif tingkat tinggi, sikap, kompetensi spesifik bidang studi, dan keterampilan abad ke-21. Sebagian besar pengembangan instrumen tes HOTS dilakukan melalui pendekatan penelitian dan pengembangan (*Research and Development*) dengan model-model Borg & Gall, ADDIE, 4D, dan Tessmer**.** Bentuk instrumen yang banyak digunakan berupa pilihan ganda, uraian, dan two-tier dengan menekankan kemampuan menganalisis, mengevaluasi, dan mencipta sesuai taksonomi Bloom revisi. Namun, ditemukan juga instrumen berbasis kinerja dan autentik.

Kualitas psikometrik instrumen masih menjadi fokus utama penelitian berupa tingkat validitas dan reliabilitas. Namun demikian, tidak semua penelitian melaporkan analisis psikometrik secara komprehensif. Sementara itu, sebagian besar penelitian masih berfokus pada konteks umum dan belum banyak mengintegrasikan konteks lokal sehingga masih terdapat peluang pengembangan lebih lanjut, terutama dalam penerapan instrumen yang lebih luas.

Meskipun terdapat keterbatasan penelitian, baik dari aspek ruang lingkup maupun tahapan pengembangan yang dilakukan, hasil penelitian diharapkan menjadi salah satu referensi untuk penelitian lanjutan yang lebih komprehensif. Peneliti selanjutnya disarankan untuk melibatkan lebih banyak basis data nasional dan internasional dan memperpanjang rentang tahun publikasi sehingga tren penelitian lebih komprehensif. Selain memperoleh perbandingan karakteristik penelitian di Indonesia dan luar negeri, juga dapat mengidentifikasi tren penelitian yang lebih jelas. Selanjutnya, agar perbandingan antar studi yang diperoleh dari data data base lebih mendalam, dapat dilakukan pengelompokan studi berdasarkan jenis instrumen dan pendekatan psikometrik yang digunakan baik metode klasik maupun modern.

DAFTAR PUSTAKA

Anderson, L. W. & Krathwohl, D. R. (2010). *Kerangka landasan untuk pembelajaran, pengajaran, dan asesmen: Revisi taksonomi Bloom*. Pustaka Pelajar.

Aryanti, N., Sari, N. K., Siregar, R. C., Putri, R. I. I., & Safitri, S. (2023). Pengembangan soal kognitif berbasis HOTS dalam mata pelajaran sejarah pada pokok bahasan perkembangan kerajaan Islam di Indonesia pada masa Islam untuk siswa kelas X. *DIAJAR: Jurnal Pendidikan dan Pembelajaran*, 2(1), 63–68. https://doi.org/10.54259/diajar.v2i1.1228

Budiastuti, E., Sugiyem, S., & Puad, F. N. A. (2023). Developing self-assessment instruments to measure students' performance characters in making dresses using a high order thinking skills approach. *Jurnal Cakrawala Pendidikan*, 42(1), 27–37. https://doi.org/10.21831/cp.v42i1.48241

Bunari, A. F., Al Fiqri, Y., Zakaria, N. B., & Jali, J. M. (2022, Januari). Development of HOTS (Higher Order Thinking Skills) Test Instrumen in the Maritime History Course based on Riau Local History. In *ICOME 2021: Proceedings of the 1st International Conference on Maritime Education, ICOME 2021, 3-5 November 2021, Tanjungpinang, Riau Islands, Indonesia* (p. 121). European Alliance for Innovation.

Bunari, Yuliantoro, & Fiqri, Y. A. (2020, Agustus). Development of Historical Learning Evaluation Instruments Based on High Order Thinking Skills for Riau History course at Departement of History Education in Riau University. In *International Conference On*

https://doi.org/10.51878/educational.v6i3.11239

Social Studies, Globalisation And Technology (ICSSGT 2019) (pp. 211-219). Atlantis Press. https://doi.org/10.2991/assehr.k.200803.027

Darmana, A., Sutiani, A., & Jasmidi. (2020, February). Development of The Thermochemistry-Hots-Tawheed Multiple Choice Instrument. In *Journal of Physics: Conference Series* (Vol. 1462, No. 1, p. 012057). IOP Publishing. https://doi.org/10.1088/1742-6596/1462/1/012020

Dewi, R. M., Sholikhah, N. M., & Fitrayati, D. (2020). High Order Thinking Skills Instrument on Microeconomics Course: A Development Research. *International Journal of Instruction*, *13*(4), 283-294. https://doi.org/10.29333/iji.2020.13418a

Fadlila, N., & Sagala, P. N. (2021, March). Development of Test Instrument Based Realistic Mathematics Education to Improve High Order Thinking Skills. In *Journal of Physics: Conference Series* (Vol. 1819, No. 1, p. 012068). IOP Publishing. https://doi.org/10.1088/1742-6596/1819/1/012061

Handayani, F., Hartono, H., & Lestari, W. (2019). Need analysis in the development of HOTS-oriented study project assessment instrument in android-based science learning. *Journal of Educational Research and Evaluation*, 8(1), 57–64. https://doi.org/10.15294/jere.v8i1.25209

Husnayayin, A., Gustina, Z., & Dewi, D. E. C. (2024). Karakteristik dan langkah-langkah metode penelitian Research and Development (Borg & Gall) dalam pendidikan. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, *9*(04), 490-501. https://doi.org/10.23969/jp.v9i04.19906

Kaningtyas, D. Y. K., Rokhman, F., & Suminar, T. (2025). Pengembangan instrumen tes diagnostik bentuk two-tier multiple choice (TTMC) pada mata pelajaran sejarah di SMA. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 10(3), 221–234. https://journal.unpas.ac.id/index.php/pendas/article/view/21045

Maxnun, L., Kristiani, K., & Sulistyaningrum, C. D. (2024). Development of HOTS-based cognitive assessment instruments: ADDIE model. *Journal of Education and Learning (EduLearn)*, 18(2), 489–498. https://doi.org/10.11591/edulearn.v18i2.21139

Melati, M. N. M., Subakti, Y. R., & Kurniawan, H. (2022). Pengembangan soal HOTS sejarah materi kerajaan-kerajaan maritim Indonesia masa Islam untuk siswa kelas XI IPS. *Historia Vitae*, 2(1), 53–66. https://doi.org/10.24071/hv.v2i1.4127

Mesra, R. (2023). *Research & development dalam pendidikan*. Mifandi Mandiri Digital.

Mildasari, I. G., & Aisiah. (2022). Pengembangan soal higher order thinking skill (HOTS) pada mata pelajaran sejarah di SMA. *Kronologi*, 1(1), 245–254. https://doi.org/10.24036/jk.v1i1.16

Nasution, A. S., Hadi, W., & Eviyanti, E. (2021, November). Development of Assessment Instruments Based on High-Level Thinking Skills in the Receptive Language Skills Course for Unimed Indonesian Language and Literature Education Students. In *6th Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2021)* (pp. 668-670). Atlantis Press. https://doi.org/10.2991/assehr.k.211110.160

Nurhaibi, N. (2025). Analisis Konsep Holistik Assesmen dalam Pembelajaran PAI: Tinjauan Literatur dari Perspektif Pendidikan Islam dan Umum: Penelitian. *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, *4*(2), 10240-10244. https://doi.org/10.31004/jerkin.v4i2.3477

Ndoen, F. A., Madu, A., Taneo, M., Ande, A., Rato, F. S., & Mali, S. J. (2025). Development of higher order thinking skills (HOTS) assessment instruments based on local history in

history learning. *Veredas Do Direito*, 22(2). https://revista.domhelder.edu.br/index.php/veredas/article/view/2854

Nursalam, Angriani, A. D., Darmawati, Baharuddin, & Aminuddin. (2018). Developing test instruments for measurement of students' high-order thinking skill on mathematics in junior high school in Makassar. *Journal of Physics: Conference Series*, 1028. https://doi.org/10.1088/1742-6596/1028/1/012134

Paksi, J. J. (2024). Pengembangan asesmen pembelajaran sejarah berbasis HOTS berbantu Wizer.me pada materi proklamasi kemerdekaan Indonesia untuk siswa SMA kelas XI. *Historia Vitae*, 4(2), 11–22. https://doi.org/10.24071/hv.v4i2.7212

Palgunadi, I. M. P. K., Pradnyanita, A. D. C., Payadnya, I. P. A. A., Wena, I. M., & Novianti, P. S. (2023). Penerapan media pembelajaran virtual exhibition berbasis RME untuk meningkatkan kemampuan high order thinking skills (HOTS) siswa. *Jurnal Pembelajaran dan Pengembangan Matematika (PEMANTIK)*, 3(1), 39–52. https://doi.org/10.61179/pemantik.v3i1.92

Partnership for 21st Century Learning. (2009). *Framework for 21st Century Learning*.

Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The development of an instrument to measure the higher order thinking skill in physics. *European Journal of Educational Research*, 8(3), 743–751. https://doi.org/10.12973/eu-jer.8.3.743

Richeal P. T. Kim How, Zulnaidi, H., & Abdul Rahim, S. S. B. (2023). Development of higher-order thinking skills test instrument on quadratic equation (HOTS-QE) for secondary school students. *Pegem Journal of Education and Instruction*, 13(1). https://doi.org/10.47750/pegegog.13.01.18

Rintayati, P., Lukitasari, H., & Syawaludin, A. (2021). Development of two-tier multiple choice test to assess Indonesian elementary students' higher-order thinking skills. *International Journal of Instruction*, 14(1), 555–566. https://doi.org/10.29333/iji.2021.14133a

Rosidin, U., Suyanta, A., & Abdurrahman, A. (2019). A combined HOTS-based assessment/STEM learning model to improve secondary students' thinking skills: A development and evaluation study. *Journal for the Education of Gifted Young Scientists*, 7(3), 435–448. https://doi.org/10.17478/jegys.610377

Sagala, P. N., & Andriani, A. (2019). Development of higher-order thinking skills (HOTS) questions of probability theory subject based on Bloom's taxonomy. *Journal of Physics: Conference Series*, 1188. https://doi.org/10.1088/1742-6596/1188/1/012025

Shalikhah, N. D., & Nugroho, I. (2023). Implementation of Higher-Order Thinking Skills in Elementary School Using Learning Model, Media, and Assessment. *Al-Ishlah: Jurnal Pendidikan, 15*(3), 3978–3990. https://doi.org/10.35445/alishlah.v15i3.3091

Sari, R. W. A., & Purwaningsih, S. M. (2024). Pengembangan soal higher order thinking skill (HOTS) berbasis classpoint pada mata pelajaran sejarah kelas X di SMAN 1 Sooko Mojokerto. *AVATARA: e-Journal Pendidikan Sejarah*, 15(1), 1–6. https://ejournal.unesa.ac.id/index.php/avatara/article/view/60432

Serevina, V., Sari, Y. P., & Maynastiti, D. (2019). Developing high order thinking skills (HOTS) assessment instrument for fluid static at senior high school. *Journal of Physics: Conference Series*, 1185. https://doi.org/10.1088/1742-6596/1185/1/012052

Setiawan, J., Sudrajat, A., Aman, A., & Kumalasari, D. (2021). Development of higher order thinking skill assessment instruments in learning Indonesian history. *International Journal of Evaluation and Research in Education (IJERE)*, 10(2), 545. https://doi.org/10.11591/ijere.v10i2.20796

https://doi.org/10.51878/educational.v6i3.11239

Widyaningsih, S. W., Yusuf, I., Prasetyo, Z. K., & Istiyono, E. (2021). The development of the HOTS test of physics based on modern test theory: Question modeling through e-learning of Moodle LMS. *International Journal of Instruction*, 14(4), 51–68. https://doi.org/10.29333/iji.2021.1444a

Yunita, Y., Wahidin, W., & Tsurayya, A. (2018, September). The development of mathematics higher order thinking skills instrument for grade VIII junior high school. In *Journal of Physics: Conference Series* (Vol. 1088, No. 1, p. 012093). IOP Publishing. https://doi.org/10.1088/1742-6596/1088/1/012094

Zaki, M., Amalia, R., & Sofyan, S. (2020, April). Development of high order thinking skills (HOTS) test instrument on exponent for junior high school students. In *Journal of Physics: Conference Series* (Vol. 1521, No. 3, p. 032096). IOP Publishing. https://doi.org/10.1088/1742-6596/1521/3/032104

Zamsiswaya, Z., Syawaluddin, S., & Syahrizul, S. (2024). Pengembangan Model ADDIE (Analisys, Design, Development, Implemetation, Evaluation). *Jurnal Pendidikan Tambusai*, *8*(3), 46363–46369. https://jptam.org/index.php/jptam/article/view/22709

Zhou, Y., Gan, L., Chen, J., Wijaya, T. T., & Li, Y. (2023). Development and validation of a higher-order thinking skills assessment scale for pre-service teachers. *Thinking Skills and Creativity*, *48*, 101272. https://doi.org/10.1016/j.tsc.2023.101278

https://doi.org/10.51878/educational.v6i3.11239