

PENGEMBANGAN BUTIR SOAL HOTS PADA PENDIDIKAN TINGGI: MODEL BERBASIS STIMULUS, PENALARAN, DAN ANALISIS BUTIR

Reva Arda Try Pradima¹, Erdhita Oktrifianty², Mohammad Zenith Zacky Ayyashri³, Zulfa Muthia Bilqis⁴, Ulfa Tuzzami⁵
revaarda4@gmail.com¹, erdhitaoktrifianty@gmail.com², zenithzackyyayashri@gmail.com³,
zulfaabel20@gmail.com⁴, ulfatuzzami2025@gmail.com⁵
Universitas Muhammadiyah Tangerang

ABSTRAK

Penguatan Higher Order Thinking Skills (HOTS) pada pendidikan tinggi menuntut asesmen yang mampu mengukur penalaran, pengambilan keputusan berbasis bukti, dan kemampuan merancang solusi dalam konteks keilmuan yang autentik. Tantangan umum di perguruan tinggi ialah penyusunan butir yang tampak kompleks namun tidak benar-benar menilai penalaran (misalnya hanya memperpanjang stem atau menaikkan kesulitan hitung). Artikel ini menyajikan model terpadu penyusunan butir HOTS untuk mahasiswa melalui kajian pustaka dan sintesis kerangka teoretik taksonomi kognitif, constructive alignment, prinsip penulisan item, serta standar kualitas tes dan interpretasi skor. Metode yang digunakan adalah kajian pustaka naratif dengan analisis tematik untuk merumuskan langkah operasional, kriteria stimulus, desain tugas (task), dan prosedur penjaminan mutu. Hasil kajian merumuskan delapan tahap pengembangan butir HOTS (penyelarasan capaian pembelajaran, indikator C4-C6, desain stimulus, perumusan task, pemilihan format, penyusunan kunci/rubrik, telaah ahli, uji coba dan analisis butir). Artikel ini juga menyediakan contoh butir pilihan ganda berbasis kasus dan soal uraian berbasis desain solusi, lengkap dengan rubrik analitik dan checklist telaah. Implikasi praktisnya adalah penyusunan asesmen HOTS yang lebih sahih, adil, dan selaras dengan capaian pembelajaran lulusan di pendidikan tinggi.

Kata Kunci: HOTS, Pendidikan Tinggi, Penyusunan Butir Soal, Stimulus, Constructive Alignment, Analisis Butir.

ABSTRACT

Strengthening Higher Order Thinking Skills (HOTS) in higher education requires assessments that can capture reasoning, evidence-based decision making, and solution design in authentic disciplinary contexts. A recurring challenge in universities is the construction of items that appear complex yet fail to elicit higher-order reasoning (e.g., lengthy stems or inflated computational burden). This paper proposes an integrated model for HOTS item development for undergraduate students through a narrative literature review and thematic synthesis of key frameworks: the revised Bloom's taxonomy, constructive alignment, item-writing principles, and testing standards for validity-oriented score interpretation. The study synthesizes operational steps, stimulus quality criteria, task design patterns, and quality assurance procedures. The proposed model comprises eight stages: alignment with learning outcomes, C4-C6 indicator specification, stimulus design, task formulation, format selection, key/rubric development, expert review, and pilot testing with item analysis. We also provide sample case-based multiple-choice items and design-oriented essay tasks with analytic rubrics and a review checklist. The model supports more valid, fair, and outcome-aligned HOTS assessment practices in higher education.

Keywords: HOTS, Higher Education, Item Development, Stimulus, Constructive Alignment, Item Analysis.

PENDAHULUAN

Asesmen di perguruan tinggi berperan ganda: (1) memotret capaian belajar mahasiswa secara akuntabel, dan (2) membentuk cara belajar mahasiswa melalui sinyal tentang apa

yang dianggap penting (assessment drives learning). Di banyak program studi, capaian pembelajaran lulusan menekankan kemampuan analitis, evaluatif, dan kreatif, misalnya kemampuan menafsirkan data, menilai argumen ilmiah, merancang intervensi, atau membuat keputusan profesional berdasarkan bukti. Karena itu, penyusunan butir soal HOTS bukan sekadar menaikkan tingkat kesulitan, tetapi merancang tugas yang memaksa mahasiswa mengolah informasi, menimbang alternatif, dan memberikan justifikasi yang dapat dipertanggungjawabkan.

Panduan nasional penulisan soal HOTS menegaskan bahwa soal HOTS umumnya mengukur proses kognitif menganalisis (C4), mengevaluasi (C5), dan mencipta (C6), sering kali pada dimensi pengetahuan metakognitif. Sementara itu, revisi Taksonomi Bloom memberikan kerangka yang lebih operasional untuk menyelaraskan tujuan pembelajaran, aktivitas belajar, dan asesmen. Di pendidikan tinggi, prinsip constructive alignment menekankan keselarasan eksplisit antara capaian pembelajaran, aktivitas belajar-mengajar, dan tugas asesmen, sehingga butir HOTS harus dibangun dari capaian yang menuntut penalaran, bukan ditempelkan di akhir sebagai variasi soal.

Artikel ini menyajikan model terpadu penyusunan butir HOTS untuk mahasiswa, dengan fokus pada: (a) desain stimulus autentik; (b) pola task yang memicu penalaran; (c) pemilihan format (pilihan ganda berbasis kasus, uraian analitis, atau tugas desain); dan (d) penjaminan mutu melalui telaah ahli serta analisis butir. Model ini ditujukan untuk dosen, tim penjaminan mutu, dan penyusun ujian program studi yang membutuhkan panduan praktis namun berbasis literatur.

METODOLOGI

Artikel ini merupakan kajian pustaka naratif dengan analisis tematik. Sumber literatur mencakup: (1) taksonomi kognitif dan HOTS (Anderson & Krathwohl, 2001; Krathwohl, 2002; Brookhart, 2010); (2) constructive alignment untuk konteks perguruan tinggi (Biggs, 1996); (3) prinsip penulisan item dan blueprint (Haladyna, 2004); (4) standar kualitas tes dan validitas interpretasi skor (AERA et al., 2014); serta (5) pedoman nasional penulisan soal HOTS sebagai referensi praktik lokal (Kemdikbud, 2019). Analisis dilakukan dengan mengode tema-tema kunci (indikator HOTS, stimulus, task, format, rubrik, telaah, dan analisis butir) untuk kemudian disintesis menjadi model operasional yang dapat diterapkan pada asesmen perkuliahan.

HASIL DAN PEMBAHASAN

1. Model Delapan Tahap Pengembangan Butir HOTS untuk Mahasiswa

Tahap 1 - Penyelarasan capaian pembelajaran

Identifikasi Capaian Pembelajaran Mata Kuliah (CPMK) dan/atau Sub-CPMK yang menuntut penalaran tinggi. Gunakan prinsip constructive alignment: asesmen harus mengukur capaian yang diklaim dan dilatihkan dalam kegiatan belajar.

Tahap 2 - Perumusan indikator butir C4-C6

Turunkan CPMK menjadi indikator butir yang spesifik dan teramati. Contoh: 'menganalisis bias pada rancangan penelitian', 'mengevaluasi kekuatan bukti dari dua artikel', atau 'merancang prosedur pengumpulan data yang etis'.

Tahap 3 - Penetapan jenis penalaran

Tentukan bentuk penalaran yang ingin ditangkap: inferensi dari data, kausalitas, argumentasi berbasis kriteria, pemilihan strategi, atau desain solusi. Tahap ini mencegah soal hanya 'berlabel' HOTS.

Tahap 4 - Desain stimulus autentik

Kembangkan stimulus yang relevan dengan disiplin: ringkasan artikel, data tabel/grafik, skenario klinis, studi kasus kebijakan, potongan kode/hasil eksperimen, atau dokumen profesional. Stimulus harus cukup informasi, tidak bias, dan sesuai literasi mahasiswa.

Tahap 5 - Perumusan task (tuntutan berpikir)

Tulis perintah yang memaksa mahasiswa mengolah stimulus (bukan sekadar menyalin). Gunakan kata kerja yang menuntut analisis, evaluasi, atau kreasi, dan jelaskan kriteria bila perlu.

Tahap 6 - Pemilihan format dan skema skor

Pilih format yang efisien sekaligus sesuai tujuan: pilihan ganda berbasis kasus (C4-C5), uraian analitis (C5), atau tugas desain/proyek mini (C6). Tentukan apakah perlu alasan (justification) dan bagaimana skornya.

Tahap 7 - Penyusunan kunci, rasional, dan rubrik

Untuk PG, susun kunci serta rasional mengapa opsi lain salah (berbasis miskonsepsi umum). Untuk uraian, gunakan rubrik analitik agar penilaian konsisten dan transparan.

Tahap 8 - Penjaminan mutu: telaah ahli, uji coba, dan analisis butir

Lakukan telaah isi (kesesuaian indikator-stimulus-task), telaah bahasa (keterbacaan), dan telaah bias. Jika memungkinkan, lakukan uji coba untuk menghitung indeks kesukaran, daya pembeda, fungsi distraktor, dan reliabilitas sebagai dasar revisi sebelum penggunaan berskala besar.

2. Pola Task HOTS yang Relevan untuk Mahasiswa

Pada pendidikan tinggi, task HOTS perlu merepresentasikan praktik berpikir dalam disiplin (disciplinary thinking). Berikut pola task yang umum dan efektif:

- 1) Analisis (C4): menafsirkan data, menemukan pola/inkonsistensi, menganalisis penyebab, mengidentifikasi asumsi tersembunyi.
- 2) Evaluasi (C5): menilai kualitas argumen/metode berdasarkan kriteria (validitas, reliabilitas, etika, efisiensi), membandingkan alternatif dan memilih yang paling dapat dipertanggungjawabkan.
- 3) Kreasi (C6): merancang prosedur, model, prototipe, atau rekomendasi kebijakan dengan batasan yang jelas (waktu, biaya, etika, sumber daya).

Kemdikbud (2019) menekankan bahwa soal HOTS mengukur kemampuan memproses dan menerapkan informasi, mengaitkan berbagai informasi, menggunakan informasi untuk menyelesaikan masalah, serta menelaah ide secara kritis; prinsip ini dapat diadaptasi untuk konteks mahasiswa dengan stimulus dan kriteria yang lebih kaya.

3. Penjaminan Mutu Empirik: Analisis Butir (CTT) sebagai Minimum Quality Gate

Untuk ujian pilihan ganda atau tes objektif, analisis butir berbasis Classical Test Theory (CTT) dapat digunakan sebagai pemeriksaan minimum mutu. Indikator yang lazim digunakan antara lain: (a) indeks kesukaran (p) sebagai proporsi jawaban benar; (b) daya pembeda, misalnya korelasi point-biserial antara skor item dan skor total; (c) fungsi distraktor (apakah dipilih oleh kelompok kemampuan rendah secara proporsional); dan (d) reliabilitas internal (mis. koefisien alpha) untuk konsistensi skor. Crocker dan Algina (1986) membahas reliabilitas, validitas, dan item analysis sebagai komponen pengembangan tes. Standards (AERA et al., 2014) menekankan pengumpulan bukti yang mendukung interpretasi skor; analisis butir membantu memberikan bukti internal structure dan mengidentifikasi item bermasalah untuk direvisi.

Contoh Butir HOTS untuk Pendidikan Tinggi

1. Contoh Pilihan Ganda Berbasis Kasus (C5 - Evaluasi Metode)

Stimulus (ringkasan studi):

Seorang peneliti ingin menguji efektivitas modul belajar daring terhadap peningkatan kemampuan literasi data mahasiswa tahun pertama. Peneliti membandingkan dua kelas yang berbeda program studi. Kelas A menggunakan modul daring selama 4 minggu, kelas B menggunakan metode biasa. Di akhir, keduanya diberi tes yang sama. Peneliti menyimpulkan modul daring lebih efektif karena skor rata-rata kelas A lebih tinggi.

Pertanyaan:

Evaluasi paling tepat terhadap kesimpulan peneliti adalah...

- A. Kesimpulan kuat karena menggunakan dua kelompok dan tes yang sama.
- B. Kesimpulan lemah karena kelompok berasal dari program studi berbeda sehingga terdapat variabel perancu (confounding).
- C. Kesimpulan pasti benar karena skor rata-rata kelas A lebih tinggi.
- D. Kesimpulan tidak dapat dibuat karena penelitian kuantitatif tidak boleh membandingkan dua kelas.

Kunci: B

Rasional: Perbedaan program studi dapat berkaitan dengan kemampuan awal, motivasi, atau kebiasaan belajar sehingga mengancam validitas internal. Item ini menuntut mahasiswa mengevaluasi kekuatan inferensi kausal berdasarkan desain studi dan potensi variabel perancu.

2. Contoh Uraian Analitis (C5 - Evaluasi Bukti) + Rubrik

Stimulus:

Mahasiswa membaca dua artikel yang membahas efektivitas pembelajaran berbasis proyek. Artikel 1 menggunakan sampel besar namun tanpa kelompok kontrol. Artikel 2 menggunakan desain eksperimen dengan kelompok kontrol tetapi sampelnya kecil.

Tugas:

Dengan kriteria yang Anda tetapkan, evaluasilah bukti dari kedua artikel tersebut. Tentukan artikel mana yang memberikan bukti lebih kuat untuk mendukung klaim efektivitas, jelaskan alasan Anda, dan tuliskan rekomendasi perbaikan desain untuk artikel yang lebih lemah.

Rubrik analitik (skor 0-4 per aspek; total 16):

Aspek	Indikator kinerja	Skor (0-4)
Kriteria evaluasi	Menetapkan kriteria yang relevan (mis. validitas internal, generalisasi, bias, ukuran efek) dan menjelaskannya.	0-4
Analisis bukti	Menganalisis kekuatan-keterbatasan masing-masing artikel berdasarkan kriteria; menyebutkan konsekuensi metodologis.	0-4
Justifikasi keputusan	Memilih artikel dengan bukti lebih kuat dan memberikan alasan logis, konsisten, dan berbasis stimulus.	0-4
Rekomendasi perbaikan	Memberikan perbaikan desain yang realistis dan terkait langsung dengan kelemahan yang diidentifikasi.	0-4

3. Checklist Telaah Butir HOTS (Siap Pakai)

- Checklist berikut dapat digunakan dosen atau tim penelaah (peer review) sebelum butir digunakan pada ujian. Setiap butir disarankan memenuhi mayoritas kriteria berikut.
- Selaras dengan CPMK/Sub-CPMK dan constructive alignment (tujuan-aktivitas-asesmen).
- Indikator jelas pada level C4/C5/C6; jawaban tidak cukup dengan hafalan.
- Stimulus autentik, relevan, cukup informasi, dan tidak bias; sumber/istilah dapat dipahami mahasiswa.
- Task memaksa pengolahan stimulus (analisis/evaluasi/desain) dan menyebutkan kriteria bila diperlukan.
- Bahasa ringkas, tidak multitafsir, tidak mengandung petunjuk jawaban terselubung.
- Untuk PG: opsi homogen, distraktor fungsional, tidak ada opsi ekstrem/aneh, dan hanya satu kunci yang paling benar.
- Untuk uraian: rubrik analitik tersedia; kriteria dan bobot skor jelas.
- Rencana uji coba/analisis butir tersedia (p, point-biserial, fungsi distraktor, reliabilitas) untuk revisi berkelanjutan.

Implikasi Implementasi di Perguruan Tinggi

Pertama, program studi perlu menyusun blueprint asesmen yang memetakan CPMK terhadap level kognitif dan jenis tugas. Blueprint mencegah dominasi butir hafalan dan memastikan proporsi butir HOTS yang sejalan dengan profil lulusan. Kedua, untuk kelas besar, pilihan ganda berbasis kasus dapat dipakai sebagai format efisien, tetapi harus memenuhi kaidah item writing dan didukung analisis butir; untuk capaian C6, asesmen berbasis produk atau desain lebih tepat. Ketiga, penjaminan mutu idealnya dilakukan melalui siklus: rancang - telaah - uji coba - analisis - revisi, sehingga bank soal program studi terus membaik dan interpretasi skor lebih dapat dipertanggungjawabkan (AERA et al., 2014).

Keterbatasan dan Agenda Riset

Artikel ini bersifat konseptual berbasis kajian pustaka, sehingga belum menyajikan data uji coba lapangan pada mata kuliah tertentu. Riset lanjutan disarankan untuk: (1) menguji efektivitas model delapan tahap pada berbagai disiplin; (2) mengembangkan indikator HOTS yang lebih spesifik pada domain keilmuan (mis. penalaran klinis, penalaran matematis, penalaran hukum); (3) membandingkan kualitas butir HOTS lintas format (PG kasus vs uraian vs proyek mini) menggunakan bukti validitas dan analisis reliabilitas.

KESIMPULAN

Pengembangan butir soal HOTS pada pendidikan tinggi menuntut desain yang sistematis dan selaras dengan capaian pembelajaran. Model delapan tahap yang diusulkan menekankan desain stimulus autentik, perumusan task yang memicu penalaran C4-C6, pemilihan format yang tepat, serta penjaminan mutu melalui telaah ahli dan analisis butir. Implementasi model ini mendukung asesmen yang lebih sahih, adil, dan berdampak pada pembelajaran mahasiswa.

DAFTAR PUSTAKA

- AERA, APA, & NCME. (2014). Standards for educational and psychological testing. American Educational Research Association.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364. <https://doi.org/10.1007/BF00138871>
- Brookhart, S. M. (2010). How to assess higher-order thinking skills in your classroom. ASCD.

- Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Holt, Rinehart and Winston.
- Haladyna, T. M. (2004). Developing and validating multiple-choice test items (3rd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203825945>
- Kementerian Pendidikan dan Kebudayaan. (2019). Penulisan soal HOTS - Higher Order Thinking Skills (keterampilan berpikir tingkat tinggi). Pusat Penilaian Pendidikan.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), 212-218. https://doi.org/10.1207/S15430421TIP4104_2