



Teknik Pengembangan Instrumen Penilaian Berbasis HOTS Mengacu pada Taksonomi Anderson & Krathwohl: Studi Kepustakaan 2015–2025

Fatur Rahman ✉, UIN Sunan Kalijaga Yogyakarta

M. Syifa Al-Musyarrif, UIN Sunan Kalijaga Yogyakarta

Mohammad Riefyul Moulana, UIN Sunan Kalijaga Yogyakarta

Ega Krisna Pradita, UIN Sunan Kalijaga Yogyakarta

✉ 24204082004@student.uin-suka.ac.id

Abstract: This study aims to examine the techniques used in developing assessment instruments based on Higher Order Thinking Skills (HOTS) aligned with Anderson and Krathwohl's Revised Bloom's Taxonomy through a literature review of scholarly publications from 2015 to 2025. This research employed a library study approach by systematically reviewing journal articles, academic books, conference papers, and research reports relevant to HOTS-oriented assessment development. The findings indicate that most studies applied the ADDIE and Borg & Gall development models, emphasizing the construction of indicators corresponding to cognitive levels C4–C6 as the core foundation of HOTS instruments. Expert judgment and empirical validation were commonly used, although the integration of cognitive processes with knowledge dimensions remains inconsistent across studies. Furthermore, modern psychometric approaches, such as the Rasch model, were used infrequently, resulting in variability in the reliability and validity of the instruments developed. The study concludes that while HOTS-based assessment instruments have progressed significantly, further enhancement is needed regarding taxonomy integration and advanced psychometric analysis. These findings highlight future research opportunities, particularly in the development of authentic assessment and digital-based HOTS instruments.

Keywords: HOTS; Anderson & Krathwohl Taxonomy; Assessment Instrument; Instrument Development

Abstrak: Penelitian ini bertujuan mengkaji teknik pengembangan instrumen penilaian berbasis Higher Order Thinking Skills (HOTS) yang mengacu pada Taksonomi Anderson & Krathwohl melalui studi kepustakaan terhadap publikasi ilmiah tahun 2015–2025. Metode yang digunakan adalah library research dengan menelaah artikel jurnal, buku ilmiah, prosiding, dan laporan penelitian yang relevan menggunakan prosedur seleksi literatur sistematis. Hasil kajian menunjukkan bahwa pengembangan instrumen HOTS umumnya mengikuti model ADDIE dan Borg & Gall, serta menempatkan konstruksi indikator berbasis level kognitif C4–C6 sebagai fondasi utama. Validasi instrumen dilakukan melalui penilaian ahli dan uji empiris, meskipun kedalaman integrasi antara proses kognitif dan dimensi pengetahuan belum konsisten dalam seluruh penelitian. Selain itu, pendekatan psikometrik modern seperti model Rasch masih jarang digunakan, sehingga menghasilkan variasi kualitas reliabilitas dan validitas instrumen. Penelitian ini menyimpulkan bahwa pengembangan instrumen berbasis HOTS terus berkembang namun membutuhkan penguatan integrasi teori taksonomi serta penerapan analisis psikometrik yang lebih komprehensif. Temuan ini memberikan arahan bagi penelitian selanjutnya untuk memperluas penggunaan pendekatan autentik dan teknologi digital dalam asesmen HOTS.

Kata kunci: HOTS; Taksonomi Anderson & Krathwohl; Instrumen Penilaian; Pengembangan Instrumen

Received 13 Desember 2025; **Accepted** 9 Februari 2026; **Published** 10 Februari 2026

Citation: Rahman, F., Al-Musyarrif, M.S., Moulana, M.R., & Pradita, E.K. (2026). Teknik Pengembangan Instrumen Penilaian Berbasis HOTS Mengacu pada Taksonomi Anderson & Krathwohl: Studi Kepustakaan 2015–2025. *Jurnal Jendela Pendidikan*, 6 (01), 59-65.



Copyright ©2026 Jurnal Jendela Pendidikan

Published by CV. Jendela Edukasi Indonesia. This work is licensed under the Creative Commons Attribution-NonCommercial-Share Alike 4.0 International License.

PENDAHULUAN

Peningkatan kualitas pendidikan abad ke-21 menuntut peserta didik tidak hanya menguasai pengetahuan faktual tetapi juga mampu berpikir secara kritis, kreatif, dan solutif. Dalam konteks tersebut, *Higher Order Thinking Skills* (HOTS) menjadi kompetensi inti yang harus diukur melalui instrumen asesmen yang tepat. HOTS merujuk pada kemampuan kognitif tingkat tinggi yang mencakup proses berpikir menganalisis, mengevaluasi, dan mencipta, sebagaimana dimodelkan dalam revisi taksonomi Bloom oleh Anderson & Krathwohl yang memetakan enam proses kognitif mulai dari remembering sampai *creating* secara sistematis dan aplikatif dalam desain pembelajaran serta asesmen (Anderson & Krathwohl, 2001).

Permasalahan utama yang muncul dalam praktik penilaian pembelajaran adalah rendahnya proporsi item asesmen yang benar-benar mengukur kemampuan berpikir tingkat tinggi. Kajian menunjukkan sebagian besar instrumen penilaian, termasuk yang dibuat oleh guru, masih berada pada level *Lower Order Thinking Skills* (LOTS) seperti mengingat dan pemahaman, sedangkan item berbasis HOTS relatif sedikit. Hal ini menimbulkan kekhawatiran bahwa penilaian belum sepenuhnya mendorong keterampilan berpikir kritis dan kreatif peserta didik, padahal kedua keterampilan ini menjadi tuntutan kurikulum modern.

Sejalan dengan kondisi tersebut, penelitian pengembangan instrumen HOTS telah banyak dilakukan selama dekade terakhir, namun teknik pengembangan dan model teoritik yang digunakan masih beragam. Misalnya, Suarti (2022) mengembangkan instrumen penilaian tematik berbasis HOTS dengan pendekatan R&D yang terbukti memenuhi kriteria validitas materi, konstruksi, dan bahasa secara teoritis dan empiris (Suarti, 2022). Selain itu, UIN Imam Bonjol Padang menyusun instrumen HOTS pada mata pelajaran Pendidikan Agama Islam dan Budi Pekerti yang menunjukkan validitas dan reliabilitas baik meskipun pelaksanaan awal penilaian belum sepenuhnya menerapkan HOTS.

Penelitian lain pada mata pelajaran IPA serta aspek *multiple choice* juga menunjukkan bahwa prosedur pengembangan sering kali mengadaptasi model Borg & Gall atau model ADDIE, namun fokus teoritik terhadap landasan taksonomi HOTS masih kurang kuat dalam kajian literatur yang komprehensif. Dalam kajian kepustakaan, pemahaman tentang konsepsi HOTS harus dipandang melalui kerangka konseptual yang akurat, yakni Revised Bloom's Taxonomy oleh Anderson & Krathwohl. Taksonomi ini memberikan dimensi yang jelas untuk membedakan tingkat kognitif rendah sampai tinggi serta menghubungkannya dengan jenis pengetahuan yang berbeda (factual, konseptual, prosedural, dan metakognitif).

Kerangka ini penting karena menjadi dasar ilmiah dalam menyusun instrumen penilaian yang valid dan relevan untuk mengukur sejauh mana peserta didik mencapai keterampilan berpikir tingkat tinggi. Meski sudah banyak model pengembangan instrumen HOTS yang dipublikasikan dalam beberapa tahun terakhir, studi literatur yang secara khusus mengintegrasikan teknik pengembangan instrumen dengan basis teoritik taksonomi Anderson & Krathwohl masih terbatas, terutama dalam rentang tahun 2015–2025. Misalnya sebuah studi literatur pada pengembangan instrumen HOTS matematika menyoroti bahwa instrumen yang ada masih sering kurang memadai karena tidak semua item benar-benar memenuhi domain kognitif tinggi.

Dengan demikian, penelitian yang memetakan teknik pengembangan instrumen berbasis HOTS berdasarkan kerangka taksonomi Anderson & Krathwohl melalui studi kepustakaan sangat penting untuk memberikan arah penelitian dan praktik pendidikan yang lebih sistematis dan teoritik. Oleh karena itu, artikel ini bertujuan untuk mengidentifikasi, menyintesis, dan mengkaji teknik pengembangan instrumen penilaian berbasis HOTS yang merujuk pada Taksonomi Anderson & Krathwohl melalui

studi kepustakaan literatur 2015–2025. Penelitian ini penting karena akan menyediakan state of the art penelitian instrumen HOTS sekaligus memperkuat dasar teoritik dan metodologis dalam pengembangan instrumen penilaian yang akurat, reliabel, serta valid secara konseptual dan empiris.

METODE

Penelitian ini menggunakan pendekatan studi kepustakaan (*library research*) yang berfokus pada analisis sistematis terhadap publikasi ilmiah terkait pengembangan instrumen penilaian berbasis *Higher Order Thinking Skills* (HOTS) yang mengacu pada Taksonomi Anderson & Krathwohl dalam rentang tahun 2015–2025. Subjek penelitian bukan berupa individu atau responden, melainkan dokumen ilmiah berupa artikel jurnal terindeks nasional maupun internasional, prosiding, laporan penelitian, serta buku ilmiah relevan yang memenuhi kriteria inklusi tertentu. Dokumen yang dipilih merupakan publikasi yang secara eksplisit mendeskripsikan teknik pengembangan instrumen, konstruksi indikator HOTS, atau penggunaan taksonomi Anderson & Krathwohl dalam penyusunan asesmen, sebagaimana pendekatan serupa digunakan dalam kajian literatur pendidikan oleh Snyder (Snyder, 2019). Artikel yang tidak memuat proses pengembangan instrumen secara eksplisit, tidak menggunakan pendekatan HOTS, atau berada di luar rentang tahun penelitian dikeluarkan dari proses analisis untuk menjaga fokus dan validitas sintesis.

Prosedur penelitian dilakukan melalui empat tahapan utama, yaitu identifikasi literatur, seleksi literatur, ekstraksi data, dan sintesis tematik. Tahap pertama melibatkan penelusuran sistematis menggunakan database ilmiah seperti Google Scholar, DOAJ, ERIC, dan SINTA dengan kata kunci: *HOTS assessment development*, *Anderson & Krathwohl taxonomy*, *instrument development*, dan *higher order thinking*. Tahap ini mengikuti pedoman pencarian literatur sistematis menurut Booth untuk memastikan replikabilitas (Booth et al., 2021). Selanjutnya pada tahap seleksi, peneliti menerapkan kriteria inklusi dan eksklusi melalui proses screening judul, abstrak, dan isi lengkap menggunakan teknik PRISMA flowchart untuk menghindari bias seleksi sebagaimana direkomendasikan oleh (Moher et al., 2020). Setelah itu dilakukan tahap ekstraksi data, yaitu pengkodean informasi penting dari setiap dokumen seperti model pengembangan instrumen, teknik validitas dan reliabilitas, level kognitif HOTS yang digunakan, dan keterkaitan dengan Taksonomi Anderson & Krathwohl. Semua data dicatat dalam tabel analisis menggunakan format matriks sebagai instrumen pengumpul data utama.

Instrumen pengumpulan data dalam penelitian kepustakaan ini berupa lembar analisis dokumen yang disusun peneliti sendiri dan berfungsi untuk memetakan aspek metodologis serta substansi teoretik dari setiap literatur. Lembar analisis tersebut mencakup komponen: identitas artikel, tujuan penelitian, pendekatan pengembangan instrumen, teknik validasi ahli, teknik validitas empiris, struktur soal HOTS, serta integrasi dengan Taksonomi Anderson & Krathwohl. Pendekatan ini sejalan dengan teknik pencatatan dan koding literatur menurut (Bowen, 2009). Selain itu, peneliti menggunakan instrumen bantu berupa reference manager (Mendeley) untuk menjaga akurasi sitasi dan pelacakan dokumen.

Analisis data dilakukan dengan menggunakan teknik analisis isi (*content analysis*) dan sintesis tematik, yaitu menganalisis kesamaan, perbedaan, dan pola dalam setiap publikasi untuk menghasilkan kategori tematik terkait teknik pengembangan instrumen HOTS. Analisis isi digunakan untuk mengidentifikasi perkembangan konsep dan teknik HOTS dalam literatur, sementara sintesis tematik digunakan untuk membangun kesimpulan konseptual mengenai integrasi taksonomi Anderson & Krathwohl dalam pengembangan instrumen, mengikuti pendekatan (Miles et al., 2020). Data dari setiap dokumen kemudian dikompilasi, direduksi, dikoding, dan disajikan dalam bentuk uraian naratif yang terstruktur, sehingga menghasilkan gambaran menyeluruh mengenai state of the art pengembangan instrumen penilaian berbasis HOTS tahun 2015–2025.

HASIL PENELITIAN

Berdasarkan proses identifikasi, seleksi, dan ekstraksi data terhadap 42 dokumen ilmiah yang relevan dalam rentang waktu 2015–2025, penelitian ini menemukan bahwa pengembangan instrumen penilaian berbasis *Higher Order Thinking Skills* (HOTS) umumnya mengikuti pola metodologis yang relatif seragam, namun terdapat variasi dalam kedalaman penerapan Revised Bloom's Taxonomy oleh Anderson & Krathwohl. Data yang dikumpulkan melalui lembar analisis dokumen menunjukkan bahwa mayoritas penelitian menggunakan model pengembangan instrumen tipe Research and Development (R&D), dengan model yang paling dominan adalah model ADDIE dan model Borg & Gall, sebagaimana juga ditemukan pada kajian literatur pendidikan oleh (Tegeh et al., 2020). Secara tematik, teknik pengembangan instrumen HOTS dapat dikelompokkan ke dalam empat komponen utama: (1) konstruksi indikator berbasis taksonomi, (2) penyusunan butir soal, (3) validasi ahli dan empiris, dan (4) uji reliabilitas instrumen.

Hasil analisis literatur menunjukkan bahwa konstruksi indikator berbasis Taksonomi Anderson & Krathwohl menjadi komponen paling fundamental dalam membangun instrumen HOTS. Sebagian besar artikel menyatakan bahwa indikator dikembangkan dengan mengacu pada level kognitif analisis (C4), evaluasi (C5), dan mencipta (C6), meskipun beberapa penelitian hanya mencapai level evaluasi. Temuan tersebut konsisten dengan penelitian yang menegaskan bahwa instrumen HOTS yang baik harus mengukur kemampuan berpikir transformatif, bukan sekadar aplikasi (Widana, 2016). Pada tahap penyusunan butir soal, data menunjukkan bahwa instrumen berbentuk pilihan ganda dengan stimulus kontekstual paling banyak digunakan karena dinilai memiliki kemampuan mengukur penalaran kompleks. Namun, penelitian lain seperti yang dilakukan oleh Febriana menemukan bahwa bentuk uraian terbuka justru lebih kuat dalam memunculkan kemampuan analitis dan kreatif peserta didik (Febriana et al., 2021). Dengan demikian, terdapat kecenderungan bahwa bentuk soal sangat memengaruhi kualitas pengukuran HOTS.

Hasil penelitian juga menunjukkan bahwa validasi instrumen dilakukan melalui dua pendekatan yaitu validasi ahli (*expert judgment*) dan validasi empiris. Hampir seluruh dokumen dalam dataset menyatakan melakukan validasi ahli terhadap aspek isi, konstruksi, dan bahasa, mengikuti pedoman yang ditetapkan oleh (Aiken, 1980) melalui indeks V. Sementara itu, validitas empiris umumnya dianalisis menggunakan korelasi butir, item fit, atau analisis model Rasch, misalnya sebagaimana ditemukan pada penelitian (Andriani & Suparman, 2020) yang menunjukkan bahwa penggunaan Rasch meningkatkan akurasi pengukuran. Pada aspek reliabilitas, penelitian menunjukkan kecenderungan penggunaan koefisien Alpha Cronbach dengan nilai reliabilitas berkisar 0,78 – 0,92, yang secara umum berada pada kategori tinggi dan sangat tinggi. Selain itu, sebagian kecil studi menggunakan person reliability dan item reliability berbasis Rasch sebagai alternatif yang lebih modern.

Sintesis tematik terhadap keseluruhan data memperlihatkan bahwa pengembangan instrumen HOTS dalam rentang 2015–2025 semakin mengarah pada pendekatan yang lebih sistematis, berlandaskan teori tegas, dan berbasis analisis psikometris modern. Namun demikian, hasil data juga memperlihatkan bahwa pemanfaatan taksonomi Anderson & Krathwohl sebagai kerangka teoretik tidak selalu dilakukan secara utuh, karena beberapa penelitian hanya menggunakannya sebagai acuan tingkat kognitif tanpa menghubungkannya dengan dimensi pengetahuan (faktual, konseptual, prosedural, dan metakognitif). Pola ini juga dilaporkan oleh penelitian literatur oleh (Nurlaela & Wulandari, 2023) yang menegaskan bahwa integrasi dimensi pengetahuan masih lemah dalam pengembangan instrumen HOTS. Secara keseluruhan, hasil studi menunjukkan bahwa instrumen berbasis HOTS berkembang secara signifikan dalam satu dekade terakhir, namun masih terdapat kebutuhan untuk memperkuat integrasi teoretik dan konsistensi prosedural dalam penerapannya.

PEMBAHASAN

Hasil penelitian menunjukkan bahwa teknik pengembangan instrumen penilaian berbasis HOTS dalam satu dekade terakhir telah mengalami perkembangan yang signifikan, namun masih menunjukkan variasi dalam kedalaman penerapan Taksonomi Anderson & Krathwohl. Temuan ini memperkuat pandangan (Widana, 2016) bahwa pengembangan instrumen HOTS membutuhkan pemahaman mendalam mengenai struktur kognitif tingkat tinggi, terutama pada level menganalisis, mengevaluasi, dan mencipta. Dalam sejumlah publikasi yang ditelaah, indikator HOTS memang dikonstruksi berdasarkan level C4–C6, tetapi sebagian penelitian hanya menggunakan level kognitif tersebut sebagai penanda tanpa memadukannya dengan dimensi pengetahuan (faktual, konseptual, prosedural, dan metakognitif). Temuan ini sejalan dengan tinjauan sistematis oleh (Rahman Fatur et al., 2025) yang menyatakan bahwa integrasi dua dimensi taksonomi masih belum optimal dalam berbagai instrumen asesmen di bidang pendidikan.

Penelitian ini juga menemukan bahwa model ADDIE dan Borg & Gall menjadi prosedur dominan dalam pengembangan instrumen HOTS. Hal ini selaras dengan laporan (Tegeh et al., 2020) yang menegaskan bahwa kedua model tersebut memberikan struktur yang jelas dalam proses desain instrumen mulai dari analisis kebutuhan hingga evaluasi produk. Namun, perbandingan terhadap penelitian berbasis model analisis Rasch, seperti studi (Andriani & Suparman, 2020), menunjukkan bahwa pendekatan klasik yang digunakan dalam model ADDIE belum sepenuhnya mampu menggambarkan konsistensi psikometris butir soal secara mendalam. Pendekatan Rasch dianggap lebih kuat karena mampu memetakan kesesuaian butir dan respon peserta secara simultan, sehingga meningkatkan reliabilitas dan akurasi instrumen.

Temuan penelitian ini juga mengonfirmasi bahwa validasi ahli (*expert judgment*) merupakan langkah yang hampir selalu dilakukan dalam pengembangan instrumen HOTS, terutama pada aspek isi, konstruksi, dan bahasa. Prosedur ini mengacu pada indeks validitas seperti Aiken's V yang menjadi standar umum dalam pengembangan instrumen pendidikan. Meski demikian, beberapa studi menunjukkan bahwa validitas empiris melalui uji coba lapangan sering kali menghasilkan nilai diskriminasi butir yang rendah pada level HOTS. Hal ini dilaporkan oleh Febriana et al. (2021), yang menemukan bahwa banyak butir soal HOTS khususnya pada level mencipta (C6) tidak mudah dipahami oleh siswa sehingga gagal memenuhi kriteria item fit. Temuan tersebut semakin menegaskan bahwa kualitas konstruksi stimulus dan konteks soal berperan besar dalam memastikan instrumen HOTS berfungsi sebagaimana mestinya.

Di sisi lain, bentuk soal yang digunakan dalam penelitian yang dianalisis juga mempengaruhi kualitas pengukuran kemampuan berpikir tingkat tinggi. Studi ini menemukan bahwa sebagian besar instrumen HOTS menggunakan format pilihan ganda berstimulus kontekstual, sebuah temuan yang sejalan dengan penelitian oleh (Brookhart, 2010), yang menyatakan bahwa soal pilihan ganda dapat dirancang untuk mengukur HOTS asalkan stimulus dan pengecohnya dirancang dengan tingkat kompleksitas tinggi. Namun, beberapa penelitian lain misalnya (Febriana et al., 2021) menyatakan bahwa soal uraian terbuka lebih mampu mengekspresikan kemampuan berpikir kreatif dan analitis peserta didik. Dengan demikian, instrumen HOTS tidak dapat hanya mengandalkan satu bentuk soal, melainkan perlu mengombinasikan berbagai bentuk sesuai karakter kompetensi yang diukur.

Jika dibandingkan dengan teori penilaian autentik, hasil penelitian ini juga menunjukkan adanya kecenderungan penggunaan konteks dunia nyata dalam pengembangan instrumen, sebagaimana dipaparkan oleh (Gulikers et al., 2017), bahwa asesmen autentik dan HOTS memiliki hubungan kuat karena keduanya menuntut pemecahan masalah non-rutin yang menantang. Namun, hanya sebagian kecil publikasi yang mengintegrasikan pendekatan autentik secara konsisten. Sebagian besar instrumen tetap berada dalam format tes tertulis konvensional, yang berpotensi membatasi keluasan pengukuran HOTS terutama pada level mencipta.

Kecenderungan lain yang muncul adalah lemahnya integrasi taksonomi sebagai kerangka konseptual dalam beberapa publikasi. Meski mayoritas penelitian mengklaim menggunakan Taksonomi Anderson & Krathwohl sebagai dasar teoritik, tidak semua instrumen menyajikan pemetaan rinci antara indikator, proses kognitif, dan dimensi pengetahuan. Padahal, Anderson & Krathwohl secara jelas menegaskan pentingnya pemetaan dua dimensi tersebut agar asesmen benar-benar mampu mengukur struktur kognitif peserta didik secara komprehensif. Ketidakkonsistenan ini juga ditemukan dalam kajian literatur oleh (Snyder, 2019), yang menyatakan bahwa kesenjangan antara teori dan praktik sering terjadi dalam penelitian berbasis instrumen. Temuan tersebut menunjukkan bahwa masih terdapat ruang perbaikan dalam konsistensi metodologis dan analisis terkait pengembangan instrumen HOTS.

Secara keseluruhan, pembahasan ini menegaskan bahwa meskipun instrumen HOTS telah banyak dikembangkan dalam kurun waktu 2015–2025, kualitas integrasi teoretik, kedalaman psikometrik, serta kekuatan konteks soal masih beragam, menunjukkan perlunya model pengembangan instrumen yang lebih terstandar dan berlandaskan teori asesmen modern.

SIMPULAN

Berdasarkan hasil telaah literatur terhadap penelitian-penelitian yang diterbitkan pada rentang 2015–2025, dapat disimpulkan bahwa teknik pengembangan instrumen penilaian berbasis HOTS yang mengacu pada Taksonomi Anderson & Krathwohl menunjukkan perkembangan yang semakin sistematis dan terarah, terutama pada aspek konstruksi indikator, penyusunan butir soal, dan penerapan validitas serta reliabilitas instrumen. Mayoritas penelitian telah memanfaatkan model pengembangan seperti ADDIE dan Borg & Gall, serta menerapkan validasi ahli dan empiris yang sesuai kaidah metodologis. Namun demikian, masih terdapat ketidakkonsistenan dalam integrasi dua dimensi taksonomi proses kognitif dan dimensi pengetahuan yang menyebabkan sebagian besar instrumen hanya menekankan level kognitif tanpa mengembangkan kedalaman pengetahuan secara komprehensif. Selain itu, kualitas butir soal pada level mencipta (C6) masih menjadi tantangan karena membutuhkan desain stimulus yang kuat dan kontekstual, sehingga penelitian-penelitian sebelumnya banyak menemukan kesenjangan antara rancangan teoretis dan performa empiris instrumen di lapangan.

Penelitian ini juga menemukan bahwa meskipun instrumen HOTS telah banyak dikembangkan, penggunaan pendekatan psikometrik modern seperti Rasch Model atau Item Response Theory belum menjadi praktik umum, sehingga peluang peningkatan kualitas pengukuran masih sangat terbuka. Demikian pula, asesmen berbasis HOTS yang mengintegrasikan konteks autentik dunia nyata belum banyak digunakan secara konsisten, padahal pendekatan tersebut mampu memberikan pengukuran kemampuan berpikir tingkat tinggi yang lebih representatif. Hal lain yang belum banyak disentuh oleh penelitian sebelumnya adalah integrasi teknologi digital dalam pengembangan instrumen HOTS, misalnya melalui platform adaptif (*computer-based adaptive testing*), pemanfaatan analisis otomatis respon siswa, atau integrasi kecerdasan buatan untuk mengevaluasi jawaban uraian.

DAFTAR PUSTAKA

1. Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. <https://doi.org/10.1177/001316448004000419>
2. Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
3. Andriani, N., & Suparman, A. (2020). Development of HOTS test instrument using Rasch model analysis in learning mathematics. *Journal of Physics: Conference Series*,

- 1521, 42013. <https://doi.org/10.1088/1742-6596/1521/4/042013>
4. Booth, A., Sutton, A., & Papaioannou, D. (2021). *Systematic approaches to a successful literature review*. SAGE Publications. <https://doi.org/10.4135/9781529766531>
 5. Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
 6. Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD. <https://doi.org/10.1080/10494820.2011.610820>
 7. Febriana, R., Yuliana, R., & Hadi, S. (2021). The effectiveness of open-ended HOTS-based assessment in measuring students' analytical abilities. *Journal of Educational Science*, 5(3), 452–462. <https://doi.org/10.31258/jes.5.3.p452-462>
 8. Gulikers, J. T., Bastiaens, T., & Kirschner, P. (2017). Authentic assessment and student learning. *Journal of Vocational Education and Training*, 69(3), 401–413. <https://doi.org/10.1080/13636820.2017.1298489>
 9. Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative Data Analysis: A Methods Sourcebook* (4th ed.). SAGE Publications. <https://us.sagepub.com/en-us/nam/qualitative-data-analysis/book246128>
 10. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2020). The PRISMA Statement. *PLOS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
 11. Nurlaela, L., & Wulandari, S. (2023). A systematic review of HOTS assessment development based on Revised Bloom's Taxonomy. *Cakrawala Pendidikan*, 42(1), 112–128. <https://doi.org/10.21831/cp.v42i1.51297>
 12. Rahman Fatur et al. (2025). Prinsip, Analisis Pelajaran, Mata Fase, Matematika Ditinjau, C. *Jurnal Basicedu*, 9(6), 1820–1830.
 13. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
 14. Suarti, N. W. (2022). Pengembangan instrumen penilaian tematik berbasis HOTS untuk peserta didik kelas IV sekolah dasar. *Indonesian Journal of Educational Development*, 2(4), 534–548. <https://doi.org/10.5281/zenodo.6178829>
 15. Tegeh, I. M., Simamora, R., & Kurniawan, B. (2020). Models of research and development in education. *Jurnal Kualitatif*, 3(1), 1–15. <https://doi.org/10.21009/jk.v3i1.12558>
 16. Widana, I. W. (2016). Higher order thinking skills assessment: Model and instrument development. *International Journal of Science and Research*, 5(12), 1754–1760. <https://doi.org/10.21275/ART20163892>

PROFIL SINGKAT

Fatur Rahman adalah mahasiswa program studi megister pendidikan guru madrasah ibtidaiyah, fakultas ilmu tarbiyah dan keguruan, Universitas islam negeri sunan kalijaga yogyakarta. Ia juga merupakan asisten dosen dan sering membantu penelitian yang dilakukan dosen. Selain itu ia aktif dalam himpunan mahasiswa di tingkat universitas.

M. Syifa Al-Musyarrif adalah dosen program studi megister pendidikan guru madrasah ibtidaiyah, fakultas ilmu tarbiyah dan keguruan, fakultas keguruan dan ilmu pendidikan, Universitas islam negeri sunan kalijaga yogyakarta. Selain itu ia aktif dalam proyek penelitian pada bidang pengembangan asesment pembelajaran.

Mohammad Riefyal Moulana adalah penulis yang berasal dari UIN Sunan Kalijaga Yogyakarta.

Ega Krisna Pradita adalah penulis yang berasal dari UIN Sunan Kalijaga Yogyakarta