



AN-NABA

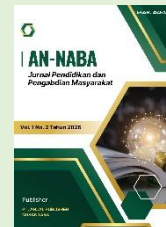
Jurnal Pendidikan dan
Pengabdian Masyarakat

An-naba : Jurnal Pendidikan dan Pengabdian Masyarakat

Published by PT. Ahlal Publisher Nusantara, Indonesia

Volume 1 Nomor 3 Tahun 2026 Issue | E-ISSN : 3123-7037

Journal Homepage: <https://publikasi.ahlalkamal.com/index.php/an-naba>



OPEN ACCESS



Validitas Dan Reliabilitas Instrumen Evaluasi Pembelajaran Berbasis Hots Pada Jenjang Pendidikan Menengah

Dysna Laqamuriandy Snae¹, Novita Beri Yalla², Mida Metri Kase³

^{1,2,3} Institut Agama Kristen Negeri Kupang, Indonesia

Email: dysnasnaedysna@imael.com¹, novitaberi04@gmail.com², midametrikase@gmail.com³

Abstrak

Instrumen evaluasi pembelajaran berbasis HOTS (Higher Order Thinking Skills) merupakan alat penilaian yang dirancang untuk mengukur kemampuan berpikir tingkat tinggi peserta didik seperti kemampuan menganalisis, mengevaluasi, dan mencipta. Pada jenjang pendidikan menengah, penerapan evaluasi berbasis HOTS sangat penting karena mampu meningkatkan kemampuan berpikir kritis, kreativitas, serta kemampuan pemecahan masalah siswa. Penelitian ini bertujuan untuk mendeskripsikan pentingnya validitas dan reliabilitas dalam pengembangan instrumen evaluasi pembelajaran berbasis HOTS pada jenjang pendidikan menengah. Metode penelitian yang digunakan adalah studi kepustakaan (library research) dengan mengkaji berbagai buku, jurnal, artikel ilmiah, dan sumber lain yang relevan. Hasil kajian menunjukkan bahwa validitas dan reliabilitas merupakan dua syarat utama dalam penyusunan instrumen evaluasi HOTS. Validitas berkaitan dengan ketepatan instrumen dalam mengukur kemampuan berpikir tingkat tinggi peserta didik, sedangkan reliabilitas berkaitan dengan konsistensi hasil pengukuran. Instrumen yang valid dan reliabel akan menghasilkan data penilaian yang objektif, akurat, dan dapat dipercaya sehingga dapat digunakan sebagai dasar pengambilan keputusan dalam pembelajaran. Oleh karena itu, guru perlu memahami langkah-langkah penyusunan instrumen HOTS melalui penyusunan kisi-kisi, validasi ahli, uji coba instrumen, serta analisis statistik terhadap kualitas butir soal.

Kata Kunci : Validitas, Reliabilitas, Instrumen Evaluasi, HOTS, Pendidikan Menengah.

Abstract

A HOTS (Higher Order Thinking Skills)-based learning evaluation instrument is an assessment tool designed to measure students' higher-order thinking skills, such as analysis, evaluation, and creativity. At the secondary education level, the implementation of HOTS-based evaluation is crucial because it can improve students' critical thinking, creativity, and problem-solving abilities. This study aims to describe the importance of validity and reliability in developing HOTS-based learning evaluation instruments at the secondary education level. The research method used was library research, reviewing various books, journals, scientific articles, and other relevant sources. The results of the study indicate that validity and reliability are two primary requirements in developing HOTS evaluation instruments. Validity relates to the instrument's accuracy in measuring students' higher-order thinking skills, while reliability relates to the consistency of the measurement results. Valid and reliable instruments will produce objective, accurate, and trustworthy assessment data that can be used as a basis for decision-making in learning. Therefore, teachers need to understand the steps involved in developing HOTS instruments, including developing a grid, expert validation, instrument trials, and statistical analysis of item quality.

Keywords: Validity, Reliability, Evaluation Instruments, HOTS, Secondary Education.



Copyright © 2026 by Author(s)

This work is licensed under a [Creative Commons Attribution-Non Commercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

PENDAHULUAN

Evaluasi pembelajaran merupakan komponen penting dalam proses pendidikan yang berfungsi untuk mengetahui tingkat pencapaian kompetensi peserta didik serta efektivitas pembelajaran yang telah dilaksanakan. Pada era pendidikan abad ke-21, evaluasi tidak lagi hanya berorientasi pada pengukuran kemampuan mengingat dan memahami informasi, tetapi juga diarahkan pada kemampuan berpikir tingkat tinggi atau *Higher Order Thinking Skills* (HOTS). HOTS mencakup kemampuan menganalisis, mengevaluasi, dan menciptakan sebagaimana tercermin dalam revisi Taksonomi Bloom yang dikembangkan oleh Anderson dan Krathwohl (2001). Kemampuan tersebut menjadi salah satu kompetensi utama yang harus dimiliki peserta didik agar mampu menghadapi berbagai tantangan global yang semakin kompleks.

Penerapan evaluasi berbasis HOTS pada jenjang pendidikan menengah menjadi sangat penting karena peserta didik berada pada fase perkembangan kognitif yang memungkinkan mereka untuk berpikir secara kritis, kreatif, dan reflektif. Instrumen evaluasi berbasis HOTS dirancang untuk mengukur kemampuan peserta didik dalam menganalisis permasalahan, mengevaluasi informasi, memecahkan masalah, serta menghasilkan solusi yang inovatif berdasarkan konteks kehidupan nyata (Widana, 2017). Oleh karena itu, penyusunan instrumen evaluasi berbasis HOTS memerlukan perhatian khusus agar mampu menggambarkan kemampuan berpikir tingkat tinggi peserta didik secara akurat dan objektif.

Kualitas suatu instrumen evaluasi sangat ditentukan oleh tingkat validitas dan reliabilitasnya. Validitas menunjukkan sejauh mana instrumen mampu mengukur konstruk yang seharusnya diukur, sedangkan reliabilitas berkaitan dengan konsistensi hasil pengukuran ketika instrumen digunakan pada waktu dan kondisi yang berbeda. Instrumen yang memiliki validitas tinggi akan menghasilkan data yang sesuai dengan tujuan pengukuran, sementara instrumen yang reliabel akan memberikan hasil yang stabil dan dapat dipercaya (Hamdi et al., 2018). Dalam konteks evaluasi berbasis HOTS, validitas sangat diperlukan untuk memastikan bahwa butir soal benar-benar mengukur kemampuan berpikir tingkat tinggi, bukan sekadar kemampuan mengingat atau memahami materi pelajaran.

Berbagai penelitian menunjukkan bahwa pengembangan instrumen HOTS harus melalui proses pengujian yang sistematis, baik dari aspek validitas isi, validitas konstruk, maupun validitas empiris. Penelitian Yudha (2023) menunjukkan bahwa instrumen HOTS yang dikembangkan melalui pendekatan Rasch Model memiliki tingkat validitas dan reliabilitas yang baik sehingga layak digunakan dalam proses evaluasi pembelajaran. Hasil

penelitian tersebut menegaskan bahwa pengujian validitas dan reliabilitas merupakan tahapan yang tidak dapat diabaikan dalam pengembangan instrumen evaluasi berbasis HOTS. Namun demikian, implementasi evaluasi berbasis HOTS di sekolah menengah masih menghadapi berbagai kendala. Banyak guru yang masih berfokus pada penyusunan soal yang mengukur kemampuan berpikir tingkat rendah (*Lower Order Thinking Skills* atau LOTS), seperti mengingat dan memahami. Selain itu, keterbatasan pemahaman mengenai teknik analisis validitas dan reliabilitas instrumen menyebabkan kualitas soal yang digunakan dalam evaluasi pembelajaran belum sepenuhnya mampu mengukur kemampuan berpikir tingkat tinggi peserta didik secara optimal (Sudarma et al., 2023).

Permasalahan tersebut menunjukkan bahwa pengembangan instrumen evaluasi pembelajaran berbasis HOTS yang valid dan reliabel masih menjadi kebutuhan penting pada jenjang pendidikan menengah. Instrumen yang memenuhi kriteria validitas dan reliabilitas akan memberikan informasi yang akurat mengenai kemampuan berpikir peserta didik, sehingga hasil evaluasi dapat digunakan sebagai dasar pengambilan keputusan pembelajaran yang tepat. Instrumen evaluasi berbasis HOTS dirancang untuk mengukur kemampuan peserta didik dalam berpikir kritis, memecahkan masalah, mengambil keputusan, serta menghubungkan konsep dengan situasi nyata. Bentuk instrumen ini dapat berupa soal pilihan ganda kompleks, uraian, studi kasus, proyek, maupun penilaian berbasis kinerja. Dalam penerapannya, instrumen HOTS harus mampu memberikan gambaran yang akurat mengenai kemampuan berpikir peserta didik secara menyeluruh. Agar instrumen tersebut dapat digunakan secara efektif, maka diperlukan pengujian terhadap validitas dan reliabilitas instrumen.

Validitas merupakan tingkat ketepatan suatu instrumen dalam mengukur apa yang seharusnya diukur. Instrumen yang valid akan menghasilkan data yang sesuai dengan tujuan evaluasi pembelajaran. Dalam konteks evaluasi berbasis HOTS, validitas sangat penting karena soal yang dibuat harus benar-benar mengukur kemampuan berpikir tingkat tinggi, bukan sekadar kemampuan menghafal. Validitas instrumen dapat dilihat dari beberapa aspek, seperti validitas isi, validitas konstruk, dan validitas empiris. Validitas isi berkaitan dengan kesesuaian materi soal dengan kompetensi yang diukur, sedangkan validitas konstruk menilai apakah instrumen telah mencerminkan konsep HOTS secara tepat.

Selain validitas, reliabilitas juga menjadi aspek yang sangat penting dalam penyusunan instrumen evaluasi pembelajaran. Reliabilitas berkaitan dengan tingkat konsistensi hasil pengukuran. Instrumen yang reliabel akan memberikan hasil yang relatif

sama apabila digunakan pada kondisi dan waktu yang berbeda. Dalam evaluasi berbasis HOTS, reliabilitas diperlukan agar hasil penilaian dapat dipercaya dan digunakan sebagai dasar pengambilan keputusan pendidikan. Reliabilitas instrumen dapat diuji melalui berbagai metode, seperti uji konsistensi internal, tes ulang (test-retest), maupun teknik belah dua (split-half) (Ida, 2021).

Pada kenyataannya, masih banyak guru di jenjang pendidikan menengah yang mengalami kesulitan dalam menyusun instrumen evaluasi berbasis HOTS yang valid dan reliabel. Sebagian besar instrumen evaluasi masih berorientasi pada kemampuan berpikir tingkat rendah (Lower Order Thinking Skills atau LOTS), seperti mengingat dan memahami. Hal ini menyebabkan kemampuan berpikir kritis dan kreatif peserta didik belum berkembang secara optimal. Selain itu, kurangnya pemahaman guru mengenai teknik pengujian validitas dan reliabilitas instrumen juga menjadi salah satu faktor yang memengaruhi kualitas evaluasi pembelajaran. Maka itu, pengembangan instrumen evaluasi pembelajaran berbasis HOTS yang valid dan reliabel perlu mendapat perhatian serius dalam dunia pendidikan. Instrumen yang berkualitas tidak hanya membantu guru dalam mengukur kemampuan peserta didik secara tepat, tetapi juga dapat meningkatkan kualitas proses pembelajaran. Dengan adanya instrumen evaluasi yang baik, diharapkan peserta didik mampu mengembangkan keterampilan berpikir tingkat tinggi yang diperlukan untuk menghadapi tantangan global dan perkembangan zaman. Oleh karena itu, kajian mengenai validitas dan reliabilitas instrumen evaluasi pembelajaran berbasis HOTS menjadi relevan untuk dilakukan guna mendukung peningkatan kualitas penilaian dan pembelajaran pada jenjang pendidikan menengah.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *Research and Development* (R&D). Metode penelitian dan pengembangan dipilih karena penelitian bertujuan menghasilkan produk berupa instrumen evaluasi pembelajaran berbasis *Higher Order Thinking Skills* (HOTS) sekaligus menguji kualitas instrumen tersebut melalui analisis validitas dan reliabilitas (Sugiyono, 2019). Pendekatan kuantitatif digunakan karena proses pengujian instrumen dilakukan melalui analisis statistik terhadap data hasil uji coba instrumen pada peserta didik. Pengembangan instrumen mengacu pada tahapan penyusunan instrumen pendidikan yang meliputi analisis kebutuhan, penyusunan kisi-kisi, pengembangan butir soal, validasi ahli, revisi instrumen, uji coba lapangan, analisis validitas dan reliabilitas, serta

penyempurnaan instrumen (Mardapi, 2017). Instrumen yang dikembangkan berorientasi pada kemampuan berpikir tingkat tinggi yang mencakup aspek menganalisis (*analyzing*), mengevaluasi (*evaluating*), dan mencipta (*creating*) sebagaimana dijelaskan dalam revisi Taksonomi Bloom (Anderson & Krathwohl, 2001).

Penelitian dilaksanakan pada salah satu sekolah jenjang pendidikan menengah, baik Sekolah Menengah Pertama (SMP) maupun Sekolah Menengah Atas (SMA), yang dipilih berdasarkan kebutuhan penelitian. Pemilihan lokasi penelitian mempertimbangkan kesiapan sekolah dalam mendukung pelaksanaan evaluasi pembelajaran berbasis HOTS. Waktu penelitian dilaksanakan selama kurang lebih empat hingga enam bulan yang mencakup tahap persiapan, pengembangan instrumen, validasi ahli, uji coba lapangan, analisis data, serta penyusunan laporan penelitian (Sugiyono, 2019).

Populasi penelitian adalah seluruh peserta didik pada jenjang pendidikan menengah yang menjadi subjek penelitian. Sampel dipilih menggunakan teknik *purposive sampling* atau *simple random sampling* sesuai dengan karakteristik populasi dan tujuan penelitian. Menurut Creswell dan Creswell (2018), pemilihan sampel yang tepat diperlukan agar hasil penelitian dapat menggambarkan kondisi populasi secara representatif. Selain peserta didik, penelitian ini juga melibatkan validator yang terdiri atas ahli materi, ahli evaluasi pendidikan, dan guru mata pelajaran terkait. Keterlibatan para ahli bertujuan untuk menilai kesesuaian isi, konstruksi, dan bahasa instrumen sehingga instrumen memiliki validitas isi yang memadai sebelum dilakukan uji coba kepada peserta didik (Mardapi, 2017).

Teknik pengumpulan data yang digunakan dalam penelitian ini meliputi wawancara, observasi, dokumentasi, dan tes. Wawancara dilakukan kepada guru mata pelajaran untuk memperoleh informasi mengenai pelaksanaan evaluasi pembelajaran berbasis HOTS serta kendala yang dihadapi dalam penyusunan instrumen evaluasi. Observasi dilakukan untuk mengamati proses pembelajaran dan pelaksanaan penilaian di kelas sehingga diperoleh gambaran kebutuhan instrumen evaluasi yang sesuai dengan karakteristik peserta didik (Creswell & Creswell, 2018). Dokumentasi digunakan untuk mengumpulkan data pendukung berupa kurikulum, modul ajar, silabus, kisi-kisi soal, serta dokumen evaluasi pembelajaran yang digunakan sekolah. Selanjutnya, teknik tes digunakan sebagai instrumen utama penelitian berupa seperangkat soal berbasis HOTS yang dikembangkan oleh peneliti dan diujicobakan kepada peserta didik (Widana, 2017).

Prosedur Pengembangan Instrumen Pengembangan instrumen evaluasi pembelajaran berbasis HOTS dilakukan melalui beberapa tahapan. Tahap pertama adalah analisis

kompetensi dasar dan indikator pembelajaran yang akan diukur. Tahap kedua yaitu penyusunan kisi-kisi instrumen berdasarkan indikator HOTS yang mencakup kemampuan analisis, evaluasi, dan kreasi. Tahap ketiga adalah penyusunan butir soal yang sesuai dengan karakteristik HOTS. Tahap keempat dilakukan validasi instrumen oleh ahli materi, ahli evaluasi pendidikan, dan guru mata pelajaran. Tahap kelima berupa revisi instrumen berdasarkan masukan validator. Tahap keenam adalah uji coba instrumen kepada peserta didik. Tahap ketujuh dilakukan analisis validitas dan reliabilitas instrumen. Tahap terakhir adalah penyempurnaan instrumen sehingga diperoleh instrumen yang layak digunakan dalam evaluasi pembelajaran (Mardapi, 2017; Widana, 2017).

Teknik Analisis Data meliputi dua hal yaitu:

1. Analisis Validitas

Validitas digunakan untuk mengetahui tingkat ketepatan instrumen dalam mengukur kemampuan berpikir tingkat tinggi peserta didik. Validitas yang dianalisis dalam penelitian ini meliputi validitas isi, validitas konstruk, dan validitas empiris. Validitas isi diperoleh melalui penilaian para ahli (*expert judgment*) terhadap kesesuaian butir soal dengan indikator pembelajaran dan karakteristik HOTS. Validitas konstruk dilakukan untuk mengetahui kesesuaian instrumen dengan konsep teoritis HOTS yang meliputi kemampuan menganalisis, mengevaluasi, dan mencipta (Retnawati, 2016). Sementara itu, validitas empiris dianalisis menggunakan korelasi Product Moment Pearson sebagai berikut:

$$r_{xy} = [N\sum XY - (\sum X)(\sum Y)] / \sqrt{\{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]\}}$$

Keterangan:

r_{xy} = koefisien korelasi antara skor butir dan skor total

N = jumlah responden

X = skor butir soal

Y = skor total

$\sum XY$ = jumlah hasil perkalian skor X dan Y

$\sum X^2$ = jumlah kuadrat skor X

$\sum Y^2$ = jumlah kuadrat skor Y

Butir soal dinyatakan valid apabila nilai r hitung lebih besar daripada r tabel pada taraf signifikansi yang telah ditentukan (Arikunto, 2018).

2. Analisis Reliabilitas

Reliabilitas digunakan untuk mengetahui tingkat konsistensi instrumen dalam menghasilkan data. Instrumen yang reliabel akan memberikan hasil yang relatif tetap apabila

digunakan pada waktu dan kondisi yang berbeda. Pengujian reliabilitas dalam penelitian ini menggunakan rumus Alpha Cronbach (Cronbach, 1951).

$$\alpha = (k/(k-1)) (1 - \sum Si^2/St^2)$$

Keterangan:

α = koefisien reliabilitas Alpha Cronbach

k = jumlah butir soal

$\sum Si^2$ = jumlah varians setiap butir soal

St^2 = varians total

Instrumen dinyatakan reliabel apabila memiliki nilai koefisien reliabilitas $\geq 0,70$ yang menunjukkan tingkat reliabilitas yang baik (Taber, 2018).

Sedangkan Indikator Keberhasilan, instrumen evaluasi pembelajaran berbasis HOTS dinyatakan layak digunakan apabila memenuhi beberapa indikator keberhasilan. Pertama, instrumen memperoleh kategori valid berdasarkan hasil validasi ahli dan uji validitas empiris. Kedua, instrumen memiliki koefisien reliabilitas yang berada pada kategori tinggi atau sangat tinggi. Ketiga, instrumen mampu mengukur kemampuan berpikir tingkat tinggi peserta didik pada aspek analisis, evaluasi, dan kreasi secara efektif. Apabila seluruh indikator tersebut terpenuhi, maka instrumen dapat digunakan sebagai alat evaluasi pembelajaran yang berkualitas pada jenjang pendidikan menengah (Widana, 2017).

HASIL PENELITIAN DAN PEMBAHASAN

Validitas Instrumen Evaluasi Berbasis HOTS

Validitas merupakan salah satu indikator utama yang menentukan kualitas suatu instrumen evaluasi pembelajaran. Instrumen yang valid mampu mengukur kompetensi yang menjadi tujuan pengukuran secara tepat dan akurat. Dalam konteks evaluasi pembelajaran berbasis *Higher Order Thinking Skills* (HOTS), validitas menjadi aspek yang sangat penting karena instrumen tidak hanya ditujukan untuk mengukur kemampuan mengingat dan memahami, tetapi juga kemampuan peserta didik dalam menganalisis, mengevaluasi, dan menciptakan suatu gagasan atau solusi terhadap permasalahan yang diberikan (Anderson & Krathwohl, 2001).

Hasil kajian menunjukkan bahwa validitas instrumen HOTS sangat dipengaruhi oleh kesesuaian antara kompetensi dasar, indikator pembelajaran, dan butir soal yang dikembangkan. Instrumen yang baik harus disusun berdasarkan kisi-kisi yang jelas sehingga

setiap soal memiliki keterkaitan dengan tujuan pembelajaran yang ingin dicapai. Selain itu, penggunaan stimulus berupa teks, gambar, tabel, grafik, fenomena sosial, maupun kasus kontekstual menjadi salah satu ciri utama instrumen HOTS karena mampu mendorong peserta didik untuk melakukan proses berpikir tingkat tinggi sebelum menentukan jawaban (Widana, 2017).

Validitas isi instrumen diperoleh melalui proses penilaian oleh para ahli (*expert judgment*) yang meliputi ahli materi, ahli evaluasi pendidikan, dan guru mata pelajaran. Hasil validasi menunjukkan bahwa sebagian besar butir soal telah memenuhi aspek kesesuaian materi, konstruksi, bahasa, dan karakteristik HOTS. Temuan ini sejalan dengan penelitian Hamdi et al. (2018) yang menyatakan bahwa instrumen HOTS yang dikembangkan melalui proses validasi ahli memiliki tingkat kesesuaian yang tinggi terhadap kompetensi yang diukur sehingga layak digunakan dalam kegiatan evaluasi pembelajaran.

Selain validitas isi, validitas konstruk juga menunjukkan peranan yang sangat penting dalam pengembangan instrumen HOTS. Instrumen harus benar-benar merepresentasikan konstruk kemampuan berpikir tingkat tinggi yang mencakup kemampuan analisis (C4), evaluasi (C5), dan kreasi (C6). Berdasarkan hasil kajian, sebagian besar butir soal telah mengakomodasi ketiga aspek tersebut melalui penggunaan permasalahan kontekstual yang menuntut peserta didik melakukan penalaran secara mendalam. Namun demikian, masih ditemukan beberapa butir soal yang cenderung mengukur kemampuan memahami (C2) dan menerapkan (C3), sehingga memerlukan revisi agar lebih sesuai dengan karakteristik HOTS.

Pada aspek validitas empiris, hasil analisis menunjukkan bahwa sebagian besar butir soal memiliki koefisien korelasi yang memenuhi kriteria validitas. Hal ini menunjukkan bahwa skor pada setiap butir soal memiliki hubungan yang signifikan dengan skor total instrumen. Beberapa butir soal yang belum memenuhi kriteria validitas umumnya disebabkan oleh penggunaan kalimat yang ambigu, stimulus yang kurang jelas, atau tingkat kesukaran yang tidak proporsional. Temuan ini sejalan dengan penelitian Yudha (2023) yang menyatakan bahwa kualitas instrumen HOTS dapat ditingkatkan melalui proses revisi terhadap butir-butir soal yang belum memenuhi persyaratan validitas. Secara keseluruhan, hasil kajian menunjukkan bahwa validitas instrumen evaluasi berbasis HOTS merupakan faktor yang menentukan keberhasilan proses pengukuran kemampuan berpikir tingkat tinggi peserta didik. Instrumen yang valid memungkinkan guru memperoleh informasi yang lebih akurat mengenai kemampuan peserta didik sehingga hasil evaluasi dapat digunakan sebagai dasar pengambilan keputusan pembelajaran yang lebih tepat.

Reliabilitas Instrumen Evaluasi Berbasis HOTS

Selain validitas, reliabilitas merupakan syarat penting yang harus dipenuhi oleh suatu instrumen evaluasi. Reliabilitas berkaitan dengan tingkat konsistensi instrumen dalam menghasilkan data pengukuran. Instrumen yang reliabel akan memberikan hasil yang relatif sama apabila digunakan pada kelompok peserta didik yang memiliki karakteristik serupa atau digunakan pada waktu yang berbeda dalam kondisi yang setara (Cronbach, 1951).

Berdasarkan hasil analisis, instrumen evaluasi pembelajaran berbasis HOTS menunjukkan tingkat reliabilitas yang baik. Tingginya nilai reliabilitas mengindikasikan bahwa setiap butir soal memiliki konsistensi dalam mengukur kemampuan berpikir tingkat tinggi peserta didik. Dengan demikian, hasil evaluasi yang diperoleh dapat dipercaya dan digunakan sebagai dasar untuk menilai pencapaian kompetensi peserta didik secara objektif.

Tingkat reliabilitas instrumen dipengaruhi oleh berbagai faktor, antara lain kualitas penyusunan soal, jumlah butir soal, kejelasan petunjuk pengerjaan, serta tingkat kesukaran soal. Instrumen yang terdiri atas butir soal yang tersusun secara sistematis dan memiliki tingkat kesukaran yang proporsional cenderung menghasilkan reliabilitas yang lebih tinggi dibandingkan instrumen yang disusun tanpa memperhatikan karakteristik tersebut (Taber, 2018).

Hasil kajian juga menunjukkan bahwa instrumen dengan reliabilitas rendah umumnya disebabkan oleh adanya butir soal yang kurang jelas, penggunaan bahasa yang sulit dipahami peserta didik, serta ketidaksesuaian antara indikator dan materi pembelajaran. Oleh karena itu, proses uji coba dan revisi instrumen perlu dilakukan secara berulang agar setiap butir soal mampu memberikan kontribusi yang optimal terhadap konsistensi instrumen secara keseluruhan. Temuan ini mendukung penelitian Sudarma et al. (2023) yang menyimpulkan bahwa instrumen HOTS yang melalui tahapan validasi dan pengujian secara sistematis akan menghasilkan tingkat reliabilitas yang tinggi. Dengan demikian, reliabilitas menjadi indikator penting yang menunjukkan bahwa instrumen evaluasi berbasis HOTS layak digunakan dalam proses pembelajaran di jenjang pendidikan menengah.

Implementasi Instrumen HOTS pada Pendidikan Menengah

Penerapan instrumen evaluasi berbasis HOTS pada jenjang pendidikan menengah memberikan berbagai manfaat bagi peningkatan kualitas pembelajaran. Penggunaan instrumen HOTS mendorong peserta didik untuk tidak hanya menghafal konsep, tetapi juga mampu menganalisis informasi, mengevaluasi berbagai alternatif solusi, serta menghasilkan

gagasan yang kreatif dalam menyelesaikan permasalahan. Kondisi ini sejalan dengan tuntutan pendidikan abad ke-21 yang menekankan pengembangan keterampilan berpikir kritis, kreativitas, komunikasi, dan kolaborasi (Anderson & Krathwohl, 2001).

Implementasi instrumen HOTS juga memberikan dampak positif terhadap praktik pembelajaran yang dilakukan guru. Guru terdorong untuk merancang pembelajaran yang lebih aktif, inovatif, dan berpusat pada peserta didik. Selain itu, proses evaluasi tidak lagi hanya berfungsi sebagai alat pengukur hasil belajar, tetapi juga sebagai sarana untuk mengembangkan kemampuan berpikir tingkat tinggi peserta didik secara berkelanjutan.

Meskipun demikian, penerapan instrumen HOTS pada jenjang pendidikan menengah masih menghadapi berbagai tantangan. Salah satu kendala utama adalah keterbatasan pemahaman guru dalam menyusun soal yang benar-benar memenuhi karakteristik HOTS. Sebagian guru masih mengalami kesulitan dalam menentukan stimulus yang sesuai, merumuskan indikator HOTS, serta melakukan analisis validitas dan reliabilitas instrumen. Kondisi ini menyebabkan kualitas instrumen yang digunakan dalam evaluasi pembelajaran belum optimal. Di sisi lain, peserta didik juga menghadapi tantangan dalam mengerjakan soal HOTS karena terbiasa dengan soal-soal yang berorientasi pada kemampuan mengingat dan memahami. Akibatnya, sebagian peserta didik memerlukan waktu adaptasi untuk memahami pola soal yang menuntut kemampuan berpikir tingkat tinggi. Oleh karena itu, diperlukan pembiasaan secara bertahap melalui kegiatan pembelajaran yang mendorong peserta didik untuk berpikir kritis dan kreatif.

Untuk mengatasi berbagai kendala tersebut, sekolah perlu memberikan dukungan melalui program pelatihan dan pendampingan bagi guru terkait penyusunan instrumen HOTS. Selain itu, pengembangan komunitas belajar guru dan penyediaan sumber belajar yang relevan dapat membantu meningkatkan kompetensi guru dalam merancang instrumen evaluasi yang valid dan reliabel. Dengan adanya dukungan tersebut, implementasi evaluasi berbasis HOTS diharapkan dapat berjalan lebih efektif dan memberikan kontribusi nyata terhadap peningkatan kualitas pendidikan.

KESIMPULAN

Berdasarkan hasil kajian, dapat disimpulkan bahwa validitas dan reliabilitas merupakan dua komponen utama yang menentukan kualitas instrumen evaluasi pembelajaran berbasis *Higher Order Thinking Skills* (HOTS) pada jenjang pendidikan menengah. Validitas memastikan bahwa instrumen benar-benar mampu mengukur kemampuan berpikir tingkat

tinggi yang meliputi aspek analisis, evaluasi, dan kreasi, sedangkan reliabilitas menunjukkan tingkat konsistensi instrumen dalam menghasilkan data pengukuran yang dapat dipercaya. Hasil pembahasan menunjukkan bahwa instrumen HOTS yang dikembangkan melalui tahapan penyusunan kisi-kisi, validasi ahli, uji coba, dan revisi secara sistematis cenderung memiliki tingkat validitas dan reliabilitas yang tinggi. Selain itu, implementasi instrumen HOTS memberikan dampak positif terhadap peningkatan kemampuan berpikir kritis dan kreatif peserta didik serta mendorong guru untuk menerapkan pembelajaran yang lebih inovatif. Meskipun masih terdapat berbagai kendala dalam penerapannya, seperti keterbatasan kompetensi guru dan rendahnya pengalaman peserta didik dalam mengerjakan soal HOTS, berbagai upaya pengembangan profesional dan pendampingan dapat menjadi solusi yang efektif. Oleh karena itu, pengembangan instrumen evaluasi HOTS yang valid dan reliabel perlu terus dilakukan untuk mendukung terciptanya sistem penilaian yang berkualitas dan relevan dengan tuntutan pendidikan abad ke-21.

DAFTAR PUSTAKA

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Arifin, Z., & Retnawati, H. (2017). Pengembangan instrumen pengukur higher order thinking skills matematika siswa SMA kelas X. *Pythagoras: Jurnal Pendidikan Matematika*, 12(1), 98–108. <https://doi.org/10.21831/pg.v12i1.14058>
- Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan* (Edisi 3). Bumi Aksara.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Hamdi, S., Suganda, I. A., & Hayati, N. (2018). Developing higher-order thinking skill (HOTS) test instrument using Lombok local cultures as contexts for junior secondary school mathematics. *Research and Evaluation in Education*, 4(2), 126–135. <https://doi.org/10.21831/reid.v4i2.22089>
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan reliabilitas dalam analisis butir soal. *Al-muarrib journal of arabic education*, 1(1), 34-44.
- Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan* (2nd ed.). Parama Publishing.
- Mujtahid, M., Assidiqi, A. H., & Sadiyah, D. (2025). Implementasi pembelajaran mendalam (deep learning) di sekolah dasar sebagai penguatan Kurikulum Merdeka. *Jurnal Ilmu Pendidikan Guru Sekolah Dasar Dan Usia Dini*, 2(2), 31-37.

- Pratama, O. (2025). Cosmopolitan Morality: A Critical Review of Peter Singer's Ethics in the Real World and Its Implications for Public Policy. *Abjad Journal of Humanities & Education*, 3(2), 159-171.
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian*. Parama Publishing.
- Saputri, H. A. S., & Larasati, N. J. (2023). Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(5), 2986-2995. <https://doi.org/10.36088/bintang.v3i2.1291>
- Sudarma, T. F., Irfandi, I., Festiyed, F., Adha, Y., & Yerimadesi, Y. (2023). Validity and reliability of higher order thinking test instrument development skills: Systematic literature reviews. *Jurnal Pelita Pendidikan*, 11(1), 1–8. <https://doi.org/10.24114/jpp.v11i1.42158>
- Sugiyono. (2019). *Metode penelitian dan pengembangan (Research and Development/R&D)*. Alfabeta.
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Widana, I. W. (2017). Higher order thinking skills assessment (HOTS). *JISAE: Journal of Indonesian Student Assessment and Evaluation*, 3(1), 32–44. <https://doi.org/10.21009/jisae.v3i1.4859>
- Yudha, R. P. (2023). Higher Order Thinking Skills (HOTS) test instrument: Validity and reliability analysis with the Rasch model. *EduMa: Mathematics Education Learning and Teaching*, 12(1), 21–38. <https://doi.org/10.24235/eduma.v12i1.9468>