

Analisis Butir Soal Ujian Tengah Semester Pada Mata Pelajaran Pendidikan Agama Islam

Nurul Hidayah Siregar¹, Remiswal², Khadijah³

Universitas Islam Negeri Imam Bonjol Padang, Sumatera Barat, Indonesia ^{1,2,3}

nurulhidayahsrg08@gmail.com,¹ remiswal@uinib.ac.id,² khadijahmpd@uinib.ac.id³

Abstract: This article aims to analyze the midterm exam questions conducted at Pertiwi 1 Padang Private High School in class XII, this is to see how far the questions play an important role in the evaluation process. Is it in accordance with the learning objectives to be achieved. So a question item analysis is needed to determine its suitability. The research method used is descriptive quantitative research which aims to describe events that occur using statistical data. data collection is done by documenting questions, answer keys and student answer sheets. The results of the research on the analysis of semester exam questions show that the validity found in the analysis of questions is 15 at the 75% level while the invalid ones are 5 at the 5% level. The reliability is 0.775. So it can be concluded that the data from the question items have high criteria. The difficulty index obtained is in the medium criteria as many as 12 questions at a percentage of 60%. The differential power of the analysis question is in the good category with a total of 8 questions at a percentage of 40%. Thus, every question made by educators is expected to have a good level of validity, reliability, difficulty index, and differentiation according to the category or criteria of the questions made.

Keyword : *Analysis, Problem Items, Islamic Religious Education*

Abstrak: *Artikel ini bertujuan untuk menganalisis soal ujian tengah semester yang dilakukan di SMA Swasta Pertiwi 1 Padang pada kelas XII, hal ini melihat sejauh mana soal berperan penting dalam proses evaluasi. Apakah sudah sesuai dengan tujuan pembelajaran yang ingin dicapai. Jadi diperlukan analisis butir soal untuk mengetahui kesesuaiannya. Adapun metode peneliti yang digunakan ialah jenis penelitian deskriptif kuantitatif yang bertujuan untuk menggambarkan peristiwa yang terjadi dengan menggunakan data statistik. pengambilan data dilakukan dengan cara dokumentasi soal, kunci jawaban dan lembar jawaban siswa. Hasil penelitian analisis soal ujian semester menunjukkan bahwa validitas yang ditemukan pada analisis soal sebanyak 15 pada taraf 75% sedangkan yang tidak valid sebanyak 5 orang pada taraf 5%. Reliabilitasnya ialah sebesar 0,775. Maka dapat disimpulkan bahwa data dari butir soal memiliki kriteria yang tinggi. Indeks kesukaran yang di dapat yaitu pada kriteria sedang sebanyak 12 soal pada persentase 60%. Daya beda soal analisis nya pada*

kategori baik dengan jumlah soal sebanyak 8 pada persentase 40 %. Dengan demikian setiap soal yang dibuat oleh pendidik diharapkan memiliki tingkat validitas, reliabilitas, indeks kesukaran soal, serta daya beda yang baik sesuai dengan kategori atau kriteria soal yang dibuat.

Kata kunci: *Analisis, Butir Soal, Pendidikan Agama Islam*

Corresponding Author:

Nurul Hidayah Siregar

Universitas Islam Negeri Imam Bonjol Padang, Sumatera Barat, Indonesia ;

nurulhidayahsrg08@gmail.com

Pendahuluan

Evaluasi dalam dunia pendidikan merupakan komponen penting yang tidak hanya membantu mengukur efektivitas proses belajar mengajar, tetapi juga menjadi alat untuk memajukan kualitas pendidikan itu sendiri. Melalui evaluasi, pendidik dapat mengetahui sejauh mana tujuan pembelajaran telah tercapai, mengidentifikasi kelemahan dan kelebihan metode pengajaran, serta merancang strategi yang lebih baik untuk meningkatkan hasil belajar siswa (Mahirah, 2017; Munandar dkk., 2023). Selain itu, evaluasi juga berfungsi sebagai sarana untuk memberikan umpan balik yang konstruktif baik bagi siswa maupun guru, sehingga mendorong perbaikan yang berkesinambungan dalam sistem pendidikan. Dalam konteks yang lebih luas, evaluasi juga membantu para pembuat kebijakan pendidikan dalam merumuskan kebijakan yang lebih efektif dan responsif terhadap kebutuhan dunia pendidikan yang terus berkembang (Zuanda dkk., 2024).

Evaluasi juga bertujuan untuk mengetahui, memperbaiki serta meningkatkan kualitas proses pembelajaran yang dilakukan. Evaluasi hasil pembelajaran tersebut nantinya akan memberikan gambaran tentang tingkat keberhasilan pembelajaran yang dilakukan, antara lain dalam hal pencapaian terhadap ketuntasan belajar peserta didik (Fatzuarni, 2022; Magdalena dkk., 2020). Evaluasi dapat dilaksanakan melalui ulangan harian, UTS, maupun UAS. Dengan melakukan analisis soal secara sistematis, sehingga mampu memberikan gambaran yang akurat mengenai kompetensi siswa. Hal ini tidak hanya meningkatkan kualitas penilaian tetapi juga membantu dalam meningkatkan proses pembelajaran secara keseluruhan (Arikunto, 2011; Bulqis, 2019).

Dalam proses evaluasi kita memerlukan suatu instrument atau alat tes. Alat tes merupakan salah satu alat yang dapat digunakan oleh guru untuk mengevaluasi seberapa besar penyerapan materi yang telah disampaikan dalam proses pembelajaran yang telah berlangsung. Soal yang baik adalah instrumen yang memenuhi syarat-syarat atau kaidah-kaidah tertentu, agar dapat memberikan data yang akurat sesuai dengan fungsinya, kemudian dalam hal proses evaluasi, tes yang baik adalah tes yang dapat menggambarkan keadaan siswa (Purwanto, 2018; Supriyadi, 2011).

Kemudian dalam membuat suatu tes yang perlu diperhatikan agar evaluasi dapat memberikan hasil yang akurat dan dapat dipercaya, kualitas soal harus diperhatikan.

Analisis soal bertujuan untuk memastikan bahwa setiap item dalam ujian valid (mengukur apa yang seharusnya diukur) dan reliabel (konsisten dalam pengukurannya) (Afiyanti, 2008; Yusup, 2018). Hal ini penting untuk menjamin bahwa hasil ujian benar-benar mencerminkan kemampuan peserta didik. Soal-soal dalam ujian harus memiliki derajat kesukaran yang bervariasi agar dapat mengukur seluruh kemampuan peserta didik. Selain itu, daya pembeda soal harus baik, yakni mampu membedakan antara siswa yang memahami materi dengan baik dan yang kurang memahami.

Maka dalam hal ini diperlukan analisis butir soal bertujuan untuk mengetahui apakah butir-butir item yang membangun tes hasil belajar tersebut sudah dapat menjalankan fungsinya sebagai alat pengukur hasil belajar yang memadai atau belum. Analisis butir soal juga bertujuan untuk mengidentifikasi soal-soal yang baik, kurang baik, dan soal yang jelek. Kegiatan analisis soal menghasilkan informasi tentang kejelekan sebuah soal dan petunjuk untuk mengadakan perbaikan. Penganalisisan terhadap butir-butir soal tes hasil belajar dapat dilakukan dari empat segi, yaitu segi validitas, reliabilitas, kesukaran itemnya, dan segi daya pembeda itemnya (Indah dkk., 2021).

Terkait dengan analisis butir soal, banyak penelitian yang melakukan analisisnya, misalnya pada penelitian butir soal yang dilakukan, peneliti menemukan bahwa analisis soal sangat penting dalam mengetahui kualitas soal dan meningkatkan kemampuan siswa (Sari dkk., 2022). Analisis butir soal juga membantu dalam menentukan kualitas soal yang valid dan reliabel, serta membedakan siswa yang pandai dan yang kurang pandai (Elviana, 2020). Analisis butir soal yang diperlukan untuk meningkatkan mutu pendidikan. Guru harus dapat mengevaluasi apakah evaluasi yang telah disusun sudah dapat menjalankan fungsinya sebagai alat pengukur hasil belajar yang memiliki kualitas yang baik (Muluki, 2020). Dengan demikian analisis butir soal menjadi sangat penting dalam memantau dan meningkatkan kualitas proses belajar mengajar di Indonesia. Guru hendaknya dapat melakukan analisis butir-butir soal secara terus-menerus untuk meningkatkan kualitas soal dan hasil belajar peserta didik.

Kemudian pada mata pelajaran Pendidikan Agama Islam (PAI) memiliki peran penting dalam pembentukan karakter dan pemahaman keagamaan siswa. Oleh karena itu, analisis butir soal pada mata pelajaran ini menjadi sangat relevan untuk memastikan bahwa evaluasi yang dilakukan benar-benar efektif dalam mengukur kompetensi siswa sesuai dengan tujuan pendidikan agama Islam. Adapun soal yang akan dianalisis yaitu pada kelas XII SMA Swasta Pertiwi 1 Padang. Dimana dalam penelitian ini yang akan dibahas yaitu, mengukur validitas setiap butir soal untuk memastikan kesesuaian dengan kompetensi yang diharapkan. Kemudian menilai reliabilitas soal untuk melihat konsistensinya hasil penilaian. Lalu mengkategorikan tingkat kesukaran butir soal untuk menilai distribusi soal. Dan menganalisis daya beda setiap butir soal untuk menilai kemampuan soal dalam membedakan kemampuan peserta didik. Dengan melakukan analisis ini diharapkan dapat memberikan rekomendasi yang jelas mengenai butir soal mana yang disimpan dan butir soal mana yang perlu direvisi atau bahkan tidak digunakan lagi untuk sebuah soal. Sehingga penelitian ini dapat menjadi kontribusi dalam peningkatan kualitas pembelajaran dan secara lebih luasnya pada pendidikan di Indonesia

Metode Penelitian

Adapun metode penelitian yang digunakan ialah jenis penelitian deskriptif kuantitatif yang bertujuan untuk menggambarkan peristiwa yang terjadi dengan menggunakan data statistik (Sugiyono, 2013). Prosedur penelitian ini melewati 2 tahap yaitu tahap persiapan dan tahap pelaksanaan. Secara garis besar tahap pelaksanaan yang dilakukan sebagai berikut: 1) pengumpulan data, 2) mempersiapkan dan menjalankan program SPSS 26, 3) mendeskripsikan hasil perhitungan, 4) membahas atau menelaah soal yang menurut perhitungan menunjukkan kualitas yang kurang baik. Teknik pengambilan data dilakukan dengan cara dokumentasi soal, kunci jawaban dan lembar jawaban siswa (Ramdhan, 2021). Metode analisis data menggunakan program SPSS 26. Analisis butir soal menggunakan validitas, reliabilitas, indeks kesukaran serta daya beda soalnya. Adapun populasi dalam penelitian ini sebanyak 4 kelas Mia, namun pada penelitian ini hanya menggunakan sampel sebanyak 1 kelas yaitu 20 orang yaitu pada kelas XII Mia I. Adapun teknik pengambilan sampel yaitu dengan menggunakan teknik *purposive sampling* digunakan dalam penelitian ini untuk memilih subjek yang memiliki karakteristik tertentu yang relevan dengan tujuan penelitian, serta untuk memastikan bahwa sampel yang diambil memiliki kualitas yang baik dan representatif. Sehingga akan lebih mudah dalam menentukan hasil analisis butir soal (Arikunto, 2013).

HASIL PENELITIAN DAN PEMBAHASAN

Hasil Penelitian

Mengenai temuan hasil penelitian analisis yang dilakukan dengan butir soal menggunakan SPSS 26. Analisis soal Ujian Tengah Semester (UTS) dilakukan dengan 20 soal objektif. Adapun yang dianalisis mengenai validitas, reliabilitas, kesukaran soal, serta daya beda soal. Untuk lebih lanjut kita lihat Tabel Berikut ini:

Tabel.1

Hasil Analisis Validitas Butir Soal

Kategori	Jumlah	Persentase	Nomor Soal
Valid	15	75%	1,2,3,4,5,6,7,8,11,13,14,15,18,19,20
Tidak Valid	5	25%	9,10,12,16,17
Jumlah	20	100%	

Validitas disini menggunakan rumus Pearson yang dapat dilihat dari SPSS 26. Berdasarkan Tabel diatas dapat disimpulkan bahwa dari 20 butir soal maka yang memenuhi nilai signifikan $< 0,005$ yaitu valid sebanyak 15 soal dengan jumlah persentase 75% sedangkan yang tidak valid sebanyak 5 soal dengan persentase 25%. Dengan Interpretasi validitas soal tinggi sebanyak 6 soal kemudian interpretasi soal cukup ada 9 soal.

Tabel.2

Hasil Analisis Reliabilitas Soal

Reliability Statistics

Cronbach's Alpha	N of Items
.775	20

Kemudian dalam hal reliabilitas untuk lebih jelasnya reliabilitas adalah ukuran seberapa konsisten dan stabil sebuah tes atau alat ukur dalam memberikan hasil yang sama jika diulang pada kondisi yang sama. Jika sebuah tes memiliki reliabilitas tinggi, itu berarti tes tersebut dapat diandalkan untuk memberikan hasil yang konsisten, bahkan jika dilakukan oleh orang yang berbeda, pada waktu dan tempat yang berbeda, terhadap subjek yang sama. Alat ukur yang memiliki reliabilitas tinggi disebut reliabel. Tujuan dari uji reliabilitas ini adalah untuk menunjukkan konsistensi skor-skor yang diberikan skor satu dengan skor lainnya (Khafidin, 2014; Purwanto, 2018) . Menurut Budiastuti (2022) dalam hal ini nilai interpretasi Korelasi reliabilitas ini kita akan melihat pada tabel.3 berikut ini:

Tabel.3

Koefisien Korelasi Reliabilitas

Koefisien Korelasi	Kriteria
$0,8 \leq r \leq 1,0$	Reliabilitas Sangat Tinggi
$0,6 \leq r < 0,8$	Reliabilitas Tinggi
$0,4 \leq r < 0,6$	Reliabilitas Cukup
$0,2 \leq r < 0,4$	Reliabilitas Rendah
$r < 0,2$	Reliabilitas Sangat Rendah

Jika dilihat dari data analisis butir soal yang dianalisis dengan menggunakan SPSS. 26, maka reliabilitasnya ialah sebesar 0,775. Maka dapat disimpulkan bahwa data dari butir soal memiliki kriteria yang tinggi.

Tabel.4

Hasil Analisis Indeks Kesukaran Butir Soal

Kriteria Soal	Nomor Soal	Jumlah	Persentase
Sukar	-	0	0
Sedang	1,2,3,4,5,6,7,11,13,14,15,18	12	60%
Mudah	8,9,10,12,16,17,19,20	8	40%
Jumlah		20	100%

Selanjutnya untuk melihat indeks Menurut Nitko dalam Arifin (2017) kesukaran soal maka kita lihat dulu kriteria penilaiannya yaitu jika 0,00-0,029 maka kriteria kesukaran soalnya mudah. Kemudian jika 0,30-0,69 kriteria kesukaran soalnya sangat mudah. Dan yang terakhir 0,70-1,00 kriteria kesukaran soalnya sukar. Dari tabel. 3 dapat dilihat bahwa sebanyak 12 soal mendapatkan indeks kesukaran dengan kriteria sedang dengan taraf 60%, kemudian pada kriteria mudah sebanyak 8 soal dengan persentase 40%, sedangkan untuk kriteria yang sukar tidak ada yang mewakili soalnya. Jadi dapat disimpulkan kategori indeks kesukaran analisis butir soal yaitu kategori sedang.

Tabel. 4
Hasil Analisis Daya Beda Soal

Kriteria Soal	Nomor Soal	Jumlah	Persentase
Jelek	7,9,10,15,16	5	25%
Cukup	1,4,11.17.19,20	6	30%
Baik	2,3,5,6,8,13,14,18	8	40%
Sangat Baik	1	1	5%
Jumlah		20	100%

Daya beda analisis butir soal adalah pengukuran sejauh mana suatu butir soal mampu membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum atau kurang menguasai kompetensi berdasarkan kriteria tertentu. Semakin tinggi koefisien daya pembeda suatu butir soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi. Dalam daya beda juga memiliki beberapa kategori yaitu: jika 0,00-0,19 maka daya bedanya pada kategori jelek, 0,20-0,39 dalam kategori cukup, kemudian 0,40-0,69 daya bedanya kategori baik, dan 0,70-1,00 pada kategori baik sekali (Arikunto, 2011; Solichin, 2017). Untuk tabel.4 maka dilihat daya beda soalnya pada kriteria jelek sebanyak 5 soal bertaraf 25%, kemudian dalam kriteria cukup sebanyak 6 soal pada persentase 30%. Selanjutnya pada kriteria baik sebanyak 8 soal dengan persentase 40%. Dan kriteria sangat baik ada 1 soal pada taraf 5%. Dapat disimpulkan daya beda analisis soal yang dilakukan terdapat pada kriteria baik.

Hasil Pembahasan

Penelitian ini menunjukkan bahwa analisis butir soal Pendidikan Agama Islam (PAI) sangat dibutuhkan. Untuk mengetahui apakah hasil pembelajaran yang sudah dipelajari sesuai dengan tujuan pembelajaran yang ingin dicapai. Adapun hasil pembahasan analisisnya sebagai berikut ini:

1. Validitas

Sebuah instrumen penilaian dapat dikategorikan baik jika memiliki validitas soal memiliki kriteria tinggi. Validitas merupakan suatu konsep sejauh mana alat ukur evaluasi dapat menilai sesuai dengan apa yang sedang dinilainya. Artinya sebuah alat ukur tersebut secara keseluruhan menggambarkan apa yang akan diukurinya (Kadir, 2015; Zainab dkk., 2017). Setelah diberikan soal Ujian Tengah Semester dalam hal validitas peneliti menemukan beberapa soal yang tidak valid, yaitu dari 20 soal objektif yang diberikan ada 5 soal yang tidak valid. Adapun soal yang tidak valid pada nomor soal 9,10,12,16, dan 17. Menurut Sukardi dalam Mahmudah dkk., (2016), menyatakan terdapat banyak faktor yang dapat mempengaruhi hasil tes evaluasi tidak valid. Beberapa faktor tersebut secara garis besar dapat dibedakan menurut sumbernya, yaitu faktor internal dari tes, faktor eksternal tes, dan faktor yang berasal dari peserta didik yang bersangkutan.

a. Faktor yang berasal dari dalam tes

- 1) Arahan tes yang disusun dengan makna tidak jelas sehingga dapat mengurangi validitas tes.
- 2) Kata-kata yang digunakan dalam struktur instrument evaluasi, tidak terlalu sulit.
- 3) Item tes dikonstruksi dengan jelas.
- 4) Tingkat kesulitan item tes tidak tepat dengan materi pembelajaran yang diterima peserta didik.
- 5) Waktu yang dialokasikan tidak tepat, hal ini termasuk kemungkinan terlalu kurang atau terlalu longgar.
- 6) Jumlah item terlalu sedikit sehingga tidak mewakili sampel
- 7) Jawaban masing-masing item evaluasi bisa diprediksi peserta didik

b. Faktor yang berasal dari administrasi dan skor tes.

- 1) Waktu pengerjaan tidak cukup sehingga peserta didik dalam memberikan jawaban dalam situasi tergesa-gesa.
- 2) Adanya kecurangan dalam tes sehingga tidak membedakan antara peserta didik yang belajar dengan melakukan kecurangan.
- 3) Pemberian petunjuk dari dari pengawas yang tidak dapat dilakukan pada semua peserta didik.
- 4) Teknik pemberian skor yang tidak konsisten.
- 5) Peserta didik tidak dapat mengikuti arahan yang diberikan dalam tes baku.
- 6) Adanya joki (orang lain bukan peserta didik) yang masuk dalam menjawab item tes yang diberikan.

c. Faktor yang berasal dari jawaban peserta didik Seringkali terjadi bahwa interpretasi terhadap item-item tes evaluasi tidak valid, karena dipengaruhi oleh jawaban peserta didik dari pada interpretasi item-item pada tes evaluasi.

Dengan demikian validitas tes evaluasi dipengaruhi oleh berbagai faktor, baik dari dalam tes itu sendiri, dari cara tes dikelola dan dinilai, maupun dari bagaimana peserta didik memahami dan menjawab item-item tes. Untuk memastikan bahwa tes

benar-benar mengukur apa yang seharusnya diukur, perlu dilakukan perbaikan pada setiap aspek ini.

2. Reliabilitas

Salah satu kriteria instrumen yang dapat dipercaya jika instrumen tersebut digunakan secara berulang-ulang, hasil pengukurannya tetap. Mistar dapat dipercaya sebagai alat ukur, karena berdasarkan pengalaman jika mistar digunakan dua kali atau lebih mengukur panjang sebuah benda, maka hasil pengukuran pertama dan selanjutnya terbukti tidak berbeda. Sebuah tes dapat dikatakan reliabel jika tes tersebut digunakan secara berulang terhadap peserta didik yang sama hasil pengukurannya relatif tetap sama (Aulia Rahman & Eva Nasyah, 2019).

Dalam hal reliabilitas ditemukan pada Jika dilihat dari data analisis butir soal yang dianalisis dengan menggunakan SPSS 26, maka reliabilitasnya ialah sebesar 0,775. Maka dapat disimpulkan bahwa data dari butir soal memiliki kriteria yang tinggi. Namun pada kenyataannya sering kita jumpai suatu soal yang tidak reliabel adapun faktor yang mempengaruhi reliabilitas suatu soal yaitu: 1) Panjang Tes. 2) Variabilitas Kelompok. 3) Objektivitas Penskoran. 4) Metode Estimasi Reliabilitas 5) Level Kelompok dan Tingkat Kesulitan Tes. 6) Homogenitas Tes (Setiyawan, 2014). Adapun faktor lainnya banyaknya sampel tes, fluktuasi dalam administrasi tes, faktor pribadi, instruksi/petunjuk dan fluktuasi dalam Penilaian Subjektivitas. Dengan melihat faktor yang mempengaruhi suatu soal itu reliabel maka kita dapat lebih mudah dalam mengidentifikasi alasan suatu soal tidak reliabel (Tosuncuoglu, 2018).

3. Indeks Kesukaran Soal

Tingkat kesukaran soal dilakukan agar soal tes dapat digunakan secara luas, penting untuk menyelidiki tingkat kesukarannya sehingga dapat membedakan soal yang tergolong mudah, sedang, dan sulit. Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sulit. Soal yang terlalu mudah tidak akan memberikan tantangan kepada peserta didik untuk meningkatkan kemampuan mereka dalam memecahkan masalah. Sebaliknya, soal yang terlalu sulit dapat membuat peserta didik merasa putus asa dan menjawab tanpa berpikir panjang berdasarkan pengetahuan mereka (Fitrianawati, 2017). Sehingga dapat disimpulkan Indeks kesukaran yang didapat yaitu pada kriteria sedang sebanyak 12 soal pada persentase 60 %.

4. Daya beda

Soal yang baik adalah soal yang mampu membedakan antara kemampuan peserta didik yang pandai dan peserta didik yang rendah. Perhitungan daya pembeda adalah pengukuran sejauh mana suatu butir soal mampu membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum atau kurang menguasai kompetensi berdasarkan kriteria tertentu. Semakin tinggi koefisien daya pembeda suatu butir soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi (Ilyas Ismail, 2020). Adapun daya beda soal yang dianalisis

dapat dikategorikan baik dengan nomor soal 2,3,5,6,8,13,14,18 dengan jumlah 8 butir soal.

Kesimpulan

Mayoritas butir soal memiliki validitas yang baik, menunjukkan bahwa soal-soal tersebut secara efektif mengukur kompetensi yang diharapkan. Namun, beberapa butir soal perlu revisi untuk memastikan mereka lebih akurat mengukur tujuan pembelajaran. Validitas yang ditemukan pada analisis soal sebanyak 15 pada taraf 75% sedangkan yang tidak valid sebanyak 5 orang pada taraf 5%. Reliabilitas tes keseluruhan berada pada kategori tinggi, menunjukkan bahwa tes ini konsisten dalam mengukur kemampuan peserta didik. Hal ini berarti bahwa hasil tes dapat dipercaya dan digunakan untuk penilaian yang berkelanjutan. Reliabilitasnya ialah sebesar 0,775. Maka dapat disimpulkan bahwa data dari butir soal memiliki kriteria yang tinggi.

Distribusi indeks kesukaran menunjukkan bahwa sebagian besar soal berada dalam kategori sedang (tidak terlalu mudah dan tidak terlalu sulit). Namun, ada beberapa soal yang terlalu mudah atau terlalu sulit yang perlu diperbaiki atau diganti untuk meningkatkan distribusi kesukaran yang lebih seimbang. Indeks kesukaran yang didapat yaitu pada kriteria sedang sebanyak 12 soal pada persentase 60%. Sebagian besar butir soal memiliki daya beda yang baik, yang berarti soal-soal tersebut mampu membedakan antara siswa yang memiliki kemampuan tinggi dan rendah. Beberapa soal dengan daya beda rendah harus direvisi atau dihapus untuk meningkatkan efektivitas tes dalam mengevaluasi kemampuan peserta didik. Kemudian daya beda soal analisisnya pada kategori baik dengan jumlah soal sebanyak 8 pada persentase 40%.

DAFTAR PUSTAKA

- Afiyanti, Y. (2008). Validitas Dan Reliabilitas Dalam Penelitian Kualitatif. *Jurnal Keperawatan Indonesia*, 12(2), 137-141.
- Arifin, Z. (2017). Kriteria Instrumen Dalam Suatu Penelitian. *Jurnal Theorems (The Original Research Of Mathematics)*, 2(1).
<https://www.unma.ac.id/jurnal/index.php/th/article/view/571>
- Arikunto, S. (2011). *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara.
- Arikunto, S. (2013). *Prosedur Penelitian Suatu Pendekatan Praktek*. Rineka Cipta.
- Aulia Rahman, A., & Eva Nasyah, C. (2019). *Evaluasi Pembelajaran*. Uwais Inspirasi Indonesia.
- Budiastuti, D. (2022). *Validitas Dan Reliabilitas Penelitian*. https://repo.stikes-ibnusina.ac.id/xmlui/bitstream/handle/123456789/147/1686032019154_validitas%20dan%20reliabilitas.pdf?sequence=1
- Bulqis, A. (2019). Analisis Butir Soal Pilihan Ganda Biasa Ulangan Tengah Semester Buatan Guru Biologi Kelas X Sma Negeri 3 Maros. *Prosiding Seminar Nasional Fkip Universitas Muslim Maros*, 1, 191-197.
<http://ejournals.umma.ac.id/index.php/prosiding/article/view/377>
- Elviana, E. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pai Menggunakan Program Anates. *Jurnal Mudarrisuna: Media Kajian Pendidikan Agama Islam*, 10(2), 209-224.
- Fatuzuarni, M. (2022). *Artikel Pentingnya Evaluasi Dalam Proses Pembelajaran*. <https://osf.io/preprints/g8h3p/>
- Fitrianawati, M. (2017). *Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik*. <https://publikasiilmiah.ums.ac.id/xmlui/handle/11617/9117>
- Ilyas Ismail, M. (2020). *Evaluasi Pembelajaran*. Raja Grafindo Persada.
- Indah, M., Karoma, K., & Rusdi, A. (2021). Analisis Tes Butir Soal Guru Dalam Mata Pelajaran Pendidikan Agama Islam (Pai) Di Sekolah Menengah Pertama Negeri 8 Palembang. *Muaddib: Islamic Education Journal*, 4(1), 21-28.
<https://doi.org/10.19109/muaddib.v4i1.8860>
- Khafidin, Z. (2014). Analisis Validitas Dan Reliabilitas Tes Mata Pelajaran Pendidikan Agama Islam Tingkat Sma. *Edukasia: Jurnal Penelitian Pendidikan Islam*, 9(2).
<http://journal.iainkudus.ac.id/index.php/edukasia/article/view/775>
- Magdalena, I., Ridwanita, A., & Aulia, B. (2020). Evaluasi Belajar Peserta Didik. *Pandawa*, 2(1), 117-127.
- Mahirah, B. (2017). Evaluasi Belajar Peserta Didik (Siswa). *Idaarah: Jurnal Manajemen Pendidikan*, 1(2). <https://journal3.uin-alauddin.ac.id/index.php/idaarah/article/view/4269>
- Mahmudah, R., Pramudya, Y., & Sulisworo, D. (2016). Analisis Validitas Butir Soal Certainty Of Respons Index (Cri) Untuk Identifikasi Miskonsepsi Materi Tata Surya Dan Fenomena Astronomi. *Prosiding Seminar Nasional Pendidikan Sains*, 579-587.

- Muluki, A. (2020). Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran Ipa Kelas Iv Mi Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 86. <https://doi.org/10.23887/jisd.v4i1.23335>
- Munandar, A., Alfian, M. R., Echa, A. J., Zora, K. A., Aprianti, A., Mulyani, G., Mujahidin, M., Rahmawati, E., Febiola, F., & Pitriani, P. (2023). Evaluasi Program Pendidikan Karakter. *Innovative: Journal Of Social Science Research*, 3(6), 682–688.
- Purwanto, P. (2018). *Teknik Penyusunan Instrumen Uji Validitas Dan Reliabilitas Penelitian Ekonomi Syariah*.
- Ramadhan, M. (2021). *Metode Penelitian*. Cipta Media Nusantara.
- Sari, V. N. I., Utomo, A. P. Y., & Sumarwati, S. (2022). Kualitas Soal Bahasa Indonesia Di Smp Muhammadiyah 1 Pontianak: Analisis Butir Soal. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 11(2), 112–119.
- Setiyawan, A. (2014). Faktor-Faktor Yang Mempengaruhi Reliabilitas Tes. *An Nur: Jurnal Studi Islam*, 6(2). <https://jurnalannur.ac.id/index.php/an-nur/article/view/53>
- Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes Dan Validitas Ramalan Dalam Evaluasi Pendidikan. *Dirasat: Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 192–213.
- Sugiyono, S. (2013). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, Dan R&D*. Alfabeta.
- Supriyadi, G. (2011). *Pengantar Dan Teknik Evaluasi Pembelajaran*. Intimedia.
- Tosuncuoglu, I. (2018). Importance Of Assessment In Elt. *Journal Of Education And Training Studies*, 6(9), 163–167.
- Yusup, F. (2018). Uji Validitas Dan Reliabilitas Instrumen Penelitian Kuantitatif. *Tarbiyah: Jurnal Ilmiah Kependidikan*, 7(1). <http://103.180.95.17/index.php/jtik/article/view/2100>
- Zuanda, S., Fahrezi, D. W., Rised, G. F., & Syaifuddin, M. (2024). Implementasi Dan Evaluasi Kebijakan Pendidikan. *Jurnal Pendidikan Tambusai*, 8(2), 27207–27218.