



Analisis Psikometrik dan Keselarasan Instruksional pada Butir Penilaian Sumatif dalam Pembelajaran Bahasa Indonesia

Kejora Padmarani¹, Wagiran²

^{1,2} Universitas Negeri Semarang, Indonesia

Corresponding Author:  kejorapadmarani5@students.unnes.ac.id

ABSTRACT

Kualitas asesmen merupakan komponen fundamental dalam keberhasilan implementasi Kurikulum Merdeka, terutama pada satuan pendidikan madrasah yang menghadapi tantangan kompleks dalam memastikan keselarasan antara pembelajaran dan penilaian. Penelitian ini bertujuan menganalisis kualitas psikometri instrument asesmen sumatif Bahasa Indonesia Fase F sekaligus mengkaji kesesuaian pembelajaran yang melandasinya. Menggunakan pendekatan *mixed method* dengan desain sequential explanatory, penelitian ini melibatkan 100 peserta didik kelas XI serta seorang guru mata pelajaran Bahasa Indonesia sebagai informan kualitatif. Data kuantitatif diperoleh melalui dokumentasi naskah soal dan hasil ujian, kemudian dianalisis berdasarkan validitas, reliabilitas, daya pembeda, dan tingkat kesukaran. Data kualitatif dikumpulkan melalui wawancara dan observasi untuk memberikan penjelasan mendalam terhadap temuan statistic. Hasil penelitian menunjukkan bahwa revisi instrument memberikan peningkatan signifikan pada kualitas tes, terutama pada aspek validitas pilihan ganda yang meningkat 72% menjadi 92% dan reliabilitas yang meningkat dari 0,486 menjadi 0,630. Pada soal esai, seluruh butir soal menunjukkan validitas sangat tinggi (0,793-0,911) dan reliabilitas unggul (0,923). Meskipun demikian, daya pembeda sebagian butir soal pilihan ganda tetap rendah dan mayoritas butir soal pindah ke kategori mudah yang menandakan berkurangnya daya seleksi instrumen. Temuan kualitatif menunjukkan bahwa penjelasan belum memberikan pengalaman membaca dan analisis teks kompleks secara memadai sehingga tidak sejalan dengan tuntutan Capaian Pembelajaran Fase F. Penelitian ini menegaskan urgensi perbaikan instrument asesmen sekaligus penguatan pembelajaran agar penilaian mampu menggambarkan kemampuan literasi kritis peserta didik secara lebih akurat, valid, dan selaras dengan prinsip Kurikulum Merdeka.

ARTICLE INFO

Article history:

Received

15 December 2025

Revised

01 January 2025

Accepted

17 January 2025

Key Word

Analisis Psikometrik, Asesmen Sumatif, Mixed Method, Evaluasi Pembelajaran.

How to cite

<https://pusdikra-publishing.com/index.php/josr>



This work is licensed under a

[Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

PENDAHULUAN

Upaya peningkatan kualitas sumber daya manusia bergantung pada keberhasilan penyelenggaraan Pendidikan, karena Pendidikan merupakan

fondasi utama dalam membentuk kompetensi individu dan daya saing bangsa (Maulansyah et al., 2023; Yasin, 2021). Mutu Pendidikan yang baik diyakini mampu melahirkan generasi yang berkualitas, sebagaimana ditegaskan pula oleh Mardhiyah et al., (2021) bahwa kualitas lulusan merupakan cerminan langsung dari kualitas proses pendidikannya. Agar mutu tersebut tercapai, proses pembelajaran di kelas harus berlangsung secara efektif serta mampu menumbuhkan kemampuan kognitif, efektif, dan psikomotor peserta didik Haris & Mustari, (2023).

Tujuan Pendidikan adalah untuk mengembangkan kemampuan peserta didik secara menyeluruh, meliputi aspek pengetahuan, keterampilan, dan karakter sesuai dengan kebutuhan perkembangan abad ke-21. Pencapaian tujuan tersebut dipengaruhi oleh berbagai komponen Pendidikan, seperti peserta didik, pendidik, proses pembelajaran, model, metode, media, dan evaluasi Setiawan & Kamalia, (2025) Pendidikan berperan sebagai fasilitator yang mengelola interaksi pembelajaran, sedangkan peserta didik menjadi subjek aktif dalam proses kontruksi pengetahuan. Keberhasilan Pendidikan memang sangat dipengaruhi oleh sinergi seluruh komponen dalam mendukung berlangsungnya pembelajaran yang efektif dan bermakna.

Dalam penyelenggaraan pembelajaran, evaluasi diposisikan sebagai proses sistematis untuk mengumpulkan, mengolah dan menafsirkan informasi mengenai capaian belajar peserta didik. Evaluasi dilakukan untuk memperoleh gambaran objektif mengenai perkembangan peserta didik serta efektivitas strategi pembelajaran yang diterapkan Suharyat et al., (2022). Panduan penilaian dalam Pendidikan mencangkup berbagai bentuk evaluasi, yakni tes maupun non-tes yang masing-masing dipilih berdasarkan tujuan dan karakteristik kompetensi yang diukur Penilaian tengah semester menjadi salah satu bentuk asesmen sumatif yang digunakan untuk memantau capaian belajar peserta didik pada pertengahan periode pembelajaran. Selain itu, dapat memberikan umpan balik kepada pendidik mengenai efektivitas proses pembelajaran Jauhari et al., (2023).

Penyusunan tes yang baik memerlukan pemenuhan sejumlah kriteria, seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta kualitas pengecoh. Tes hanya dapat dinyatakan berkualitas apabila mampu mengukur kompetensi secara tepat dan konsisten. Oleh karena itu, analisis psikometrik diperlukan untuk memastikan bahwa setiap butir soal juga berperan sebagai instrumen refleksi pendidik untuk memperbaiki penyusunan tes dan memastikan kesesuaian antara instrument asesmen dengan tujuan dan proses pembelajaran Penyediaan instrument asesmen yang valid dan reliabel telah menjadi tuntutan dalam upaya mewujudkan Pendidikan berkualitas tinggi di

era globalisasi saat ini. Pendidikan tidak hanya bertujuan untuk metransfer pengetahuan, tetapi juga membentuk kompetensi peserta didik yang siap menghadapi tantangan dunia kerja dan masyarakat yang kompleks. Rendahnya kualitas asesmen dapat menghambat pencapaian Sustainable Development Goals (SDGs), khususnya dalam aspek literasi yang meliputi kemampuan membaca, menulis, dan berpikir kritis; kualitas pembelajaran yang mencangkup metode pengajaran inovatif dan inklusif; serta pemerataan Pendidikan yang memastikan akses yang adil bagi semua lapisan masyarakat UNESCO, (2020) Pentingnya asesmen yang bermutu juga telah ditekankan melalui pandangan bahwa pelaksanaan asesmen tidak hanya berfungsi sebagai alat pengukuran, tetapi juga sebagai mekanisme yang berperan aktif dalam meningkatkan kualitas proses pembelajaran secara berkelanjutan, seperti memberikan feedback konstruktif kepada pendidik dan peserta didik untuk perbaikan terus-menerus Adriantoni et al., (2025)

Di tingkat nasional, kewajiban penyusunan asesmen sumatif yang autentik untuk mengukur Capaian Pembelajaran (CP) telah diatur melalui kebijakan Kurikulum Merdeka. Pada fase F, tuntutan kemampuan literasi kritis tingkat lanjut, seperti mengevaluasi gagasan dalam teks kompleks, menilai argumentasi, dan menyajikan informasi secara analitis telah ditetapkan sebagai kompetensi esensial Kemendikbudristek, (2024). Rendahnya literasi serta kemampuan berpikir kritis peserta didik di Indonesia turut dilaporkan sebagai salah satu faktor yang memperburuk kondisi tersebut Anisa et al., (2021). Selain itu, ketidakterbiasaan peserta didik dalam menghadapi soal-soal yang menuntut kemampuan berpikir tingkat tinggi juga ditemukan sebagai salah satu penyebab rendahnya kemampuan mereka dalam menyelesaikan soal HOTS, terutama ketika pengalaman berlatih menghadapi soal kompleks tidak diberikan secara memadai selama pembelajaran Septriansyah & Imamuddin, (2024)

Pada tingkat madrasah, tantangan implementasi Kurikulum Merdeka telah ditemukan semakin nyata dan kompleks, terutama karena adanya variasi pemahaman pendidik terhadap substansi Capaian Pembelajaran (CP) Fase F yang sering kali berbeda antara satu madrasah dengan yang lainnya, sebagaimana yang diatur dalam panduan resmi kurikulum Kemdikbudristek, (2022). Variasi pemahaman tersebut telah diperkuat oleh temuan yang menunjukkan bahwa pendidik madrasah masih menghadapi kesulitan dalam menerjemahkan kurikulum ke dalam rancangan pembelajaran dan asesmen, sehingga pelaksanaan Kurikulum Merdeka di madrasah beran tidak merata dan tidak konsisten Alfikri & Handayani, (2024). Kondisi ini sejalan dengan laporan yang menunjukkan bahwa kesiapan pendidik dalam

mengimplementasikan Kurikulum Merdeka masih rendah akibat keterbatasan pelatihan, pemahaman konsep, dan kemampuan merancang asesmen yang sesuai tuntutan kompetensi Romadhon et al., (2023). Tantangan tersebut kemudian diperberat oleh rendahnya literasi dan kemampuan berpikir kritis peserta didik yang telah dilaporkan sebagai masalah nasional yang berpengaruh langsung terhadap efektivitas pelaksanaan kurikulum (Anisa et al., 2021).

Fenomena awal di MA Al-Asror juga telah menunjukkan bahwa Sebagian butir soal asesmen tengah semester belum mencerminkan tuntutan proses berpikir tingkat tinggi, seperti analisis struktur teks, evaluasi argumentasi, dan interpretasi makna implisit. Ketidaksesuaian antara kompleksitas butir soal dan aktivitas pembelajaran yang berlangsung telah ditemukan berpotensi menghasilkan pengukuran kompetensi yang tidak akurat serta menurunkan validitas interpretasi skor peserta didik. Selain itu, ketidaksiapan peserta didik dalam menghadapi soal-soal kompleks telah dijelaskan sebagai akibat dari minimnya pembiasaan terhadap Latihan membaca teks kompleks dan menyelesaikan soal berbasis penalaran tingkat lanjut selama proses pembelajaran, sebagaimana ditunjukkan oleh temuan bahwa kemampuan menyelesaikan soal HOTS sering terhambat oleh kurangnya pengalaman menghadapi tugas yang menuntut penalaran mendalam (Sepriansyah & Imamuddin, 2024).

Beberapa penelitian sebelumnya telah mengkaji analisis butir soal dan pelaksanaan asesmen bahasa Indonesia, tetapi belum secara terpadu menghubungkan kualitas psikometris soal dengan kesesuaian pembelajaran. Penelitian Kholida DjS et al., (2024) misalnya, menelaah pelaksanaan asesmen sumatif fase E di MA Al-Kahiriyyah dan menemukan bahwa Sebagian besar peserta didik belum mencapai standar pembelajaran. Namun, kajian tersebut belum menyentuh aspek validitas, reliabilitas, daya pembeda, dan tingkat kesukaran butir soal secara kuantitatif. Di sisi lain, penelitian Rahayu dan Sukenti (2024) telah menganalisis kualitas butir soal berdasarkan tingkat kesukaran, daya pembeda, dan keefektifan pengecoh. Namun, belum mengaitkannya dengan implementasi pembelajaran atau kesesuaian butir soal terhadap tuntutan kurikulum. Dengan demikian, terdapat kesenjangan penelitian berupa minimnya kajian yang memadukan analisis psikometris butir soal dengan analisis kesesuaian pembelajaran, khususnya pada konteks madrasah dan Kurikulum Merdeka Fase F. kesenjangan ini menunjukkan perlunya penelitian yang secara menyeluruh dalam menilai kualitas butir soal sekaligus menilai kecocokannya dengan praktik pembelajaran di kelas.

Urgensi penelitian ini muncul dari kebutuhan untuk memastikan bahwa asesmen yang digunakan di MA Al-Asror dapat memberikan gambaran yang akurat mengenai kemampuan peserta didik, khususnya dalam hal literasi kritis dan analisis teks. Dalam konteks Kurikulum Merdeka yang menekankan kebebasan satuan Pendidikan, evaluasi berbasis data terhadap instrument asesmen menjadi sangat penting agar keputusan pembelajaran dapat diambil secara tepat. Selain itu, penguatan asesmen di madrasah juga selaras dengan rekomendasi Lembaga nasional maupun internasional yang mendorong pentingnya asesmen factual yang dijadikan sebagai pendukung pembelajaran bermakna dan peningkatan kualitas Pendidikan keagamaan.

Kebaruan penelitian ini terletak pada integrasi dua fokus analisis, yakni analisis psikometris butir soal dan analisis kesesuaian pembelajaran yang dikaji secara bersamaan dalam konteks asesmen tengah semester fase F pada madrasah Aliyah. Selain itu, kebaruan dari penelitian ini juga terletak pada penggunaan metode mixed method yang mengintegrasikan analisis statistic dengan temuan kualitatif mengenai proses dan kesesuaian pembelajaran. Pendekatan ini masih jarang digunakan dalam penelitian madrasah, karena Sebagian besar studi hanya menggunakan metode kuantitatif atau deskriptif semata.

Dengan demikian, penelitian ini dilakukan untuk menganalisis kualitas psikometris butir soal asesmen tengah semester bahasa Indonesia fase F di MA Al-Asror serta mengevaluasi kesesuaianya dengan praktik pembelajaran. Rekomendasi berbasis data melalui penelitian ini diharapkan dapat memberikan kontribusi dalam peningkatan kualitas asesmen dan pembelajaran secara berkelanjutan di madrasah.

METODE PENELITIAN

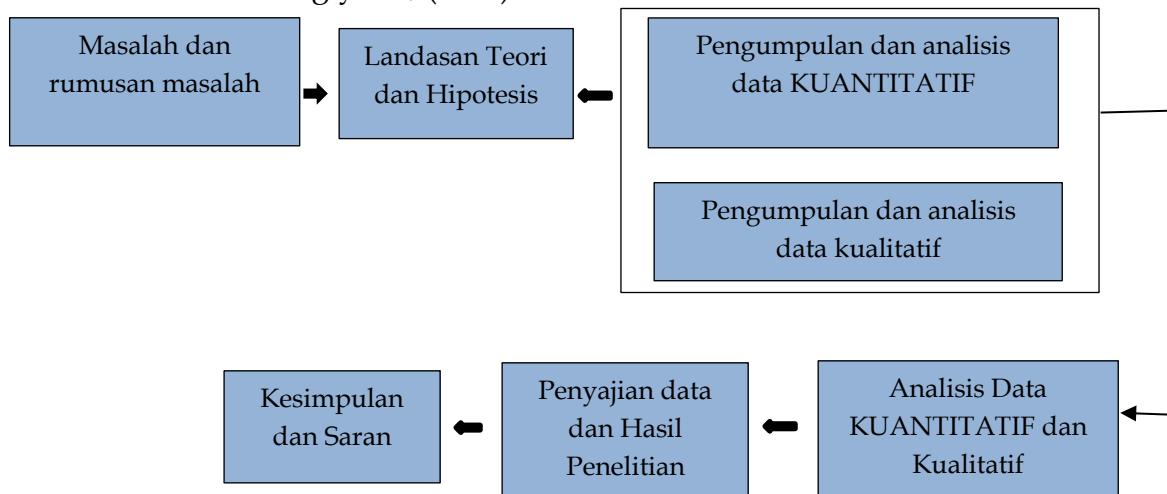
Penelitian ini menggunakan pendekatan metode campuran atau mixed method dengan desain sequential explanatory. Desain ini mengombinasikan metode kuantitatif dan kualitatif yang dilaksanakan secara berurutan. Tahap kuantitatif dilakukan terlebih dahulu untuk memperoleh gambaran objektif mengenai kualitas instrumen asesmen, kemudian dilanjutkan dengan tahap kualitatif untuk memberikan penjelasan yang lebih mendalam terhadap hasil analisis kuantitatif tersebut. Pendekatan ini dipilih agar penelitian mampu menghasilkan pemahaman yang komprehensif terkait kualitas psikometrik asesmen dan kesesuaian antara pembelajaran dan penilaian (Manjato & Novita, 2020; Nitko & Brookhart, 2011).

Penelitian dilaksanakan di Madrasah Aliyah Al-Asror. Subjek penelitian terdiri atas 100 peserta didik kelas XI Fase F yang mengikuti asesmen sumatif

tengah semester mata pelajaran Bahasa Indonesia. Selain itu, satu orang guru Bahasa Indonesia dilibatkan sebagai informan utama pada tahap kualitatif. Pemilihan subjek dilakukan secara purposive karena kelas tersebut merupakan kelas yang secara langsung menerapkan asesmen sumatif sesuai dengan Capaian Pembelajaran Fase F.

Data penelitian dibedakan menjadi data kuantitatif dan data kualitatif. Data kuantitatif diperoleh melalui teknik dokumentasi berupa naskah soal asesmen sumatif dan hasil ujian peserta didik. Data ini dianalisis untuk menilai kualitas butir soal yang meliputi validitas, reliabilitas, daya pembeda, dan tingkat kesukaran. Uji validitas dilakukan dengan membandingkan nilai korelasi butir dengan nilai r tabel. Uji reliabilitas menggunakan koefisien Cronbach's Alpha. Daya pembeda dianalisis melalui nilai corrected item-total correlation, sedangkan tingkat kesukaran dihitung berdasarkan proporsi jawaban benar peserta didik.

Data kualitatif dikumpulkan melalui wawancara terstruktur dengan guru Bahasa Indonesia dan observasi selama proses pembelajaran berlangsung. Wawancara bertujuan untuk menggali informasi mengenai perencanaan pembelajaran, penyusunan asesmen, serta pemahaman guru terhadap Capaian Pembelajaran Fase F. Observasi dilakukan untuk melihat secara langsung aktivitas pembelajaran, keterlibatan peserta didik, dan kesempatan yang diberikan untuk membaca serta menganalisis teks kompleks. Teknik ini digunakan untuk memperoleh gambaran kontekstual yang mendukung hasil analisis kuantitatif Sugiyono, (2017).



HASIL DAN PEMBAHASAN

Tabel 1.
Validitas Soal Pilihan Ganda

Kondisi Intrumen	Jumlah Soal	Butir Valid	Butir Tidak Valid	Presentase Valid
Sebelum revisi	25	18	7	72%
Setelah revisi	25	23	2	92%

Hasil analisis pada tabel 1 menunjukkan bahwa kualitas instrumen asesmen sumatif Bahasa Indonesia Fase F di Madrasah Aliyah Al-Asror mengalami peningkatan yang jelas setelah dilakukan revisi butir soal. Peningkatan ini terlihat pada aspek validitas, reliabilitas, daya pembeda, dan tingkat kesukaran, baik pada soal pilihan ganda maupun soal esai. Namun demikian, temuan juga mengungkap sejumlah keterbatasan yang berkaitan dengan kesesuaian pembelajaran dan kemampuan literasi kritis peserta didik.

Pada aspek validitas soal pilihan ganda, sebelum revisi terdapat 18 dari 25 butir soal atau 72 persen yang dinyatakan valid, sedangkan 7 butir lainnya tidak valid karena nilai r hitung berada di bawah r tabel sebesar 0,195. Setelah dilakukan revisi terhadap redaksi, kesesuaian indikator, dan kejelasan stimulus, jumlah butir valid meningkat menjadi 23 dari 25 butir atau 92 persen. Hanya dua butir soal yang masih dinyatakan tidak valid. Peningkatan ini menunjukkan bahwa revisi berhasil memperkuat keterkaitan antara butir soal dengan konstruk kemampuan yang diukur, sehingga instrumen menjadi lebih representatif dalam mengukur capaian pembelajaran Bahasa Indonesia Fase F.

Tabel 2.
Validitas Uraian

Kondisi Instrumen	Jumlah Soal	Rentang r hitung	Keputusan
Sebelum revisi	5	0,283-0,566	Valid
Setelah revisi	5	0,793-0,911	Sangat valid

Validitas soal uraian pada tabel 2 menunjukkan hasil yang lebih kuat dibandingkan pilihan ganda. Sebelum revisi, seluruh butir soal esai sudah dinyatakan valid dengan nilai r hitung berkisar antara 0,283 hingga 0,566. Setelah revisi, nilai validitas meningkat secara signifikan dengan rentang 0,793 hingga 0,911. Nilai ini termasuk kategori sangat tinggi dan menunjukkan bahwa perbaikan pada perintah soal, indikator, serta rubrik penilaian membuat soal esai semakin selaras dengan kompetensi yang diukur. Kondisi ini mengindikasikan bahwa soal esai lebih mampu merepresentasikan

kemampuan analisis dan pemahaman peserta didik dibandingkan soal pilihan ganda.

Tabel 3.
Reliabilitas Asesmen

Jenis Soal	Cronbach Alpha Sebelum	Keputusan	Cronbach Alpha Sesudah	Keputusan
Pilihan ganda	0,486	Tidak reliabel	0,630	Reliabel
Uraian	0,414	Tidak reliabel	0,923	Reliabel

Hasil uji reliabilitas pada tabel 3 juga menunjukkan peningkatan yang substansial. Sebelum revisi, nilai Cronbach Alpha untuk soal pilihan ganda sebesar 0,486 dan soal esai sebesar 0,414, yang berarti instrumen belum reliabel karena berada di bawah batas minimal 0,60. Setelah revisi, reliabilitas soal pilihan ganda meningkat menjadi 0,630 dan reliabilitas soal esai meningkat tajam hingga 0,923. Peningkatan ini menunjukkan bahwa revisi berhasil memperbaiki konsistensi internal antarbutir soal. Instrumen setelah revisi mampu memberikan hasil pengukuran yang lebih stabil dan dapat dipercaya dalam menilai kemampuan peserta didik.

Tabel 4.
Daya Pembeda Pilihan Ganda (Sesudah Revisi)

Kategori Daya Pembeda	Jumlah Butir
Baik	0
Cukup Baik	5
Perlu Revisi	6
Buruk	14
Total	25

Berbeda dengan validitas dan reliabilitas, hasil analisis daya pembeda menunjukkan temuan yang lebih beragam. Pada soal pilihan ganda, sebagian besar butir masih berada pada kategori buruk atau perlu revisi, baik sebelum maupun sesudah revisi. Meskipun beberapa butir mengalami peningkatan, seperti soal nomor 16 yang meningkat dari nilai -0,103 menjadi 0,396 dan berpindah ke kategori cukup baik, serta soal nomor 21, 22, dan 23 yang meningkat ke kategori cukup baik dengan nilai di atas 0,330, sebagian besar butir tetap memiliki daya pembeda rendah. Hal ini menunjukkan bahwa banyak soal belum optimal dalam membedakan peserta didik berkemampuan tinggi dan rendah. Revisi yang hanya bersifat teknis, seperti perbaikan redaksi,

belum cukup untuk meningkatkan daya pembeda apabila tidak disertai perbaikan substansi dan logika pengecoh.

Tabel 5.

Daya Pembeda Uraian

Kondisi Instrumen	Kategori Dominan	Jumlah Butir
Sebelum revisi	Perlu revisi	5
Sesudah revisi	Baik	5

Sebaliknya, daya pembeda soal esai pada tabel 5 menunjukkan peningkatan yang sangat signifikan. Sebelum revisi, sebagian besar soal esai berada pada kategori buruk, dengan nilai daya pembeda terendah mencapai $-0,250$. Setelah revisi, seluruh soal esai masuk ke kategori baik dengan nilai daya pembeda antara $0,678$ hingga $0,850$. Hal ini menunjukkan bahwa revisi pada rubrik penilaian dan kejelasan tuntutan soal membuat soal esai mampu membedakan kemampuan peserta didik secara lebih proporsional.

Tabel 6.

Tingkat Kesukaran Pilihan Ganda

Kondisi Instrumen	Mudah	Sedang	Sulit	Total
Sebelum revisi	3	13	9	25
Sesudah revisi	18	7	0	25

Pada aspek tingkat kesukaran, hasil analisis menunjukkan kecenderungan peningkatan ke arah kategori mudah, terutama pada soal pilihan ganda. Sebelum revisi, distribusi tingkat kesukaran relatif bervariasi antara sulit, sedang, dan mudah. Setelah revisi, sebagian besar soal berpindah ke kategori mudah. Misalnya, soal nomor 3 meningkat dari nilai $0,25$ yang tergolong sulit menjadi $0,89$ yang tergolong mudah, dan soal nomor 10 meningkat dari $0,19$ menjadi $0,73$. Pola ini menunjukkan bahwa revisi meningkatkan keterpahaman soal, tetapi sekaligus menurunkan daya seleksi instrumen. Dominasi soal kategori mudah berpotensi membuat tes kurang mampu membedakan kemampuan peserta didik secara optimal.

Tabel 7.

Tingkat Kesukaran Soal Uraian

Kondisi Instrumen	Mudah	Sedang	Total
Sebelum revisi	1	4	5
Sesudah revisi	5	0	5

Pola serupa juga terjadi pada tabel 7 soal uraian. Sebagian besar soal yang sebelumnya berada pada kategori sedang berpindah ke kategori mudah setelah

revisi, dengan nilai mean meningkat hingga di atas 0,77. Hal ini menunjukkan bahwa peserta didik lebih mampu menjawab soal dengan benar setelah perbaikan redaksi dan kejelasan tuntutan soal. Namun, kondisi ini juga perlu dicermati karena asesmen yang terlalu mudah tidak sepenuhnya mencerminkan kemampuan literasi kritis yang diharapkan pada Fase F.

Temuan kuantitatif ini diperkuat oleh hasil analisis kualitatif yang menunjukkan bahwa proses pembelajaran di MA Al-Asror belum sepenuhnya selaras dengan tuntutan Capaian Pembelajaran Fase F. Pembelajaran masih didominasi penyampaian materi dan latihan dasar, sementara aktivitas membaca teks kompleks, analisis mendalam, dan evaluasi argumentasi belum dilakukan secara rutin. Akibatnya, peserta didik belum terbiasa menghadapi soal dengan tuntutan berpikir tingkat tinggi. Kondisi ini menjelaskan mengapa peningkatan nilai setelah revisi instrumen tidak selalu mencerminkan peningkatan kemampuan literasi kritis, melainkan lebih menunjukkan peningkatan keterpahaman soal.

Secara keseluruhan, hasil penelitian menunjukkan bahwa revisi instrumen asesmen memberikan dampak positif terhadap validitas dan reliabilitas, terutama pada soal esai. Namun, rendahnya daya pembeda pada sebagian soal pilihan ganda serta dominasi tingkat kesukaran kategori mudah menegaskan bahwa perbaikan asesmen perlu diiringi dengan perbaikan proses pembelajaran. Asesmen yang berkualitas hanya dapat berfungsi optimal apabila didukung pembelajaran yang memberi ruang cukup bagi peserta didik untuk berlatih membaca, menganalisis, dan mengevaluasi teks kompleks sesuai tuntutan Kurikulum Merdeka Fase F.

Pembahasan

Asesmen sumatif memiliki peran strategis dalam merefleksikan ketercapaian tujuan pembelajaran, khususnya pada mata pelajaran Bahasa Indonesia Fase F yang menekankan penguasaan literasi kritis. Oleh karena itu, kualitas instrumen asesmen perlu dikaji tidak hanya dikaji dari sisi hasil belajar peserta didik, tetapi juga karakteristik psikometris butir soal yang digunakan. Analisis dalam penelitian ini menunjukkan bahwa revisi instrumen asesmen sumatif berkontribusi terhadap peningkatan kualitas butir soal, terutama pada aspek validitas dan reliabilitas. Peningkatan tersebut mengindikasikan adanya kesesuaian yang lebih baik antara indikator pembelajaran, konstruk kompetensi, dan bentuk soal.

Berdasarkan teori evaluasi pendidikan, validitas dan reliabilitas merupakan syarat utama agar instrumen asesmen dapat digunakan secara tepat dalam pengambilan keputusan pembelajaran (Crocker & Algina, 1986). Temuan penelitian ini selaras dengan pandangan tersebut, karena perbaikan

redaksi soal dan penyesuaian indikator terbukti memperkuat hubungan antara butir soal dan kemampuan yang diukur. Penelitian terdahulu oleh (Pratama & Guspa, 2022) menunjukkan bahwasanya revisi instrumen berbasis analisis butir dapat meningkatkan kualitas asesmen secara signifikan. Namun, penelitian ini melangkah lebih jauh dengan mengaitkan peningkatan kualitas instrumen tersebut dengan kesesuaian praktik pembelajaran di kelas.

Meskipun demikian, kualitas asesmen tidak dapat hanya dinilai berdasarkan aspek validitas dan reliabilitas. Analisis daya pembeda dan tingkat kesukaran dalam penelitian ini mengungkapkan bahwa sebagian besar soal pilihan ganda masih berada pada kategori mudah dan memiliki daya pembeda rendah. Kondisi ini menunjukkan bahwa instrumen belum sepenuhnya mampu membedakan tingkat kemampuan peserta didik secara optimal. Fenomena tersebut menguatkan argumen bahwa asesmen yang baik secara statistik belum tentu efektif secara pedagogis apabila tingkat kompleksitas kognitif soal tidak selaras dengan tuntutan capaian pembelajaran. Temuan ini sejalan dengan (Setiawan & Kamalia, 2025) yang menekankan pentingnya keseimbangan antara kualitas teknis instrumen dan tuntutan berpikir tingkat tinggi.

Perbedaan karakteristik hasil analisis antara soal pilihan ganda dan soal uraian juga menjadi temuan penting dalam penelitian ini. Soal uraian menunjukkan peningkatan yang lebih signifikan pada aspek validitas, reliabilitas, dan daya pembeda setelah dilakukannya revisi. Hal ini memperkuat pandangan bahwa soal uraian lebih efektif dalam mengukur kemampuan analisis dan reflektif peserta didik dibandingkan soal pilihan ganda. Menurut (Adriantoni et al., 2025) menegaskan bahwa asesmen berbasis uraian memiliki potensi lebih besar dalam menggambarkan kemampuan berpikir kritis, terutama dalam konteks literasi Bahasa Indonesia. Temuan penelitian ini mendukung argumen tersebut dengan bukti empiris dari analisis psikometris.

Analisis kuantitatif tersebut kemudian diperdalam melalui temuan kualitatif yang menyoroti kesesuaian antara pembelajaran dan asesmen. Praktik pembelajaran di kelas masih didominasi oleh metode ceramah dan latihan soal berlevel rendah, sementara aktivitas yang menuntut analisis teks kompleks, interpretasi makna implisit, dan evaluasi argumentasi belum dilaksanakan secara konsisten. Kondisi ini menunjukkan meningkatnya keterpahaman peserta didik terhadap bentuk dan redaksi soal. Temuan ini menguatkan hasil penelitian (Anisa et al., 2021) dan (Sepriansyah & Imamuddin, 2024) yang menyatakan bahwa rendahnya pembiasaan menghadapi soal berpikir tingkat tinggi menjadi salah satu faktor utama lemahnya performa peserta didik pada asesmen berbasis HOTS.

Apabila dibandingkan dengan penelitian terdahulu, penelitian ini memiliki perbedaan yang cukup signifikan. Penelitian (Manjato & Novita, 2020) menitikberatkan pada analisis statistik butir soal tanpa mengaitkannya dengan kesesuaian pembelajaran dan tuntutan kurikulum. Sementara itu, Penelitian (Kholida DjS et al., 2024) lebih berfokus pada capaian hasil asesmen tanpa mengkaji psikometris instrumen yang digunakan, sehingga belum mampu menjelaskan akar permasalahan secara komprehensif. Penelitian ini mengisi celah tersebut dengan mengintegrasikan analisis psikometris dan evaluasi kesesuaian pembelajaran dalam satu kerangka kajian. Dengan demikian, pembeda utama penelitian ini terletak pada pendekatan integratif yang memadukan analisis kuantitatif kualitas butir soal dengan analisis kualitatif kesesuaian pembelajaran dalam konteks Kurikulum Merdeka Fase F di madrasah. Penelitian ini menegaskan bahwa upaya peningkatan kualitas asesmen tidak dapat dilakukan secara parsial melalui revisi instrumen semata, tetapi harus disertai perbaikan proses pembelajaran yang berorientasi pada pengembangan literasi kritis dan kemampuan berpikir tingkat tinggi. Integrasi antara pembelajaran dan asesmen menjadi kunci dalam menghasilkan evaluasi yang bermakna dan berkedilan.

KESIMPULAN

Penelitian ini menegaskan bahwa kualitas instrument asesmen sumatif Bahasa Indonesia Fase F di MA Al-Asror mengalami peningkatan substansial setelah dilakukan proses revisi, terutama pada aspek validitas dan reliabilitas. Pada soal pilihan ganda, proporsi butir soal valid meningkat dari 72% menjadi 92%, sedangkan reliabilitas meningkat dari 0,486 menjadi 0,630. Pada soal esai, seluruh butir soal menunjukkan validitas sangat tinggi dan reliabilitas mencapai 0,923 sehingga instrument tersebut telah memenuhi standar pengukuran psikometri yang baik. Meskipun demikian, temuan mengenai rendahnya daya pembeda pada beberapa butir soal pilihan ganda serta dominasi tingkat kesukaran kategori mudah menunjukkan bahwa instrument belum sepenuhnya optimal dalam melakukan diferensiasi kemampuan peserta didik. Hasil analisis kualitatif memperjelas bahwa ketidaksesuaian antara praktik pembelajaran dan tuntutan Capaian Pembelajaran Fase F menjadi salah satu faktor utama yang menghambat efektivitas asesmen, terutama terkait kurangnya Latihan membaca dan menganalisis teks kompleks.

Secara keseluruhan, penelitian ini memberikan kontribusi penitng berupa bukti empiris bahwa peningkatan kualitas asesmen tidak dapat berdiri sendiri, melainkan harus diintegrasikan dengan perbaikan desain pembelajaran. Kolaborasi antara revisi instrument, penguatan literasi kritis, dan peningkatan

kompetensi pendidik dalam merancang asesmen berbasis psikometri menjadi prasyarat untuk menghasilkan proses evaluasi yang valid, adil, dan selaras dengan prinsip Kurikulum Merdeka.

DAFTAR PUSTAKA

- Adriantoni, Yanre, Gusneti, I., & Angraini, S. (2025). Menilai bukan sekedar menghitung: Peran asesmen formatif dalam meningkatkan kualitas pembelajaran. *Didaktik: Jurnal Pendidikan*, 11, 221–229. <https://journal.stkipsubang.ac.id/index.php/didaktik/article/view/5976/3748>
- Alfikri, M. Y., & Handayani, S. (2024). Tantangan dan peluang implementasi Kurikulum Merdeka di madrasah: Menuju madrasah unggul yang berdaya saing. *Jurnal Pendidikan Islam*, 1(2), 698–702.
- Anisa, A. R., Ipungkarti, A. A., & Saffanah, N. (2021). *Pengaruh kurangnya literasi serta kemampuan berpikir kritis dalam pendidikan di Indonesia*. 1(1), 1–12.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.
- Haris, M. A., & Mustari, M. (2023). Meningkatkan kualitas pembelajaran melalui peningkatan pengelolaan sarana dan prasarana pendidikan di SMPN 4 Narmada. *Khazanah Pendidikan*, 97–105.
- Jauhari, Z. A., Sholihin, M., Agisna, R., Zuar, M. S., & Khusnul, A. (2023). *Evaluasi pembelajaran*. Citapustaka Media.
- Kemdikbudristek. (2022). *Panduan pembelajaran dan asesmen pendidikan anak usia dini, pendidikan dasar, dan menengah*. <https://kurikulum.kemdikbud.go.id/wp-content/uploads/2022/06/Panduan-Pembelajaran-dan-Asesmen.pdf>
- Kemendikbudristek. (2024). *Implementasi Kurikulum Merdeka dan peningkatan kemampuan literasi dan numerasi peserta didik* (L. Solihin & D. N. Rakhmah (eds.)). Pusat Standar dan Kebijakan Pendidikan. https://pskp.kemendikdasmen.go.id/file/kebijakan/1729135165_file.pdf
- Kholida DjS, I., Suryani, I., & Abrar, M. (2024). Analisis asesmen sumatif pada pembelajaran Bahasa Indonesia Fase E. *Semantik*, 13(2). <https://doi.org/10.22460/semantik.v13i2.p239-255>
- Manjato, A., & Novita, R. (2020). *Analisis butir soal ujian akhir semester genap mata pelajaran Bahasa Indonesia siswa kelas X IIS 4 SMAN 8 Kota Bengkulu*. 8, 33–37.
- Mardhiyah, R. H., Aldriani, S. N. F., Chitaa, F., & Zulfikar, M. R. (2021). Pentingnya keterampilan belajar di abad 21 sebagai tuntutan dalam pengembangan sumber daya manusia. *Lectural: Jurnal Pendidikan*.

- Maulansyah, R., Febrianty, D., & Asbari, M. (2023). Peran guru dalam meningkatkan mutu pendidikan. *Journal of Information System and Management*, 31–35.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Pearson.
- Pratama, M., & Guspa, A. (2022). Analisis properti psikometrik skala student engagement versi bahasa Indonesia. *Psychoidea*, 20, 108–117. <https://doi.org/10.30595/psychoidea.v20i2.13310>
- Romadhon, K., Rokhimawan, M. A., & Ayuningtyas, D. R. (2023). Analisis kesiapan guru dalam implementasi Kurikulum Merdeka. *Al-Mudarris*, 7(3), 1049–1063. <https://doi.org/10.35931/am.v7i2.2239>
- Sepriansyah, A., & Imamuddin, M. (2024). Kemampuan menyelesaikan soal HOTS ditinjau dari hasil belajar siswa SDI Ibnu Syam. *TJIE*, 4(2), 98–106. <https://doi.org/10.61456/tjie.v4i2.179>
- Setiawan, A. N., & Kamalia, P. U. (2025). *Evaluating test item quality: A comprehensive analysis of economics multiple-choice questions in Indonesian high schools*. 6(3), 287–297.
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R\&D*. Alfabeta.
- Suharyat, Y., Muthi, I., & Haswani, H. (2022). Evaluating a mid school program in Bekasi: The building learning. *Al-Ishlah: Jurnal Pendidikan*, 14, 1297–1310. <https://doi.org/10.35445/alishlah.v14i1.1058>
- UNESCO. (2020). *Inklusi dan pendidikan: Semua berarti semua* (1st ed.). United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000373721_ind
- Yasin, I. (2021). Problem kultural peningkatan mutu pendidikan di Indonesia: Perspektif total quality management. *Ainara Journal*.