

# INTEGRASI KUALITAS PSIKOMETRIK, KOMPETENSI GURU, DAN TEKNOLOGI DIGITAL DALAM PENGEMBANGAN SOAL PILIHAN GANDA BERBASIS HIGHER-ORDER THINKING SKILLS (HOTS) A SYSTEMATIC LITERATURE REVIEW

Isa Faizul Haq<sup>1</sup>, Aji Joko Budi Pramono<sup>2</sup>

[faizulisa20@gmail.com](mailto:faizulisa20@gmail.com)<sup>1</sup>, [ajijoko@gmail.com](mailto:ajijoko@gmail.com)<sup>2</sup>

UIN Raden Mas Said Surakarta

## ABSTRAK

Soal pilihan ganda masih menjadi instrumen asesmen yang paling banyak digunakan dalam pendidikan, namun sebagian besar belum dirancang untuk mengukur Higher-Order Thinking Skills (HOTS) secara optimal. Penelitian ini bertujuan mensintesis temuan empiris mengenai penggunaan soal pilihan ganda sebagai instrumen asesmen formatif dan sumatif serta implikasinya terhadap pengembangan HOTS. Penelitian menggunakan metode Systematic Literature Review (SLR) dengan mengikuti pedoman PRISMA 2020. Sebanyak 16 artikel ilmiah nasional dan internasional yang diterbitkan pada periode 2015–2026 dipilih melalui proses identifikasi, penyaringan, penilaian kelayakan, dan inklusi berdasarkan kriteria yang telah ditetapkan. Data dianalisis menggunakan thematic analysis untuk mengidentifikasi pola dan tema utama penelitian. Hasil sintesis menunjukkan lima tema utama, yaitu: (1) pengembangan dan validasi soal pilihan ganda HOTS, (2) analisis kualitas butir soal, (3) kompetensi guru dalam pengembangan instrumen, (4) implementasi soal pilihan ganda dalam asesmen formatif dan sumatif berbasis digital, dan (5) penggunaan three-tier multiple choice untuk mengidentifikasi miskonsepsi siswa. Temuan menunjukkan bahwa soal pilihan ganda yang dirancang berdasarkan indikator HOTS, didukung validitas dan reliabilitas yang memadai, serta diintegrasikan dengan teknologi digital mampu mendukung keterampilan berpikir kritis, kreatif, analitis, dan pemecahan masalah. Kajian ini menegaskan pentingnya pendekatan holistik yang mengintegrasikan kualitas instrumen, kompetensi guru, dan inovasi teknologi dalam pengembangan asesmen berbasis HOTS.

**Kata Kunci:** Multiple-Choice Questions, Higher-Order Thinking Skills, Item Analysis, Educational Assessment.

## PENDAHULUAN

### Latar Belakang

Asesmen dalam pembelajaran memiliki peran strategis untuk mengukur pencapaian peserta didik sekaligus memberikan umpan balik yang mendukung proses belajar. Salah satu bentuk asesmen yang paling banyak digunakan di sekolah adalah soal pilihan ganda (multiple-choice questions/MCQ), karena praktis dan mudah dianalisis. Namun, berbagai penelitian menunjukkan bahwa pembuatan soal pilihan ganda di berbagai jenjang pendidikan masih didominasi oleh level kognitif rendah (C1–C2), sehingga belum sepenuhnya mendorong pengembangan keterampilan berpikir tingkat tinggi atau Higher-Order Thinking Skills (HOTS) (Yuliantaningrum & Sunarti, 2020; Warju et al., 2020). Kondisi ini memperlihatkan kesenjangan antara praktik asesmen saat ini dengan tuntutan kurikulum dan prinsip pembelajaran berbasis deep learning.

HOTS merupakan kemampuan berpikir non-algoritmik yang kompleks, melibatkan analisis, evaluasi, dan kreasi dalam menyelesaikan masalah. Kompetensi ini menjadi bagian

integral dari keterampilan abad ke-21, termasuk berpikir kritis, kreatif, dan problem solving (Handoyo et al., 2025). Studi empiris menunjukkan bahwa soal PG yang dikembangkan tanpa memperhatikan level kognitif HOTS cenderung hanya menilai kemampuan mengingat dan memahami, sehingga gagal mengembangkan potensi berpikir kritis peserta didik (Kejora & Padmarani, 2026; Sri Nurhalimah et al., 2022).

Selain itu, penelitian terdahulu juga menyoroti kesenjangan antara kurikulum, praktik asesmen di sekolah, dan penerapan asesmen formatif yang efektif. Banyak soal PG belum mampu menstimulasi peserta didik untuk melakukan refleksi, analisis, atau mengaitkan pengetahuan dengan konteks nyata. Padahal, asesmen formatif yang diterapkan secara sistematis dapat mendorong pembelajaran yang berpusat pada murid, memperkuat pemahaman mendalam, dan meningkatkan keterampilan berpikir tingkat tinggi (Suryaningsih et al., 2024; Wuryanti et al., 2017).

Lebih lanjut, validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distractor menjadi aspek penting dalam pengembangan soal PG berbasis HOTS. Butir soal yang memenuhi kriteria tersebut mampu mengakomodasi asesmen formatif maupun sumatif secara bermakna, sekaligus menjadi alat diagnostik untuk mengidentifikasi miskonsepsi siswa (Anggraeni et al., 2018; Mania et al., 2020). Selain itu, integrasi teknologi digital, seperti Google Forms dan sistem berbasis AI, terbukti mendukung proses evaluasi, mempercepat validasi, dan meningkatkan kualitas umpan balik bagi siswa (Handoyo et al., 2025; Suryaningsih et al., 2024).

Penelitian ini hadir untuk mengisi kesenjangan literatur dengan meninjau secara sistematis studi terkait soal PG sebagai instrumen asesmen formatif dan sumatif, serta implikasinya terhadap pengembangan HOTS. Dengan menggabungkan analisis kualitas butir soal, kompetensi guru, dan integrasi teknologi, kajian ini bertujuan memberikan rekomendasi praktis yang dapat meningkatkan efektivitas PG dalam mendukung kompetensi abad ke-21.

Meskipun berbagai penelitian telah membahas pengembangan soal pilihan ganda, sebagian besar kajian masih berfokus pada aspek tertentu, seperti validitas instrumen, analisis butir soal, kompetensi guru, atau pemanfaatan teknologi asesmen secara terpisah. Kajian yang mensintesis secara komprehensif keterkaitan antara kualitas psikometrik instrumen, kompetensi guru, digitalisasi asesmen, dan kontribusinya terhadap pengembangan Higher-Order Thinking Skills (HOTS) masih relatif terbatas. Selain itu, belum banyak penelitian yang mengintegrasikan fungsi formatif, sumatif, dan diagnostik soal pilihan ganda dalam satu kerangka analisis yang utuh. Oleh karena itu, diperlukan tinjauan literatur sistematis untuk memperoleh pemahaman yang lebih komprehensif mengenai peran soal pilihan ganda dalam mendukung pengembangan HOTS pada berbagai konteks pembelajaran.

## **METODOLOGI**

Penelitian ini menggunakan metode Systematic Literature Review (SLR) dengan pendekatan kualitatif-deskriptif. SLR dipilih karena memungkinkan peneliti untuk mengidentifikasi, meninjau, dan mensintesis literatur ilmiah yang relevan dengan topik pengembangan soal pilihan ganda dan implikasinya terhadap Higher-Order Thinking Skills (HOTS) serta asesmen formatif dan sumatif (Snyder, 2019).

Data penelitian bersumber dari 16 artikel ilmiah peer-reviewed, baik nasional maupun internasional, yang membahas: pengembangan soal pilihan ganda, analisis butir soal, validitas, reliabilitas, tingkat kesukaran, daya pembeda, efektivitas pengecoh, dan keterkaitannya dengan HOTS. Analisis dilakukan secara deskriptif dan interpretatif untuk menemukan pola, kecenderungan, kendala, serta gap penelitian terkait asesmen dan evaluasi

HOTS.

Proses penelitian mengikuti tahapan PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) agar pengumpulan, seleksi, dan sintesis literatur dilakukan secara sistematis, transparan, dan dapat direplikasi (Page et al., 2021).

## **HASIL DAN PEMBAHASAN**

### **1. Gambaran Umum Artikel yang Dianalisis**

Kajian ini menganalisis 16 artikel ilmiah yang memiliki keterkaitan langsung dengan penggunaan soal pilihan ganda, pengembangan instrumen HOTS, evaluasi pembelajaran, analisis butir soal, serta pemanfaatan teknologi dalam asesmen. Artikel-artikel yang dianalisis mencakup rentang publikasi dari tahun 2015 hingga 2026 dan merepresentasikan beberapa jenjang pendidikan, mulai dari SMP, SMA, SMK, hingga konteks pengembangan instrumen oleh guru. Secara umum, literatur menunjukkan bahwa soal pilihan ganda masih menjadi salah satu bentuk instrumen evaluasi yang banyak digunakan karena praktis, efisien, dan mudah dianalisis. Namun, efektivitas soal pilihan ganda tidak hanya ditentukan oleh bentuk soalnya, melainkan oleh kualitas konstruksi butir, kesesuaian dengan indikator pembelajaran, validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh atau distractor (Mania et al., 2020; Nurhalimah et al., 2022; Warju et al., 2020).

Dari aspek jenjang pendidikan, artikel yang dianalisis menunjukkan bahwa soal pilihan ganda digunakan dalam berbagai konteks pembelajaran. Pada jenjang SMP, penelitian menyoroti hubungan validitas item dengan daya pembeda dan tingkat kesukaran soal PAS IPA, serta rekonstruksi soal pilihan ganda dalam asesmen formatif berbasis deep learning (Nurhalimah et al., 2022; Kejora et al., 2026). Pada jenjang SMA, beberapa studi berfokus pada pengembangan soal pilihan ganda untuk mengukur berpikir kritis, pemanfaatan Google Forms untuk meningkatkan HOTS, serta penggunaan tes pilihan ganda tiga tingkat dalam mengidentifikasi miskonsepsi siswa (Aripin, 2018; Suryaningsih et al., 2024; Anggraeni et al., 2018; Wuryanti et al., 2017). Pada jenjang SMK, penelitian Warju et al. (2020) memperlihatkan pentingnya analisis kualitas butir soal tipe HOTS berdasarkan validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Sementara itu, studi tentang kompetensi guru menunjukkan bahwa kualitas soal pilihan ganda sangat dipengaruhi oleh kemampuan guru dalam menyusun kisi-kisi, menulis butir soal, melakukan telaah, serta melakukan analisis butir soal (Vitiarti, 2022; Gusnita et al., 2021).

Berdasarkan fokus kajiannya, artikel-artikel tersebut dapat dikelompokkan ke dalam beberapa kecenderungan utama. Pertama, terdapat artikel yang menempatkan soal pilihan ganda sebagai instrumen untuk mengukur keterampilan berpikir tingkat tinggi, seperti berpikir kritis, berpikir kreatif, dan pemecahan masalah (Yuliantaningrum & Sunarti, 2020; Aripin, 2018). Kedua, terdapat artikel yang menekankan pentingnya analisis kualitas butir soal melalui indikator validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distractor (Mania et al., 2020; Nurhalimah et al., 2022; Warju et al., 2020). Ketiga, terdapat artikel yang memperluas fungsi soal pilihan ganda sebagai instrumen diagnostik, terutama melalui penggunaan three-tier multiple choice untuk mengidentifikasi miskonsepsi siswa (Anggraeni et al., 2018; Wuryanti et al., 2017). Keempat, muncul pula kecenderungan penggunaan teknologi dalam evaluasi, seperti sistem ujian daring, Google Forms, perangkat lunak evaluasi soal, serta Retrieval-Augmented Generation dan Large Language Model dalam pembuatan soal otomatis (Utomo & Kustijono, 2015; Suryaningsih et al., 2024; Handoyo et al., 2025).

Dengan demikian, artikel-artikel yang dianalisis memperlihatkan bahwa penelitian mengenai soal pilihan ganda tidak lagi terbatas pada fungsi evaluatif konvensional. Fokus

kajian telah berkembang ke arah pengembangan soal berbasis HOTS, peningkatan kualitas psikometrik butir soal, penguatan kompetensi guru, pemanfaatan asesmen formatif dan sumatif, identifikasi miskonsepsi, serta integrasi teknologi digital dalam evaluasi pembelajaran. Pemetaan ini menunjukkan bahwa pengembangan soal pilihan ganda yang relevan dengan HOTS memerlukan keterpaduan antara desain pedagogis, validasi instrumen, analisis butir soal, kompetensi penyusun soal, dan dukungan teknologi asesmen. Rincian artikel yang menjadi dasar kajian ini disajikan pada Tabel 1 berikut.

Tabel 1. Daftar Artikel yang Dianalisis dalam Systematic Literature Review

No	Judul Artikel	Penulis	Tahun	Vol/Issue	Fokus Utama
1	Optimalisasi Google Forms untuk Meningkatkan Kemampuan HOTS Siswa SMA Melalui Soal Pilihan Ganda Beralasan	Yeni Suryaningsih, Mohamad Gilar Jatisunda, Abdur Rasyid	2024	SANISKALA, Vol. 2 No. 1	Google Forms sebagai media evaluasi untuk meningkatkan HOTS/berpikir kritis siswa SMA melalui soal pilihan ganda beralasan.
2	Pengembangan Soal-Soal Pilihan Ganda untuk Mengukur Kemampuan Berpikir Kritis Siswa pada Konsep Sistem Regulasi Manusia untuk Jenjang SMA	Ipin Aripin	2018	Mangifera Edu, Vol. 3 No. 1	Pengembangan soal pilihan ganda untuk mengukur kemampuan berpikir kritis pada konsep sistem regulasi manusia.
3	Keefektifan Tes Pilihan Ganda Tiga Tingkat dan Pilihan Ganda-Wawancara dalam Mengidentifikasi Miskonsepsi Asam Basa	Maya Erliza Anggraeni, Effendy, Munzil	2018	EduChemia, Vol. 3 No. 1	Perbandingan efektivitas three-tier multiple choice dan pilihan ganda-wawancara untuk mengidentifikasi miskonsepsi asam-basa.
4	Hubungan antara Validitas Item dengan Daya Pembeda dan Tingkat Kesukaran Soal Pilihan Ganda PAS	Sri Nurhalimah, Yunin Hidayati, Irsad Rosidi, Wiwin Puspita Hadi	2022	Jurnal Natural Science Educational Research, Vol. 4 No. 3	Hubungan validitas item, daya pembeda, dan tingkat kesukaran pada soal pilihan ganda PAS IPA SMP.
5	Penerapan Retrieval-Augmented Generation untuk Pembuatan Soal Pilihan Ganda dari Materi Berbasis PDF	K. S. Handoyo, M. L. Klau, R. J. Wijaya, R. P. Kristianto	2025	Volume X, Nomor X, 1–10.	Penerapan RAG dan LLM untuk membuat soal pilihan ganda otomatis dari dokumen pembelajaran berbasis PDF.

No	Judul Artikel	Penulis	Tahun	Vol/Issue	Fokus Utama
6	Peningkatan Kompetensi Pedagogik Guru dalam Menyusun Kisi-Kisi Soal Pilihan Ganda Melalui Supervisi Akademik dengan Metode In-House Training	Vitiarti	2022	DWIJA CENDEKIA, Vol. 6 No. 3	Peningkatan kompetensi pedagogik guru dalam menyusun kisi-kisi soal pilihan ganda melalui supervisi akademik dan in-house training.
7	Pengembangan Sistem Ujian Online Soal Pilihan Ganda dengan Menggunakan Software Wondershare Quiz Creator	Dwi Wiji Utomo, Rudy Kustijono	2015	JIPF, Vol. 04 No. 03	Pengembangan sistem ujian online pilihan ganda berbasis Wondershare Quiz Creator ditinjau dari validitas, kepraktisan, dan efektivitas.
8	Pengembangan Instrumen Soal HOTS untuk Mengukur Keterampilan Berpikir Kritis, Berpikir Kreatif, dan Pemecahan Masalah Materi Gerak Lurus pada Peserta Didik SMA	Lina Yuliantaningrum, Titin Sunarti	2020	IPF, Vol. 09 No. 02	Pengembangan instrumen HOTS materi gerak lurus untuk mengukur berpikir kritis, kreatif, dan pemecahan masalah.
9	Analisis Kualitas Butir Soal Tipe HOTS pada Kompetensi Sistem Rem Siswa di Sekolah Menengah Kejuruan	Warju, Sudirman Rizki Ariyanto, Soeryanto, Rio Adi Trisna	2020	Jurnal Pendidikan Teknologi dan Kejuruan, Vol. 17 No. 1	Analisis kualitas soal pilihan ganda tipe HOTS pada kompetensi sistem rem berdasarkan validitas, reliabilitas, kesukaran, daya pembeda, dan pengecoh.
10	Analisis Butir Soal Ujian Akhir Sekolah	Sitti Mania, Fitriani, Ahmad Farham Majid, Nidya Nina Ichiana, Andi Ika Prasasti Abrar	2020	Al asma: Journal of Islamic Education, Vol. 2 No. 2	Analisis kualitas butir soal UAS matematika menggunakan program Anates untuk menentukan kategori baik, cukup baik, dan tidak baik.
11	Analisis Miskonsepsi Siswa pada Materi Dinamika Gerak Menggunakan Tes Diagnostik Pilihan Ganda Tiga Tingkat	Sitri Wuryanti, Yennita, Fakhruddin	2017	Jurnal Geliga Sains, Vol. 5 No. 2	Identifikasi tingkat miskonsepsi siswa pada dinamika gerak menggunakan tes diagnostik three-tier multiple choice.

No	Judul Artikel	Penulis	Tahun	Vol/Issue	Fokus Utama
12	Cara Guru dalam Mengembangkan Butir Tes Pilihan Ganda sebagai Instrumen Evaluasi Hasil Belajar	Gusnita Efrina, Ayu Rahma Nengsi, Yona Syaida Oktira	2021	Volume X No. 1	Kajian proses guru dalam mengembangkan butir tes pilihan ganda, mulai dari kisi-kisi, penulisan butir, review, hingga kendala uji coba dan analisis.
13	Mampu Memilih Soal Berdasarkan Tingkat Kesukaran	Meyke Joice Tresia Rajagukguk, Dorlan Naibaho	2023	Pediaqu, Vol. 2 No. 4	Pembahasan kemampuan memilih atau menyusun soal berdasarkan tingkat kesukaran agar sesuai dengan kemampuan siswa.
14	Evaluasi dan Rekonstruksi Butir Soal Pilihan Ganda sebagai Asesmen Formatif dalam Pembelajaran Berbasis Deep Learning	Kejora Padmarani, Deby Luriawati, Wagiran	2026	JMIA, Vol. 3 No. 2	Evaluasi dan rekonstruksi butir soal pilihan ganda agar selaras dengan HOTS, asesmen formatif, dan pembelajaran deep learning.
15	Strategi Pengembangan Tes Objektif (Pilihan Ganda)	Wartoni, Priskila Issak Benyamin	2021	Volume X, Nomor X, 1–8.	Kajian strategi pengembangan tes objektif, meliputi analisis materi, penyusunan distraktor, review instrumen, uji coba, dan analisis butir soal.
16	Penggunaan Bentuk Soal Pilihan Ganda dalam Ujian	Wiwin Arbaini W	2020	Volume X, Nomor X, 1–11.	Pembahasan bentuk soal pilihan ganda, variasi, keunggulan, keterbatasan, dan analisis soal.

Berdasarkan Tabel 1 yang memuat daftar 16 artikel yang dianalisis, dapat dilihat bahwa penelitian mengenai soal pilihan ganda tidak hanya menekankan aspek evaluatif tradisional, tetapi juga mulai mengintegrasikan berbagai pendekatan yang mendukung pengembangan Higher-Order Thinking Skills (HOTS). Setiap artikel memberikan kontribusi spesifik terhadap pemahaman tentang kualitas butir soal, validitas dan reliabilitas instrumen, kompetensi guru, serta pemanfaatan teknologi digital dalam asesmen. Penelaahan literatur ini mengidentifikasi bahwa beberapa studi fokus pada rekonstruksi soal agar selaras dengan HOTS, sementara yang lain menekankan analisis statistik untuk menilai kesukaran, daya pembeda, dan efektivitas distraktor. Selain itu, terdapat penelitian yang menyoroti penggunaan PG tiga tingkat untuk identifikasi miskonsepsi, pengembangan ujian daring, serta pemanfaatan Retrieval-Augmented Generation (RAG) dan Large Language Model (LLM) dalam pembuatan soal otomatis.

Dengan latar belakang tersebut, klasifikasi tematik menjadi langkah berikutnya untuk mengelompokkan literatur berdasarkan fokus penelitian dan temuan utama. Hal ini bertujuan untuk mempermudah sintesis tematik dan analisis lintas studi, serta memberikan struktur yang sistematis bagi pembahasan lebih lanjut. File tabel klasifikasi tematik yang menyajikan lima tema utama dari 16 artikel tersebut akan ditempatkan setelah paragraf ini untuk memperjelas pengelompokan fokus penelitian, mulai dari pengembangan dan validasi

soal PG HOTS, analisis butir soal, strategi pengembangan guru, implementasi PG dalam asesmen formatif dan sumatif, hingga diagnostik dan identifikasi miskonsepsi siswa. Dengan demikian, pembaca dapat memahami hubungan antar tema dan pola temuan yang muncul di seluruh literatur sebelum memasuki analisis kritis dan sintesis tematik.

Tabel 2. Klasifikasi Tematik Artikel

Tema Utama	Sub-Tema / Fokus	Jurnal Terkait	Temuan Kunci
<b>1. Pengembangan dan Validasi Soal PG HOTS</b>	Pengembangan instrumen, validitas, reliabilitas, konstruksi soal, kesesuaian dengan HOTS, berpikir kritis, berpikir kreatif, dan pemecahan masalah.	Yuliantaningrum & Sunarti (2020); Warju et al. (2020); Aripin (2018); Handoyo et al. (2025).	Soal pilihan ganda dapat dikembangkan sebagai instrumen HOTS apabila disusun berdasarkan indikator berpikir tingkat tinggi, divalidasi secara teoritik maupun empiris, serta diuji kualitas butirnya. Tema ini mencakup pengembangan soal untuk berpikir kritis, kreatif, pemecahan masalah, analisis validitas-reliabilitas, dan otomasi pembuatan soal berbasis RAG/LLM.
<b>2. Analisis Butir Soal</b>	Validitas item, reliabilitas, indeks kesukaran, daya pembeda, efektivitas pengecoh, evaluasi kualitas soal, dan perangkat lunak analisis soal.	Mania et al. (2020); Sri Nurhalimah et al. (2022); Rajagukguk & Naibaho (2023); Wiwin Arbaini W.	Kualitas soal pilihan ganda sangat dipengaruhi oleh analisis butir. Artikel dalam tema ini menekankan pentingnya validitas item, daya pembeda, tingkat kesukaran, efektivitas pengecoh, penggunaan Anates/perangkat lunak evaluasi, serta pemilihan soal berdasarkan tingkat kesukaran agar sesuai dengan kemampuan siswa.
<b>3. Strategi Pengembangan dan Kompetensi Guru</b>	Kompetensi pedagogik guru, penyusunan kisi-kisi, penulisan butir, pengembangan distractor, review instrumen, uji coba, supervisi, dan pelatihan.	Vitiarti (2022); Wartoni & Benyamin; Gusnita et al. (2021).	Kualitas soal pilihan ganda sangat ditentukan oleh kompetensi guru atau pengembang instrumen. Guru membutuhkan panduan sistematis mulai dari analisis materi, penyusunan kisi-kisi, penulisan butir, pengembangan distractor, review instrumen, uji coba, sampai analisis butir. Supervisi dan pelatihan terbukti memperkuat kompetensi penyusun soal.
<b>4. Implementasi PG dalam Asesmen Formatif dan Sumatif</b>	Asesmen formatif, asesmen sumatif, ujian daring, Google Forms, evaluasi pembelajaran, digital assessment, dan deep learning.	Kejora Padmarani et al. (2026); Suryaningsih et al. (2024); Utomo & Kustijono (2015); Wiwin Arbaini W.; Rajagukguk & Naibaho (2023).	Soal pilihan ganda digunakan secara luas dalam asesmen formatif maupun sumatif. Artikel dalam tema ini menunjukkan bahwa PG dapat direkonstruksi menjadi asesmen formatif berbasis HOTS dan deep learning, digunakan dalam Google Forms dan sistem ujian online, serta diterapkan pada ujian sekolah, ujian nasional, dan seleksi masuk perguruan tinggi.
<b>5. Diagnostik dan Identifikasi</b>	Three-tier multiple choice, pilihan ganda-wawancara, tes	Wuryanti et al. (2017);	Soal pilihan ganda dapat diperluas menjadi instrumen diagnostik. Three-tier multiple choice dan pilihan ganda-

Tema Utama	Sub-Tema / Fokus	Jurnal Terkait	Temuan Kunci
<b>Miskonsepsi Siswa</b>	diagnostik, identifikasi miskonsepsi, dan pemahaman konseptual.	Anggraeni et al. (2018).	wawancara membantu guru mengidentifikasi miskonsepsi siswa, menelusuri kesalahan konsep, dan menyesuaikan strategi pembelajaran berdasarkan pemahaman konseptual siswa.

## 2. Klasifikasi Tematik dan Fokus Kajian Literatur

### a. Pengembangan dan Validasi Soal PG HOTS

Analisis literatur menunjukkan bahwa pengembangan soal pilihan ganda berbasis HOTS menekankan pentingnya validitas, reliabilitas, konstruksi soal, dan kesesuaian instrumen dengan indikator keterampilan berpikir tingkat tinggi. Analisis literatur menunjukkan bahwa pengembangan soal PG berbasis HOTS menekankan pentingnya validitas, reliabilitas, konstruksi soal, dan kesesuaian instrumen dengan indikator keterampilan berpikir tingkat tinggi (Maryani et al., 2021; Nugrahnastiti, 2024). Yuliantaningrum dan Sunarti (2020) menunjukkan bahwa instrumen soal HOTS dapat dikembangkan untuk mengukur keterampilan berpikir kritis, berpikir kreatif, dan pemecahan masalah. Temuan tersebut sejalan dengan Aripin (2018) yang mengembangkan soal pilihan ganda untuk mengukur kemampuan berpikir kritis siswa SMA pada konsep sistem regulasi manusia. Kedua penelitian ini memperlihatkan bahwa soal pilihan ganda tidak hanya berfungsi untuk mengukur penguasaan materi, tetapi juga dapat diarahkan untuk menilai kemampuan berpikir tingkat tinggi apabila dirancang berdasarkan indikator kognitif yang tepat.

Warju et al. (2020) memperkuat temuan tersebut melalui analisis kualitas butir soal tipe HOTS pada kompetensi sistem rem di SMK. Penelitian tersebut menunjukkan bahwa kelayakan soal HOTS perlu dilihat melalui validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Dengan demikian, pengembangan soal pilihan ganda berbasis HOTS tidak cukup hanya dilakukan melalui penyusunan pertanyaan yang tampak sulit, tetapi harus disertai pengujian kualitas butir secara sistematis.

Selain itu, Handoyo et al. (2025) memperluas arah pengembangan soal pilihan ganda melalui pemanfaatan Retrieval-Augmented Generation (RAG) dan Large Language Model (LLM) untuk menghasilkan soal otomatis dari dokumen pembelajaran berbasis PDF. Temuan ini menunjukkan bahwa pengembangan soal pilihan ganda mulai bergerak ke arah digital dan berbasis kecerdasan buatan. Namun, penggunaan teknologi tetap memerlukan kontrol pedagogis agar soal yang dihasilkan tetap sesuai dengan materi, tujuan pembelajaran, dan indikator HOTS. Dengan demikian, tema pengembangan dan validasi soal PG HOTS menegaskan bahwa kualitas instrumen sangat bergantung pada integrasi antara desain pedagogis, validasi empiris, dan pemanfaatan teknologi secara terarah.

### b. Analisis Butir Soal

Tema analisis butir soal menekankan bahwa kualitas soal pilihan ganda sangat dipengaruhi oleh validitas item, reliabilitas, indeks kesukaran, daya pembeda, dan efektivitas pengecoh. Tema analisis butir soal menekankan bahwa kualitas soal PG sangat dipengaruhi oleh validitas item, reliabilitas, indeks kesukaran, daya pembeda, dan efektivitas pengecoh (Rahmaniasan et al., 2022; Purwati et al., 2021). Mania et al. (2020) menunjukkan bahwa analisis butir soal Ujian Akhir Sekolah dapat digunakan untuk menentukan kategori soal yang baik, cukup baik, dan tidak baik. Temuan ini penting karena hasil analisis butir dapat menjadi dasar untuk mempertahankan, merevisi, atau mengganti soal dalam bank soal.

Sri Nurhalimah et al. (2022) menambahkan bahwa validitas item memiliki hubungan dengan daya pembeda soal, sedangkan tingkat kesukaran tidak selalu menunjukkan hubungan yang kuat dengan validitas item. Temuan ini memperlihatkan bahwa kualitas soal tidak dapat dinilai hanya dari sulit atau mudahnya soal, tetapi juga dari kemampuan butir soal dalam membedakan peserta didik yang telah menguasai materi dan yang belum menguasai materi. Oleh karena itu, analisis butir soal menjadi bagian penting dalam memastikan bahwa soal pilihan ganda benar-benar berfungsi sebagai alat ukur yang adil dan akurat.

Rajagukguk dan Naibaho (2023) menekankan pentingnya pemilihan soal berdasarkan tingkat kesukaran agar instrumen sesuai dengan kemampuan siswa. Sementara itu, Wiwin Arbaini W. menjelaskan bahwa soal pilihan ganda memiliki keunggulan dalam mencakup ruang lingkup materi yang luas, tetapi tetap memiliki keterbatasan, terutama dalam penyusunan pengecoh yang efektif dan kecenderungan soal hanya mengukur kemampuan kognitif rendah. Dengan demikian, analisis butir soal menjadi fondasi penting untuk meningkatkan kualitas soal PG, baik dalam asesmen formatif maupun sumatif.

#### c. Strategi Pengembangan dan Kompetensi Guru

Literatur yang termasuk dalam tema ini menunjukkan bahwa kualitas soal pilihan ganda sangat ditentukan oleh kompetensi guru dalam merancang instrumen evaluasi. Vitiarti (2022) menunjukkan bahwa supervisi akademik dengan metode in-house training mampu meningkatkan kompetensi pedagogik guru dalam menyusun kisi-kisi soal pilihan ganda. Hal ini mengindikasikan bahwa peningkatan kualitas soal tidak hanya berkaitan dengan instrumen, tetapi juga dengan kapasitas guru sebagai penyusun soal.

Wartoni dan Benyamin menjelaskan bahwa pengembangan tes objektif perlu dilakukan melalui beberapa langkah sistematis, yaitu analisis karakteristik materi, pengembangan distractor, pengaturan format jawaban, review instrumen, uji coba, analisis butir, dan perbandingan hasil tes dengan instrumen lain. Strategi tersebut menjadi penting karena soal pilihan ganda yang baik tidak dapat disusun secara spontan, melainkan memerlukan proses perencanaan, penelaahan, dan evaluasi yang berulang.

Gusnita et al. (2021) memperlihatkan bahwa dalam praktiknya guru masih menghadapi beberapa kendala dalam mengembangkan butir tes pilihan ganda. Kendala tersebut meliputi penggunaan kata kerja operasional yang masih terbatas pada aspek pengetahuan dan pemahaman, penggunaan bentuk soal negatif, pengecoh yang mudah ditebak, bahasa soal yang ambigu, serta belum dilakukannya uji coba dan analisis kualitas butir secara optimal. Temuan ini menunjukkan bahwa kompetensi guru menjadi faktor penting dalam keberhasilan pengembangan soal PG HOTS. Oleh karena itu, pelatihan, supervisi, review soal, dan pembiasaan analisis butir perlu menjadi bagian dari budaya evaluasi di sekolah.

#### d. Implementasi PG dalam Asesmen Formatif dan Sumatif

Tema implementasi PG dalam asesmen formatif dan sumatif menunjukkan bahwa soal pilihan ganda masih digunakan secara luas dalam berbagai bentuk evaluasi pembelajaran. Kejora Padmarani et al. (2026) menekankan bahwa soal pilihan ganda dapat direkonstruksi menjadi instrumen asesmen formatif yang lebih bermakna apabila diselaraskan dengan HOTS dan prinsip deep learning. Penelitian tersebut memperlihatkan bahwa asesmen formatif tidak hanya berfungsi untuk mengukur capaian belajar, tetapi juga dapat digunakan untuk mendorong pemahaman mendalam dan kemampuan berpikir kritis siswa.

Suryaningsih et al. (2024) menunjukkan bahwa penggunaan Google Forms dalam evaluasi pembelajaran dapat meningkatkan kemampuan berpikir kritis siswa SMA. Teknologi digital dalam konteks ini tidak hanya mempermudah proses penilaian, tetapi juga

memungkinkan guru memberikan umpan balik yang lebih cepat. Pemanfaatan teknologi digital, seperti Quizizz, menunjukkan bahwa instrumen PG HOTS dapat dikembangkan lebih efisien dan adaptif terhadap kebutuhan pembelajaran (Amaranto et al., 2022). Utomo dan Kustijono (2015) juga menunjukkan bahwa sistem ujian online berbasis Wondershare Quiz Creator layak digunakan karena memenuhi aspek validitas, kepraktisan, dan efektivitas. Kedua penelitian tersebut memperlihatkan bahwa digitalisasi asesmen dapat memperkuat implementasi soal pilihan ganda, baik dalam konteks formatif maupun sumatif.

Wiwin Arbaini W. menjelaskan bahwa soal pilihan ganda banyak digunakan dalam ujian berskala besar karena mampu mencakup materi yang luas, mudah dikoreksi, dan dapat digunakan secara objektif. Rajagukguk dan Naibaho (2023) juga menegaskan bahwa soal pilihan ganda digunakan dalam tes formatif, sumatif, ujian sekolah, ujian nasional, dan seleksi masuk perguruan tinggi. Namun, penggunaan yang luas tersebut perlu diimbangi dengan kualitas konstruksi soal agar tidak hanya mengukur hafalan, tetapi juga mampu menilai kemampuan analisis, evaluasi, dan pemecahan masalah. Dengan demikian, implementasi soal PG dalam asesmen formatif dan sumatif memerlukan keseimbangan antara efisiensi, kualitas instrumen, dan kesesuaian dengan tujuan pembelajaran.

#### e. Diagnostik dan Identifikasi Miskonsepsi Siswa

Tema diagnostik dan identifikasi miskonsepsi siswa menunjukkan bahwa soal pilihan ganda dapat dikembangkan lebih lanjut sebagai instrumen diagnostik. Wuryanti et al. (2017) menggunakan tes diagnostik pilihan ganda tiga tingkat untuk mengidentifikasi miskonsepsi siswa pada materi dinamika gerak. Melalui format tiga tingkat, guru dapat mengetahui tidak hanya jawaban siswa, tetapi juga alasan dan tingkat keyakinan siswa terhadap jawabannya. Hal ini membuat instrumen pilihan ganda menjadi lebih informatif dibandingkan bentuk pilihan ganda biasa.

Anggraeni et al. (2018) membandingkan keefektifan tes pilihan ganda tiga tingkat dan pilihan ganda yang diikuti wawancara dalam mengidentifikasi miskonsepsi asam-basa. Penelitian tersebut menunjukkan bahwa instrumen pilihan ganda dapat digunakan untuk mendeteksi pemahaman konseptual siswa, meskipun dalam beberapa kasus wawancara dapat memberikan informasi yang lebih mendalam. Temuan ini menunjukkan bahwa pilihan ganda tiga tingkat dapat menjadi alternatif praktis untuk membantu guru mengenali kesalahan konsep siswa secara lebih sistematis.

Secara keseluruhan, tema ini memperlihatkan bahwa soal pilihan ganda tidak hanya berfungsi sebagai alat ukur hasil belajar, tetapi juga dapat digunakan sebagai alat diagnostik untuk memperbaiki proses pembelajaran. Dengan mengetahui pola miskonsepsi siswa, guru dapat merancang intervensi pembelajaran yang lebih tepat, menyesuaikan strategi pengajaran, dan memperbaiki instrumen evaluasi berikutnya. Oleh karena itu, pengembangan soal PG berbasis diagnostik menjadi salah satu arah penting dalam penguatan asesmen berbasis HOTS.

### **3. Sintesis Integratif**

Berdasarkan analisis tematik terhadap 16 artikel yang dikaji, dapat disimpulkan bahwa pengembangan soal pilihan ganda yang mendukung Higher-Order Thinking Skills (HOTS) membutuhkan keterpaduan antara kualitas konstruksi soal, validitas dan reliabilitas instrumen, analisis butir, kompetensi guru, serta pemanfaatan teknologi evaluasi. Instrumen yang valid dan reliabel dapat membantu menstimulasi kemampuan berpikir analitis, evaluatif, kreatif, dan pemecahan masalah. “Temuan dari literatur terbaru menunjukkan bahwa pengembangan PG HOTS harus memperhatikan kualitas konstruksi, validitas, reliabilitas, analisis butir, kompetensi guru, serta integrasi teknologi digital untuk mendukung asesmen formatif dan sumatif (Maryani et al., 2021; Nugrahnastiti, 2024; Rahmaniasan et al., 2022; Purwati et al., 2021; Amaranto et al., 2022). Temuan

Yuliantaningrum dan Sunarti (2020), Aripin (2018), serta Warju et al. (2020) memperlihatkan bahwa soal pilihan ganda dapat diarahkan untuk mengukur kemampuan berpikir tingkat tinggi apabila disusun berdasarkan indikator kognitif yang tepat dan diuji melalui prosedur validasi yang memadai. Dengan demikian, kualitas soal PG HOTS tidak hanya ditentukan oleh tingkat kesulitan soal, tetapi juga oleh kesesuaian antara materi, indikator, konstruksi butir, dan tujuan asesmen.

Sintesis lintas tema juga menunjukkan bahwa analisis butir soal menjadi fondasi penting dalam pengembangan instrumen pilihan ganda. Mania et al. (2020) dan Sri Nurhalimah et al. (2022) menunjukkan bahwa validitas item, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh perlu dianalisis agar soal dapat berfungsi sebagai alat ukur yang akurat. Rajagukguk dan Naibaho (2023) memperkuat hal tersebut dengan menekankan pentingnya pemilihan soal berdasarkan tingkat kesukaran agar sesuai dengan kemampuan siswa. Sementara itu, Wiwin Arbaini W. menegaskan bahwa soal pilihan ganda memiliki keunggulan dalam mencakup materi yang luas dan mudah dikoreksi, tetapi tetap memiliki kelemahan apabila pengecoh tidak berfungsi dan soal hanya mengukur kemampuan kognitif rendah.

Pengembangan soal PG HOTS juga tidak dapat dipisahkan dari kompetensi guru. Guru perlu memiliki kemampuan menyusun kisi-kisi, memilih kata kerja operasional yang sesuai, membuat distractor yang berfungsi, melakukan review soal, serta menganalisis kualitas butir. Vitiarti (2022) menunjukkan bahwa supervisi akademik dan in-house training dapat meningkatkan kompetensi pedagogik guru dalam menyusun kisi-kisi soal pilihan ganda. Wartoni dan Benyamin juga menekankan perlunya strategi sistematis dalam pengembangan tes objektif, mulai dari analisis materi, penyusunan pengecoh, review instrumen, uji coba, hingga analisis butir. Temuan Gusnita et al. (2021) memperlihatkan bahwa guru masih menghadapi kendala dalam penulisan soal, seperti penggunaan kata kerja operasional yang rendah, pengecoh yang mudah ditebak, dan belum optimalnya uji coba butir soal. Hal ini menunjukkan bahwa penguatan kompetensi guru merupakan prasyarat penting dalam pengembangan soal PG berbasis HOTS.

Selain itu, pemanfaatan teknologi seperti Google Forms, sistem ujian daring, serta RAG dan LLM menunjukkan bahwa digitalisasi asesmen mulai menjadi arah penting dalam pengembangan soal pilihan ganda. Suryaningsih et al. (2024) menunjukkan bahwa Google Forms dapat digunakan sebagai media evaluasi untuk meningkatkan kemampuan berpikir kritis siswa. Utomo dan Kustijono (2015) memperlihatkan bahwa sistem ujian daring pilihan ganda dapat dikembangkan secara layak dari aspek validitas, kepraktisan, dan efektivitas. Handoyo et al. (2025) memperluas arah pengembangan tersebut melalui pemanfaatan Retrieval-Augmented Generation dan Large Language Model untuk menghasilkan soal pilihan ganda otomatis berbasis dokumen. Akan tetapi, teknologi tidak dapat menggantikan peran pedagogis guru, karena kualitas soal tetap memerlukan kesesuaian dengan tujuan pembelajaran, indikator HOTS, karakteristik materi, dan kebutuhan siswa.

Temuan dari kajian ini juga memperlihatkan bahwa soal pilihan ganda dapat diperluas fungsinya sebagai instrumen diagnostik, terutama melalui model three-tier multiple choice. Wuryanti et al. (2017) dan Anggraeni et al. (2018) menunjukkan bahwa format pilihan ganda bertingkat dapat membantu guru mengidentifikasi miskonsepsi siswa secara lebih mendalam. Dengan model ini, guru tidak hanya mengetahui jawaban benar atau salah, tetapi juga dapat menelusuri alasan dan tingkat keyakinan siswa. Oleh karena itu, pengembangan soal pilihan ganda berbasis HOTS perlu dilakukan melalui pendekatan holistik yang menggabungkan desain pedagogis, kualitas psikometrik, kompetensi guru, analisis butir,

pemanfaatan teknologi, serta fungsi diagnostik.

Tabel 3. Ringkasan Sintesis Integratif

Aspek Sintesis	Artikel Pendukung	Temuan Integratif	Implikasi terhadap PG HOTS
<b>Pengembangan dan validasi soal PG HOTS</b>	Yuliantaningrum & Sunarti (2020); Aripin (2018); Warju et al. (2020)	Soal pilihan ganda dapat diarahkan untuk mengukur berpikir kritis, kreatif, analitis, evaluatif, dan pemecahan masalah apabila disusun berdasarkan indikator kognitif yang tepat serta diuji melalui validasi yang memadai.	Pengembangan PG HOTS harus memperhatikan kesesuaian materi, indikator, konstruksi butir, tujuan asesmen, validitas, dan reliabilitas instrumen.
<b>Analisis butir soal</b>	Mania et al. (2020); Sri Nurhalimah et al. (2022); Rajagukguk & Naibaho (2023); Wiwin Arbaini W.	Validitas item, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh perlu dianalisis agar soal berfungsi sebagai alat ukur yang akurat. Soal PG juga memiliki keunggulan mencakup materi luas, tetapi lemah apabila pengecoh tidak berfungsi dan hanya mengukur kognitif rendah.	Analisis butir menjadi dasar untuk mempertahankan, merevisi, atau mengganti soal sehingga instrumen lebih adil, akurat, dan sesuai kemampuan siswa.
<b>Kompetensi guru dan strategi pengembangan soal</b>	Vitiarti (2022); Wartoni & Benyamin; Gusnita et al. (2021)	Pengembangan soal PG HOTS membutuhkan kemampuan guru dalam menyusun kisi-kisi, memilih kata kerja operasional, membuat distractor, mereview soal, melakukan uji coba, dan menganalisis kualitas butir.	Pelatihan, supervisi akademik, in-house training, dan pembiasaan review soal perlu diperkuat agar guru mampu menghasilkan soal PG HOTS yang berkualitas.
<b>Digitalisasi asesmen dan teknologi evaluasi</b>	Suryaningsih et al. (2024); Utomo & Kustijono (2015); Handoyo et al. (2025)	Google Forms, sistem ujian daring, serta RAG dan LLM menunjukkan bahwa teknologi dapat meningkatkan efisiensi evaluasi, mempercepat analisis, dan mendukung pembuatan soal otomatis berbasis dokumen.	Teknologi mendukung asesmen PG HOTS, tetapi tetap memerlukan kontrol pedagogis agar soal sesuai tujuan pembelajaran, indikator HOTS, karakteristik materi, dan kebutuhan siswa.
<b>Fungsi diagnostik soal pilihan ganda</b>	Wuryanti et al. (2017); Anggraeni et al. (2018)	Soal pilihan ganda dapat diperluas sebagai instrumen diagnostik melalui model three-tier multiple choice untuk mengidentifikasi miskonsepsi, alasan jawaban, dan tingkat keyakinan siswa.	PG diagnostik membantu guru memahami kesalahan konsep siswa, memperbaiki strategi pembelajaran, dan merancang intervensi yang lebih tepat.

#### **4. Keterbatasan Penelitian**

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, kajian ini hanya mencakup 16 artikel yang memenuhi kriteria inklusi dan eksklusi, sehingga generalisasi temuan tetap perlu dilakukan secara hati-hati. Keterbatasan jumlah artikel tersebut membuat hasil kajian lebih tepat dipahami sebagai sintesis terhadap literatur terpilih, bukan sebagai representasi menyeluruh dari seluruh penelitian mengenai soal pilihan ganda, HOTS, dan asesmen pendidikan.

Kedua, artikel yang dianalisis berasal dari jenjang pendidikan dan konteks mata pelajaran yang beragam, mulai dari pendidikan dasar, menengah, hingga konteks pengembangan instrumen oleh guru. Keragaman ini memperkaya cakupan kajian, tetapi juga menuntut kehati-hatian dalam menafsirkan temuan lintas jenjang dan lintas bidang studi. Dengan kata lain, temuan pada satu konteks pembelajaran belum tentu dapat diterapkan secara langsung pada konteks lain tanpa penyesuaian terhadap karakteristik peserta didik, mata pelajaran, dan tujuan asesmen.

Ketiga, sintesis yang dilakukan dalam penelitian ini masih berbasis pada kajian literatur, sehingga belum melibatkan data lapangan secara langsung, seperti observasi kelas, wawancara guru, atau uji coba instrumen kepada peserta didik. Oleh karena itu, beberapa temuan masih bersifat konseptual dan interpretatif, sehingga memerlukan pengujian lebih lanjut melalui penelitian empiris.

Keempat, meskipun beberapa artikel telah membahas pemanfaatan teknologi digital dan kecerdasan buatan dalam pengembangan soal pilihan ganda, jumlah literatur yang secara khusus menguji efektivitas implementasinya dalam pembelajaran masih terbatas. Hal ini menunjukkan bahwa penelitian lanjutan perlu memperluas cakupan kajian, menambahkan data empiris, serta menguji penerapan instrumen PG berbasis HOTS dalam berbagai konteks pendidikan, khususnya pada evaluasi nasional, asesmen digital, dan pembelajaran berbasis deep learning.

#### **5. Diskusi**

Interpretasi hasil menunjukkan bahwa literatur yang dianalisis secara umum mendukung teori Higher-Order Thinking Skills (HOTS) yang menekankan keterampilan berpikir kritis, kreatif, analitis, evaluatif, dan pemecahan masalah. Studi Yuliantaningrum dan Sunarti (2020) serta Aripin (2018) menunjukkan bahwa soal pilihan ganda dapat dikembangkan untuk mengukur kemampuan berpikir kritis, berpikir kreatif, dan pemecahan masalah apabila dirancang berdasarkan indikator HOTS. Warju et al. (2020) juga memperlihatkan bahwa soal PG tipe HOTS yang dianalisis secara empiris dapat digunakan untuk menilai kemampuan siswa secara lebih terukur. Temuan ini memperkuat pandangan bahwa soal pilihan ganda tidak selalu identik dengan pengukuran hafalan, tetapi dapat diarahkan untuk mengukur keterampilan kognitif tingkat tinggi apabila konstruksinya dirancang secara tepat. Hasil sintesis literatur menunjukkan bahwa efektivitas soal pilihan ganda dalam mendukung pengembangan HOTS tidak hanya ditentukan oleh bentuk instrumen yang digunakan, tetapi oleh kualitas perancangan, validitas konstruk, reliabilitas instrumen, kompetensi guru, serta konteks implementasi asesmen. Temuan dari berbagai penelitian memperlihatkan bahwa soal pilihan ganda dapat berfungsi sebagai instrumen untuk mengukur kemampuan berpikir kritis, kreatif, analitis, evaluatif, dan pemecahan masalah apabila dirancang berdasarkan indikator kognitif tingkat tinggi dan melalui proses validasi yang memadai. Dengan demikian, anggapan bahwa soal pilihan ganda hanya mampu mengukur kemampuan berpikir tingkat rendah perlu ditinjau kembali.

Namun, kajian ini juga menemukan bahwa penggunaan soal pilihan ganda masih menghadapi sejumlah kelemahan. Kejora Padmarani et al. (2026) menunjukkan bahwa soal pilihan ganda yang digunakan dalam asesmen formatif masih sering didominasi oleh level

kognitif rendah dan belum sepenuhnya mendukung pembelajaran berbasis deep learning. Temuan ini sejalan dengan Gusnita et al. (2021) yang menunjukkan bahwa guru masih sering menyusun soal dengan kata kerja operasional yang terbatas pada aspek pengetahuan dan pemahaman, pengecoh yang mudah ditebak, serta bahasa soal yang ambigu. Dengan demikian, tantangan utama bukan terletak pada bentuk pilihan gandanya, melainkan pada kualitas desain, validasi, dan analisis butir soal. Meskipun sebagian besar penelitian menunjukkan potensi soal pilihan ganda dalam mendukung HOTS, hasil sintesis juga mengindikasikan adanya variasi kualitas implementasi di lapangan. Beberapa penelitian menunjukkan keberhasilan pengembangan soal HOTS, sedangkan penelitian lain menemukan dominasi soal pada level kognitif rendah. Perbedaan temuan ini mengindikasikan bahwa keberhasilan implementasi lebih ditentukan oleh kualitas konstruksi soal dan kompetensi guru dibandingkan oleh bentuk soal itu sendiri.

Dampak kualitas soal terhadap pencapaian kompetensi abad ke-21 terlihat dari hubungan antara konstruksi soal dan kemampuan berpikir siswa. Butir soal yang valid, reliabel, memiliki tingkat kesukaran yang proporsional, daya pembeda yang baik, serta pengecoh yang berfungsi dapat mendukung pengukuran berpikir kritis, kreatif, dan pemecahan masalah. Mania et al. (2020), Sri Nurhalimah et al. (2022), serta Rajaguguk dan Naibaho (2023) menunjukkan bahwa analisis butir menjadi langkah penting untuk memastikan soal mampu membedakan kemampuan siswa secara akurat. Tanpa analisis butir, soal pilihan ganda berisiko digunakan secara administratif semata, tetapi tidak memberikan informasi yang bermakna bagi perbaikan pembelajaran. Temuan ini menunjukkan bahwa analisis butir soal tidak hanya berfungsi sebagai prosedur teknis evaluasi, tetapi juga sebagai mekanisme pengendalian mutu instrumen asesmen. Instrumen yang tidak melalui proses analisis empiris berpotensi menghasilkan informasi yang kurang akurat mengenai kemampuan peserta didik dan mengurangi fungsi asesmen sebagai dasar pengambilan keputusan pembelajaran.

Perbandingan dengan penelitian yang berfokus pada implementasi teknologi menunjukkan bahwa digitalisasi asesmen dapat memperkuat efektivitas soal pilihan ganda. Suryaningsih et al. (2024) menunjukkan bahwa Google Forms dapat digunakan untuk meningkatkan kemampuan berpikir kritis siswa melalui evaluasi berbasis digital. Utomo dan Kustijono (2015) menunjukkan bahwa sistem ujian online pilihan ganda dapat meningkatkan kepraktisan dan efektivitas evaluasi. Handoyo et al. (2025) memperlihatkan bahwa teknologi berbasis RAG dan LLM dapat membantu menghasilkan soal pilihan ganda secara otomatis dari materi berbasis PDF. Meskipun demikian, pemanfaatan teknologi tetap memerlukan kontrol kualitas agar soal yang dihasilkan tidak hanya relevan secara konten, tetapi juga memenuhi indikator HOTS dan kaidah penyusunan soal yang baik. Meskipun teknologi digital menawarkan berbagai kemudahan dalam penyusunan dan pengelolaan asesmen, penggunaannya juga memerlukan perhatian terhadap aspek validitas dan kualitas pedagogis. Pemanfaatan teknologi berbasis kecerdasan buatan, seperti RAG dan LLM, berpotensi menghasilkan soal secara cepat, tetapi tetap memerlukan proses verifikasi oleh guru untuk memastikan kesesuaian dengan indikator HOTS, karakteristik peserta didik, dan tujuan pembelajaran. Oleh karena itu, teknologi perlu diposisikan sebagai alat pendukung profesionalisme guru, bukan sebagai pengganti pertimbangan pedagogis.

Dimensi diagnostik juga menjadi temuan penting dalam kajian ini. Wuryanti et al. (2017) dan Anggraeni et al. (2018) menunjukkan bahwa pilihan ganda tiga tingkat dapat digunakan untuk mengidentifikasi miskonsepsi siswa. Temuan ini memperluas fungsi soal pilihan ganda dari sekadar alat ukur hasil belajar menjadi instrumen diagnostik yang dapat membantu guru memahami kesalahan konsep siswa. Dengan demikian, soal pilihan ganda dapat menjadi bagian dari asesmen formatif yang lebih bermakna apabila hasilnya

digunakan untuk memberikan umpan balik dan memperbaiki strategi pembelajaran. Berdasarkan keseluruhan temuan, penelitian ini mengidentifikasi adanya kesenjangan penelitian yang masih perlu dieksplorasi lebih lanjut. Sebagian besar penelitian yang dianalisis berfokus pada pengembangan instrumen dan analisis kualitas butir, sedangkan penelitian mengenai efektivitas jangka panjang soal pilihan ganda berbasis HOTS terhadap pengembangan kompetensi abad ke-21 masih relatif terbatas. Selain itu, kajian empiris mengenai integrasi kecerdasan buatan dalam pengembangan soal HOTS juga masih memerlukan pengujian yang lebih mendalam.

Implikasi pedagogis dari temuan ini menunjukkan bahwa guru dan pembuat soal perlu memperhatikan kualitas butir, validitas, reliabilitas, distribusi kesukaran, daya pembeda, serta kualitas distractor. Strategi desain soal perlu diselaraskan dengan taksonomi Bloom revisi, terutama pada level C3 sampai C6, menggunakan stimulus kontekstual, dan menghindari soal yang hanya menuntut ingatan faktual. Selain itu, guru perlu mendapatkan pelatihan berkelanjutan dalam penyusunan kisi-kisi, penulisan butir HOTS, penggunaan teknologi asesmen, serta analisis hasil tes. Dengan cara tersebut, soal pilihan ganda dapat digunakan secara lebih optimal sebagai instrumen asesmen formatif dan sumatif yang mendukung pengembangan HOTS. Secara konseptual, hasil sintesis menunjukkan bahwa pengembangan soal pilihan ganda berbasis HOTS merupakan proses yang melibatkan keterkaitan antara kompetensi guru, perancangan kisi-kisi, pengembangan instrumen, validasi dan analisis butir, implementasi asesmen berbasis teknologi, serta pengembangan keterampilan berpikir tingkat tinggi peserta didik. Dengan demikian, keberhasilan asesmen HOTS tidak bergantung pada satu komponen tertentu, tetapi merupakan hasil integrasi berbagai aspek pedagogis, psikometrik, dan teknologi yang saling mendukung.

## **KESIMPULAN**

Berdasarkan sintesis 16 artikel, soal pilihan ganda memiliki potensi untuk mendukung pengembangan Higher-Order Thinking Skills (HOTS) apabila dikembangkan dengan memperhatikan validitas, reliabilitas, kualitas konstruksi butir, efektivitas pengecoh, tingkat kesukaran, daya pembeda, kompetensi guru, serta integrasi teknologi digital. Soal pilihan ganda yang dirancang secara sistematis tidak hanya mampu mengukur pengetahuan faktual, tetapi juga keterampilan analisis, evaluasi, berpikir kritis, berpikir kreatif, dan pemecahan masalah. Selain itu, model three-tier multiple choice dapat dimanfaatkan sebagai instrumen diagnostik untuk mengidentifikasi miskonsepsi siswa. Temuan ini menunjukkan bahwa efektivitas soal pilihan ganda berbasis HOTS ditentukan oleh keterpaduan aspek pedagogis, psikometrik, dan teknologi, bukan oleh bentuk soal semata.

Kontribusi utama penelitian ini adalah menghasilkan sintesis integratif yang memperlihatkan hubungan antara pengembangan instrumen, analisis butir soal, kompetensi guru, digitalisasi asesmen, dan fungsi diagnostik soal pilihan ganda dalam mendukung HOTS. Hasil kajian menunjukkan bahwa pengembangan asesmen HOTS merupakan proses yang melibatkan perencanaan indikator pembelajaran, penyusunan butir soal, validasi instrumen, analisis kualitas butir, pemanfaatan teknologi asesmen, serta evaluasi dan perbaikan berkelanjutan. Dengan demikian, kualitas asesmen HOTS dapat dipahami sebagai hasil integrasi berbagai komponen yang saling mendukung dalam menghasilkan pengukuran kemampuan berpikir tingkat tinggi yang lebih akurat dan bermakna.

Secara praktis, guru dan pengembang instrumen perlu merancang soal berbasis HOTS dengan mengacu pada indikator kognitif tingkat tinggi, menggunakan konteks yang relevan,

serta melakukan telaah, uji coba, dan analisis butir sebelum instrumen digunakan dalam asesmen formatif maupun sumatif. Pemanfaatan teknologi digital dan kecerdasan buatan dapat meningkatkan efisiensi dan fleksibilitas evaluasi, tetapi tetap memerlukan validasi pedagogis agar sesuai dengan tujuan pembelajaran. Penelitian selanjutnya perlu menguji efektivitas jangka panjang soal pilihan ganda berbasis HOTS, mengembangkan asesmen digital dan adaptif, serta mengevaluasi validitas instrumen yang dihasilkan melalui teknologi kecerdasan buatan pada berbagai jenjang pendidikan dan bidang studi.

## DAFTAR PUSTAKA

- Amaranto, et al. (2022). Pengembangan instrumen soal HOTS berbasis Quizizz dengan tema hipertensi. *Jurnal Eksakta Pendidikan*.
- Anggraeni, M. E., Effendy, & Munzil. (2018). Keefektifan three-tier MCQ untuk identifikasi miskonsepsi asam basa. *EduChemia*, 5(2), 110–118.
- Aripin, I. (2018). Pengembangan soal-soal pilihan ganda untuk mengukur kemampuan berpikir kritis siswa pada konsep sistem regulasi manusia untuk jenjang SMA. *Mangifera Edu*, 3(1), 26–39.
- Gusnita, E., Rahma, N., & Oktira, Y. S. (2021). Cara guru dalam mengembangkan butir tes pilihan ganda sebagai instrumen evaluasi hasil belajar. *Academia Edu*, 10(1), X, 1–10.
- Handoyo, K. S., Klau, M. L., Wijaya, R. J., & Kristianto, R. P. (2025b). Penerapan retrieval-augmented generation untuk pembuatan soal pilihan ganda dari materi berbasis PDF. *Jurnal Universitas Atma Jaya Yogyakarta*, Volume X, 10(10), 1–10.
- Kejora, P., Luriawati, D., & Wagiran. (2026). Rekonstruksi butir soal pilihan ganda berbasis deep learning. *Jurnal Multidisiplin Ilmu Akademik*, 3(2), 541–550.
- Mania, S., Fitriani, Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis butir soal ujian akhir sekolah. *Al Asma: Journal of Islamic Education*, 2(2), 274–284.
- Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitriyanawati, M. (2021). HOTS multiple choice and essay questions: A validated instrument to measure higher-order thinking skills of prospective teachers. *TUSED Educational Sciences Journal*, 2(1).
- Nugrahnastiti, N. (2024). Development of a higher-order thinking skills (HOTS) test instrument on electrochemistry. *Jurnal Ilmu Kependidikan dan Pembelajaran*.
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan validitas item dengan daya pembeda dan tingkat kesukaran soal PG PAS. *Jurnal Natural Science Educational Research*, 4(3), 249–258.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
- Purwati, et al. (2021). Analysis of multiple choice questions based on HOTS in geography final semester exam. *Jurnal Evaluasi Pendidikan*.
- Rahmaniasan, V., Berlian, L., & Suryani, D. I. (2022). Pengembangan instrumen tes two-tier multiple choice berbasis HOTS tema pemanfaatan gelombang. *Jurnal Pendidikan Matematika dan IPA*.
- Rajagukguk MJT, Naibaho D. Mampu memilih soal berdasarkan tingkat kesukaran. *Jurnal Pendidikan Sosial dan Humaniora*. 2023;2(4):12736–12747.
- Snyder H. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*. 2019;104:333–339.
- Suryaningsih, Y., Jatisunda, M. G., & Rasyid, A. (2024). Optimalisasi Google Forms untuk asesmen HOTS siswa SMA. *SANISKALA: Jurnal Pengabdian Masyarakat*, 2(1), 15–19.
- Utomo, D. W., & Kustijono, R. (2015). Pengembangan sistem ujian online soal pilihan ganda. *Jurnal Inovasi Pendidikan Fisika*, 4(3), 1–6.
- Vitiarti. (2022). Peningkatan kompetensi pedagogik guru dalam menyusun soal pilihan ganda. *Dwija Cendekia: Jurnal Riset Pedagogik*, 6(3), 613–620.
- Warju, Ariyanto, S. R., Soeryanto, & Trisna, R. A. (2020). Analisis kualitas butir soal tipe HOTS kompetensi sistem rem. *Jurnal Pendidikan Teknologi dan Kejuruan*, 17(1), 95–102.

- Wartoni, & Benyamin, P. I. (2021). Strategi pengembangan tes objektif pilihan ganda. *Jurnal Teologi*, 10(10), 1–8.
- Wiwin Arbaini, W. (2020). Penggunaan bentuk soal pilihan ganda dalam ujian. *Jurnal Widyasari Press*, 10(10), 1–11.
- Wuryanti, S., Yennita, & Fakhruddin. (2017). Analisis miskonsepsi siswa pada materi dinamika gerak menggunakan tes diagnostik pilihan ganda tiga tingkat. *Jurnal Geliga Sains*, 5(2), 110–118.
- Yuliantaningrum, L., & Sunarti, T. (2020). Pengembangan instrumen PG HOTS materi gerak lurus. *Inovasi Pendidikan Fisika*, 9(2), 76–82.