



This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2026 by the author

**Submitted:** 01-02-2026 **Revised:** 12-04-2026 **Accepted:** 15-05-2026 **Published:** 01-06-2026

## **Analisis Kualitas Instrumen Evaluasi Pembelajaran Pendidikan Agama Islam dalam Mengukur Kompetensi Peserta Didik di Sekolah Dasar**

**Syifa Khoirun Nisa<sup>1\*</sup>, Ahmad Hidayat<sup>2</sup>, Laila Sabrina<sup>3</sup>, Dzikra Fu'adiah<sup>4</sup>**

<sup>1234</sup>STAI Siliwangi Bandung, Indonesia

\*Correspondence: [ssyifaakn@gmail.com](mailto:ssyifaakn@gmail.com)

**Abstrak:** Evaluasi pembelajaran Pendidikan Agama Islam (PAI) pada jenjang Sekolah Dasar memerlukan instrumen yang mampu mengukur kompetensi peserta didik secara valid, reliabel, dan proporsional. Penelitian ini bertujuan untuk menganalisis kualitas instrumen evaluasi PAI berdasarkan validitas, reliabilitas, daya pembeda, dan tingkat kesukaran butir soal, serta menentukan kelayakan penggunaannya. Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif-evaluatif. Data berupa respons peserta didik terhadap 25 butir soal, terdiri atas 20 soal pilihan ganda dan 5 soal uraian, yang kemudian dianalisis menggunakan program ANATES. Hasil penelitian menunjukkan bahwa validitas butir soal bervariasi pada kategori tinggi, cukup, dan rendah, sedangkan reliabilitas instrumen mencapai 0,78 dan berada pada kategori tinggi. Dari aspek tingkat kesukaran, terdapat 7 butir (28%) berkategori sukar, 9 butir (36%) berkategori sedang, dan 9 butir (36%) berkategori mudah, yang menunjukkan distribusi relatif yang seimbang. Daya pembeda juga menunjukkan variasi dari kategori baik, cukup, hingga jelek. Berdasarkan integrasi hasil analisis, 12 butir (48%) dapat digunakan, 10 butir (40%) memerlukan revisi, dan 3 butir (12%) harus dibuang. Kebaruan penelitian ini terletak pada pengintegrasian empat indikator analisis butir dengan keputusan pedagogis terhadap instrumen evaluasi PAI pada jenjang Sekolah Dasar. Penelitian ini berkontribusi dalam memberikan dasar empiris bagi guru untuk melakukan seleksi, revisi, dan pengembangan instrumen evaluasi PAI secara sistematis sehingga mampu mendukung pengukuran kompetensi peserta didik secara lebih akurat.

**Kata Kunci:** *evaluasi pembelajaran; pendidikan agama islam; kualitas instrumen; analisis butir soal; kompetensi peserta didik.*

**Abstract:** Islamic Religious Education (IRE) assessment at the elementary school level requires instruments capable of measuring students' competencies in a valid, reliable, and proportional manner. This study aims to analyse the quality of an IRE assessment instrument based on item validity, reliability, discrimination power, and difficulty level, and to determine its suitability for use. The study employed a quantitative approach with a descriptive-evaluative method. The data consisted of students' responses to 25 test items, comprising 20 multiple-choice items and 5 essay items, which were analysed using the ANATES program. The findings indicate that item validity varied across high, moderate, and low categories, while the overall instrument reliability coefficient was 0.78, categorised as high. In terms of difficulty level, 7 items (28%) were categorised as difficult, 9 items (36%) as moderate, and 9 items (36%) as easy, indicating a relatively proportional distribution. The

discrimination power of the items also ranged from good to moderate to poor. Based on the integrated analysis, 12 items (48%) were considered usable, 10 items (40%) required revision, and 3 items (12%) had to be discarded. The novelty of this study lies in integrating four item-analysis indicators with pedagogical decisions regarding the quality of an IRE assessment instrument at the elementary school level. The study provides empirical evidence to help teachers systematically select, revise, and develop IRE assessment instruments, thereby supporting more accurate measurement of students' competencies.

**Keywords:** *learning evaluation; islamic religious education; assessment instruments; item analysis; student competencies.*

---

## Pendahuluan

Evaluasi pembelajaran merupakan bagian integral dari proses pendidikan karena berfungsi untuk mengetahui ketercapaian tujuan pembelajaran sekaligus menyediakan informasi mengenai perkembangan kemampuan peserta didik.<sup>1</sup> Dalam Pendidikan Agama Islam (PAI), fungsi evaluasi tidak berhenti pada pengukuran penguasaan materi, tetapi juga berkaitan dengan upaya untuk mengetahui ketercapaian kompetensi keagamaan yang menjadi tujuan pembelajaran. Mahmudi dan Ahmad menegaskan bahwa evaluasi PAI diperlukan untuk memantau proses dan kemajuan belajar peserta didik, meningkatkan efektivitas pembelajaran, mengidentifikasi kekuatan dan kelemahan peserta didik, serta menyediakan data bagi pengambilan keputusan pembelajaran.<sup>2</sup> Karena itu, instrumen evaluasi PAI harus memiliki kualitas pengukuran yang memadai agar informasi yang dihasilkan dapat dipertanggungjawabkan secara akademik dan pedagogis.

Kualitas instrumen evaluasi menjadi persoalan penting karena hasil penilaian sangat bergantung pada kualitas butir soal yang digunakan. Instrumen yang tidak valid, tidak reliabel, memiliki daya pembeda rendah, atau memiliki tingkat kesukaran yang tidak sesuai dapat menghasilkan informasi yang kurang akurat tentang kemampuan peserta didik. Boini, Sahrul, Asmanah, dan Sumin menempatkan validitas, reliabilitas, daya pembeda, dan tingkat kesukaran sebagai karakteristik penting pada instrumen penilaian PAI.<sup>3</sup> Mereka juga menjelaskan bahwa validitas berkaitan dengan kesesuaian instrumen dengan kurikulum, materi, dan tujuan pembelajaran, sedangkan karakteristik empiris butir diperlukan agar data hasil evaluasi dapat digunakan secara tepat.

Permasalahan tersebut menjadi semakin relevan pada jenjang sekolah dasar. Pitriani, Nugraha, dan Hilmiyanti, melalui analisis terhadap instrumen tes PAI pada tingkat sekolah dasar di Kecamatan Tigaraksa, menemukan bahwa 47,5% butir soal berada pada level kognitif C1, sementara secara umum masih terdapat banyak butir yang tidak valid dan indeks reliabilitas instrumen dikategorikan tidak reliabel.<sup>4</sup> Temuan tersebut memperlihatkan bahwa kualitas instrumen PAI pada jenjang dasar masih menjadi persoalan empiris, dan bahwa analisis butir diperlukan untuk memastikan bahwa soal yang digunakan benar-benar mampu mengukur kompetensi peserta didik.

---

<sup>1</sup> Dzikra Fuadiah, "Pengembangan Instrumen Tes Kemampuan Penalaran Kuantitatif Bagi Siswa Kelas VI SD/MI," *Jurnal Ilmiah Pendidikan Dasar* 9, no. 1 (January 31, 2022): 45, <https://doi.org/10.30659/pendas.9.1.45-67>.

<sup>2</sup> Ihwan Mahmudi and Raden Rizky Fahrial Ahmad, "Item Analysis of Islamic Education for Class VIII at Junior High School 1 Jetis Ponorogo," *Educan: Jurnal Pendidikan Islam* 4, no. 1 (June 28, 2020): 118, <https://doi.org/10.21111/educan.v4i1.4565>.

<sup>3</sup> Boini Boini et al., "Analisis Instrumen Tes Penilaian Pendidikan Agama Islam Menggunakan Tes Teori Klasik," *At Turots: Jurnal Pendidikan Islam* 5, no. 1 (February 11, 2024): 391–96, <https://doi.org/10.51468/jpi.v5i001.445>.

<sup>4</sup> Pipit Pitriani, Enung Nugraha, and Fitri Hilmiyanti, "PAI SD Test Quality Analysis: Validity and Reliability Review," *Mimbar Kampus: Jurnal Pendidikan Dan Agama Islam* 23, no. 1 (March 15, 2024): 85–97, <https://doi.org/10.47467/mk.v23i1.1372>.

Persoalan kualitas butir juga ditemukan pada jenjang pendidikan lainnya. Mulyani et al., menganalisis sepuluh soal evaluasi PAI dan Budi Pekerti pada buku ajar kelas X SMKN 3 Kota Serang. Hasilnya menunjukkan 8 soal sangat valid dan 2 soal tidak valid, dengan reliabilitas sebesar 0,74. Dari aspek tingkat kesukaran, 6 soal berkategori mudah dan 4 sedang, sementara dari aspek daya pembeda terdapat 5 soal sangat baik, 3 cukup dan perlu diperbaiki, serta 2 soal buruk.<sup>5</sup> Temuan tersebut menunjukkan bahwa dalam satu perangkat evaluasi dapat terjadi perbedaan kualitas antarbutir, sehingga penilaian terhadap instrumen tidak cukup dilakukan secara menyeluruh.

Kondisi serupa terlihat dalam penelitian Ningsih dan Karnadi mengenai kualitas soal Penilaian Akhir Tahun mata pelajaran PAI dan Budi Pekerti. Penelitian tersebut menggunakan analisis kuantitatif terhadap peserta didik di dua sekolah dan mengkaji tingkat kesukaran, daya pembeda, efektivitas pengecoh, serta reliabilitas menggunakan ANATES. Hasil penelitian menunjukkan bahwa pada salah satu sekolah terdapat 20 soal valid yang dapat disimpan dalam bank soal dan 20 soal tidak valid yang perlu diperbaiki, sedangkan pada sekolah lainnya terdapat 15 soal yang dinilai baik dan 15 soal yang perlu direvisi.<sup>6</sup> Temuan ini menunjukkan bahwa analisis butir dapat menghasilkan keputusan praktis mengenai soal yang dapat dipertahankan dan soal yang perlu diperbaiki.

Selain validitas dan reliabilitas, tingkat kesukaran dan daya pembeda menjadi aspek penting dalam menentukan kemampuan instrumen untuk menggambarkan variasi kompetensi peserta didik. Surul, dalam penelitian terhadap soal PAI pada jenjang sekolah dasar, menempatkan tingkat kesukaran, daya pembeda, dan efektivitas pengecoh sebagai indikator penting untuk menentukan kualitas soal. Analisis tersebut memperlihatkan bahwa setiap butir perlu ditelaah secara individual agar dapat diketahui apakah butir tersebut benar-benar mampu memberikan informasi mengenai kemampuan peserta didik.<sup>7</sup> Dengan demikian, analisis butir bukan hanya prosedur teknis, tetapi juga merupakan bagian dari proses penjaminan mutu instrumen penilaian PAI.

Kebutuhan tersebut secara konkret ditemukan dalam penelitian ini di SDN 02 Gunung Bentang. Instrumen evaluasi PAI yang dianalisis terdiri atas 25 butir soal, yaitu 20 soal pilihan ganda dan 5 soal uraian. Analisis menggunakan program ANATES menunjukkan adanya variasi pada karakteristik butir, terutama pada validitas dan daya pembeda. Beberapa butir memiliki koefisien korelasi pada kategori tinggi dan cukup sehingga dinyatakan valid, sedangkan sebagian lainnya berada pada kategori rendah dan belum mampu merepresentasikan kompetensi peserta didik secara optimal.

Hasil pengujian reliabilitas menunjukkan koefisien sebesar 0,78 yang berada pada kategori tinggi. Temuan ini menunjukkan bahwa secara keseluruhan instrumen memiliki konsistensi yang baik dalam mengukur hasil belajar peserta didik. Namun, reliabilitas tersebut tidak serta-merta menunjukkan bahwa seluruh butir memiliki kualitas yang sama karena masih ditemukan variasi dalam validitas dan daya pembeda. Kondisi tersebut

---

<sup>5</sup> Siti Ina Mulyani et al., "Analisis Kualitas Soal Evaluasi pada Buku Ajar PAI dan Budi Pekerti SMKN 3 Kota Serang," *ALFIHRIS: Jurnal Inspirasi Pendidikan* 3, no. 1 (December 21, 2024): 119–31, <https://doi.org/10.59246/alfihris.v3i1.1163>.

<sup>6</sup> Rahayu Ningsih and Karnadi Karnadi, "Quality of End-Year Assessment Question Items for Islamic Religious Education & Character Subjects," *Scaffolding: Jurnal Pendidikan Islam Dan Multikulturalisme* 4, no. 3 (December 18, 2023): 690–701, <https://doi.org/10.37680/scaffolding.v4i3.4168>.

<sup>7</sup> Rofiatul Surul, "Analisis Daya Pembeda Pengecoh dan Tingkat Kesukaran Soal Pendidikan Agama Islam dalam Ujian Akhir Semester Tingkat Sekolah Dasar: Studi Kasus Sdn Kaliwinig 07 Rambipuji Jember," *PIONIR: JURNAL PENDIDIKAN* 12, no. 3 (December 26, 2023): 200, <https://doi.org/10.22373/pjp.v12i3.20012>.

memperkuat pandangan Azwar bahwa reliabilitas merupakan karakteristik penting dalam suatu pengukuran, tetapi interpretasi kualitas instrumen tetap membutuhkan pengujian terhadap karakteristik item.<sup>8</sup>

Persoalan tersebut juga tampak pada tingkat kesukaran. Dari 25 butir soal, terdapat 7 butir (28%) berkategori sukar, 9 butir (36%) berkategori sedang, dan 9 butir (36%) berkategori mudah. Distribusi tersebut menunjukkan komposisi yang relatif seimbang sehingga instrumen tidak terkonsentrasi pada satu tingkat kesukaran. Kondisi ini penting dalam evaluasi PAI karena instrumen perlu memberikan informasi mengenai peserta didik dengan tingkat penguasaan kompetensi yang bervariasi. Dengan demikian, keberadaan soal mudah, sedang, dan sukar perlu dipahami sebagai bagian dari desain pengukuran, bukan semata-mata sebagai kategori statistik.

Permasalahan yang paling signifikan terlihat pada keputusan akhir terkait kualitas butir. Dari 25 butir, hanya 12 butir (48%) yang dapat digunakan, 10 butir (40%) digunakan dengan revisi, dan 3 butir (12%) harus dibuang. Pada soal pilihan ganda, terdapat 9 soal yang dapat digunakan, 9 soal yang perlu direvisi, dan 2 soal yang harus dibuang. Pada soal uraian, terdapat 3 soal yang dapat digunakan, 1 soal memerlukan revisi, dan 1 soal harus dibuang. Dengan demikian, 52% butir soal belum dapat digunakan secara langsung, meskipun reliabilitas instrumen secara keseluruhan berada dalam kategori tinggi.

Berdasarkan kondisi tersebut, penelitian ini memiliki posisi yang berbeda dari penelitian sebelumnya. Penelitian terdahulu umumnya menganalisis kualitas instrumen berdasarkan validitas, reliabilitas, tingkat kesukaran, daya pembeda, atau efektivitas pengecoh. Penelitian ini mengintegrasikan empat indikator utama: validitas, reliabilitas, daya pembeda, dan tingkat kesukaran. Dengan keputusan pedagogis terhadap butir soal, yaitu mempertahankan, merevisi, atau mengeliminasi butir. Dengan demikian, kebaruan penelitian tidak terletak pada penciptaan teknik analisis baru, melainkan pada pemanfaatan analisis empiris untuk memetakan kelayakan instrumen evaluasi PAI dan menghubungkannya dengan fungsi pengukuran kompetensi peserta didik sekolah dasar.

Berdasarkan persoalan tersebut, penelitian ini bertujuan untuk menganalisis kualitas instrumen evaluasi pembelajaran Pendidikan Agama Islam dalam mengukur kompetensi peserta didik di sekolah dasar berdasarkan validitas, reliabilitas, daya pembeda, dan tingkat kesukaran butir soal. Penelitian ini diharapkan memberikan kontribusi akademik bagi pengembangan kajian evaluasi dalam Pendidikan Agama Islam sekaligus kontribusi praktis bagi guru dalam mengembangkan instrumen penilaian yang lebih valid, reliabel, proporsional, dan mampu memberikan informasi yang akurat mengenai capaian kompetensi peserta didik.

## Metodologi

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif-evaluatif yang bertujuan untuk memperoleh gambaran empiris mengenai kualitas instrumen evaluasi pembelajaran Pendidikan Agama Islam (PAI) pada peserta didik sekolah dasar.<sup>9</sup> Penelitian dilaksanakan di SDN 02 Gunung Bentang dengan objek penelitian berupa instrumen evaluasi PAI yang terdiri atas 25 butir soal, meliputi 20 soal pilihan ganda dan 5 soal uraian. Data penelitian berupa respons atau jawaban peserta didik terhadap seluruh butir soal yang

---

<sup>8</sup> Saifududin Azwar, *Reliabilitas dan Validitas* (Yogyakarta: Pustaka Pelajar, 2017), h.76.

<sup>9</sup> Sugiyono, *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, Dan R&D* (Bandung: Alfabeta, 2019), h. 39.

kemudian dianalisis menggunakan program ANATES. Analisis dilakukan melalui 4 indikator utama, yaitu validitas butir, reliabilitas instrumen, daya pembeda, dan tingkat kesukaran. Validitas dianalisis berdasarkan hubungan skor setiap butir dengan skor total untuk mengetahui kemampuan butir dalam mengukur kompetensi yang menjadi sasaran evaluasi, sedangkan reliabilitas digunakan untuk mengetahui konsistensi instrumen secara menyeluruh. Daya pembeda dianalisis untuk mengetahui kemampuan setiap butir dalam membedakan peserta didik berdasarkan tingkat penguasaan kompetensinya, sementara tingkat kesukaran digunakan untuk mengetahui proporsi peserta didik yang mampu menjawab setiap butir dengan benar. Hasil analisis selanjutnya diinterpretasikan secara deskriptif dengan mengelompokkan butir berdasarkan karakteristik dan kelayakannya, kemudian digunakan untuk menentukan tindak lanjut berupa butir yang dapat digunakan, butir yang perlu direvisi, dan butir yang harus dieliminasi.<sup>10</sup> Pendekatan tersebut memungkinkan penelitian tidak hanya menggambarkan kualitas statistik instrumen, tetapi juga menghasilkan rekomendasi empiris untuk perbaikan dan pengembangan instrumen evaluasi PAI di sekolah dasar.

## Hasil Penelitian dan Pembahasan

### Validitas dan Reliabilitas Instrumen Evaluasi PAI

Analisis validitas dilakukan terhadap 25 butir soal menggunakan program ANATES<sup>11</sup> melalui korelasi antara skor setiap butir dan skor total tes. Hasil analisis menunjukkan bahwa koefisien korelasi butir berada pada kategori tinggi, cukup, dan rendah. Butir dengan korelasi tinggi dan cukup dinyatakan valid karena memiliki hubungan yang memadai dengan kemampuan yang diukur, sedangkan butir dengan korelasi rendah menunjukkan keterbatasan dalam merepresentasikan kompetensi peserta didik.

Temuan tersebut relevan dengan Boini et al. yang menjelaskan bahwa validitas instrumen PAI tidak hanya berkaitan dengan perhitungan empiris, tetapi juga dengan kesesuaian isi instrumen dengan kurikulum, materi, dan tujuan pembelajaran.<sup>12</sup> Dengan demikian, butir yang secara empiris memiliki validitas rendah perlu dikaji kembali dari sisi konstruksi, materi, dan indikator kompetensi. Dalam penelitian ini, butir yang tidak valid tidak seharusnya digunakan kembali karena dapat mengurangi ketepatan informasi yang diberikan oleh instrumen mengenai kompetensi peserta didik. Dari aspek reliabilitas, instrumen memperoleh koefisien 0,78 dan termasuk dalam kategori tinggi. Nilai tersebut menunjukkan konsistensi internal instrumen yang baik. Temuan ini relatif lebih baik apabila dibandingkan dengan penelitian Pitriani et al., yang menemukan bahwa instrumen PAI sekolah dasar yang diteliti masih memiliki banyak butir tidak valid dan indeks reliabilitas yang dikategorikan tidak reliabel.<sup>13</sup> Perbedaan tersebut menunjukkan bahwa kualitas instrumen PAI pada jenjang sekolah dasar tidak seragam dan sangat bergantung pada proses penyusunan serta pengujian instrumen.

---

<sup>10</sup> Nasrudin Nasrudin and Nina Nursari, *Pengantar Ilmu Hukum (Membedah Logika Dan Etika Keadilan)* (Bandung: CV. Widina Media Utama, 2025), <https://repository.penerbitwidina.com/publications/618115/pengantar-ilmu-hukum-membedah-logika-dan-etika-keadilan#cite>.

<sup>11</sup> Program komputer yang digunakan guru atau peneliti untuk menganalisis kualitas butir soal ujian atau tes secara otomatis. Aplikasi ini dapat menghitung tingkat kesukaran, daya pembeda, validitas, reliabilitas, dan efektivitas pengecoh (distractor) untuk jenis soal pilihan ganda maupun uraian

<sup>12</sup> Boini et al., "Analisis Instrumen Tes Penilaian Pendidikan Agama Islam Menggunakan Tes Teori Klasik."

<sup>13</sup> Pipit Pitriani, Enung Nugraha, and Fitri Hilmiyanti, "PAI SD Test Quality Analysis: Validity and Reliability Review."

Temuan penelitian ini juga dapat dibandingkan dengan penelitian Mulyani et al., yang memperoleh reliabilitas 0,74 pada instrumen PAI dan Budi Pekerti. Meskipun sama-sama berada pada kategori tinggi, penelitian tersebut masih menemukan 2 dari 10 soal tidak valid dan 2 soal memiliki daya pembeda yang rendah.<sup>14</sup> Hal tersebut memperkuat temuan penelitian ini bahwa reliabilitas instrumen secara keseluruhan tidak cukup dijadikan dasar untuk menyatakan bahwa semua butir berkualitas. Evaluasi harus tetap dilanjutkan dengan analisis karakteristik pada setiap item.

### **Daya Pembeda sebagai Indikator**

#### **Kemampuan Instrumen Mengukur Perbedaan Kompetensi**

Daya pembeda merupakan salah satu indikator penting dalam menentukan kualitas butir soal karena menunjukkan kemampuan butir dalam membedakan peserta didik berdasarkan tingkat penguasaan kompetensi yang diukur.<sup>15</sup> Hasil analisis terhadap 25 butir soal evaluasi PAI di SDN 02 Gunung Bentang menunjukkan bahwa daya pembeda soal berada pada kategori baik, cukup, dan jelek. Variasi tersebut menunjukkan bahwa kemampuan setiap butir dalam mengidentifikasi perbedaan tingkat penguasaan peserta didik belum sepenuhnya seragam. Butir dengan daya pembeda baik memiliki kemampuan yang lebih optimal dalam membedakan peserta didik yang memiliki penguasaan kompetensi relatif tinggi dan rendah, sedangkan butir dengan daya pembeda rendah atau jelek belum memberikan informasi yang memadai mengenai perbedaan tersebut. Dengan demikian, hasil analisis daya pembeda memberikan informasi penting bahwa kualitas instrumen tidak dapat ditentukan hanya berdasarkan jumlah soal atau reliabilitas instrumen secara keseluruhan, tetapi perlu dilihat hingga karakteristik masing-masing butir.

Temuan tersebut secara konseptual sejalan dengan Mulyani et al., yang menempatkan differentiation atau daya pembeda sebagai salah satu karakteristik penting instrumen evaluasi Pendidikan Agama Islam, bersama dengan validitas, reliabilitas, dan tingkat kesukaran.<sup>16</sup> Daya pembeda digunakan untuk melihat kemampuan butir dalam membedakan kelompok peserta didik dengan tingkat kemampuan yang berbeda. Sehingga, suatu soal yang memiliki daya pembeda yang baik lebih potensial digunakan untuk menghasilkan informasi yang dapat membedakan capaian kompetensi peserta didik. Sebaliknya, butir dengan daya pembeda rendah perlu ditelaah kembali karena dapat menurunkan ketepatan interpretasi hasil evaluasi. Dalam konteks pembelajaran PAI, kemampuan membedakan tersebut memiliki konsekuensi pedagogis yang penting. Evaluasi tidak semestinya hanya menghasilkan skor akhir, tetapi juga membantu guru memperoleh informasi mengenai peserta didik yang telah menguasai kompetensi dan peserta didik yang masih membutuhkan penguatan. Oleh sebab itu, butir dengan daya pembeda baik dapat dipertahankan sebagai bagian dari instrumen atau bank soal, sedangkan butir dengan daya pembeda rendah perlu dianalisis lebih lanjut dari aspek materi, konstruksi, redaksi, dan kesesuaiannya dengan indikator pembelajaran. Posisi ini sejalan dengan Fajariani dan Hanum yang menegaskan bahwa analisis butir pada instrumen sekolah dasar perlu mencakup validitas, reliabilitas,

---

<sup>14</sup> Siti Ina Mulyani et al., "Analisis Kualitas Soal Evaluasi pada Buku Ajar PAI dan Budi Pekerti SMKN 3 Kota Serang."

<sup>15</sup> Dessy Trisnawati, Ririn Widiyarsari, and Viarti Eminita, "Pengaruh Realistic Mathematics Education (RME) Berbasis Multisensori Terhadap Pemahaman Konsep Dan Disposisi Matematis Siswa Sekolah Dasar," *Pendas: Jurnal Ilmiah Pendidikan Dasar* 11, no. 01 (January 17, 2026): 169–85, <https://doi.org/10.23969/jp.v11i01.40469>.

<sup>16</sup> Siti Ina Mulyani et al., "Analisis Kualitas Soal Evaluasi pada Buku Ajar PAI dan Budi Pekerti SMKN 3 Kota Serang."

tingkat kesukaran, daya pembeda, dan efektivitas pengecoh karena kualitas butir yang rendah dapat memengaruhi akurasi penilaian terhadap kompetensi peserta didik.<sup>17</sup>

Daya pembeda juga tidak dapat ditafsirkan secara terpisah dari tingkat kesukaran. Butir yang terlalu mudah dapat dijawab dengan benar oleh hampir seluruh peserta didik, sehingga kemampuan diskriminasi menjadi rendah. Sebaliknya, butir yang terlalu sulit dapat menyebabkan sebagian besar peserta didik memberikan jawaban yang salah dan juga menghasilkan daya pembeda yang kurang optimal. Oleh karena itu, variasi daya pembeda dalam penelitian ini perlu dibaca bersamaan dengan distribusi tingkat kesukaran yang menunjukkan bahwa 7 butir (28%) berada pada kategori sukar, 9 butir (36%) pada kategori sedang, dan 9 butir (36%) pada kategori mudah. Komposisi tersebut menunjukkan bahwa instrumen memiliki variasi tingkat kesukaran, sehingga rendahnya daya pembeda pada beberapa butir tidak dapat langsung dianggap sebagai persoalan tunggal pada aspek diskriminasi, tetapi perlu ditelusuri melalui hubungan antara tingkat kesukaran, konstruksi soal, dan kemampuan pengecoh.

Perspektif tersebut diperkuat oleh penelitian Faqih, Maghfiroh, dan Qomari terkait evaluasi akhir semester di Madrasah Ibtidaiyah Thoriqul Huda Malang.<sup>18</sup> Penelitian tersebut menggunakan ANATES untuk menganalisis reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh. Hasilnya menunjukkan bahwa meskipun instrumen memiliki reliabilitas tinggi sebesar 0,78, karakteristik daya pembeda antarbutir tetap bervariasi: 22 butir memiliki daya pembeda baik hingga sangat baik, 14 butir berkategori kurang, dan 2 butir memiliki daya pembeda negatif. Penelitian tersebut juga menghasilkan keputusan bahwa 19 butir dapat digunakan kembali, sedangkan 21 butir memerlukan revisi dan pengembangan lebih lanjut. Temuan tersebut menunjukkan bahwa reliabilitas tinggi pada tingkat instrumen tidak secara otomatis menjamin bahwa seluruh butir memiliki kemampuan diskriminasi yang baik. Hal yang sama terlihat pada penelitian ini, ketika reliabilitas instrumen mencapai 0,78, tetapi daya pembeda antarbutir masih menunjukkan variasi.

Perbandingan tersebut memberikan posisi yang menarik bagi hasil penelitian di SDN 02 Gunung Bentang. Nilai reliabilitas yang sama-sama mencapai 0,78 menunjukkan adanya kemiripan konsistensi instrumen dengan penelitian Faqih et al., tetapi karakteristik item tetap perlu dianalisis secara individual.<sup>19</sup> Perbedaan karakteristik tersebut menunjukkan bahwa reliabilitas sebaiknya dipahami sebagai informasi mengenai konsistensi instrumen secara keseluruhan, sedangkan daya pembeda memberikan informasi yang lebih spesifik mengenai kemampuan masing-masing butir dalam mengidentifikasi perbedaan kompetensi peserta didik. Dengan demikian, kedua indikator tersebut saling melengkapi dan tidak dapat saling menggantikan.

Temuan penelitian ini juga dapat dibandingkan dengan Putri et al., yang menganalisis kualitas soal ujian Akidah Akhlak pada jenjang TPQ/TQA.<sup>20</sup> Penelitian tersebut secara eksplisit menggunakan pendekatan kuantitatif deskriptif dan menganalisis validitas, reliabilitas, daya pembeda, serta tingkat kesukaran soal yang dikerjakan oleh peserta didik.

---

<sup>17</sup> Kiki Fajariani and Erlia Hanum, "Evaluasi Kualitas Instrumen Tes Pilihan Ganda Pada Mata Pelajaran IPA Kelas V Sekolah Dasar Berdasarkan Analisis Butir Soal," *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan* 4, no. 2 (October 31, 2025): 9224–29, <https://doi.org/10.31004/jerkin.v4i2.2984>.

<sup>18</sup> Faqih Zainal Abidin Faqih, Maghfiroh, and Nur Qomari, "Item Analysis of the Final Semester Evaluation in Madrasah Ibtidaiyah Thotiqul Huda Malang," *Innovasi: Jurnal Inovasi Pendidikan* 11, no. 2 (August 3, 2025): 101–13, <https://doi.org/10.64540/82p9tn94>.

<sup>19</sup> Faqih, Maghfiroh, and Qomari.

<sup>20</sup> Irma Rezkia Putri et al., "Analysis of the Quality of Akidah Akhlak Exam Questions at TPQ/TQA," *Ri'ayatu Al-Qur'an* 7, no. 2 (July 28, 2025): 7–13, <https://doi.org/10.62990/riqu.v7i2.139>.

Kesamaan indikator analisis tersebut menunjukkan bahwa daya pembeda merupakan komponen yang relevan dalam evaluasi pembelajaran keagamaan, bukan hanya pada mata pelajaran umum.<sup>21</sup> Dalam konteks penelitian ini, daya pembeda menjadi penting karena instrumen PAI dituntut mampu memberikan gambaran mengenai variasi penguasaan kompetensi keagamaan peserta didik. Dari sisi pengembangan instrumen, butir dengan daya pembeda rendah tidak selalu harus langsung dibuang. Butir tersebut terlebih dahulu perlu ditelaah untuk mengetahui sumber kelemahannya. Apabila persoalan berasal dari redaksi yang ambigu, materi yang kurang sesuai dengan indikator, pengecoh yang tidak berfungsi, atau tingkat kesukaran yang tidak proporsional, maka butir tersebut masih dapat diperbaiki melalui revisi. Sebaliknya, apabila kelemahan bersifat mendasar dan setelah ditelaah tidak lagi mampu mengukur kompetensi yang dituju, butir tersebut lebih tepat untuk dieliminasi. Pendekatan semacam ini sejalan dengan Fajariani dan Hanum yang merekomendasikan analisis butir secara rutin serta pemanfaatan teknologi analisis untuk membantu guru menentukan butir yang dapat dipertahankan, direvisi, atau dibuang.<sup>22</sup>

Dalam penelitian ini, hasil daya pembeda kemudian menjadi salah satu dasar untuk menentukan tindak lanjut instrumen. Dari keseluruhan 25 butir soal, terdapat 12 butir (48%) yang dapat digunakan, 10 butir (40%) yang memerlukan revisi, dan 3 butir (12%) yang harus dibuang. Data tersebut menunjukkan bahwa kualitas daya pembeda tidak hanya memiliki arti statistik, tetapi juga berimplikasi langsung pada pengembangan instrumen evaluasi PAI. Butir yang memiliki karakteristik baik dapat dipertahankan sebagai bagian dari bank soal, sedangkan butir yang memiliki kelemahan perlu diperbaiki agar pada penggunaan berikutnya dapat memberikan informasi yang lebih akurat mengenai tingkat penguasaan kompetensi peserta didik. Sehingga, daya pembeda dalam penelitian ini dapat diposisikan sebagai jembatan antara kualitas statistik soal dan fungsi pedagogis evaluasi PAI. Analisis daya pembeda memungkinkan guru mengetahui apakah suatu soal benar-benar mampu membedakan peserta didik berdasarkan tingkat kompetensi mereka. Temuan penelitian juga memperlihatkan bahwa instrumen dengan reliabilitas tinggi sekalipun masih membutuhkan evaluasi pada tingkat item. Oleh karena itu, pengembangan instrumen evaluasi PAI perlu dilakukan melalui siklus penyusunan soal, uji coba, analisis validitas dan reliabilitas, analisis daya pembeda dan tingkat kesukaran, revisi, serta seleksi soal untuk membentuk bank soal. Dengan mekanisme tersebut, analisis daya pembeda tidak berhenti sebagai prosedur statistik, melainkan menjadi bagian dari penjaminan mutu evaluasi PAI yang berorientasi pada ketepatan dalam mengukur kompetensi peserta didik.

### **Tingkat Kesukaran dan Proporsionalitas Instrumen**

Analisis terhadap 25 butir menghasilkan 7 soal sukar (28%), 9 soal sedang (36%), dan 9 soal mudah (36%). Distribusi tersebut menunjukkan bahwa instrumen memiliki komposisi tingkat kesukaran yang relatif seimbang. Tidak terdapat dominasi ekstrem pada kategori mudah maupun sukar, sehingga instrumen berpotensi memberikan informasi yang lebih beragam mengenai tingkat penguasaan kompetensi peserta didik. Dalam evaluasi PAI, soal mudah dapat digunakan untuk mengetahui penguasaan kompetensi dasar, soal sedang untuk mengukur penguasaan materi secara lebih representatif, dan soal sukar untuk

---

<sup>21</sup> Dzikra Fu'adiah, Ayu Alfia, and Pipih Nurhayati, "Penggunaan Jarimatika Pada Siswa Madrasah Ibtidaiyah An-Na'im Sebagai Metode Pembelajaran Menarik Dalam Meningkatkan Hasil Belajar Matematika," *ISLAMICA* 8, no. 1 (June 19, 2024): 1–10, <https://doi.org/10.59908/islamica.v8i1.83>.

<sup>22</sup> Kiki Fajariani dan Erlia Hanum, "Evaluasi Kualitas Instrumen Tes Pilihan Ganda pada Mata Pelajaran IPA Kelas V Sekolah Dasar Berdasarkan Analisis Butir Soal."

mengidentifikasi peserta didik dengan tingkat penguasaan yang lebih tinggi. Dengan demikian, soal sukar tidak dapat langsung dianggap sebagai soal yang buruk, sebagaimana soal mudah tidak otomatis harus dibuang. Penilaian harus mempertimbangkan hubungan antara tingkat kesukaran, indikator, dan kompetensi yang ingin dicapai.

Temuan ini dapat dibandingkan dengan Mulyani et al., yang menemukan bahwa dari 10 soal PAI yang dianalisis terdapat 6 soal mudah dan 4 soal sedang, tanpa soal sukar.<sup>23</sup> Penelitian tersebut kemudian merekomendasikan penambahan soal yang lebih menantang untuk mengembangkan kemampuan berpikir peserta didik. Jika dibandingkan dengan temuan tersebut, komposisi penelitian di SDN 02 Gunung Bentang relatif lebih beragam karena terdapat 28% soal yang sukar. Temuan Ningsih dan Karnadi juga menunjukkan bahwa analisis tingkat kesukaran merupakan bagian penting dari penilaian kualitas soal PAI dan Budi Pekerti, bersama dengan daya pembeda, efektivitas pengecoh, dan reliabilitas.<sup>24</sup> Penggunaan ANATES dalam penelitian mereka menunjukkan bahwa karakteristik empiris soal dapat digunakan untuk menentukan apakah soal layak disimpan dalam bank soal atau perlu diperbaiki. Hal tersebut sejalan dengan penelitian ini bahwa proporsi tingkat kesukaran perlu dipertahankan melalui analisis berkala.

### **Kelayakan Instrumen dan Kontribusinya terhadap Peningkatan Evaluasi PAI**

Hasil integrasi analisis validitas, reliabilitas, daya pembeda, dan tingkat kesukaran menghasilkan keputusan yang konkret untuk 25 butir soal. Sebanyak 12 butir (48%) dapat digunakan, 10 butir (40%) digunakan dengan revisi, dan 3 butir (12%) dibuang. Dengan demikian, meskipun instrumen memiliki reliabilitas tinggi sebesar 0,78, masih terdapat 52% butir yang belum dapat digunakan secara langsung. Jika dilihat berdasarkan bentuk soal, dari 20 soal pilihan ganda, sebanyak 9 butir digunakan, 9 butir direvisi, dan 2 butir dibuang. Pada 5 soal uraian, 3 digunakan, 1 direvisi, dan 1 dibuang. Data tersebut menunjukkan bahwa persoalan kualitas tidak hanya terdapat pada soal objektif, tetapi juga pada soal uraian. Karena itu, pengembangan instrumen PAI perlu memperhatikan karakteristik kedua bentuk soal tersebut.

Temuan ini memiliki kesesuaian yang kuat dengan penelitian Ningsih dan Karnadi.<sup>25</sup> Penelitian mereka juga menghasilkan keputusan praktis berupa soal yang dapat disimpan dalam bank soal dan soal yang perlu diperbaiki. Pada dua sekolah yang diteliti, masing-masing ditemukan kelompok soal yang valid/baik dan kelompok soal yang perlu direvisi. Dengan demikian, penelitian SDN 02 Gunung Bentang tidak hanya menghasilkan deskripsi statistik, tetapi juga peta tindak lanjut instrumen yang dapat langsung dimanfaatkan oleh guru. Hal serupa terlihat dalam penelitian Mahmudi dan Ahmad terkait PAI.<sup>26</sup> Mereka menempatkan kualitas pembuatan soal, pelaksanaan tes, serta pengolahan hasil sebagai bagian dari keberhasilan evaluasi. Evaluasi PAI juga diarahkan untuk mengidentifikasi kekuatan dan kelemahan peserta didik serta menyediakan data bagi pengambilan keputusan pembelajaran. Berdasarkan perspektif tersebut, keputusan untuk mempertahankan,

---

<sup>23</sup> Siti Ina Mulyani et al., "Analisis Kualitas Soal Evaluasi pada Buku Ajar PAI dan Budi Pekerti SMKN 3 Kota Serang."

<sup>24</sup> Ningsih and Karnadi, "Quality of End-Year Assessment Question Items for Islamic Religious Education & Character Subjects."

<sup>25</sup> Ningsih and Karnadi.

<sup>26</sup> Mahmudi and Ahmad, "Item Analysis of Islamic Education For Class VIII at Junior High School 1 Jetis Ponorogo."

merevisi, atau membuang butir dalam penelitian ini bukan sekadar keputusan statistik, melainkan bagian dari proses peningkatan mutu pembelajaran PAI.

Secara konseptual, hasil penelitian menunjukkan bahwa reliabilitas yang tinggi tidak identik dengan seluruh butir yang berkualitas tinggi. Instrumen penelitian memiliki reliabilitas 0,78, tetapi hasil analisis item menunjukkan bahwa 48% soal dapat digunakan langsung, 40% memerlukan revisi, dan 12% harus dibuang. Temuan ini memperkuat prinsip analisis instrumen yang dikemukakan Boini et al., bahwa karakteristik instrumen harus dilihat melalui beberapa indikator, bukan hanya satu ukuran kualitas. Dengan demikian, kontribusi utama penelitian ini terletak pada integrasi antara analisis statistik butir dan keputusan pedagogis. Analisis tidak berhenti pada pertanyaan apakah soal valid atau tidak, tetapi diteruskan pada pertanyaan tentang apa yang harus dilakukan terhadap soal tersebut. Butir yang memenuhi kriteria dipertahankan sebagai bagian dari bank soal, butir yang memiliki kelemahan tertentu direvisi, sedangkan butir yang tidak memenuhi kriteria dieliminasi. Pendekatan ini menjadikan analisis butir sebagai mekanisme *quality assurance* dalam evaluasi PAI.

Kebaruan tersebut menjadi semakin jelas jika dibandingkan dengan penelitian Pitriani et al., yang menyoroti masih rendahnya kualitas sebagian instrumen PAI sekolah dasar,<sup>27</sup> penelitian Mulyani et al. yang menunjukkan variasi validitas dan daya pembeda,<sup>28</sup> serta penelitian Ningsih dan Karnadi yang menghasilkan klasifikasi soal yang baik dan perlu diperbaiki.<sup>29</sup> Penelitian ini melanjutkan temuan-temuan tersebut dalam konteks SDN 02 Gunung Bentang dengan menggabungkan empat indikator kualitas dan menerjemahkan hasilnya menjadi keputusan tindak lanjut untuk seluruh 25 butir. Dengan demikian, instrumen evaluasi PAI di SDN 02 Gunung Bentang dapat dikatakan memiliki fondasi kualitas yang cukup baik, terutama dari aspek reliabilitas sebesar 0,78 dan distribusi tingkat kesukaran yang relatif seimbang. Namun, kualitas tersebut belum merata pada tingkat butir karena 52% soal masih memerlukan revisi atau penghapusan. Temuan ini menunjukkan bahwa peningkatan mutu evaluasi PAI perlu dilakukan secara berkelanjutan melalui siklus penyusunan–uji coba–analisis–revisi–seleksi–pemanfaatan kembali instrumen.

## Simpulan

Penelitian ini menunjukkan bahwa kualitas instrumen evaluasi pembelajaran Pendidikan Agama Islam (PAI) di SDN 02 Gunung Bentang cukup baik secara keseluruhan, tetapi belum seluruh butir memenuhi kriteria kelayakan secara optimal. Analisis terhadap 25 butir soal, yang terdiri atas 20 soal pilihan ganda dan 5 soal uraian, menunjukkan bahwa validitas dan daya pembeda butir masih bervariasi, sedangkan reliabilitas instrumen mencapai 0,78 yang berada pada kategori tinggi. Dari aspek tingkat kesukaran, terdapat 7 butir (28%) berkategori sukar, 9 butir (36%) berkategori sedang, dan 9 butir (36%) berkategori mudah, sehingga distribusinya relatif seimbang. Berdasarkan integrasi hasil analisis validitas, reliabilitas, daya pembeda, dan tingkat kesukaran, diperoleh 12 butir (48%) yang dapat digunakan, 10 butir (40%) yang memerlukan revisi, dan 3 butir (12%) yang harus dibuang. Temuan ini menegaskan bahwa reliabilitas instrumen yang tinggi tidak secara otomatis menunjukkan bahwa seluruh butir memiliki kualitas yang baik, sehingga analisis

---

<sup>27</sup> Pipit Pitriani, Enung Nugraha, and Fitri Hilmiyanti, "PAI SD Test Quality Analysis: Validity and Reliability Review."

<sup>28</sup> Siti Ina Mulyani et al., "Analisis Kualitas Soal Evaluasi Pada Buku Ajar PAI Dan Budi Pekerti SMKN 3 Kota Serang."

<sup>29</sup> Ningsih and Karnadi, "Quality of End-Year Assessment Question Items for Islamic Religious Education & Character Subjects."

item secara komprehensif tetap diperlukan. Kebaruan penelitian terletak pada pemanfaatan hasil analisis butir tidak hanya untuk mendeskripsikan karakteristik statistik instrumen, tetapi juga sebagai dasar pengambilan keputusan pedagogis berupa mempertahankan, merevisi, atau mengeliminasi butir. Dengan demikian, analisis kualitas instrumen secara berkala dapat menjadi bagian dari penjaminan mutu evaluasi PAI dan membantu guru mengembangkan instrumen yang lebih valid, reliabel, proporsional, serta mampu mengukur kompetensi peserta didik dengan lebih akurat.

## Referensi

- Azwar, Saifudin. *Reliabilitas Dan Validitas*. Yogyakarta: Pustaka Pelajar, 2017.
- Boini, Boini, Sahrul Sahrul, Asmanah Asmanah, and Sumin Sumin. "Analisis Instrumen Tes Penilaian Pendidikan Agama Islam Menggunakan Tes Teori Klasik." *At Turots: Jurnal Pendidikan Islam* 5, no. 1 (February 11, 2024): 391–96. <https://doi.org/10.51468/jpi.v5i001.445>.
- Dessy Trisnawati, Ririn Widiyarsari, and Viarti Eminita. "Pengaruh Realistic Mathematics Education (RME) Berbasis Multisensori Terhadap Pemahaman Konsep Dan Disposisi Matematis Siswa Sekolah Dasar." *Pendas : Jurnal Ilmiah Pendidikan Dasar* 11, no. 01 (January 17, 2026): 169–85. <https://doi.org/10.23969/jp.v11i01.40469>.
- Faqih, Faqih Zainal Abidin, Maghfiroh, and Nur Qomari. "Item Analysis of the Final Semester Evaluation in Madrasah Ibtidaiyah Thotiqul Huda Malang." *Innovasi: Jurnal Inovasi Pendidikan* 11, no. 2 (August 3, 2025): 101–13. <https://doi.org/10.64540/82p9tn94>.
- Fu'adiah, Dzikra, Ayu Alfia, and Pipih Nurhayati. "Penggunaan Jarimatika Pada Siswa Madrasah Ibtidaiyah An-Na'im Sebagai Metode Pembelajaran Menarik Dalam Meningkatkan Hasil Belajar Matematika." *ISLAMICA* 8, no. 1 (June 19, 2024): 1–10. <https://doi.org/10.59908/islamica.v8i1.83>.
- Fuadiah, Dzikra. "Pengembangan Instrumen Tes Kemampuan Penalaran Kuantitatif Bagi Siswa Kelas VI SD/MI." *Jurnal Ilmiah Pendidikan Dasar* 9, no. 1 (January 31, 2022): 45. <https://doi.org/10.30659/pendas.9.1.45-67>.
- Kiki Fajariani, and Erlia Hanum. "Evaluasi Kualitas Instrumen Tes Pilihan Ganda Pada Mata Pelajaran IPA Kelas V Sekolah Dasar Berdasarkan Analisis Butir Soal." *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan* 4, no. 2 (October 31, 2025): 9224–29. <https://doi.org/10.31004/jerkin.v4i2.2984>.
- Mahmudi, Ihwan, and Raden Rizky Fahrial Ahmad. "Item Analysis of Islamic Education For Class VIII at Junior High School 1 Jetis Ponorogo." *Educan : Jurnal Pendidikan Islam* 4, no. 1 (June 28, 2020): 118. <https://doi.org/10.21111/educan.v4i1.4565>.
- Nasrudin, Nasrudin, and Nina Nursari. *Pengantar Ilmu Hukum (Membedah Logika Dan Etika Keadilan)*. Bandung: CV. Widina Media Utama, 2025. <https://repository.penerbitwidina.com/publications/618115/pengantar-ilmu-hukum-membedah-logika-dan-etika-keadilan#cite>.
- Ningsih, Rahayu, and Karnadi Karnadi. "Quality of End-Year Assessment Question Items for Islamic Religious Education & Character Subjects." *Scaffolding: Jurnal Pendidikan Islam Dan Multikulturalisme* 4, no. 3 (December 18, 2023): 690–701. <https://doi.org/10.37680/scaffolding.v4i3.4168>.
- Pipit Pitriani, Enung Nugraha, and Fitri Hilmiyanti. "PAI SD Test Quality Analysis: Validity and Reliability Review." *Mimbar Kampus: Jurnal Pendidikan Dan Agama Islam* 23, no. 1 (March 15, 2024): 85–97. <https://doi.org/10.47467/mk.v23i1.1372>.

- Putri, Irma Rezkie, Ad Dhiya At Thariq, Cindhy Yuliarnis Putri, and Monica Apriyuni. "Analysis of the Quality of Akidah Akhlak Exam Questions at TPQ/TQA." *Ri'ayatu Al-Qur'an* 7, no. 2 (July 28, 2025): 7–13. <https://doi.org/10.62990/riqu.v7i2.139>.
- Siti Ina Mulyani, Umi Syahidah, Salsabila Oktavia, Saefudin Zuhri, and Wahyu Hidayat. "Analisis Kualitas Soal Evaluasi Pada Buku Ajar PAI Dan Budi Pekerti SMKN 3 Kota Serang." *ALFIHRIS: Jurnal Inspirasi Pendidikan* 3, no. 1 (December 21, 2024): 119–31. <https://doi.org/10.59246/alfihris.v3i1.1163>.
- Sugiyono. *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, Dan R&D*. Bandung: Alfabeta, 2019.
- Surul, Rofiatus. "Analisis Daya Pembeda Pengecoh Dan Tingkat Kesukaran Soal Pendidikan Agama Islam Dalam Ujian Ahir Semester Tingkat Sekolah Dasar: Studi Kasus Sdn Kaliwinig 07 Rambipuji Jember." *PIONIR: JURNAL PENDIDIKAN* 12, no. 3 (December 26, 2023): 200. <https://doi.org/10.22373/pjp.v12i3.20012>.