

# Analisis Butir Soal Akhir Semester I Mata Pelajaran Bahasa Indonesia Kelas II Madrasah Ibtidaiyah

Uswatun Ni'mah<sup>1</sup>, Ifada Novikasari<sup>1</sup>

<sup>1</sup>Universitas Islam Negeri Prof. K. H. Saifuddin Zuhri Purwokerto

✉ uswatunnimah180@gmail.com

## Abstrak

### ARTICLE INFO

Article history:

Received February 15, 2024

Revised May 22, 2024

Accepted June 20, 2024

Penelitian ini bertujuan untuk melakukan analisis butir soal pada ujian akhir semester I mata pelajaran Bahasa Indonesia untuk siswa kelas II di Madrasah Ibtidaiyah. Metode analisis kuantitatif digunakan dengan memanfaatkan program Anates versi 4 sebagai alat evaluasi. Hasil penelitian menunjukkan beberapa temuan kunci: Pertama, tes menunjukkan tingkat reliabilitas yang cukup tinggi (0.63), menandakan konsistensi yang baik dalam hasil pengukuran. Korelasi antara skor butir dan skor total tes bervariasi, menunjukkan kebutuhan untuk evaluasi dan perbaikan pada struktur tes guna meningkatkan validitas prediktif. Kedua, kelompok unggul menunjukkan kinerja yang baik, sementara kelompok asor memberikan indikasi potensi perluasan pemahaman pada beberapa peserta. Analisis ini memberikan wawasan tentang diversitas kemampuan di antara siswa. Ketiga, mayoritas soal tergolong dalam kategori "sangat mudah" dan "mudah," menunjukkan tingkat kesukaran yang dapat diatasi oleh responden. Namun, terdapat variasi tingkat kesulitan, memberikan informasi berharga tentang tingkat kesulitan relatif setiap butir soal. Keempat, kualitas pengecoh soal tergolong tinggi, tanpa adanya opsi yang dinilai buruk atau sangat buruk. Variasi tingkat kesulitan pilihan jawaban memberikan tantangan seimbang kepada peserta, menambah validitas tes. Penelitian ini memberikan gambaran komprehensif tentang kualitas tes dan karakteristik butir soal, memberikan landasan bagi pengembangan penelitian lebih lanjut dan perbaikan instrumen pengukuran. Temuan ini dapat digunakan sebagai acuan untuk meningkatkan validitas, reliabilitas, dan keberlanjutan tes di konteks pendidikan Madrasah Ibtidaiyah.

**Kata kunci:** analisis, butir soal

## Abstract

This research aims to conduct an item analysis on the final semester I examination of the Indonesian language for second-grade students at Madrasah Ibtidaiyah. A quantitative analysis method was employed using the Anates version 4 program as an evaluation tool. The research findings revealed several key insights. First, the test exhibited a relatively high level of reliability (0.63), indicating good consistency in measurement results. The correlation between item scores and the total test score varied, suggesting the need for evaluation and improvement in the test structure to enhance predictive validity. Second, the high-achieving group demonstrated good performance, while the low-achieving group indicated potential for expanding understanding among some participants. This analysis provided insights into the diversity of abilities among students. Third, the majority of items fell into the categories of "very easy" and "easy," indicating a manageable level of difficulty for respondents. However, there was variability in the difficulty levels, offering valuable information about the relative difficulty of each item. Fourth, the quality of distractors was considered high, with no options rated as poor or very poor. The variation in difficulty levels of answer choices presented a balanced challenge to participants, contributing to the test's validity. This research offers a comprehensive overview of test quality and item characteristics, providing a foundation for further research development and measurement instrument improvement. These findings can serve as a reference for enhancing the validity, reliability, and sustainability of tests in the context of Madrasah Ibtidaiyah education.

*Keywords:* item analysis, test quality

## **A. Pendahuluan**

Pendidikan merupakan salah satu pilar utama dalam pembentukan karakter dan kualitas sumber daya manusia suatu bangsa. (Ni'mah, 2023b) Oleh karena itu, penyelenggaraan pendidikan di tingkat dasar, seperti Madrasah Ibtidaiyah, memiliki peran strategis dalam membentuk dasar pengetahuan, keterampilan, dan sikap positif pada peserta didik. Salah satu mata pelajaran yang memegang peranan penting dalam pembentukan kemampuan berbahasa dan literasi adalah Bahasa Indonesia.

Madrasah Ibtidaiyah adalah lembaga pendidikan Islam tingkat dasar yang memberikan pendidikan formal di tingkat dasar. Kelas II Madrasah Ibtidaiyah memiliki karakteristik siswa yang berada pada tahap perkembangan kognitif awal. Dalam pembelajaran Bahasa Indonesia, pemahaman siswa terhadap keterampilan dasar, seperti membaca, menulis, dan berbicara, sangat penting untuk ditingkatkan. Pembelajaran bahasa tersebut dapat dilakukan dengan mengelompokkan kemampuan siswa berdasarkan kecerdasan yang dimiliki oleh setiap siswa, agar perkembangan siswa sesuai dengan kemampuan yang mereka miliki. (Ni'mah, 2023a)

Evaluasi merupakan komponen integral dalam sistem pendidikan yang perlu dilakukan secara terencana dan sistematis. Evaluasi berperan sebagai alat untuk mengukur keberhasilan atau pencapaian target dalam proses pendidikan dan pembelajaran. (Ismail, 2020) Dalam rangka mengukur pemahaman dan kemampuan siswa, penggunaan instrumen penilaian yang valid dan reliabel menjadi suatu keharusan. Oleh karena itu, penelitian ini akan mengambil fokus pada analisis butir soal ujian akhir semester I Mata Pelajaran Bahasa Indonesia kelas II Madrasah Ibtidaiyah. Analisis butir soal menjadi langkah awal yang krusial dalam mengevaluasi kualitas instrumen penilaian, sekaligus memberikan gambaran mengenai tingkat kesulitan, daya pembeda, serta kesesuaian butir soal dengan kurikulum yang berlaku.

Evaluasi akhir semester mata pelajaran Bahasa Indonesia untuk siswa kelas 2 penting dilakukan untuk evaluasi pemahaman materi: Ulangan akhir semester memberikan kesempatan untuk mengevaluasi sejauh mana pemahaman siswa terhadap materi yang telah diajarkan selama semester tersebut. (Dwipayani, 2013) Ini memungkinkan guru untuk menilai sejauh mana siswa telah menyerap dan memahami konsep-konsep kunci dalam mata pelajaran Bahasa Indonesia. Penilaian kemajuan belajar: Ulangan akhir semester juga berfungsi sebagai alat untuk menilai kemajuan belajar siswa selama satu semester. Dengan demikian, guru dapat memantau perkembangan siswa dari waktu ke waktu dan mengidentifikasi area di mana siswa telah berhasil serta area di mana mereka mungkin memerlukan bantuan lebih lanjut. Memberikan umpan balik: Hasil ulangan dapat digunakan untuk memberikan umpan balik kepada siswa mengenai prestasi belajar mereka. Informasi ini dapat membantu siswa untuk memahami kekuatan dan kelemahan mereka dalam mata pelajaran Bahasa Indonesia, sehingga mereka dapat melakukan perbaikan dan meningkatkan kinerja mereka di masa mendatang. (Safira, 2016) Persiapan ujian lebih Lanjut: Ulangan akhir semester juga dapat dianggap sebagai persiapan awal untuk ujian-ujian berikutnya atau ujian yang lebih besar di masa depan. Ini membantu siswa untuk terbiasa dengan format ujian dan memberikan pengalaman yang berguna dalam menghadapi tantangan belajar. Evaluasi metode pengajaran: Hasil ulangan juga dapat memberikan masukan kepada

guru tentang efektivitas metode pengajaran yang telah digunakan selama semester. Jika hasilnya kurang memuaskan, guru dapat mempertimbangkan penyesuaian metode pengajaran untuk meningkatkan pemahaman dan kinerja siswa.

Analisis butir soal merupakan langkah penting dalam mengukur kualitas setiap pertanyaan dan rangkaian soal secara menyeluruh. Proses ini dapat dilakukan dengan pendekatan kualitatif atau kuantitatif. Tujuan utamanya adalah untuk menggali informasi mengenai karakteristik masing-masing pertanyaan baik melalui pemeriksaan mendalam maupun analisis berbasis data empiris. Hasil dari analisis butir soal dapat dimanfaatkan untuk menilai kualitas soal dan mengevaluasi kemampuan pembelajaran peserta didik melalui hasil ujian.(Elviana, 2020) Melalui analisis butir soal, diharapkan dapat ditemukan informasi yang mendalam terkait dengan sejauh mana instrumen penilaian tersebut mampu mencapai tujuan pembelajaran, sejauh mana siswa mampu mengaplikasikan pengetahuan yang telah diperoleh, dan sejauh mana instrumen penilaian tersebut dapat memberikan gambaran objektif terkait dengan pencapaian kompetensi siswa dalam mata pelajaran Bahasa Indonesia.

Pendidikan pada tingkat Madrasah Ibtidaiyah (MI) memiliki karakteristik khusus dalam membentuk landasan keilmuan bagi siswa. Teori pendidikan Islam, seperti yang dijelaskan oleh Ibn Sina (Avicenna), menekankan pentingnya pendidikan sebagai sarana untuk mencapai kebahagiaan dan kesempurnaan manusia. Evaluasi pembelajaran menjadi instrumen penting untuk mengukur pencapaian siswa dalam mencapai tujuan pendidikan tersebut. konsep perkembangan bahasa anak, sebagaimana dijelaskan oleh Jean Piaget dan Lev Vygotsky. Teori kognitif Piaget menunjukkan bahwa anak di kelas 2 MI mungkin berada pada tahap operasional konkret, di mana mereka mulai mampu memahami konsep-konsep yang lebih kompleks(Piaget, 1952). Teori Zona Pembangunan Proximal Vygotsky menggarisbawahi peran bimbingan dan dukungan dalam mengembangkan keterampilan bahasa anak (Vygotsky., 1978).

Teori Evaluasi Butir Soal. Aspek ini melibatkan teori evaluasi butir soal, seperti yang dijelaskan oleh David C. Berliner dan Thomas S. Popkewitz. Teori ini fokus pada konsep validitas, reliabilitas, dan keobjektifan dalam merancang dan mengukur keberhasilan butir soal sebagai alat evaluasi pembelajaran. Pengembangan Soal Ujian. Teori ini mencakup panduan pengembangan soal ujian yang baik dan berfokus pada tujuan pembelajaran. Prinsip-prinsip pengembangan soal ujian yang efektif menekankan pada kemampuan mengukur pemahaman dan aplikasi pengetahuan siswa.

Dengan demikian, artikel ilmiah ini bertujuan untuk memberikan pemahaman yang lebih mendalam mengenai kualitas instrumen penilaian pada mata pelajaran Bahasa Indonesia di kelas II Madrasah Ibtidaiyah. Analisis butir soal menjadi sarana untuk mengevaluasi dan meningkatkan kualitas pengukuran hasil belajar siswa, sekaligus memberikan kontribusi dalam upaya peningkatan mutu pendidikan di tingkat dasar.

## **B. Metode**

Penelitian ini memanfaatkan pendekatan kuantitatif karena data yang dihimpun bersifat numerik, dan analisis dilakukan melalui metode statistik(Sugiyono, 2019). Dalam mencapai

tujuan penelitian, analisis statistik yang diterapkan adalah statistik deskriptif, yang bertujuan untuk mengklasifikasi jenis kesalahan, menentukan persentase setiap jenis kesalahan, dan menggunakan nilai maksimum untuk mengidentifikasi kesalahan yang paling sering terjadi.

Pendekatan statistik deskriptif digunakan untuk menganalisis data dengan cara menggambarkan data yang terkumpul sebagaimana adanya, tanpa bermaksud membuat kesimpulan umum atau generalisasi. (Sugiyono, 2012)

Penelitian berbasis kuantitatif ini memanfaatkan analisis statistik deskriptif, yang berarti tidak bertujuan untuk mengadopsi kesimpulan umum atau generalisasi terhadap seluruh populasi sebagai responden penelitian. Populasi ini adalah semua siswa kelas 2 MI Ma'arif Caruy tahun pelajaran 2023/2024, yang berjumlah 34 siswa semuanya akan menjadi responden.

Data dikumpulkan dengan menggunakan dokumentasi, yaitu dokumen hasil ulangan semester I mata pelajaran Bahasa Indonesia sebanyak 34 siswa. Pertanyaan ulangan akhir semester I bahasa Indonesia terdiri dari 25 jenis tes pilihan ganda yang digunakan untuk menjelaskan variabel penelitian. Kesalahan dalam pengerjaan ulangan akhir semester I menjadi variabel dalam penelitian ini. Ini adalah kesalahan dalam menjalankan Ulangan Akhir Semester I Bahasa Indonesia, sehingga analisis dilakukan pada pemilihan jawaban yang diambil oleh siswa dalam memberikan jawaban. Hasil jawaban tersebut dimasukkan dalam aplikasi yang bernama anates, yang akan menghitung otomatis dan akan langsung menghitung realibilitas, daya pembeda, tingkat kesukaran, korelasi sekor, dan kualitas pengecoh.

### **C. Literatur Review**

Pendidikan di Indonesia mencakup berbagai jenjang, termasuk Madrasah Ibtidaiyah (MI), yang merupakan tingkat pendidikan dasar untuk siswa muslim. Evaluasi pembelajaran di MI penting untuk memastikan efektivitas pengajaran dan pemahaman siswa. Berikut adalah beberapa aspek penting dalam pendidikan dan evaluasi pembelajaran di MI: Kurikulum yang Berbasis Islam; Pendidikan di MI didasarkan pada nilai-nilai Islam, dengan kurikulum yang mencakup pemahaman agama, bahasa Arab, dan pengetahuan umum. Kurikulum ini dirancang untuk membentuk karakter islami dan memberikan pemahaman holistik kepada siswa mengenai ajaran agama Islam (Indonesia, n.d.). Metode Pengajaran Interaktif; Metode pengajaran di MI melibatkan interaksi aktif antara guru dan siswa. Guru diharapkan untuk menggunakan pendekatan yang menarik dan melibatkan siswa dalam pembelajaran aktif. Pemanfaatan teknologi pendidikan juga dapat mendukung proses pembelajaran interaktif di MI (Mustofa, 2019). Evaluasi Pembelajaran Berkelanjutan; Evaluasi pembelajaran di MI sebaiknya bersifat berkelanjutan untuk memantau perkembangan siswa secara terus-menerus. Ini dapat melibatkan penilaian formatif, seperti ujian harian, tugas, dan proyek, yang membantu guru memahami tingkat pemahaman siswa dan memberikan umpan balik sepanjang proses pembelajaran (Sukmadinata, 2013). Ujian Akhir Semester dan Evaluasi Sumatif; Selain evaluasi berkelanjutan, MI juga melakukan ujian akhir semester dan evaluasi sumatif untuk menilai pencapaian akademis siswa secara menyeluruh. Ini mencakup ujian tertulis, wawancara, dan penugasan proyek yang lebih besar untuk mengevaluasi

pemahaman konsep dan kemampuan praktis siswa (Mulyasa, 2013). Pengembangan Keterampilan Karakter; Selain aspek akademis, evaluasi di MI juga harus memperhatikan pengembangan keterampilan karakter siswa, seperti akhlak, tanggung jawab, dan kerjasama. Penilaian sikap dan perilaku dapat diintegrasikan ke dalam evaluasi pembelajaran untuk memastikan pendidikan yang seimbang secara holistik (Zainuddin, 2018).

Pembelajaran dan pengembangan bahasa anak melibatkan berbagai teori yang membantu memahami bagaimana anak-anak memperoleh, mengembangkan, dan menggunakan bahasa. Beberapa teori yang relevan dalam konteks ini termasuk teori nativisme, behaviorisme, konstruktivisme, dan interaksionisme. Berikut adalah penjelasan singkat tentang masing-masing teori: Nativisme; Teori ini berpendapat bahwa kemampuan bahasa bawaan (innate) dan anak-anak secara alami memiliki kemampuan untuk mengembangkan bahasa tanpa instruksi formal. Contoh: Chomsky mengemukakan gagasan tentang "gramatika universal," yaitu kecenderungan bawaan manusia untuk memahami dan menggunakan bahasa (Chomsky, 1957). Behaviorisme; Teori ini menekankan pengaruh lingkungan dan pengalaman dalam pembentukan perilaku bahasa anak-anak. Proses pembelajaran bahasa dipandang sebagai respons terhadap rangsangan dan penguatan. Contoh: Teori penguatan oleh Skinner menyoroti pentingnya umpan balik positif atau negatif dalam pembentukan perilaku bahasa (Skinner, 1957). Konstruktivisme; Teori ini berfokus pada peran aktif anak dalam membangun pemahaman mereka terhadap bahasa melalui pengalaman dan interaksi dengan lingkungan. Contoh: Vygotsky menekankan peran penting kerjasama sosial dan dukungan dari lingkungan dalam pengembangan bahasa anak (Vygotsky, 1978). Interaksionisme; Teori ini menyoroti peran interaksi sosial dalam pembelajaran bahasa, di mana anak-anak belajar melalui interaksi dengan orang dewasa dan teman sebaya. Contoh: Bruner menekankan pentingnya lingkungan yang merangsang dan berinteraksi dalam pembentukan kemampuan bahasa anak (Bruner, 1983).

Kurikulum Bahasa Indonesia; Kurikulum Bahasa Indonesia dirancang untuk memberikan landasan pembelajaran yang menyeluruh terkait bahasa dan sastra Indonesia. Ini mencakup keterampilan berbicara, mendengarkan, membaca, menulis, serta pemahaman sastra. Contoh: Kurikulum 2013 di Indonesia, yang mengintegrasikan pembelajaran berbasis kompetensi dan karakter. Standar Kompetensi Bahasa Indonesia; Standar kompetensi merinci keterampilan dan pengetahuan yang diharapkan siswa capai dalam pembelajaran Bahasa Indonesia. Ini mencakup kompetensi dasar untuk setiap tingkat pendidikan. Contoh: Standar Kompetensi Lulusan (SKL) Bahasa Indonesia yang disusun oleh Kementerian Pendidikan dan Kebudayaan, yang menjelaskan pencapaian yang diharapkan dari siswa setelah menyelesaikan tingkat pendidikan tertentu. Pendekatan Pengajaran; Pengajaran Bahasa Indonesia dapat menggunakan pendekatan beragam, seperti pendekatan komunikatif yang menekankan penggunaan bahasa dalam konteks komunikatif nyata. Contoh: Pendekatan Genre-Based yang memfokuskan pembelajaran pada pemahaman struktur dan fungsi berbagai jenis teks (Derewianka, 1990).

Teori Evaluasi Butir Soal. Teori Klasik; Teori klasik dalam evaluasi butir soal menekankan validitas dan reliabilitas. Validitas mencakup sejauh mana sebuah tes mengukur apa yang seharusnya diukur, sementara reliabilitas mencakup konsistensi hasil tes. Contoh: Pada

teori klasik, korelasi antara skor siswa pada dua tes yang berbeda dapat digunakan untuk mengukur reliabilitas (Hambleton & Jones, 1993). Teori Respon Butir (IRT); IRT memodelkan probabilitas jawaban benar atau salah berdasarkan kemampuan siswa dan kesulitan butir soal. IRT memberikan gambaran lebih mendalam tentang kualitas butir soal. Contoh: Model Rasch, salah satu model IRT, dapat digunakan untuk mengevaluasi butir soal dan mengestimasi tingkat kesulitan butir serta kemampuan siswa (Rasch, 1960). Teori Keandalan Butir (IRT); Deskripsi: Teori ini mengevaluasi keandalan butir soal berdasarkan karakteristik tes dan respons siswa. Hal ini membantu dalam mengidentifikasi butir soal yang konsisten atau tidak konsisten. Contoh: Penggunaan Kurva Informasi, yang merupakan grafik fungsi keandalan butir soal terhadap tingkat kemampuan siswa (Embretson & Reise, 2000). Teori Kesulitan Distraktor; Fokus pada evaluasi pilihan jawaban (distraktor) dalam soal pilihan ganda. Distraktor yang baik seharusnya mampu membedakan antara siswa yang paham materi dan yang tidak. Contoh: Analisis indeks daya membedakan (D) untuk mengevaluasi sejauh mana distraktor dapat membedakan antara siswa yang paham materi dan yang tidak (Haladyna et al., 2002).

Pengembangan Soal Ujian, meliputi analisis Materi; Tahap awal pengembangan soal ujian melibatkan analisis materi yang akan diujikan. Ini mencakup identifikasi konsep-konsep kunci, tujuan pembelajaran, dan cakupan materi. Contoh: Jika tujuan pembelajaran mencakup pemahaman konsep fisika mekanika, analisis materi akan fokus pada konsep-konsep seperti hukum Newton dan gerak benda (Gronlund & Linn, 1990). Penentuan Jenis Soal; Berdasarkan analisis materi, penulis soal memilih jenis soal yang paling sesuai, seperti pilihan ganda, isian singkat, atau esai, untuk mengukur pemahaman siswa. Contoh: Jika tujuan pembelajaran mencakup mengingat fakta-fakta tertentu, pilihan ganda dapat menjadi pilihan yang tepat (Nitko, 2001). Penyusunan Butir Soal; Proses ini melibatkan penulisan butir soal dengan jelas dan terukur. Butir soal seharusnya menguji pemahaman konsep secara efektif. Contoh: Untuk menguji pemahaman hukum gerak, sebuah pertanyaan mungkin dapat dirumuskan seperti "Apa yang akan terjadi pada gerakan sebuah benda jika gaya yang bekerja pada benda menjadi nol?" (Haladyna et al., 2002). Uji Kelayakan dan Keandalan; Sebelum penggunaan yang sebenarnya, soal ujian diuji kelayakannya dengan mengumpulkan tanggapan dari kelompok uji coba. Keandalan soal juga dievaluasi untuk memastikan konsistensi hasil. Contoh: Proses ini melibatkan uji coba soal pada sekelompok siswa untuk mengidentifikasi potensi kebingungan atau ambiguitas (American Educational Research Association & Education., 2014).

#### **D. Hasil Penelitian**

Analisis butir soal dilakukan untuk mengidentifikasi butir soal mana yang baik dan yang buruk. Dari hasil analisis ini, dapat diperoleh informasi yang berguna untuk menciptakan soal-soal yang lebih berkualitas, sehingga memberikan informasi yang lebih akurat terkait dengan peserta tes. Analisis soal dapat dilakukan melalui dua pendekatan, yaitu analisis kualitatif dan analisis kuantitatif. Penelitian ini memanfaatkan analisis kuantitatif untuk menilai kinerja butir soal dalam sebuah tes. Salah satu metode untuk melakukan analisis soal adalah dengan menggunakan program Anates versi 4. Program ini merupakan aplikasi komputer sederhana yang

mudah diimplementasikan, cepat, dan akurat. Anates dirancang oleh Karno dan Yudi Wibisono sebagai perangkat lunak khusus yang dikembangkan untuk menganalisis tes pilihan ganda.

Manfaat dari Anates adalah kemampuannya untuk melakukan analisis butir soal secara otomatis, memeriksa jawaban yang benar dan salah dengan cepat dan praktis. Keunggulan program ini terletak pada kemudahan pemahaman karena instruksi penggunaan program disampaikan dalam Bahasa Indonesia. Selain itu, hasil analisis yang diperoleh dapat diimpor ke Microsoft Excel untuk perhitungan nilai. Anates memiliki kemampuan untuk melakukan analisis butir soal, termasuk evaluasi reliabilitas, identifikasi kelompok unggul dan rendah, daya pembeda, tingkat kesulitan, korelasi skor butir dengan skor total, dan kualitas pilihan jawaban yang disediakan (Ida & Musyarofah, 2021).

Berikut adalah hasil analisis butir soal yang diperoleh melalui penggunaan program Anates versi 4:

#### 1. Reliabilitas

Reliabilitas suatu tes mengacu pada sejauh mana tes tersebut memberikan hasil yang konsisten dan dapat diandalkan. Dengan kata lain, jika seseorang menjawab tes pada dua waktu yang berbeda, seberapa besar kemungkinan bahwa jawaban yang diberikan akan sama atau setidaknya sangat mirip. Ada beberapa metode untuk mengukur reliabilitas, dan salah satu yang umum digunakan adalah koefisien korelasi. Koefisien korelasi antar-tes digunakan untuk mengevaluasi sejauh mana hasil tes satu sama lain berkorelasi. Semakin tinggi koefisien korelasi, semakin tinggi reliabilitasnya (American Educational Research Association & Education., 2014). Namun, reliabilitas suatu tes tidak hanya tergantung pada metodenya tetapi juga pada faktor-faktor seperti konsistensi internal pertanyaan (internal consistency), stabilitas hasil dari waktu ke waktu (test-retest reliability), atau seberapa baik tes tersebut mengukur apa yang seharusnya diukur (validitas) (Nunnally & Bernstein, 1994).

Dalam penelitian ini, rata-rata hasil tes yang diperoleh mencapai 17.56, sedangkan simpang baku sebesar 3.69. Hasil korelasi antara variabel X dan Y menunjukkan angka sebesar 0.46, mengindikasikan adanya hubungan positif antara kedua variabel tersebut. Dengan demikian, dapat disimpulkan bahwa terdapat keterkaitan positif antara faktor X dan Y dalam konteks penelitian ini.

Selain itu, reliabilitas tes yang digunakan dalam penelitian ini mencapai angka sebesar 0.63. Nilai reliabilitas yang cukup tinggi ini menunjukkan bahwa instrumen pengukuran yang digunakan dapat diandalkan dan memberikan hasil yang konsisten.

Hasil penelitian ini memberikan gambaran yang jelas tentang hubungan antarvariabel dan keandalan instrumen pengukuran yang digunakan. Implikasi dari temuan ini dapat digunakan sebagai dasar untuk mengembangkan lebih lanjut penelitian di masa depan atau sebagai landasan untuk pengembangan kebijakan yang relevan dalam konteks tertentu. Secara umum, hasil tes menunjukkan adanya hubungan positif antara variabel X dan Y, dengan rata-rata skor di atas nilai tengah, dan tes memiliki tingkat keandalan yang cukup baik.

## 2. Kelompok Unggul dan Asor

**Kelompok Unggul:** Kelompok unggul merujuk pada kumpulan peserta yang memiliki tingkat kemampuan tinggi dan cenderung menjawab butir soal dengan benar. Dalam analisis butir soal, butir yang cenderung dijawab dengan benar oleh peserta berkinerja tinggi termasuk dalam kelompok ini. Identifikasi kelompok unggul penting untuk memastikan bahwa tes dapat membedakan dengan baik antara peserta yang memiliki kemampuan tinggi dan rendah. Pemahaman mendalam terhadap butir soal dalam kelompok ini dapat membantu meningkatkan kecerdasan dan ketajaman tes (Hambleton & Swaminathan, 1985). **Kelompok Asor:** Kelompok asor merujuk pada kumpulan peserta yang memiliki tingkat kemampuan rendah dan mampu menjawab butir soal dengan benar. Butir yang cenderung dijawab dengan benar oleh peserta berkinerja rendah termasuk dalam kelompok ini. Identifikasi kelompok asor penting untuk memastikan bahwa tes tidak terlalu sulit bagi peserta dengan tingkat kemampuan rendah, yang dapat menghambat validitas tes. Analisis lebih lanjut terhadap butir soal dalam kelompok asor dapat membantu meningkatkan kualitas keseluruhan tes (Embretson & Reise, 2000).

<b>Kelompok Unggul</b>				<b>Kelompok Asor</b>		
No.	Nama	24	25	Nama	24	25
	Adel	0	1	Nafal	1	1
	Adeva	1	1	Anam	1	0
	Aleza	1	1	Hasna	1	1
	Alya	1	1	Sabi	0	0
	Sifa	1	1	Inayah	1	0
	Azka	1	1	Rifka	0	1
	Nadin	1	1	Hafidz	1	0
	Kamal	1	1	Lesti	1	0
	Zafira	1	1	Salsa	0	0
	Jumlah Benar	8	9		6	3

Hasil perhitungan tes kelompok unggul menunjukkan bahwa sebanyak delapan peserta di antara sembilan peserta yang diuji berhasil menjawab dengan benar pada tes tersebut. Secara rinci, setiap subyek diwakili oleh angka 1 atau "-" dalam kolom 24 dan 25, yang masing-masing mengindikasikan jawaban benar atau salah. Dari keseluruhan tes, terdapat delapan jawaban benar untuk subyek Adel, Adeva, Aleza, Alya, Sifa, Azka, Nadin, Kamal, dan Zafira. Dengan demikian, dapat disimpulkan bahwa mayoritas peserta (8 dari 9) menunjukkan kinerja yang unggul dalam menghadapi tes tersebut. Hasil ini mencerminkan pemahaman dan kemampuan jawaban yang baik dari sebagian besar peserta dalam kelompok tersebut.

Hasil perhitungan tes kelompok asor menunjukkan bahwa dari sembilan peserta yang diuji, sebanyak enam peserta berhasil menjawab dengan benar pada kolom 24, dan tiga peserta menjawab benar pada kolom 25. Secara rinci, subyek Nafal, Anam, Hasna, Inayah, Hafidz, dan Lesti berhasil memberikan jawaban benar pada kolom 24, sedangkan subyek Nafal, Hasna, dan Rifka berhasil menjawab dengan benar pada kolom 25. Dalam konteks ini, tes kelompok asor menunjukkan variasi dalam pencapaian peserta. Meskipun sebagian besar peserta menunjukkan

kinerja yang baik pada kolom 24, terdapat variasi yang lebih besar pada kolom 25. Meskipun demikian, hasil ini mencerminkan adanya pemahaman yang baik pada beberapa peserta, sementara yang lain mungkin perlu perhatian lebih lanjut untuk meningkatkan kinerja mereka pada aspek tertentu dari tes.

### 3. Daya Pembeda

Daya pembeda (*differential item functioning/DIF*) pada butir soal mengacu pada kemampuan butir soal untuk membedakan antara peserta dengan tingkat kemampuan yang berbeda. Sebuah butir dikatakan memiliki daya pembeda yang baik jika peserta yang memiliki tingkat kemampuan yang tinggi cenderung menjawabnya dengan benar, sementara peserta dengan tingkat kemampuan yang rendah cenderung menjawabnya dengan salah (Camilli & Shepard, 1994).

Faktor-faktor yang Mempengaruhi Daya Pembeda, antara lain; Sensitivitas Terhadap Kemampuan: Butir yang sensitif terhadap tingkat kemampuan peserta dapat memiliki daya pembeda yang baik. Pemahaman butir terhadap rentang kemampuan peserta membantu dalam memastikan butir dapat membedakan dengan baik antara tingkat kemampuan yang berbeda. Ketidaksetaraan Dalam Kelompok Peserta: Penting untuk memastikan bahwa daya pembeda tetap konsisten di berbagai kelompok peserta (misalnya, berdasarkan jenis kelamin, latar belakang budaya). Jika terdapat perbedaan dalam cara berbagai kelompok menjawab butir, hal ini dapat mengindikasikan adanya bias dan perlu diperhatikan. Ketidaksetaraan Dalam Kelompok Waktu: Butir yang memiliki daya pembeda yang baik seharusnya tetap relevan dan membedakan dengan baik antara peserta, terlepas dari kapan tes diambil. Analisis daya pembeda dapat membantu memastikan konsistensi ini (Lord, 1980).

Dari hasil analisis yang telah dilakukan dapat disimpulkan sebagai berikut: Soal dengan No Butir 2, 10, dan 25 memiliki daya pembeda tertinggi (indeks DP 66.67%). Soal dengan No Butir 14 memiliki daya pembeda tinggi (indeks DP 88.89%), menunjukkan bahwa soal ini sangat efektif dalam membedakan antara kelompok atas dan bawah. Soal dengan No Butir 5 dan 15 memiliki indeks DP negatif (-11.11%), yang menunjukkan bahwa soal tersebut mungkin perlu direvisi atau diperiksa lebih lanjut karena tidak efektif dalam membedakan antara kelompok atas dan bawah.

Analisis Kelebihan dan Kelemahan Setiap Soal: 1. No Butir 2 (Indeks DP 66.67%): Kelebihan: Daya pembeda yang tinggi (66.67%) menunjukkan bahwa soal ini mampu dengan baik membedakan antara siswa yang berkinerja tinggi dan rendah. Soal ini dapat memberikan informasi yang berharga tentang kemampuan siswa. Manfaat: Soal ini dapat digunakan sebagai alat evaluasi yang efektif untuk mengukur tingkat pemahaman siswa secara akurat. Meningkatkan validitas dan reliabilitas instrumen tes. 2. No Butir 10 (Indeks DP 66.67%): Kelebihan: Seperti No Butir 2, soal ini memiliki daya pembeda yang tinggi (66.67%). Mampu membedakan antara siswa yang berkinerja tinggi dan rendah. Manfaat: Menambah keandalan instrumen tes karena mampu memberikan informasi yang signifikan tentang kemampuan siswa. Meningkatkan keberagaman dalam pengukuran hasil pembelajaran. 3. No Butir 25 (Indeks DP 66.67%): Kelebihan: Mempunyai daya pembeda yang tinggi (66.67%), serupa dengan No Butir 2 dan 10. Berpotensi memberikan informasi yang kuat tentang kemampuan siswa. Manfaat: Menjadi komponen yang penting

dalam menyusun instrumen tes yang kaya dan bervariasi. Memungkinkan guru atau penilai untuk membuat keputusan yang lebih tepat dalam mengevaluasi kemampuan siswa. 4. No Butir 14 (Indeks DP 88.89%): Kelebihan: Mempunyai daya pembeda yang sangat tinggi (88.89%), menunjukkan kemampuannya yang luar biasa dalam membedakan antara kelompok atas dan bawah. Manfaat: Soal ini dapat dijadikan acuan utama dalam mengukur kemampuan siswa dengan tingkat ketelitian yang tinggi. Meningkatkan akurasi penilaian dan memberikan informasi mendalam tentang pemahaman siswa. 5. No Butir 5 dan 15 (Indeks DP -11.11%): Kelemahan: Kedua soal ini memiliki indeks DP negatif (-11.11%), menunjukkan ketidakmampuannya dalam membedakan antara kelompok atas dan bawah. Daya pembeda yang rendah mengindikasikan kurangnya kontribusi informasi signifikan dari soal ini. Manfaat: Diperlukan revisi atau perbaikan dalam perumusan soal untuk meningkatkan daya pembeda. Evaluasi kembali tujuan dan kejelasan soal agar sesuai dengan kemampuan siswa yang diinginkan.

#### 4. Tingkat Kesukaran

Tingkat kesukaran suatu soal mengacu pada sejauh mana soal tersebut dianggap sulit oleh peserta. Penentuan tingkat kesukaran dapat memberikan wawasan tentang seberapa baik soal dapat membedakan antara peserta yang memiliki tingkat kemampuan yang berbeda. Biasanya, tingkat kesukaran dinyatakan dalam bentuk persentase peserta yang menjawab soal dengan benar (Crocker & Algina, 1986).

Faktor-faktor yang Mempengaruhi Tingkat Kesukaran, Kompleksitas Butir: Soal yang memerlukan pemikiran mendalam atau pemahaman konsep yang rumit cenderung memiliki tingkat kesukaran yang tinggi. Pemilihan kata-kata dan struktur kalimat juga dapat mempengaruhi sejauh mana peserta dapat memahami pertanyaan. Konteks dan Penyajian Soal: Cara suatu soal disajikan dan konteksnya dapat memengaruhi tingkat kesukaran. Keterbacaan, kejelasan, dan relevansi soal terhadap materi yang diuji memainkan peran penting. Kemampuan Peserta Rata-rata: Tingkat kesukaran suatu soal relatif terhadap kemampuan peserta rata-rata yang diharapkan dalam populasi (Hambleton & Swaminathan, 1985). Soal yang dianggap sulit oleh sebagian besar peserta mungkin tidak memberikan informasi yang baik dalam membedakan antara peserta yang memiliki tingkat kemampuan yang berbeda.

Dari hasil analisis butir soal ulangan akhir semester I kelas 2 menggunakan aplikasi anates diperoleh hasil tingkat kesukaran di bawah ini:

<b>Hasil Analisis Tingkat Kesukaran Tes</b>	
sangat mudah	70%
mudah	50%-70%
sedang	30%-50%
sukar	30%

Berdasarkan hasil analisis tingkat kesukaran dari tes, dapat disimpulkan bahwa sebagian besar soal tes diklasifikasikan sebagai "sangat mudah," dengan tingkat kesukaran sebesar 70%. Sejumlah besar soal juga termasuk dalam kategori "mudah," dengan rentang tingkat kesukaran antara 50% hingga 70%. Adanya sejumlah soal dalam kategori "sedang" dengan tingkat kesukaran

antara 30% hingga 50%, menunjukkan variasi dalam tingkat kesulitan soal. Namun demikian, tidak ada soal yang tergolong dalam kategori "sukar" dengan tingkat kesukaran di bawah 30%.

Hasil ini menandakan bahwa mayoritas soal tes cenderung memiliki tingkat kesukaran yang dapat diatasi oleh responden, dengan sebagian besar soal termasuk dalam kategori tingkat kesukaran yang lebih rendah. Meskipun terdapat variasi tingkat kesulitan, keseluruhan tes menunjukkan sejumlah signifikan soal-soal yang dianggap relatif mudah oleh responden. Informasi ini dapat memberikan wawasan penting bagi penilai dalam mengevaluasi kualitas dan efektivitas tes, serta merancang strategi pembelajaran yang lebih baik untuk peserta tes.

#### 5. Korelasi Skor Butir dengan Skor Total

Korelasi skor butir dengan skor total pada suatu tes merujuk pada sejauh mana skor yang diperoleh dari menjawab setiap butir soal dalam tes berkorelasi dengan skor total keseluruhan. Analisis ini dapat memberikan wawasan tentang seberapa baik suatu butir soal dapat memprediksi performa keseluruhan peserta dalam tes tersebut.

Faktor-faktor yang Mempengaruhi Korelasi, Validitas Butir Soal: Korelasi yang tinggi antara skor butir dan skor total menunjukkan bahwa butir tersebut memiliki validitas prediktif yang baik terhadap keseluruhan kinerja peserta. Validitas butir dapat dipengaruhi oleh sejauh mana butir mengukur konsep atau keterampilan yang diinginkan. Konsistensi Internal: Tingginya korelasi dapat juga mencerminkan konsistensi internal dalam tes, yaitu sejauh mana seluruh butir soal saling berkorelasi satu sama lain (Anastasi & Urbina, 1997).

Konsistensi internal yang tinggi dapat meningkatkan validitas tes, Konteks Pengukuran: Korelasi dapat bervariasi tergantung pada konteks pengukuran dan jenis tes yang digunakan. Penting untuk mempertimbangkan apakah suatu tes dirancang untuk mengukur keterampilan spesifik atau konsep secara umum (Crocker & Algina, 1986).

Korelasi skor butir dengan skor total dalam penelitian ini adalah sebagai berikut

<b>Korelasi antar soal</b>	<b>Nomer soal</b>
sangat signifikan	10, 12, 14, 25
signifikan	2, 3, 6, 8, 10, 12, 13, 14, 19, 25
tidak signifikan	1, 4, 5, 7, 9, 11, 15, 16, 17, 18, 20, 21, 22, 23, 24

Berdasarkan hasil penelitian korelasi antara skor butir dan skor total, temuan menunjukkan adanya hubungan yang beragam antara jawaban pada setiap butir soal dengan skor total tes. Temuan yang paling menonjol adalah bahwa butir-butir soal nomor 2, 3, 6, 8, 10, 12, 13, 14, 19, dan 25 menunjukkan korelasi yang signifikan atau bahkan sangat signifikan dengan skor total tes, menandakan bahwa jawaban benar pada butir-butir tersebut memiliki dampak positif yang signifikan terhadap peningkatan skor total. Meskipun demikian, ditemukan juga sejumlah butir soal yang tidak menunjukkan korelasi yang signifikan dengan skor total tes, seperti pada butir-butir nomor 1, 4, 5, 7, 9, 11, 15, 16, 17, 18, 20, 21, 22, 23, dan 24. Hal ini menunjukkan bahwa jawaban pada butir-butir tersebut mungkin tidak secara konsisten mencerminkan kemampuan keseluruhan responden.

Oleh karena itu, penelitian ini mengindikasikan perlunya evaluasi lebih lanjut terhadap struktur dan validitas tes, serta kemungkinan revisi pada butir-butir tertentu untuk memastikan bahwa tes secara akurat mencerminkan kemampuan yang diukur. Dengan memahami korelasi antara skor butir dan skor total, langkah-langkah perbaikan dapat diarahkan pada peningkatan keakuratan dan keandalan pengukuran dalam konteks penilaian tersebut.

## 6. Kualitas Pengecoh

Kualitas pengecoh atau pilihan jawaban yang salah pada sebuah soal tes sangat penting dalam mengukur sejauh mana peserta tes memahami materi. Pengecoh yang baik harus dirancang sedemikian rupa sehingga hanya peserta yang tidak menguasai materi yang akan tertarik padanya. Berikut adalah beberapa faktor yang dapat mempengaruhi kualitas pengecoh dalam sebuah soal tes: Ketidakjelasan dan Ambigu; pengecoh sebaiknya tidak terlalu mirip dengan jawaban yang benar. Jika pengecoh terlalu ambigu atau mirip dengan jawaban yang benar, peserta dapat tertipu dan jawaban yang benar menjadi sulit diidentifikasi (Brown & Race, 2002). Gaya Bahasa dan Format; Kelengkapan Kalimat: Pengecoh harus ditulis dengan bahasa yang tepat dan lengkap. Pilihan jawaban yang tidak sesuai dengan format kalimat atau tata bahasa dapat memberikan petunjuk kepada peserta tentang jawaban yang benar (Haladyna et al., 2002). Keseimbangan Antara Pilihan Jawaban; Distribusi Kesalahan: Pengecoh sebaiknya tersebar merata di antara pilihan jawaban. Jika terdapat keseimbangan yang tidak merata, peserta yang memahami materi dapat dengan mudah mengidentifikasi jawaban yang benar (Rodriguez, 2005). Konteks dan Relevansi: Keterkaitan dengan Materi: Pengecoh sebaiknya tetap relevan dengan konteks dan materi yang diuji. Pilihan jawaban yang tidak terkait dengan materi dapat menyulitkan peserta dalam mengidentifikasi jawaban yang benar (Nitko & Brookhart, 2011). Pengecoh yang Menarik Perhatian: Kesulitan yang Seragam: Pengecoh sebaiknya memiliki tingkat kesulitan yang seragam. Jika satu pilihan jawaban jelas lebih sulit atau lebih mudah daripada yang lain, hal itu dapat memberikan petunjuk yang tidak adil (Haladyna & Downing, 1989). Korelasi dengan Materi Pemahaman: Korelasi dengan Materi Pemahaman: Pengecoh dapat dirancang dengan mempertimbangkan kesalahan umum yang mungkin dilakukan oleh peserta. Ini membantu mengidentifikasi peserta yang sebenarnya tidak menguasai materi (Downing & Haladyna, 2006).

Berdasarkan hasil analisis kualitas pengecoh soal, ditemukan beberapa pola dalam penilaian pilihan jawaban pada setiap butir soal. Berikut adalah kesimpulan yang dapat diambil dari hasil analisis tersebut: Distribusi Kualitas Jawaban: Mayoritas pilihan jawaban dinilai sebagai sangat baik (\*\*), baik (++), atau baik (+), yang menunjukkan bahwa penulis soal telah berusaha untuk menyusun pilihan jawaban yang dapat mengecoh peserta dengan mempertahankan tingkat kualitas yang tinggi. Pilihan Jawaban Buruk atau Sangat Buruk: Tidak terdapat pilihan jawaban yang dinilai sebagai kurang baik (-) atau sangat buruk (---). Hal ini menunjukkan bahwa pilihan jawaban yang disediakan pada setiap butir soal dianggap cukup memadai dalam memberikan variasi tingkat kesulitan. Variasi dalam Tingkat Kesulitan Pilihan Jawaban: Terdapat variasi dalam tingkat kesulitan pilihan jawaban, dengan beberapa butir soal menampilkan pilihan jawaban yang sangat baik atau baik, sementara yang lain dapat memiliki tingkat kesulitan yang lebih rendah (++ atau +). Hal ini dapat memberikan tantangan yang seimbang kepada peserta

tes dengan kemampuan berbeda. Korelasi dengan Kunci Jawaban: Pilihan jawaban yang diberi nilai sangat baik atau baik secara konsisten terkait dengan kunci jawaban yang benar (\*\*). Ini menunjukkan bahwa pilihan jawaban yang lebih baik cenderung lebih sesuai dengan jawaban yang diharapkan.

Dengan demikian, kesimpulan umumnya adalah bahwa pilihan jawaban pada setiap butir soal telah dirancang dengan baik, dengan menunjukkan kualitas yang tinggi dan variasi tingkat kesulitan yang cukup. Selain itu, tidak ditemukan kelemahan signifikan dalam penyusunan pilihan jawaban yang dapat merugikan peserta. Oleh karena itu, dapat disimpulkan bahwa kualitas pengecoh soal dalam ujian tersebut dapat dianggap memadai dan mendukung validitas tes.

#### E. SIMPULAN

Dalam penelitian ini, ditemukan beberapa hasil yang jelas diantaranya adalah Validitas dan Reliabilitas Tes: Tes menunjukkan reliabilitas tinggi (0.63), menandakan konsistensi hasil pengukuran. Korelasi antara skor butir dan skor total bervariasi, memerlukan evaluasi lebih lanjut terhadap struktur tes untuk meningkatkan validitas prediktif. Kelompok Unggul dan Asor: Kelompok unggul menunjukkan kinerja baik, sementara kelompok asor memberikan wawasan tentang potensi perluasan pemahaman bagi beberapa peserta. Tingkat Kesukaran Soal: Mayoritas soal termasuk dalam kategori "sangat mudah" dan "mudah," menunjukkan tingkat kesukaran yang dapat diatasi oleh responden, namun dengan variasi. Kualitas Pengecoh Soal: Pilihan jawaban menunjukkan kualitas tinggi tanpa adanya opsi yang dinilai buruk atau sangat buruk. Variasi tingkat kesulitan pilihan jawaban memberikan tantangan seimbang kepada peserta. Implikasi Penelitian: Temuan ini dapat digunakan sebagai dasar untuk pengembangan lebih lanjut atau perbaikan instrumen pengukuran, meningkatkan validitas, reliabilitas, dan keberlanjutan tes.

#### F. Daftar Pustaka

- American Educational Research Association, A. P. A., & Education., N. C. on M. in. (2014). *Standards for Educational and Psychological Testing* (Washington). American Educational Research Association.
- Anastasi, A., & Urbina, S. (1997). *Psychological Testing*. Upper Saddle River, NJ: Prentice Hall
- Brown, G., & Race, P. (2002). *Using effective questions*. In *The Lecturer's Toolkit: A Practical Guide to Assessment, Learning and Teaching*, 121.
- Bruner, J. (1983). *Child's Talk: Learning to Use Language*. Norton.
- Camilli, G., & Shepard, L. A. (1994). *Methods for Identifying Biased Test Items*. CA: Sage Publications.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. FL: Holt, Rinehart and Winston.
- Derewianka, B. (1990). *A Grammar for Writing*. Primary English Teaching Association Australia.

- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Routledge.
- Dwipayani, A. A. (2013). Analisis Validitas dan Reliabilitas Butir Soal Ulangan Akhir Semester Bidang Studi Bahasa Indonesia Kelas X.D SMA N 1 Terhadap Pencapaian Kompetensi. *Jurnal Pendidikan Bahasa Dan Sastra UNDIKSHA*, 1(5).
- Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. *Jurnal Mudarirsuna*, 10(2), 58–74.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. NJ: Lawrence Erlbaum Associates.
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and Assessment in Teaching*. Macmillan.
- Haladyna, T. M., & Downing, S. M. (1989). *Validity of a taxonomy of multiple-choice item-writing rules*. *Applied Measurement in Education*, 2(1), 51–78.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). *A review of multiple-choice item-writing guidelines for classroom assessment*. *Applied Measurement in Education*, 15(3), 309–334.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Kluwer-Nijhoff Publishing.
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Journal Of Arabic Education*, 1(1). <https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Indonesia, D. A. R. (n.d.). *Kurikulum Pendidikan Agama Islam (PAI) di Madrasah*. [www.kemenag.go.id](http://www.kemenag.go.id)
- Ismail, M. I. (2020). Evaluasi pembelajaran. In Remaja Rosdakarya.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. NJ: Lawrence Erlbaum Associates.
- Mulyasa, E. (2013). *Kurikulum Tingkat Satuan Pendidikan: Konsep, Desain, dan Implementasi*. Remaja Rosdakarya.

- Mustofa, S. (2019). Penerapan Teknologi Informasi dalam Pembelajaran di Madrasah Ibtidaiyah. *Jurnal Pendidikan Indonesia*.
- Ni'mah, U. (2023a). *Improve Students ' Literacy With Multiple Intelligence ( MI ) Approach*. EDUL (*European Union Digital Library*), Mi. <https://doi.org/10.4108/eai.26-11-2022.2339550>
- Ni'mah, U. (2023b). Pendidikan Multikultural Dalam Membentuk Karakter Toleransi Di Mi Ma'arif Caruy Cilacap. *Jurnal Pendidikan Guru MI (J-PGMI)*, 6(1).
- Nitko, A. J. (2001). *Educational Assessment of Students*. NJ: Merrill Prentice Hall.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students (6th ed.)*. Pearson.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. NY: McGraw-Hill.
- Piaget, J. (1952). The Origins of Intelligence in Children.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.
- Rodriguez, M. C. (2005). *Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research*. *Educational Measurement: Issues and Practice*, 24(2), 3-13.
- Safira, P. (2016). *Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Bahasa Indonesia SMP Negeri 2 Magelang Tahun Pelajaran 2015/2016*. In Universitas Negeri Yogyakarta.
- Skinner, B. F. (1957). *Verbal Behavior*. Appleton-Century-Crofts.
- Sugiyono. (2012). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, R&D*. Alfabeta
- Sugiyono. (2019). *Metode Penelitian Kuantitatif*. Alfabeta.
- Sukmadinata, N. S. (2013). *Evaluasi Kurikulum: Teori dan Konsep Dasar*. Remaja Rosdakarya.
- Vygotsky., L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Zainuddin, M. (2018). Pendidikan Karakter: Konsep dan Implementasinya di Madrasah Ibtidaiyah. *Jurnal Ilmu Pendidikan Madrasah*.

This page is intentionally left blank