Didaktik : Jurnal Ilmiah PGSD FKIP Universtas Mandiri,
ISSN Cetak : 2477-5673 ISSN Online : 2614-722X
Volume 12 Nomor 02, Juni 2026

ANALISIS KUALITAS TES HASIL BELAJAR HADIS BERDASARKAN VALIDITAS, RELIABILITAS, TINGKAT KESUKARAN, DAN DAYA PEMBEDA

Rifald Rizal[1], Muhammad Hafid[2], Miftahir Rizqa[3]
[1,2,3]Universitas Islam Negeri Sultan Syarif Kasim Riau
[1]22590112734@students.uin-suska.ac.id , [2]22590113011@students.uin-suska.ac.id [3]miftahir.rizqa@uin-suska.ac.id

ABSTRACT

Learning evaluation is an essential component in assessing the level of students' competency achievement, particularly in the Hadith subject. The accuracy of evaluation results largely depends on the quality of the instrument employed; therefore, an instrument with strong validity and reliability is indispensable in the assessment process. This study aimed to examine the validity, reliability, difficulty level, and discrimination index of a Hadith learning achievement test at Darul Hadits Al-Aidarusiyah Islamic Boarding School. The study employed a quantitative approach with a descriptive research design. The participants consisted of 15 students, while the instrument analyzed comprised 10 essay items. Data were analyzed manually using Microsoft Excel through the Pearson Product Moment validity test, Cronbach's Alpha reliability test, item difficulty analysis, and item discrimination index analysis. The findings revealed that 9 items met the validity criteria, whereas 1 item was found to be invalid. The reliability analysis produced a Cronbach's Alpha coefficient of 0.973, indicating a very high level of reliability. The difficulty level analysis showed that most items fell within the moderate category, while the discrimination index ranged from poor to very good. Based on these findings, the Hadith learning achievement test instrument is considered appropriate for use as a learning evaluation tool. Nevertheless, several items require improvement to further enhance the quality and accuracy of measurement.

Keywords: Validity, Reliability, Difficulty Level, Discrimination Index, Hadith.

ABSTRAK

Evaluasi pembelajaran merupakan komponen penting dalam menilai tingkat pencapaian kompetensi santri, terutama pada mata pelajaran Hadis. Ketepatan hasil evaluasi sangat bergantung pada kualitas instrumen yang digunakan, sehingga instrumen yang memiliki validitas dan reliabilitas yang baik menjadi kebutuhan utama dalam proses penilaian. Penelitian ini bertujuan untuk mengkaji validitas, reliabilitas, tingkat kesukaran, dan daya pembeda instrumen tes hasil belajar Hadis di Pondok Pesantren Darul Hadits Al-Aidarusiyah. Penelitian dilaksanakan dengan pendekatan kuantitatif menggunakan jenis penelitian deskriptif. Subjek penelitian terdiri atas 15 santri, sedangkan instrumen yang dianalisis berupa 10 butir soal esai. Proses analisis data dilakukan secara manual dengan bantuan Microsoft Excel melalui pengujian validitas Product Moment, reliabilitas Cronbach's Alpha, analisis tingkat kesukaran, serta analisis daya pembeda soal. Temuan penelitian menunjukkan bahwa 9 butir soal memenuhi kriteria valid, sedangkan 1 butir soal dinyatakan tidak valid. Hasil pengujian reliabilitas memperoleh nilai Cronbach's Alpha sebesar 0,973 yang

mengindikasikan tingkat reliabilitas sangat tinggi. Analisis tingkat kesukaran memperlihatkan bahwa sebagian besar soal berada pada kategori sedang, sementara daya pembeda soal berada pada rentang kategori kurang baik hingga sangat baik. Berdasarkan hasil tersebut, instrumen tes hasil belajar Hadis dinilai layak digunakan sebagai alat evaluasi pembelajaran, meskipun beberapa butir soal masih memerlukan penyempurnaan agar kualitas pengukuran yang dihasilkan menjadi lebih optimal.

Kata Kunci: Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, Hadis.

A. Pendahuluan

Evaluasi menempati posisi yang sangat penting dalam kegiatan pembelajaran karena berfungsi sebagai sarana untuk menilai tingkat pencapaian tujuan pendidikan yang telah dirumuskan. Melalui proses evaluasi, pendidik dapat memperoleh gambaran mengenai kemampuan peserta didik setelah mengikuti pembelajaran sehingga aspek keberhasilan maupun kelemahan yang masih ada dapat diidentifikasi secara lebih jelas. Dalam konteks pendidikan Islam, evaluasi tidak hanya diarahkan pada pengukuran kemampuan kognitif, tetapi juga mencakup penilaian terhadap perkembangan sikap serta implementasi nilai-nilai keagamaan dalam kehidupan sehari-hari. Dengan demikian, evaluasi merupakan bagian integral dari proses pembelajaran karena hasil yang diperoleh dapat digunakan sebagai landasan untuk meningkatkan mutu pembelajaran sekaligus menentukan tindak lanjut yang diperlukan. Oleh sebab itu, keberhasilan pendidikan sangat bergantung pada kualitas sistem evaluasi yang diterapkan agar mampu menghasilkan informasi yang objektif, akurat, dan dapat dipertanggungjawabkan (Rohmah & Fakhriyah, 2026).

Dalam pembelajaran Hadis, kegiatan evaluasi memiliki kedudukan yang strategis karena pembelajaran tidak hanya berorientasi pada kemampuan menghafal teks hadis, melainkan juga pada pemahaman terhadap makna, isi, dan penerapannya dalam kehidupan sehari-hari. Melalui evaluasi, guru atau ustaz dapat mengetahui tingkat pemahaman santri terhadap materi yang telah dipelajari serta menilai kemampuan mereka dalam mengimplementasikan ajaran hadis. Selain itu, hasil evaluasi dapat dimanfaatkan untuk mengenali berbagai kendala belajar yang

dihadapi santri sehingga upaya perbaikan pembelajaran dapat dilakukan secara tepat. Oleh karena itu, evaluasi pada mata pelajaran Hadis berperan sebagai alat untuk mengukur keberhasilan pembelajaran sekaligus sebagai instrumen dalam meningkatkan mutu pendidikan Islam secara berkesinambungan. Pelaksanaannya pun perlu dilakukan secara terencana dan sistematis dengan menggunakan instrumen yang selaras dengan tujuan pembelajaran yang telah ditetapkan (Fadilah & Nugraha, 2026).

Salah satu instrumen yang umum digunakan dalam proses evaluasi adalah tes. Instrumen tes berfungsi untuk mengumpulkan data mengenai tingkat penguasaan peserta didik terhadap materi yang telah diajarkan. Ketepatan hasil evaluasi sangat dipengaruhi oleh kualitas instrumen yang digunakan karena instrumen yang baik akan menghasilkan informasi yang mencerminkan kemampuan peserta didik secara akurat. Sebaliknya, instrumen yang kurang berkualitas berpotensi menghasilkan data yang bias sehingga tidak mampu menggambarkan kemampuan peserta didik secara tepat. Oleh karena itu, penyusunan instrumen tes perlu memperhatikan berbagai indikator kualitas agar hasil penilaian yang diperoleh benar-benar merepresentasikan tingkat penguasaan materi peserta didik. Di samping itu, instrumen yang berkualitas dapat membantu pendidik dalam mengambil keputusan yang tepat berkaitan dengan proses pembelajaran maupun pencapaian hasil belajar peserta didik (Saputri et al., 2023).

Kualitas suatu instrumen tes ditentukan oleh beberapa kriteria pokok, yaitu validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Validitas berkaitan dengan kemampuan instrumen dalam mengukur aspek yang memang menjadi sasaran pengukuran, sedangkan reliabilitas menunjukkan konsistensi hasil yang diperoleh dari proses pengukuran tersebut. Selain itu, tingkat kesukaran digunakan untuk mengetahui tingkat kemudahan atau kesulitan suatu soal bagi peserta didik. Adapun daya pembeda mengacu pada kemampuan butir soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan kemampuan rendah. Keempat aspek tersebut merupakan indikator utama

dalam menentukan mutu instrumen tes. Apabila seluruh kriteria tersebut terpenuhi, instrumen akan mampu menghasilkan data yang lebih akurat, objektif, dan dapat dipercaya sebagai dasar untuk menilai hasil belajar peserta didik (Anshari et al., 2024; Rachmawati & Pradana, 2025).

Berbagai hasil penelitian menunjukkan bahwa analisis terhadap validitas, reliabilitas, tingkat kesukaran, dan daya pembeda merupakan prosedur yang penting dalam menjamin mutu instrumen evaluasi pembelajaran. Penelitian Saputri et al. (2023) mengungkapkan bahwa kualitas instrumen dapat diidentifikasi secara lebih tepat melalui analisis butir soal yang mencakup keempat aspek tersebut. Temuan Anshari et al. (2024) juga memperlihatkan bahwa sebagian besar butir soal Pendidikan Agama Islam telah memenuhi kriteria valid dan reliabel, walaupun masih ditemukan beberapa soal yang memerlukan penyempurnaan berdasarkan hasil analisis tingkat kesukaran dan daya pembeda. Sementara itu, Rohmah dan Fakhriyah (2026) menegaskan bahwa analisis kualitas instrumen memiliki peran penting dalam meningkatkan ketepatan pengukuran hasil belajar peserta didik. Meskipun demikian, sebagian besar penelitian tersebut masih berfokus pada mata pelajaran umum maupun Pendidikan Agama Islam di lingkungan sekolah formal. Penelitian yang secara khusus membahas kualitas instrumen tes hasil belajar pada mata pelajaran Hadis di pondok pesantren masih tergolong terbatas. Padahal, pembelajaran Hadis memiliki karakteristik tersendiri karena menuntut penguasaan materi sekaligus pemahaman dan penerapan kandungannya dalam kehidupan sehari-hari. Oleh sebab itu, diperlukan penelitian yang secara spesifik mengkaji kualitas instrumen tes hasil belajar Hadis di lingkungan pesantren guna memperoleh gambaran empiris mengenai validitas, reliabilitas, tingkat kesukaran, dan daya pembeda instrumen evaluasi yang digunakan.

Berdasarkan hasil observasi awal terhadap instrumen tes hasil belajar Hadis di Pondok Pesantren Darul Hadits Al-Aidarusiyah, diketahui bahwa instrumen yang digunakan dalam kegiatan evaluasi belum pernah dianalisis secara menyeluruh berdasarkan aspek validitas, reliabilitas, tingkat kesukaran, dan

daya pembeda. Kondisi tersebut berpotensi menimbulkan berbagai permasalahan, seperti adanya butir soal yang kurang sesuai dengan tujuan pembelajaran, tingkat kesukaran yang tidak proporsional, atau kemampuan soal yang belum optimal dalam membedakan tingkat kemampuan santri. Jika instrumen yang kualitasnya belum teruji digunakan secara berulang dalam proses evaluasi, hasil pengukuran yang diperoleh berisiko kurang akurat dan belum sepenuhnya mencerminkan kemampuan santri yang sesungguhnya. Oleh karena itu, analisis kualitas butir soal perlu dilakukan untuk mengetahui tingkat kelayakan instrumen sekaligus menjadi dasar dalam melakukan perbaikan terhadap butir-butir soal yang belum memenuhi standar instrumen yang baik (Rohmah & Fakhriyah, 2026; Saputri et al., 2023).

B. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian deskriptif. Pendekatan kuantitatif dipilih karena penelitian berfokus pada pengolahan data numerik hasil tes santri untuk mengetahui kualitas instrumen yang digunakan dalam pembelajaran Hadis, sedangkan penelitian deskriptif bertujuan menggambarkan secara sistematis karakteristik instrumen berdasarkan aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal. Penelitian dilaksanakan di Pondok Pesantren Darul Hadits Al-Aidarusiyah dengan subjek penelitian sebanyak 15 santri kelas VII MTs yang mengikuti tes hasil belajar Hadis. Instrumen yang dianalisis berupa tes hasil belajar yang terdiri atas 10 butir soal. Data penelitian diperoleh melalui teknik tes dan dokumentasi. Teknik tes digunakan untuk memperoleh data hasil belajar santri melalui instrumen yang telah disusun berdasarkan tujuan pembelajaran, sedangkan dokumentasi digunakan untuk mengumpulkan data berupa skor setiap butir soal, skor total santri, dan dokumen pendukung lainnya. Seluruh data kemudian disusun dalam lembar kerja Microsoft Excel untuk memudahkan proses pengolahan dan analisis data secara sistematis (Sugiyono, 2019).

Analisis data dilakukan dengan bantuan program Microsoft Excel melalui beberapa tahapan. Uji validitas instrumen dilakukan

menggunakan korelasi Pearson Product Moment untuk mengetahui hubungan antara skor setiap butir soal dengan skor total, dengan kriteria butir soal dinyatakan valid apabila nilai r hitung lebih besar daripada r tabel (0,553) dan memiliki korelasi positif (Azwar, 2017). Selanjutnya, reliabilitas instrumen dianalisis menggunakan rumus Cronbach Alpha, dengan kriteria reliabel apabila koefisien reliabilitas lebih besar dari 0,60 (Ghozali, 2021). Selain itu, dilakukan analisis tingkat kesukaran dengan kategori sukar (0,00–0,30), sedang (0,31–0,70), dan mudah (0,71–1,00) (Sudjana, 2017). Kemudian analisis daya pembeda untuk mengetahui kemampuan soal dalam membedakan santri berkemampuan tinggi dan rendah. Hasil analisis tersebut digunakan untuk menentukan kualitas instrumen tes hasil belajar Hadis sehingga dapat diketahui kelayakannya sebagai alat evaluasi pembelajaran yang valid, reliabel, dan efektif (Arikunto, 2018).

C.Hasil Penelitian dan Pembahasan

Instrumen yang menjadi objek analisis dalam penelitian ini berupa tes hasil belajar pada mata pelajaran Hadis yang terdiri atas 10 butir soal dan diberikan kepada 15 santri di Pondok Pesantren Darul Hadits Al-Aidarusiyah. Analisis dilakukan untuk mengkaji kualitas setiap butir soal berdasarkan aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Kegiatan analisis ini penting karena bertujuan memastikan bahwa instrumen yang digunakan mampu mengukur hasil belajar santri secara tepat, objektif, dan konsisten. Proses pengolahan data dilakukan melalui perhitungan manual dengan bantuan Microsoft Excel sehingga diperoleh informasi yang menggambarkan kualitas masing-masing butir soal yang digunakan dalam evaluasi pembelajaran Hadis.

A. Uji Validitas Instrumen

Hasil uji validitas instrumen terhadap 10 butir soal disajikan pada Tabel 1.

Tabel 1. Hasil Uji Validitas Instrumen

No	Butir Soal	r Hitung	Keterangan
1	S1	0,663	Valid
2	S2	0,946	Valid
3	S3	0,670	Valid
4	S4	0,955	Valid
5	S5	0,740	Valid
6	S6	0,955	Valid

No	Butir Soal	r Hitung	Keterangan
7	S7	0,663	Valid
8	S8	0,606	Valid
9	S9	0,651	Valid
10	S10	0,545	Tidak Valid

Berdasarkan hasil yang disajikan pada Tabel 1, diketahui bahwa dari 10 butir soal yang dianalisis terdapat 9 butir yang memenuhi kriteria validitas karena memiliki nilai r hitung yang lebih besar daripada nilai r tabel sebesar 0,553. Butir soal yang termasuk kategori valid mencakup soal nomor 1 hingga nomor 9. Temuan tersebut menunjukkan bahwa sebagian besar soal telah mampu mengukur kompetensi yang menjadi sasaran pengukuran sesuai dengan tujuan pembelajaran Hadis. Nilai korelasi tertinggi ditemukan pada soal nomor 4 dengan nilai r hitung sebesar 0,955. Adapun nilai korelasi terendah di antara butir soal yang masih memenuhi kriteria validitas terdapat pada soal nomor 8 dengan nilai sebesar 0,606. Hasil tersebut mengindikasikan bahwa butir-butir soal yang dinyatakan valid memiliki hubungan yang kuat dengan skor total sehingga layak digunakan sebagai instrumen untuk mengukur hasil belajar santri.

Di sisi lain, soal nomor 10 memperoleh nilai r hitung sebesar 0,545, yang berada di bawah nilai r tabel sebesar 0,553, sehingga dinyatakan tidak memenuhi kriteria validitas. Kondisi tersebut menunjukkan bahwa butir soal tersebut belum mampu mengukur kompetensi yang sama dengan keseluruhan instrumen secara optimal. Dengan demikian, soal nomor 10 memerlukan revisi atau penyempurnaan sebelum digunakan kembali dalam proses evaluasi pembelajaran. Secara keseluruhan, hasil pengujian validitas menunjukkan bahwa instrumen tes hasil belajar Hadis yang digunakan telah memiliki tingkat validitas yang baik, mengingat 90% dari seluruh butir soal yang dianalisis dinyatakan valid.

B. Uji Reliabilitas Instrumen

Hasil uji reliabilitas instrumen disajikan pada Tabel 2.

Tabel 2. Hasil Uji Reliabilitas Instrumen

Komponen	Nilai
Varians Total	1308,31
Jumlah Varians Butir	163,028
Cronbach's Alpha	0,973
Standar Reliabilitas	0,60
Keterangan	Reliabel

Berdasarkan hasil analisis, diperoleh nilai Cronbach's Alpha sebesar 0,973. Nilai tersebut berada jauh di atas batas minimum reliabilitas sebesar 0,60 sehingga instrumen dinyatakan reliabel. Nilai reliabilitas yang mendekati angka 1 menunjukkan bahwa instrumen memiliki tingkat konsistensi internal yang sangat tinggi. Hal ini mengindikasikan bahwa instrumen tes hasil belajar Hadis mampu menghasilkan pengukuran yang stabil dan konsisten apabila digunakan pada kelompok santri yang memiliki karakteristik yang relatif sama.

Tingginya nilai reliabilitas juga menunjukkan bahwa setiap butir soal memiliki keterkaitan yang baik dalam mengukur kompetensi yang sama. Dengan demikian, seluruh butir soal memberikan kontribusi positif terhadap kualitas instrumen secara keseluruhan. Temuan ini mengindikasikan bahwa instrumen tes hasil belajar Hadis layak digunakan sebagai alat evaluasi pembelajaran karena memiliki tingkat keandalan yang sangat baik dalam menghasilkan data hasil belajar santri yang dapat dipercaya.

C. Uji Tingkat Kesukaran Soal

Hasil analisis tingkat kesukaran butir soal disajikan pada Tabel 3.

Tabel 3. Hasil Uji Tingkat Kesukaran

Soal	Indeks Kesukaran	Kategori
S1	0,31	Sedang
S2	0,64	Sedang
S3	0,54	Sedang
S4	0,71	Mudah
S5	0,54	Sedang
S6	0,60	Sedang
S7	0,31	Sedang
S8	0,75	Mudah
S9	0,23	Sukar
S10	0,28	Sukar

Hasil analisis tingkat kesukaran menunjukkan bahwa mayoritas butir soal berada pada kategori sedang. Sebanyak enam butir soal, yaitu nomor 1, 2, 3, 5, 6, dan 7, termasuk dalam kategori tersebut. Soal dengan tingkat kesukaran sedang dipandang

sebagai bentuk soal yang ideal karena tidak terlalu mudah maupun terlalu sulit bagi santri. Kondisi ini memungkinkan instrumen memberikan pengukuran yang lebih seimbang sehingga kemampuan peserta didik dapat dinilai secara lebih objektif dan proporsional.

Di samping itu, terdapat dua butir soal yang tergolong mudah, yaitu soal nomor 4 dan 8. Kategori mudah mengindikasikan bahwa sebagian besar santri mampu memberikan jawaban yang benar pada kedua soal tersebut. Sebaliknya, dua butir lainnya, yakni soal nomor 9 dan 10, termasuk dalam kategori sukar. Tingginya tingkat kesukaran pada soal tersebut dapat dipengaruhi oleh kompleksitas materi yang diujikan atau penggunaan redaksi yang kurang mudah dipahami oleh santri. Secara umum, sebaran tingkat kesukaran instrumen dapat dinilai cukup baik karena didominasi oleh soal berkategori sedang. Meskipun demikian, butir soal yang berada pada kategori terlalu mudah maupun terlalu sukar tetap perlu ditelaah lebih lanjut untuk meningkatkan kualitas instrumen secara keseluruhan.

D. Uji Daya Pembeda Soal

Hasil analisis daya pembeda butir soal disajikan pada Tabel 4.

Tabel 4. Hasil Uji Daya Pembeda

Soal	Daya Pembeda	Kategori
S1	0,25	Cukup
S2	0,74	Sangat Baik
S3	0,13	Kurang Baik
S4	0,82	Sangat Baik
S5	0,13	Kurang Baik
S6	0,70	Sangat Baik
S7	0,25	Cukup
S8	0,10	Kurang Baik
S9	0,44	Baik
S10	0,03	Kurang Baik

Berdasarkan hasil analisis, butir soal nomor 2, 4, dan 6 termasuk dalam kategori memiliki daya pembeda sangat baik. Temuan ini mengindikasikan bahwa ketiga butir soal tersebut mampu membedakan secara efektif antara santri yang memiliki kemampuan tinggi dan santri yang memiliki kemampuan rendah. Selain itu, soal nomor 9 memperoleh kategori daya pembeda baik sehingga

tetap layak digunakan sebagai bagian dari instrumen evaluasi. Butir-butir soal tersebut dapat dipertahankan karena telah menjalankan fungsi pengukuran dengan optimal dalam mengidentifikasi perbedaan kemampuan peserta didik.

Sementara itu, soal nomor 1 dan 7 berada pada kategori daya pembeda cukup. Meskipun kedua butir tersebut masih dapat dimanfaatkan dalam kegiatan evaluasi, penyempurnaan tetap diperlukan agar kemampuannya dalam membedakan tingkat kemampuan peserta didik menjadi lebih optimal. Di sisi lain, soal nomor 3, 5, 8, dan 10 tergolong dalam kategori kurang baik karena memiliki nilai daya pembeda di bawah 0,20. Hasil tersebut menunjukkan bahwa keempat butir soal tersebut belum mampu membedakan kemampuan santri secara efektif. Oleh sebab itu, diperlukan revisi terhadap butir-butir soal tersebut, baik pada aspek materi, tingkat kesukaran, maupun redaksi soal, sehingga kualitas instrumen dapat ditingkatkan. Secara umum, instrumen tes hasil belajar Hadis menunjukkan kualitas yang cukup baik karena sebagian besar butir soal telah memenuhi kriteria validitas dan reliabilitas yang tinggi. Meskipun demikian, masih terdapat beberapa butir yang memerlukan perbaikan berdasarkan hasil analisis daya pembeda dan tingkat kesukaran.

D. Kesimpulan

Berdasarkan temuan penelitian yang telah dilakukan, dapat dinyatakan bahwa instrumen tes hasil belajar Hadis yang digunakan di Pondok Pesantren Darul Hadits Al-Aidarusiyah pada umumnya telah memenuhi persyaratan sebagai instrumen yang berkualitas. Hasil pengujian validitas memperlihatkan bahwa 9 dari 10 butir soal termasuk dalam kategori valid, sementara 1 butir soal belum memenuhi kriteria validitas sehingga perlu dilakukan penyempurnaan. Selain itu, pengujian reliabilitas menghasilkan koefisien Cronbach’s Alpha sebesar 0,973 yang mengindikasikan bahwa instrumen memiliki tingkat reliabilitas yang sangat tinggi, sehingga hasil pengukurannya dapat dinilai konsisten dan memiliki tingkat kepercayaan yang baik.

Hasil analisis tingkat kesukaran menunjukkan bahwa mayoritas butir soal berada pada kategori sedang,

sedangkan dua butir soal tergolong mudah dan dua butir lainnya termasuk kategori sukar. Kondisi tersebut mengindikasikan bahwa tingkat kesukaran instrumen telah tersusun secara cukup seimbang untuk mengukur kemampuan santri. Di sisi lain, analisis daya pembeda memperlihatkan bahwa beberapa soal memiliki kemampuan yang sangat baik dalam membedakan santri dengan kemampuan tinggi dan santri dengan kemampuan rendah. Meskipun demikian, masih ditemukan sejumlah butir soal yang memiliki daya pembeda rendah sehingga memerlukan perbaikan lebih lanjut.

Secara keseluruhan, instrumen tes hasil belajar Hadis yang menjadi objek penelitian dapat digunakan sebagai alat evaluasi pembelajaran karena secara umum telah memenuhi kriteria validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Namun demikian, beberapa butir soal, terutama yang belum valid dan yang memiliki daya pembeda rendah, perlu direvisi baik dari aspek substansi, redaksi, maupun tingkat kesukarannya agar kualitas instrumen semakin meningkat. Untuk penelitian berikutnya, disarankan agar melibatkan jumlah responden yang lebih banyak serta mengkaji instrumen pada materi Hadis yang lebih luas sehingga diperoleh informasi mengenai kualitas instrumen yang lebih menyeluruh dan representatif.

DAFTAR PUSTAKA

Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis validitas dan reliabilitas butir soal sumatif akhir semester ganjil mata pelajaran PAI. *Edukatif: Jurnal Ilmu Pendidikan, 6*(1), 964-975.

Arifin, Z. (2017). *Evaluasi pembelajaran*. Bandung, Indonesia: Remaja Rosdakarya.

Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan* (Edisi revisi). Jakarta, Indonesia: Bumi Aksara.

Azwar, S. (2017). *Reliabilitas dan validitas* (Edisi ke-4). Yogyakarta, Indonesia: Pustaka Pelajar.

Boini, B., Sahrul, S., Asmanah, A., & Sumin, S. (2023). Analisis instrumen tes penilaian Pendidikan Agama Islam menggunakan tes teori klasik. *At Turots: Jurnal Pendidikan Islam, 5*(1), 42-53.

Daryanto. (2012). *Evaluasi pendidikan*. Jakarta, Indonesia: Rineka Cipta.

Fadilah, N. A., & Nugraha, M. S. (2026). Rekonstruksi sistem penilaian pendidikan Islam berbasis hadis: Analisis model penilaian niat, akhlak, kinerja, sosial-emosional, iman-islam-ihsan. *Educan, 10*(1), 1-18.

Ghozali, I. (2021). *Aplikasi analisis multivariate dengan program IBM SPSS*. Semarang, Indonesia: Badan Penerbit Universitas Diponegoro.

Rachmawati, D., & Pradana, A. B. (2025). Analisis butir soal mata pelajaran ekonomi: Validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. *Jurnal Pendidikan Ekonomi (JUPE), 13*(3), 273-284.

Rohmah, A. N., & Fakhriyah, F. (2026). Analisis butir uji validitas, reliabilitas, tingkat kesukaran, daya pembeda soal sumatif Pendidikan Pancasila. *Jurnal Ilmiah Profesi Pendidikan, 11*(1), 756-764.

Saputri, H. A., Larasati, N. J., & Shaleh. (2023). Analisis instrumen asesmen: Validitas, reliabilitas, tingkat kesukaran dan daya beda butir soal. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang, 9*(5), 2231-2243.

Sudijono, A. (2015). *Pengantar evaluasi pendidikan*. Jakarta, Indonesia: Rajawali Pers.

Sudjana, N. (2017). *Penilaian hasil proses belajar mengajar*. Bandung, Indonesia: Remaja Rosdakarya.

Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Bandung, Indonesia: Alfabeta.