ABDIRA Volume 6 Nomor 4 Tahun 2026 Halaman 1272-1283
JURNAL PENGABDIAN MASYARAKAT
Research & Learning in Faculty of Education
ISSN: 2798-0847 (Printed); 2798-4591 (Online)

Penerapan Prinsip Psikometri untuk Analisis dan Pengembangan Butir Soal TKA yang Berkualitas di Ma'had An-Nahla Al-Islamy

Tritiyatma Hadinugrahaningsih[1], Hayyun Lisdiana[2], Rika Siti Syaadah[3], Vida Zenitha Sudariasri[4], Achmad Ridwan[5], Lia Marsya Fazliana[6], Rifky Maulana Dimyati[7], Alfati Ridho Saputra[8], Almanda Nazwa Putri Suryana[9], Rizky Putri Sumpari[10]

Program Studi Pendidikan Kimia, Universitas Negeri Jakarta
e-mail: hayyunlisdiana@unj.ac.id tritiyatma@unj.ac.id

Abstrak

Kualitas Tes Kemampuan Akademik (TKA) tidak hanya ditentukan oleh kesesuaian materi dan konstruksi soal, tetapi juga oleh karakteristik psikometriknya. Kegiatan pengabdian kepada masyarakat ini bertujuan meningkatkan pemahaman dan keterampilan guru Ma'had An-Nahl Al-Islamy dalam menerapkan prinsip psikometri untuk analisis dan pengembangan butir soal TKA. Kegiatan dilaksanakan menggunakan pendekatan partisipatif melalui sosialisasi, pelatihan, praktik analisis butir, pendampingan, dan evaluasi. Evaluasi dilakukan melalui pretest, posttest, observasi praktik, dan kuesioner umpan balik. Hasil menunjukkan peningkatan rata-rata skor dari 54,55 pada pretest menjadi 70,45 pada posttest, dengan peningkatan sebesar 15,91 poin dan nilai normalized gain sebesar 0,41. Umpan balik peserta juga menunjukkan respons positif terhadap relevansi materi dan pelaksanaan kegiatan. Kegiatan ini memberikan dasar bagi guru untuk mengembangkan praktik evaluasi yang lebih sistematis dan berbasis data.

Kata Kunci: *Psikometri, Analisis Butir Soal, Tes Kemampuan Akademik, Kualitas Soal, Pengabdian Kepada Masyarakat.*

Abstract

The quality of the Academic Competency Test (TKA) is determined not only by content relevance and item construction but also by its psychometric characteristics. This community service program aimed to improve teachers' understanding and skills in applying psychometric principles to analyze and develop TKA items at Ma'had An-Nahl Al-Islamy. The program employed a participatory approach through socialization, training, item analysis practice, mentoring, and evaluation. Evaluation was conducted using pretests, posttests, practical observations, and participant feedback questionnaires. The results indicated an increase in the mean score from 54.55 in the pretest to 70.45 in the posttest, representing an improvement of 15.91 points and a normalized gain of 0.41. Participant feedback also indicated positive responses regarding the relevance of the training materials and implementation. The program provides a foundation for teachers to develop more systematic and data-driven assessment practices.

Kata Kunci: *Psychometrics, Item Analysis, Academic Competency Test, Item Quality, Community Service.*

PENDAHULUAN

Pendidikan tidak hanya berorientasi pada proses pembelajaran, tetapi juga pada kemampuan sistem pendidikan dalam memperoleh informasi yang akurat mengenai capaian kompetensi peserta didik. Informasi tersebut diperoleh melalui kegiatan asesmen yang dirancang secara sistematis untuk menghasilkan interpretasi terhadap kemampuan yang diukur. Kualitas informasi yang dihasilkan berkaitan dengan kualitas instrumen penilaian yang digunakan. Oleh karena itu, penyusunan instrumen perlu disertai dengan evaluasi untuk memastikan bahwa setiap butir soal mampu memberikan informasi yang memadai mengenai kemampuan peserta didik. Ali dan Ruit (2015) menunjukkan bahwa kelemahan dalam penulisan soal dan pengecoh yang tidak berfungsi dapat memengaruhi kualitas soal pilihan ganda. Ali et al. (2016) juga menunjukkan bahwa pengecoh yang berfungsi dengan baik berperan dalam mendukung validitas dan reliabilitas skor yang dihasilkan dari soal pilihan ganda.

Urgensi kualitas instrumen semakin mengemuka seiring dengan berkembangnya kebijakan asesmen pendidikan di Indonesia. Salah satu bentuk asesmen yang menjadi perhatian adalah Tes Kemampuan Akademik (TKA), yang memiliki landasan regulasi melalui Peraturan Menteri Pendidikan Dasar dan Menengah Nomor 9 Tahun 2025 tentang Tes Kemampuan Akademik. Regulasi tersebut mengatur penyelenggaraan TKA, hasil tes, serta pemantauan, evaluasi, dan pelaporannya (Kementerian Pendidikan Dasar dan Menengah, 2025). Dalam konteks ini, kesiapan satuan pendidikan dan pendidik tidak hanya berkaitan dengan pemahaman terhadap pelaksanaan TKA, tetapi juga dengan kemampuan menyusun dan mengevaluasi soal yang digunakan untuk mengukur capaian akademik secara tepat.

Kualitas suatu tes tidak cukup ditentukan oleh kesesuaian materi soal dengan kurikulum atau ketepatan kunci jawaban. Butir soal juga perlu dievaluasi berdasarkan karakteristik empirisnya setelah diujikan kepada peserta didik. Evaluasi tersebut diperlukan untuk mengetahui apakah butir soal memiliki tingkat kesukaran yang sesuai, mampu membedakan tingkat kemampuan peserta, dan menyediakan pilihan jawaban yang berfungsi dengan baik. Quaigrain dan Arhin (2017) membahas penggunaan analisis butir dan reliabilitas untuk mengevaluasi tes yang disusun oleh pendidik. Brown dan Abdulnabi (2017) juga menunjukkan pentingnya evaluasi psikometrik terhadap tes pilihan ganda yang disusun oleh pengajar karena kualitas soal dapat berkaitan dengan ketepatan interpretasi hasil belajar. Dengan demikian, penyusunan soal perlu diikuti proses analisis yang sistematis agar keputusan mengenai kualitas instrumen tidak hanya didasarkan pada pertimbangan subjektif.

Analisis butir merupakan salah satu penerapan prinsip psikometri dalam evaluasi instrumen penilaian. Melalui analisis ini, pengembang tes dapat memperoleh informasi mengenai karakteristik setiap soal berdasarkan respons peserta didik. Pada pendekatan teori tes klasik, evaluasi dapat mencakup tingkat kesukaran, daya pembeda, dan reliabilitas tes. Hayat (2021) menjelaskan

penerapan teori tes klasik dalam analisis butir dan reliabilitas melalui perangkat lunak Klasika, termasuk tahapan penyiapan data, analisis, dan interpretasi hasil. Pratama (2020) juga membahas analisis instrumen kemampuan akademik dengan pendekatan teori tes klasik. Kedua sumber tersebut relevan untuk menjelaskan bahwa hasil analisis statistik dapat digunakan sebagai dasar dalam mengevaluasi karakteristik butir dan kualitas tes secara keseluruhan.

Tingkat kesukaran merupakan salah satu karakteristik yang dapat digunakan untuk mengetahui proporsi peserta didik yang menjawab suatu butir dengan benar. Akan tetapi, tingkat kesukaran tidak dapat dijadikan satu-satunya dasar untuk menentukan kualitas soal. Interpretasinya perlu mempertimbangkan daya pembeda dan karakteristik pilihan jawaban. Namdeo dan Sahoo (2016) menunjukkan bahwa analisis tingkat kesukaran, daya pembeda, dan pengecoh dapat membantu mengidentifikasi soal yang layak digunakan, perlu diperbaiki, atau sebaiknya tidak digunakan kembali. Kheyami et al. (2018) juga menunjukkan bahwa analisis terhadap sejumlah besar soal pilihan ganda dapat mengungkap variasi tingkat kesukaran, daya pembeda, serta efektivitas pengecoh. Temuan tersebut memperkuat pentingnya mengevaluasi butir berdasarkan lebih dari satu karakteristik.

Daya pembeda memberikan informasi mengenai kemampuan suatu butir untuk membedakan peserta didik dengan tingkat kemampuan yang berbeda. Butir dengan daya pembeda rendah perlu ditinjau lebih lanjut karena belum tentu memberikan informasi yang memadai mengenai perbedaan kemampuan peserta. Ali dan Ruit (2015) membahas pengaruh kelemahan penulisan soal dan karakteristik pengecoh terhadap kualitas soal pilihan ganda. Sementara itu, Brown dan Abdulnabi (2017) mengevaluasi kualitas tes pilihan ganda yang disusun oleh pengajar dengan mempertimbangkan pendekatan teori tes klasik dan teori respons butir. Hasil kajian semacam ini menunjukkan bahwa evaluasi butir dapat memberikan informasi yang lebih terarah untuk memperbaiki instrumen penilaian.

Pada soal pilihan ganda, efektivitas pengecoh juga menjadi komponen penting. Pengecoh merupakan pilihan jawaban selain kunci yang dirancang agar tampak masuk akal bagi peserta yang belum menguasai materi. Pengecoh yang tidak berfungsi dapat mengindikasikan bahwa pilihan tersebut kurang menarik atau memiliki kelemahan dalam penyusunannya. Ali et al. (2016) menunjukkan bahwa perbaikan pengecoh yang tidak berfungsi dapat meningkatkan kualitas pengukuran. Kumar et al. (2021) juga menggunakan tingkat kesukaran, daya pembeda, efektivitas pengecoh, dan reliabilitas untuk mengevaluasi soal pilihan ganda serta mengidentifikasi butir yang perlu diperbaiki atau dikeluarkan dari bank soal. Penelitian Rezigalla et al. (2024) selanjutnya membahas hubungan efektivitas pengecoh dengan indeks kesukaran dan daya pembeda. Oleh sebab itu, analisis pengecoh dapat memberikan informasi tambahan untuk mengevaluasi kualitas soal, bukan hanya dari sisi pokok soal dan kunci jawaban, tetapi juga dari pilihan jawaban yang disediakan.

Penelitian pada berbagai konteks pendidikan menunjukkan bahwa analisis butir dapat diterapkan untuk mengevaluasi soal yang digunakan dalam asesmen. Jannah et al. (2021), misalnya, menganalisis soal pilihan ganda dalam tes uji coba Bahasa Inggris tingkat SMP di Indonesia dengan memperhatikan tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasilnya menunjukkan bahwa karakteristik soal bervariasi dan sebagian butir memiliki pengecoh yang kurang efektif. Marsevani (2022) juga mengevaluasi soal pilihan ganda untuk peserta didik usia sekolah dengan menganalisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Penelitian tersebut menunjukkan bahwa sebagian butir perlu ditinjau dan diperbaiki sebelum digunakan kembali. Temuan-temuan ini memperlihatkan bahwa evaluasi empiris relevan tidak hanya untuk tes berskala besar, tetapi juga untuk soal yang dikembangkan dan digunakan dalam lingkungan pendidikan.

Analisis butir juga dapat membantu pendidik membangun bank soal yang lebih terstruktur. Butir yang memenuhi kriteria dapat dipertimbangkan untuk digunakan kembali, sedangkan butir yang memiliki kelemahan perlu direvisi atau dievaluasi lebih lanjut. Yahia (2022) membahas analisis butir setelah validasi untuk mengevaluasi validitas dan reliabilitas soal pilihan ganda serta mendukung pemilihan butir bagi bank soal. Bhat dan Prasad (2021) juga meneliti analisis dan optimalisasi soal pilihan ganda untuk membangun bank soal yang layak digunakan. Selain itu, Bhattacherjee et al. (2022) menunjukkan bahwa analisis butir dapat membantu penguji menentukan karakteristik soal berdasarkan tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Dengan demikian, hasil analisis dapat dimanfaatkan sebagai bahan pertimbangan dalam mempertahankan, memperbaiki, atau mengeluarkan butir dari perangkat penilaian.

Penguatan kompetensi pendidik dalam analisis butir memerlukan pendekatan yang tidak hanya menekankan pemahaman konsep, tetapi juga memberikan kesempatan untuk menerapkannya secara langsung. Pendidik perlu memahami makna hasil analisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh, kemudian menghubungkannya dengan keputusan mengenai butir soal. Przymuszała et al. (2020) mengevaluasi penggunaan panduan penulisan soal pilihan ganda sebagai bentuk pengembangan kompetensi pengajar dan menemukan bahwa intervensi tersebut berkontribusi pada penurunan kelemahan dalam penulisan soal. Penelitian mengenai pengembangan dosen oleh Brown dan Abdulnabi (2017) serta evaluasi kualitas soal setelah program pengembangan fakultas juga memperlihatkan pentingnya penguatan kapasitas penyusun soal. Oleh karena itu, pelatihan berbasis praktik dapat membantu pendidik menghubungkan pengetahuan psikometri dengan kegiatan penyusunan dan perbaikan instrumen penilaian.

Berdasarkan kebutuhan tersebut, kegiatan pengabdian kepada masyarakat di Ma'had An-Nahla Al-Islamy diarahkan untuk memperkuat kompetensi pendidik dalam menerapkan prinsip psikometri pada analisis dan pengembangan

butir soal TKA. Kegiatan ini berfokus pada pengenalan prinsip dasar pengukuran pendidikan dan penerapannya dalam mengevaluasi karakteristik butir, meliputi tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Melalui kegiatan tersebut, pendidik diharapkan memperoleh pemahaman yang lebih baik mengenai penggunaan data respons peserta didik untuk menilai kualitas soal. Hasil analisis selanjutnya dapat dimanfaatkan sebagai dasar untuk mempertahankan, merevisi, atau menyempurnakan butir sesuai dengan karakteristik dan tujuan pengukurannya.

Kegiatan ini bertujuan untuk meningkatkan pemahaman dan keterampilan pendidik di Ma'had An-Nahla Al-Islamy dalam menerapkan prinsip psikometri untuk menganalisis dan mengembangkan butir soal TKA yang berkualitas. Secara khusus, kegiatan diarahkan agar peserta mampu memahami konsep dasar psikometri dan karakteristik butir, melakukan analisis terhadap respons peserta didik, menginterpretasikan hasil analisis berdasarkan kriteria yang relevan, serta menggunakan hasil tersebut sebagai dasar untuk memperbaiki soal. Tujuan ini menempatkan analisis butir sebagai proses yang menghubungkan pengetahuan mengenai pengukuran pendidikan dengan praktik penyusunan instrumen penilaian.

Penerapan prinsip psikometri dalam kegiatan ini diharapkan memberikan manfaat pada tingkat individual maupun kelembagaan. Pada tingkat individu, pendidik diharapkan memiliki kemampuan yang lebih baik dalam mengevaluasi kualitas soal secara objektif berdasarkan data empiris. Pada tingkat kelembagaan, hasil analisis dapat dimanfaatkan sebagai bahan untuk membangun dan memperbaiki bank soal secara lebih terarah. Penggunaan data analisis juga dapat membantu pendidik mendokumentasikan keputusan terhadap butir yang dipertahankan atau direvisi. Dengan demikian, pengembangan instrumen tidak hanya bergantung pada penilaian subjektif penyusun, tetapi turut mempertimbangkan karakteristik soal berdasarkan respons peserta didik.

Pada akhirnya, penerapan prinsip psikometri dalam pengembangan butir soal TKA merupakan bagian dari upaya membangun praktik asesmen yang berbasis bukti. Analisis tingkat kesukaran, daya pembeda, efektivitas pengecoh, dan reliabilitas dapat memberikan informasi yang berguna bagi pendidik dalam mengevaluasi dan menyempurnakan instrumen penilaian. Apabila dilakukan secara berkelanjutan, proses tersebut berpotensi mendukung pengembangan bank soal yang lebih terstruktur dan meningkatkan kapasitas pendidik dalam menyusun instrumen penilaian. Oleh karena itu, kegiatan di Ma'had An-Nahla Al-Islamy diharapkan dapat menjadi langkah penguatan kompetensi pendidik dalam menerapkan prinsip psikometri untuk mendukung pengembangan butir soal TKA yang lebih berkualitas.

METODE

Kegiatan pengabdian kepada masyarakat ini dilaksanakan dengan pendekatan partisipatif dan kolaboratif yang melibatkan tim pengabdi, guru, dan mahasiswa. Kegiatan dilaksanakan di Ma'had An-Nahl Al-Islamy dengan sasaran

guru yang terlibat dalam penyusunan dan evaluasi Tes Kemampuan Akademik (TKA). Pendekatan ini dipilih berdasarkan hasil identifikasi awal yang menunjukkan bahwa penyusunan soal TKA di mitra belum didukung oleh analisis psikometri secara sistematis, khususnya terkait tingkat kesukaran, daya pembeda, efektivitas pengecoh, dan reliabilitas instrumen.

Pelaksanaan kegiatan terdiri atas empat tahapan utama, yaitu sosialisasi dan analisis kebutuhan, pelatihan, praktik dan pendampingan, serta evaluasi dan keberlanjutan. Tahap sosialisasi dilakukan untuk mengidentifikasi kebutuhan mitra sekaligus memberikan pemahaman mengenai pentingnya kualitas instrumen asesmen. Tahap pelatihan mencakup konsep dasar pengukuran pendidikan, validitas, reliabilitas, Classical Test Theory (CTT), tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Selanjutnya, peserta melakukan praktik analisis menggunakan data respons peserta didik dari soal TKA mitra dengan pendampingan tim pengabdi.

Bahan yang digunakan meliputi modul pelatihan, contoh butir soal TKA, lembar kerja analisis, data respons peserta didik yang telah dianonimkan, serta panduan analisis menggunakan perangkat lunak. Analisis butir menggunakan pendekatan CTT dengan parameter tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Tingkat kesukaran digunakan untuk mengidentifikasi proporsi peserta yang mampu menjawab butir dengan benar, sedangkan daya pembeda digunakan untuk mengetahui kemampuan butir dalam membedakan peserta berdasarkan tingkat kemampuan. Analisis pengecoh dilakukan untuk menilai keberfungsian alternatif jawaban pada soal pilihan ganda.

Evaluasi keberhasilan kegiatan dilakukan melalui pretest dan posttest, observasi selama praktik, serta kuesioner umpan balik peserta. Pretest dan posttest digunakan untuk mengetahui perubahan pemahaman peserta mengenai prinsip psikometri dan analisis butir soal. Observasi digunakan untuk menilai keterampilan peserta dalam melakukan pengolahan dan interpretasi hasil analisis, sedangkan kuesioner digunakan untuk mengetahui persepsi peserta terhadap pelaksanaan dan manfaat kegiatan. Data pretest-posttest dan umpan balik dianalisis secara deskriptif menggunakan nilai rata-rata, persentase, dan distribusi respons, sedangkan hasil analisis butir diinterpretasikan berdasarkan karakteristik psikometriknya.

Tahap akhir diarahkan pada keberlanjutan program melalui penyusunan panduan analisis butir dan pengembangan bank soal TKA yang dilengkapi informasi karakteristik psikometrik. Guru didorong untuk menerapkan prosedur analisis pada pelaksanaan asesmen berikutnya secara mandiri. Dengan demikian, kegiatan tidak hanya menghasilkan peningkatan pengetahuan dan keterampilan peserta, tetapi juga mendukung terbentuknya praktik evaluasi yang lebih sistematis dan berbasis data di Ma'had An-Nahl Al-Islamy.

HASIL DAN PEMBAHASAN

1. Peningkatan Pemahaman Peserta terhadap Prinsip Psikometri

Evaluasi pemahaman peserta dilakukan melalui pretest dan posttest yang diberikan sebelum dan setelah kegiatan pelatihan. Berdasarkan data yang diperoleh, terdapat 12 peserta yang mengisi pretest dengan nilai rata-rata sebesar 54,55, sedangkan posttest diikuti oleh 11 peserta dengan nilai rata-rata sebesar 70,45. Untuk melihat perubahan pada peserta yang mengikuti kedua tahap evaluasi, dilakukan pencocokan data berdasarkan identitas peserta sehingga diperoleh 11 pasangan data. Hasil analisis menunjukkan bahwa rata-rata skor peserta meningkat dari 54,55 menjadi 70,45, atau mengalami peningkatan sebesar 15,91 poin.

Tabel 1. Analisis *Pretest* dan *Posstest* Peserta

Evaluasi	N	Rata-Rata	Nilai Maksimal	Nilai Minimal
Pretest	11	54.55	25	90
Posttest	11	70.45	45	100

Sumber: Data hasil kegiatan pengabdian masyarakat, 2026

Peningkatan rata-rata tersebut menunjukkan adanya perubahan pemahaman peserta setelah memperoleh pelatihan mengenai TKA dan prinsip psikometri. Secara individual, 10 dari 11 peserta mengalami peningkatan skor, sedangkan satu peserta mengalami penurunan skor sebesar 5 poin. Besarnya peningkatan individu berada pada rentang 10–35 poin, dengan peningkatan terbesar sebesar 35 poin. Jika dihitung menggunakan normalized gain pada 11 pasangan data, diperoleh rata-rata 0,41, yang menunjukkan adanya peningkatan pemahaman setelah kegiatan. Hasil ini memperlihatkan bahwa pelatihan tidak hanya memberikan informasi konseptual, tetapi juga membantu peserta memahami penerapan prinsip psikometri dalam konteks pengembangan soal.

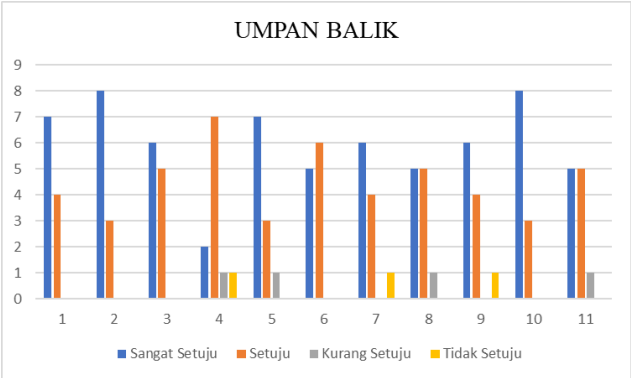
Peningkatan tersebut relevan dengan karakteristik materi pelatihan yang tidak hanya membahas konsep TKA, tetapi juga mencakup validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Pemahaman terhadap karakteristik tersebut penting karena kualitas soal tidak dapat ditentukan hanya berdasarkan kesesuaian materi atau kebenaran kunci jawaban. Analisis butir memberikan informasi empiris mengenai bagaimana suatu soal berfungsi ketika diberikan kepada peserta tes. Kumar et al. (2021) menjelaskan bahwa analisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh dapat digunakan untuk mengidentifikasi butir yang layak dipertahankan maupun butir yang membutuhkan revisi.

Dengan demikian, peningkatan skor setelah pelatihan dapat dipandang sebagai indikasi awal bahwa pendekatan pelatihan yang memadukan pemaparan konsep dan praktik analisis dapat membantu meningkatkan penguasaan peserta terhadap materi psikometri. Meskipun demikian, hasil tersebut perlu ditafsirkan secara proporsional karena kegiatan ini tidak menggunakan kelompok kontrol dan jumlah peserta relatif terbatas. Oleh karena itu, peningkatan skor lebih tepat diposisikan sebagai indikator keberhasilan pembelajaran peserta selama kegiatan,

bukan sebagai bukti kausal bahwa pelatihan secara eksperimental menyebabkan peningkatan kemampuan.

2. Hasil Umpan Balik Peserta terhadap Pelaksanaan Kegiatan

Selain mengevaluasi perubahan pemahaman, kegiatan juga memperoleh umpan balik dari peserta untuk mengetahui persepsi terhadap pelaksanaan pelatihan. Sebanyak 11 peserta mengisi lembar evaluasi kegiatan. Instrumen umpan balik mencakup 11 aspek yang meliputi kepuasan terhadap penyampaian materi, pengetahuan baru yang diperoleh, relevansi materi, keterterapan materi dalam pembelajaran, munculnya ide baru, penguasaan materi oleh pemateri, kecukupan waktu, kejelasan penyampaian, kecukupan waktu pelatihan, kemampuan pemateri menjawab pertanyaan, serta kelancaran teknis kegiatan.



Gambar 1. Hasil Umpan Balik Peserta

Secara umum, respons peserta menunjukkan kecenderungan positif terhadap pelaksanaan kegiatan. Pada aspek kepuasan terhadap penyampaian materi (butir soal 1), sebanyak 7 peserta menyatakan sangat setuju dan 4 peserta menyatakan setuju. Pola positif juga terlihat pada aspek perolehan wawasan baru (butir soal 2), dengan 8 peserta menyatakan sangat setuju dan 3 peserta menyatakan setuju. Sementara itu, pada aspek relevansi materi dengan harapan peserta (butir soal 3), 6 peserta menyatakan sangat setuju dan 5 peserta menyatakan setuju. Data tersebut menunjukkan bahwa materi pelatihan dipandang relevan dan memberikan tambahan pengetahuan bagi peserta.

Respons terhadap keterserapan materi menunjukkan hasil yang relatif positif meskipun lebih beragam. Pada aspek bahwa materi dapat diterapkan dalam pembelajaran di sekolah (butir soal 4), 2 peserta menyatakan sangat setuju, 7 peserta menyatakan setuju, 1 peserta kurang setuju, dan 1 peserta tidak setuju. Temuan ini penting karena menunjukkan bahwa sebagian besar peserta melihat potensi penerapan materi dalam praktik pendidikan, tetapi masih terdapat peserta yang membutuhkan penguatan lebih lanjut agar penerapan prinsip psikometri dapat dilakukan secara mandiri.

Aspek penguasaan materi oleh pemateri (butir soal 6) juga memperoleh respon positif, yaitu 5 peserta menyatakan sangat setuju dan 6 peserta menyatakan setuju. Pada aspek penyampaian materi yang mudah dipahami serta pemberian contoh implementasi yang relevan, 5 peserta menyatakan sangat setuju dan 5 peserta menyatakan setuju, sementara 1 peserta menyatakan kurang setuju. Hasil

tersebut mengindikasikan bahwa strategi penyampaian materi dan contoh praktis telah cukup mendukung pemahaman peserta. Hal ini sejalan dengan karakteristik pelatihan analisis butir yang membutuhkan kombinasi antara pemahaman konseptual dan pengalaman praktik karena peserta perlu memahami makna statistik sebelum mengambil keputusan terhadap kualitas suatu butir.

3. Persepsi Peserta terhadap Intensitas dan Keberlanjutan Kegiatan

Umpan balik juga digunakan untuk mengidentifikasi kebutuhan peserta terhadap keberlanjutan program. Berdasarkan hasil evaluasi, mayoritas peserta menghendaki kegiatan pelatihan dilaksanakan dua kali dalam satu tahun. Data tersebut menunjukkan bahwa peserta tidak memandang kegiatan sebagai intervensi yang cukup dilakukan satu kali, tetapi terdapat kebutuhan untuk memperoleh kegiatan lanjutan secara berkala.

Kebutuhan terhadap pelaksanaan berkelanjutan dapat dipahami karena kompetensi analisis butir tidak hanya berkaitan dengan penguasaan konsep, tetapi juga membutuhkan latihan berulang dalam menginterpretasikan data dan mengambil keputusan terhadap kualitas soal. Analisis butir secara empiris memungkinkan pendidik mengidentifikasi soal yang dapat dipertahankan, direvisi, atau dieliminasi sehingga dapat digunakan untuk membangun bank soal yang lebih berkualitas. Kumar et al. (2021) juga menempatkan analisis butir sebagai bagian dari proses penjaminan mutu instrumen dan pengembangan bank soal.

4. Kebutuhan Peserta terhadap Pelatihan Lanjutan

Hasil evaluasi juga menunjukkan adanya kebutuhan terhadap pengembangan topik pelatihan berikutnya. Berdasarkan pilihan peserta, topik penyusunan soal berbasis HOTS memperoleh jumlah pilihan tertinggi, yaitu 6 peserta. Selanjutnya, penyusunan modul ajar dipilih oleh 1 peserta, persiapan TKA oleh 2 peserta, dan pelatihan untuk meningkatkan deep learning peserta didik juga dipilih oleh 2 peserta.

Temuan tersebut menunjukkan bahwa kebutuhan peserta tidak berhenti pada kemampuan menganalisis soal yang telah digunakan, tetapi berkembang menuju kemampuan menghasilkan soal yang berkualitas sejak tahap konstruksi. Hal tersebut merupakan temuan yang relevan bagi keberlanjutan program karena analisis psikometri pada dasarnya merupakan bagian dari siklus pengembangan instrumen. Butir yang telah dianalisis dapat digunakan sebagai dasar untuk memperbaiki konstruksi soal berikutnya, terutama dalam aspek tingkat kesukaran, kemampuan membedakan peserta, dan kualitas pengecoh. Rezigalla et al. (2024) menunjukkan bahwa kualitas pengecoh memiliki keterkaitan dengan karakteristik tingkat kesukaran dan daya pembeda, sehingga perhatian terhadap konstruksi opsi jawaban menjadi bagian penting dalam pengembangan soal pilihan ganda.

Secara keseluruhan, hasil kegiatan menunjukkan tiga capaian utama, yaitu peningkatan pemahaman peserta mengenai prinsip psikometri, respons positif terhadap pelaksanaan pelatihan, serta munculnya kebutuhan untuk melanjutkan

pengembangan kompetensi asesmen. Peningkatan rata-rata skor sebesar 15,91 poin menunjukkan adanya perubahan kemampuan setelah pelatihan, sementara umpan balik peserta memperlihatkan bahwa materi dinilai memberikan wawasan baru dan relevan. Di sisi lain, keinginan peserta untuk memperoleh pelatihan secara berkala dan kebutuhan terhadap materi penyusunan soal HOTS menunjukkan bahwa penguatan kompetensi asesmen perlu dilaksanakan secara berkelanjutan. Dengan demikian, kegiatan penerapan prinsip psikometri di Ma'had An-Nahl Al-Islamy dapat menjadi tahap awal dalam membangun praktik pengembangan soal yang lebih sistematis dan berbasis data.

SIMPULAN

Kegiatan penerapan prinsip psikometri untuk analisis dan pengembangan butir soal TKA di Ma'had An-Nahl Al-Islamy menunjukkan hasil yang positif dalam meningkatkan pemahaman peserta terhadap prinsip dasar psikometri dan analisis butir soal. Berdasarkan hasil evaluasi terhadap 11 peserta yang mengikuti pretest dan posttest, rata-rata skor meningkat dari 54,55 menjadi 70,45, dengan peningkatan sebesar 15,91 poin dan nilai normalized gain sebesar 0,41. Selain itu, umpan balik dari 11 peserta menunjukkan respons yang secara umum positif terhadap materi, penyampaian pemateri, relevansi kegiatan, serta manfaat pelatihan dalam menambah wawasan peserta.

Kegiatan ini juga menunjukkan adanya kebutuhan untuk melanjutkan penguatan kompetensi asesmen secara berkelanjutan. Mayoritas peserta mengusulkan pelaksanaan kegiatan sebanyak dua kali dalam satu tahun, sedangkan topik yang paling banyak diharapkan pada kegiatan berikutnya adalah penyusunan soal berbasis HOTS. Dengan demikian, penerapan prinsip psikometri dapat menjadi langkah awal dalam membangun kemampuan guru untuk mengembangkan instrumen TKA yang lebih sistematis, berbasis data, dan berorientasi pada peningkatan kualitas asesmen di Ma'had An-Nahl Al-Islamy.

Ucapan terima kasih disampaikan kepada Rektor dan Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Negeri Jakarta yang telah memfasilitasi dan mendanai penelitian ini melalui hibah pelaksanaan program pengabdian kepada masyarakat kuliah kerja nyata sesuai dengan kontrak nomor 19/PPM-KKN/LPPM/III/2026.

DAFTAR PUSTAKA

Ali, S. H., Carr, P. A., & Ruit, K. G. (2016). Validity and reliability of scores obtained on multiple-choice questions: Why functioning distractors matter. *Journal of the Scholarship of Teaching and Learning, 16*(1), 1–14. https://doi.org/10.14434/josotl.v16i1.19106

Ali, S. H., & Ruit, K. G. (2015). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Medical Education, 15*, Article 40. https://doi.org/10.1186/s12909-015-0319-0

Bhat, S. K., & Prasad, K. H. L. (2021). Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: A cross-sectional

study. *Indian Journal of Ophthalmology, 69*(2), 343–346. https://doi.org/10.4103/ijo.IJO_1610_20

Bhattacherjee, S., Mukherjee, A., Bhandari, K., & Rout, A. J. (2022). Evaluation of multiple-choice questions by item analysis, from an online internal assessment of 6th semester medical students in a rural medical college, West Bengal. *Indian Journal of Community Medicine*, 47 (1), 92–95. https://doi.org/10.4103/ijcm.ijcm_1156_21

Brown, G. T. L., & Abdulnabi, H. H. A. (2017). Evaluating the quality of higher education instructor-constructed multiple-choice tests: Impact on student grades. *Frontiers in Education*, 2, Article 24. https://doi.org/10.3389/feduc.2017.00024

DeCarlo, L. T. (2023). Classical item analysis from a signal detection perspective. *Journal of Educational Measurement*. https://doi.org/10.1111/jedm.12358

Hayat, B. (2021). Klasika: Program analisis item dan tes dengan pendekatan klasik. *JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)*, 10 (1), 1–11. https://doi.org/10.15408/jp3i.v10i1.20551

Jannah, R., Hidayat, D. N., Husna, N., & Khasbani, I. (2021). An item analysis on multiple-choice questions: A case of a junior high school English try-out test in Indonesia. *Leksika: Jurnal Bahasa, Sastra dan Pengajarannya*, 15 (1), 9–17. https://doi.org/10.30595/lks.v15i1.8768

Kheyami, D., Jaradat, A., Al-Shibani, T., & Ali, F. A. (2018). Item analysis of multiple choice questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos University Medical Journal*, 18 (1), e68–e74. https://doi.org/10.18295/squmj.2018.18.01.011

Kulshreshtha, S., Gupta, G., Goyal, G., Gupta, K., & Davda, K. (2024). Item analysis of single best response type multiple choice questions for formative assessment in obstetrics and gynaecology. *The Journal of Obstetrics and Gynecology of India*, 74 (3), 256–264. https://doi.org/10.1007/s13224-023-01904-2

Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, 77 (Suppl. 1), S85–S89. https://doi.org/10.1016/j.mjafi.2020.11.007

Marsevani, M. (2022). Item analysis of multiple-choice questions: An assessment of young learners. *English Review: Journal of English Education*, 10 (2). https://doi.org/10.25134/erjee.v10i2.6241

Namdeo, S. K., & Sahoo, B. (2016). Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India. *International Journal of Research in Medical Sciences*, 4 (5), 1716–1719. https://doi.org/10.18203/2320-6012.ijrms20161256

Pratama, D. (2020). Analysis of Classical Test Theory (CTT) approach on academic ability test instrument. *JISAE: Journal of Indonesian Student Assessment and Evaluation*, 5 (2), 43–54. https://doi.org/10.21009/jisae.v5i2.12701

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4 (1), Article 1301013.

https://doi.org/10.1080/2331186X.2017.1301013

Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaifi, J., Alghamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & Adam, M. I. E. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, 24, Article 445. https://doi.org/10.1186/s12909-024-05433-y

Yahia, A. I. O. (2021). Post-validation item analysis to assess the validity and reliability of multiple-choice questions at a medical college with an innovative curriculum. *The National Medical Journal of India*, 34 (6), 359–362. https://doi.org/10.25259/NMJI_414_20

Zaini, Z., Hardika, R. C., Abrori, M. A. M., & Anggara, F. (2025). Pelatihan analisis tingkat kesukaran soal dan daya pembeda soal untuk identifikasi kualitas butir soal bagi guru di SMP Bahrul Ulum. *Journal of Empowerment Community*, 7(2). https://doi.org/10.36423/jec.v7i2.2449

Permendikdasmen Republik Indonesia Nomor 9 Tahun 2025 tentang Tes Kemampuan Akademik. (2025).