

ANALISIS KUALITAS BUTIR SOAL SUMATIF AKHIR SEMESTER MATA PELAJARAN PENDIDIKAN AGAMA ISLAM

Agus Rifai¹, Riski Ramadhan², Miftahir Rizqa³

^{1,2,3}Pasca Sarjana UIN Suska Riau

¹agusrifai01@gmail.com, ²rizkiramdhanpku8@gmail.com

³Miftahir.rizqa@uin-suska.ac.id

ABSTRACT

This study aims to analyze the quality of the Final Semester Summative test items for Islamic Religious Education in 9th grade at SMP Negeri 2 Ukui, looking at aspects of validity, reliability, discriminating power, and difficulty level. The study uses a descriptive quantitative approach. The subjects of the study consisted of 31 9th-grade students, and the instrument used was 50 multiple-choice questions. Data collection was done through documentation of students' answer sheets and test papers, while data analysis was carried out using item analysis calculations with the help of SPSS. The research results show that out of 50 questions, 35 questions are valid and 15 questions are invalid. The reliability test obtained a Cronbach's Alpha value of 0.912, which falls into the very high category. For the distinguishing power aspect, 14 questions are categorized as very good, 13 questions as good, 16 questions as fair, 6 questions as poor, and 1 question as very poor. Meanwhile, in terms of difficulty level, 40 questions are categorized as easy, 8 questions as medium, and 2 questions as hard. Based on this analysis, it can be concluded that the quality of the End-of-Semester Summative Exam questions for Islamic Religious Education is generally quite good and feasible to use.

Keywords: Item analysis, Validity, Reliability, Discriminative power, Difficulty level, Islamic Religious Education

ABSTRAK

Penelitian ini bertujuan untuk menganalisis kualitas butir soal Sumatif Akhir Semester Genap mata pelajaran Pendidikan Agama Islam kelas IX SMP Negeri 2 Ukui ditinjau dari aspek validitas, reliabilitas, daya pembeda, dan tingkat kesukaran. Penelitian ini menggunakan pendekatan kuantitatif deskriptif. Subjek dalam penelitian ini terdiri dari 31 murid kelas IX dengan instrument 50 butir soal pilihan ganda. Teknik pengumpulan data dilakukan melalui dokumentasi lembar jawaban murid dan naskah soal, sedangkan analisis data dilakukan dengan perhitungan analisis soal dan bantuan SPSS. Adapun hasil penelitian menunjukkan bahwa dari 50 butir soal terdapat 35 butir soal valid dan 15 butir soal tidak valid. Uji reliabilitas memperoleh nilai *Cronbach's Alpha* sebesar 0,912 yang termasuk kategori sangat tinggi. Pada aspek daya pembeda diperoleh 14 butir soal berkategori sangat baik, 13 butir soal berkategori baik, 16 butir soal berkategori cukup, 6 butir soal berkategori jelek, dan 1 butir soal berkategori sangat jelek. Sementara pada aspek tingkat kesukaran menunjukkan 40 butir soal berkategori mudah, 8 butir soal berkategori sedang, dan 2 butir soal berkategori sukar. Berdasarkan hasil analisis ini, dapat ditarik benang merah bahwa kualitas soal Sumatif Akhir Semester Genap

mata pelajaran Pendidikan Agama Islam secara umum sudah cukup baik dan layak digunakan.

Kata Kunci: Analisis butir soal, Validitas, Reliabilitas, Daya pembeda, Tingkat kesukaran, Pendidikan Agama Islam

A. Pendahuluan

Kualitas pembelajaran yang telah dilakukan atau sedang dilakukan oleh murid menjadi salah satu hal yang krusial dalam melihat *output* yang akan datang. Salah satu instrument yang dapat digunakan dalam melihat ketuntasan murid dalam belajar adalah menggunakan tes. Pada kurikulum merdeka, salah satu tes yang dapat digunakan adalah Sumatif Akhir Semester (SAS) yang dilakukan di akhir semester. Kualitas murid juga dapat dilihat bagaimana ia dapat menyelesaikan dan memahami setiap butir soal yang paparkan dalam ujian.

Salah satu instrument yang dapat diberikan kepada murid untuk mengukur kemampuan kognitifnya adalah tes. Penilaian (*assessment*) dimaknakan sebagai Langkah yang dipakai untuk mendapatkan informasi guna mengukur taraf pengetahuan dan keterampilan murid yang hasilnya akan digunakan untuk keperluan evaluasi (Subali, 2010).

Mengingat pentingnya kualitas pembuatan butir soal pada suatu tes,

maka perlu adanya untuk di analisis butir soal tersebut. Analisis butir soal merupakan proses evaluasi kuantitatif terhadap kualitas butir-butir soal yang digunakan dalam suatu tes (Pardomuan, dkk., 2024). Hasil analisis tes yang mumpuni akan memberikan pencapaian belajar yang optimal. Salah satu cara pencapaian hasil belajar adalah dengan memberikan soal tes dalam bentuk ujian.

Hasil kriteria soal tes didapatkan dari analisis kualitas tes. Analisis tes bertujuan untuk: (1) memberikan pengetahuan tentang kelebihan dan kekurangan suatu lembar soal; (2) memberikan informasi lengkap mengenai spesifikasi butir soal; (3) mengetahui permasalahan yang terdapat pada butir soal; (4) penilaian kumpulan bank soal; (5) referensi untuk menyusun item soal (Hendra, dkk., 2022).

Selain itu, kualitas suatu butir soal juga dapat menentukan tingkat keakuratan hasil evaluasi pembelajaran. Butir soal yang baik harus mampu mengukur kemampuan

murid secara objektif, memiliki tingkat kesukaran yang proporsional, serta mampu membedakan murid yang memiliki kemampuan tinggi dan rendah. Oleh karena itu, analisis butir soal menjadi bagian penting dalam proses evaluasi pembelajaran agar instrumen yang digunakan benar-benar layak dan sesuai dengan tujuan pembelajaran yang ingin dicapai.

Dalam konteks Kurikulum Merdeka, evaluasi pembelajaran tidak hanya berorientasi pada hasil akhir, tetapi juga pada proses pembentukan kompetensi murid. Guru dituntut untuk mampu menyusun instrumen evaluasi yang berkualitas agar hasil asesmen dapat menggambarkan kemampuan murid secara menyeluruh. Dengan demikian, analisis butir soal pada Sumatif Akhir Semester menjadi langkah strategis dalam meningkatkan mutu evaluasi pembelajaran, khususnya pada mata pelajaran Pendidikan Agama Islam.

Berdasarkan uraian tersebut, analisis butir soal Sumatif Akhir Semester perlu dilakukan untuk mengetahui kualitas instrumen yang digunakan dalam proses evaluasi pembelajaran. Melalui analisis ini, dapat diketahui tingkat validitas, reliabilitas, daya pembeda, dan tingkat

kesukaran pada setiap butir soal sehingga instrumen yang digunakan benar-benar mampu mengukur pencapaian hasil belajar murid secara optimal.

B. Metode Penelitian

Jenis dan Pendekatan Penelitian

Penelitian ini adalah jenis penelitian deskriptif dengan menggunakan pendekatan kuantitatif. Penelitian deskriptif bertujuan untuk menggambarkan secara sistematis dan valid mengenai fakta serta karakteristik objek yang diteliti tanpa memberikan perlakuan tertentu terhadap variable penelitian (Arikunto, 2018). Sedangkan pendekatan kuantitatif merupakan penelitian yang bermula dari satu teori menuju bentuk perolehan suatu data numerik di lapangan (Yeni, 2021). Pendekatan kuantitatif digunakan untuk melihat dan menganalisis data yang berupa skor hasil tes murid yang kemudian diolah secara statistic dengan tujuan mengetahui kualitas butir soal dengan mengacu validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan analisis distractor.

Subjek Penelitian

Subjek penelitian ini adalah murid kelas IX SMP Negeri 2 Ukui yang mengikuti Sumatif Akhir Semester mata pelajaran Pendidikan Agama Islam pada semester genap Tahun Pelajaran 2025/2026. Jumlah murid yang menjadi subjek penelitian sebanyak 31 siswa. Objek penelitian ini adalah butir soal Sumatif Akhir Semester mata pelajaran Pendidikan Agama Islam kelas IX yang terdiri atas 50 butir soal pilihan ganda beserta lembar jawaban siswa. Analisis dilakukan terhadap kualitas setiap butir soal berdasarkan respons jawaban murid.

Bentuk tes yang digunakan adalah pilihan ganda dengan empat opsional jawaban. Instrumen tes adalah metode yang digunakan untuk menilai pemahaman, pengetahuan, atau keterampilan individu dalam bidang tertentu (Afrilianti, 2025). Data yang diperoleh dari lembar jawaban murid selanjutnya akan dianalisis setiap butir soalnya secara kuantitatif.

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah teknik dokumentasi. Teknik dokumentasi dilakukan dengan mengumpulkan dan mengkaji dokumen berupa naskah soal, kunci jawaban, serta lembar jawaban murid

yang telah mengikuti Sumatif Akhir Semester. Seluruh dokumen tersebut kemudian digunakan sebagai sumber data utama dalam analisis butir soal.

Teknik Analisis Data

Analisis Validitas Butir Soal

digunakan untuk mengetahui sejauh mana suatu item soal memberikan kontribusi terhadap skor total tes. Untuk menentukan validitas setiap butir soal dilakukan dengan menggunakan rumus korelasi product moment pearson (Arikunto, 2018). Rumus yang digunakan dalam mengukur validitas adalah:

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Adapun interpretasi besarnya koefisien korelasi sebagai berikut:

Tabel 1. Kategori Validitas Soal

No	Skala	Kategori
1	0,81 – 1,00	Sangat tinggi
2	0,61 – 0,80	Tinggi
3	0,41 – 0,60	Cukup
4	0,21 – 0,40	Rendah
5	0,00 – 0,20	Sangat rendah

Analisis Reliabilitas Tes dilakukan untuk mengetahui tingkat konsistensi instrumen tes. Reliabilitas terkait dengan konsistensi, keandalan, keajegan, ataupun stabilitas. Suatu alat ukur dikatakan reliabel bila

memberikan hasil yang sama pada berkali-kali pengulangan (Subali, 2010). Reliabilitas dihitung menggunakan koefisien Alpha Cronbach dengan bantuan SPSS. Adapun tabel yang digunakan dalam mengukur tingkat reliabilitas adalah:

Tabel 2. Kategori Reliabilitas

No	Skala	Kategori
1	$0,80 < r_{11} \leq 1,00$	Sangat tinggi
2	$0,60 < r_{11} \leq 0,80$	Tinggi
3	$0,40 < r_{11} \leq 0,60$	Cukup
4	$0,20 < r_{11} \leq 0,40$	Rendah
5	$-1,00 < r_{11} \leq 0,20$	Sangat rendah

Analisis Daya Pembeda

dilakukan untuk mengetahui kemampuan suatu butir soal dalam membedakan murid yang memiliki kemampuan tinggi dan kemampuan rendah. Adapun rumus yang digunakan dalam menghitung daya beda adalah:

$$(D) = \frac{BA}{JA} - \frac{BB}{JB} = PA - PB$$

Keterangan:

D = Daya Beda

BA = Jumlah jawaban benar kelompok atas

BB = Jumlah jawaban benar kelompok bawah

JA = Banyak murid kelompok atas

JB = Banyak murid kelompok bawah

PA = Taraf kesukaran kelompok atas

PB = Taraf kesukaran kelompok bawah

Dengan kriteria daya pembeda sebagai berikut:

Tabel 3. Daya Pembeda

No	Skala	Kategori
1	0,71 – 1,00	Sangat baik
2	0,41 – 0,70	Baik
3	0,21 – 0,40	Cukup
4	0,00 – 0,20	Jelek
5	- (negatif)	Sangat jelek

Analisis Tingkat Kesukaran

dilakukan untuk mengetahui kategori soal berdasarkan tingkat kemudahan atau kesukarannya. Indeks kesukaran dihitung dengan membandingkan jumlah siswa yang menjawab benar terhadap jumlah seluruh peserta tes. Adapun rumus yang digunakan dalam menghitung tingkat kesukaran adalah:

$$P = \frac{\text{Jumlah jawaban benar setiap butir soal}}{\text{Jumlah seluruh siswa}}$$

Dengan kategori sebagai berikut:

Tabel 4. Indeks Kesukaran

No	Skala	Kategori
1	0,71 – 1,00	Mudah
2	0,31 – 0,70	Sedang
3	0,00 – 0,30	Sukar

Hasil analisis kemudian disajikan dalam bentuk tabel, persentase, dan kategori interpretasi sehingga dapat memberikan gambaran mengenai kualitas butir soal Sumatif Akhir Semester mata pelajaran Pendidikan Agama Islam kelas IX di SMP Negeri 2 Ukui.

C. Hasil Penelitian dan Pembahasan

A. Hasil Penelitian

Berdasarkan hasil analisis 50 butir soal pilihan ganda Sumatif Akhir Semester Genap kelas IX SMP Negeri 2 Ukui, diperoleh data bahwa terdapat variasi kualitas soal yang dilihat dari validitas soal, reliabilitas, daya pembeda, dan tingkat kesukaran. Hasil analisis menunjukkan bahwa beberapa soal memiliki kualitas yang sangat baik, namun beberapa item soal masih memerlukan revisi agar lebih mampu mengukur kemampuan murid secara maksimal.

Validitas butir soal digunakan untuk mengetahui sejauh mana suatu item soal memberikan kontribusi

terhadap skor total tes. Untuk menentukan validitas setiap butir soal dilakukan dengan menggunakan rumus korelasi product moment pearson (Arikunto, 2018).

Tabel 5. Hasil Uji Validitas

N o	Ite m	rHitun g	rTabe l	Keputusa n
1	Soal 1	0.444	0.35	Valid
2	Soal 2	0.500	0.35	Valid
3	Soal 3	0.506	0.35	Valid
4	Soal 4	-0.329	0.35	Tidak Valid
5	Soal 5	0.504	0.35	Valid
6	Soal 6	0.295	0.35	Tidak Valid
7	Soal 7	0.000	0.35	Tidak Valid
8	Soal 8	0.472	0.35	Valid
9	Soal 9	0.299	0.35	Tidak Valid
10	Soal 10	0.799	0.35	Valid
11	Soal 11	0.227	0.35	Tidak Valid
12	Soal 12	0.202	0.35	Tidak Valid
13	Soal 13	0.103	0.35	Tidak Valid
14	Soal 14	0.347	0.35	Tidak Valid
15	Soal 15	0.624	0.35	Valid
16	Soal 16	0.472	0.35	Valid
17	Soal 17	0.000	0.35	Tidak Valid
18	Soal 18	0.414	0.35	Valid
19	Soal 19	0.390	0.35	Valid
20	Soal 20	0.486	0.35	Valid
21	Soal 21	0.779	0.35	Valid
22	Soal 22	0.431	0.35	Valid

23	Soal 23	0.445	0.35	Valid
24	Soal 24	0.691	0.35	Valid
25	Soal 25	0.759	0.35	Valid
26	Soal 26	0.155	0.35	Tidak Valid
27	Soal 27	0.165	0.35	Tidak Valid
28	Soal 28	0.434	0.35	Valid
29	Soal 29	0.487	0.35	Valid
30	Soal 30	0.331	0.35	Tidak Valid
31	Soal 31	0.674	0.35	Valid
32	Soal 32	0.691	0.35	Valid
33	Soal 33	0.359	0.35	Valid
34	Soal 34	0.318	0.35	Tidak Valid
35	Soal 35	0.773	0.35	Valid
36	Soal 36	0.582	0.35	Valid
37	Soal 37	0.384	0.35	Valid
38	Soal 38	0.670	0.35	Valid
39	Soal 39	0.305	0.35	Tidak Valid
40	Soal 40	0.718	0.35	Valid
41	Soal 41	0.549	0.35	Valid
42	Soal 42	0.180	0.35	Tidak Valid
43	Soal 43	0.598	0.35	Valid
44	Soal 44	0.725	0.35	Valid
45	Soal 45	0.696	0.35	Valid
46	Soal 46	0.670	0.35	Valid
47	Soal 47	0.779	0.35	Valid
48	Soal 48	0.779	0.35	Valid
49	Soal 49	0.705	0.35	Valid
50	Soal 50	1.000	0.35	Valid

Dari table di atas, dapat diketahui bahwa dari 50 butir soal terdapat 35 butir soal yang dapat dipakai dan 15 butir soal perlu direvisi dan pada butir soal nomor 4 perlu dibuang. Dengan demikian, Sebagian besar butir soal Sumatif Akhir Semester Genap mata pelajaran Pendidikan Agama Islam berada dalam kategori dapat dipakai dan direvisi.

Reliabilitas terkait dengan konsistensi, keandalan, keajegan, ataupun stabilitas. Suatu alat ukur dikatakan reliabel bila memberikan hasil yang sama pada berkali-kali pengulangan (Subali, 2010).

Cronbach's Alpha	N of Items	Keterangan
0,912	31	Sangat Reliabel

Berdasarkan tabel di atas, setelah diuji menggunakan bantuan SPSS didapatkan nilai Cronbach's Alpha sebesar 0.912, artinya jika merujuk pada table reliabilitas nilai Cronbach's Alpha besar dari 0.80, maka tingkat reliabilitas memiliki kategori sangat tinggi. Sehingga, melihat hasil ini dapat disimpulkan bahwa butir soal Sumatif Akhir Semester Genap mata pelajaran Pendidikan Agama Islam

kelas IX memiliki kepercayaan yang tinggi.

Di bagian daya pembeda, ditemukan bahwa terdapat 13 butir soal berkategori sangat bai, 12 butir soal berkategori baik, 15 butir soal berkategori cukup, dan 10 butir soal berkategori jelek hingga sangat jelek.

Tabel 7. Daya Pembeda

No	BA	BB	DP	Kategori
1	7	3	0,50	Sangat Baik
2	8	6	0,25	Cukup
3	8	3	0,62	Sangat Baik
4	0	3	- 0,38	Sangat Jelek
5	7	4	0,38	Baik
6	8	5	0,38	Baik
7	8	8	0,00	Jelek
8	8	3	0,62	Sangat Baik
9	8	6	0,25	Cukup
10	8	6	0,25	Cukup
11	8	3	0,62	Sangat Baik
12	6	2	0,50	Sangat Baik
13	3	1	0,25	Cukup
14	7	6	0,12	Jelek
15	8	6	0,25	Cukup
16	8	6	0,25	Cukup
17	8	8	0,00	Jelek
18	6	3	0,38	Baik
19	8	5	0,38	Baik
20	8	5	0,38	Baik
21	8	6	0,25	Cukup
22	8	6	0,25	Cukup
23	8	6	0,25	Cukup
24	8	6	0,25	Cukup
25	8	5	0,38	Baik
26	3	2	0,12	Jelek
27	8	5	0,38	Baik
28	7	3	0,50	Sangat Baik
29	6	3	0,38	Baik
30	5	4	0,12	Jelek
31	8	1	0,88	Sangat Baik
32	8	6	0,25	Cukup
33	8	5	0,38	Baik
34	8	3	0,62	Sangat Baik
35	8	5	0,38	Baik
36	8	6	0,25	Cukup
37	8	5	0,38	Baik
38	8	4	0,50	Sangat Baik
39	8	4	0,50	Sangat Baik

40	8	2	0,75	Sangat Baik
41	8	5	0,38	Baik
42	8	6	0,25	Cukup
43	8	6	0,25	Cukup
44	8	6	0,25	Cukup
45	8	4	0,50	Sangat Baik
46	8	4	0,50	Sangat Baik
47	8	4	0,50	Sangat Baik
48	8	6	0,25	Cukup
49	8	6	0,25	Cukup
50	8	5	0,38	Baik

Berdasarkan data daya pembeda (DP) 50 butir soal di atas, dapat diinterpretasikan bahwa kualitas soal Sumatif Akhir Semester (SAS) secara umum berada pada kategori cukup baik hingga sangat baik, meskipun masih terdapat beberapa butir yang perlu direvisi bahkan dibuang. Dengan rincian 14 butir berkategori Sangat Baik, 13 butir berkategori Baik, 16 butir berkategori Cukup, 6 butir berkategori Jelek, 1 butir berkategori Sangat Jelek.

Secara keseluruhan, mayoritas butir soal telah memenuhi kriteria daya pembeda yang baik karena lebih dari separuh butir berada pada kategori baik dan sangat baik. Hal ini menunjukkan bahwa instrumen Sumatif Akhir Semester mata pelajaran Pendidikan Agama Islam kelas IX sudah cukup efektif dalam mengukur kemampuan murid dan membedakan tingkat penguasaan materi murid. Namun demikian,

beberapa soal dengan kategori jelek dan sangat jelek tetap perlu diperbaiki agar kualitas tes menjadi lebih valid dan reliabel.

Tingkat kesukaran diperoleh dari murid yang menjawab benar dalam setiap butir soal dengan hasil sebagai berikut:

Tabel 8. Tingkat Kesukaran

No Soal	Jumlah Benar	Tingkat Kesukaran	Kategori
1	17	0,55	Sedang
2	28	0,90	Mudah
3	24	0,77	Mudah
4	5	0,16	Sukar
5	18	0,58	Sedang
6	28	0,90	Mudah
7	31	1,00	Mudah
8	25	0,81	Mudah
9	27	0,87	Mudah
10	29	0,94	Mudah
11	19	0,61	Sedang
12	11	0,35	Sedang
13	7	0,23	Sukar
14	27	0,87	Mudah
15	27	0,87	Mudah
16	28	0,90	Mudah
17	31	1,00	Mudah
18	18	0,58	Sedang
19	25	0,81	Mudah
20	28	0,90	Mudah
21	29	0,94	Mudah
22	28	0,90	Mudah
23	28	0,90	Mudah
24	28	0,90	Mudah
25	28	0,90	Mudah
26	12	0,39	Sedang
27	25	0,81	Mudah
28	22	0,71	Mudah
29	22	0,71	Mudah
30	16	0,52	Sedang
31	22	0,71	Mudah
32	28	0,90	Mudah
33	27	0,87	Mudah
34	22	0,71	Mudah
35	28	0,90	Mudah
36	29	0,94	Mudah
37	26	0,84	Mudah
38	26	0,84	Mudah
39	20	0,65	Sedang

40	24	0,77	Mudah
41	26	0,84	Mudah
42	23	0,74	Mudah
43	29	0,94	Mudah
44	29	0,94	Mudah
45	26	0,84	Mudah
46	27	0,87	Mudah
47	26	0,84	Mudah
48	29	0,94	Mudah
49	29	0,94	Mudah
50	28	0,9	Mudah

Tingkat kesukaran digunakan untuk mengetahui sejauh mana suatu soal mudah atau sulit dikerjakan oleh murid. Semakin tinggi nilai indeks kesukaran, maka soal semakin mudah dikerjakan murid. Berdasarkan hasil analisis tingkat kesukaran 50 butir soal Sumatif Akhir Semester di atas, dapat diketahui bahwa sebagian besar soal berada pada kategori mudah, dengan rincian 40 butir berkategori mudah, 8 butir berkategori sedang, 2 butir berkategori sukar.

Secara umum, dominasi soal mudah menunjukkan bahwa sebagian besar murid mampu menjawab soal dengan benar. Hal ini dapat mengindikasikan bahwa materi telah cukup dipahami oleh murid atau tingkat kesulitan soal relatif rendah. Secara keseluruhan, kualitas tingkat kesukaran soal SAS PAI kelas IX cenderung belum proporsional karena didominasi oleh soal mudah. Idealnya, sebuah tes memiliki komposisi soal

yang seimbang antara mudah, sedang, dan sukar agar dapat mengukur kemampuan murid secara lebih akurat. Meskipun demikian, keberadaan beberapa soal sedang menunjukkan bahwa instrumen tes masih memiliki kemampuan yang cukup baik dalam mengevaluasi hasil belajar murid.

B. Pembahasan

Hasil Uji Validitas

Uji validitas dilakukan menggunakan dengan bantuan program SPSS. Pengujian dilakukan terhadap 50 butir soal dengan jumlah responden sebanyak 31 orang pada taraf signifikansi 5%, sehingga diperoleh nilai r tabel sebesar 0,35. Hasil pengujian menunjukkan bahwa sebagian besar butir soal memiliki nilai r_{hitung} lebih besar daripada r_{tabel} sehingga dinyatakan valid. Adapun item yang memiliki nilai r_{hitung} lebih kecil dari r_{tabel} dinyatakan tidak valid dan tidak digunakan dalam penelitian. Berdasarkan uji validitas terhadap butir soal Sumatif Akhir Semester Genap mata pelajaran Pendidikan Agama Islam, terdapat 35 butir soal yang valid dan 15 butir soal tidak valid. Adapun hasil analisis butir soal pilihan ganda Sumatif Akhir Semester Genap kelas IX SMP Negeri 2 Ukui diketahui

35 soal valid, yaitu soal nomor 1, 2, 3, 5, 8, 10, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 28, 29, 31, 32, 33, 35, 36, 37, 38, 40, 41, 43, 44, 45, 46, 47, 48, 49, dan 50. Sedangkan nomor butir soal yang tidak valid adalah 4, 6, 7, 9, 11, 12, 13, 14, 17, 26, 27, 30, 34, 39, dan 42.

Melihat hal ini, Sumatif Akhir Semester mata pelajaran Pendidikan Agama Islam kelas IX berada pada kategori dapat dipakai atau perlu direvisi. Hal ini menunjukkan bahwa 35 butir soal menggambarkan keberhasilan yang dapat menjadi barometer dalam mengukur pemahaman murid terhadap materi Pendidikan Agama Islam. Namun, perlu diakui bahwa terdapat 14 butir soal tidak valid yang perlu direvisi atau bahkan dibuang.

Hal ini seirama dengan penelitian yang dilakukan oleh Fitrialin Elizanoviyanti dan Asep Saepurokhman (2025), dari 30 butir soal terdapat 14 butir soal yang dapat dipakai dengan presentase 47%, dan butir soal yang dibuang terdiri dari satu soal dengan presentase 3%. Sama halnya dengan penelitian yang dilakukan oleh Anshari et al. (2024), hasil penelitiannya menunjukkan bahwa dari 50 butir soal Sumatif Akhir

Semester mata pelajaran PAI terdapat 36 butir soal valid dan 14 butir soal tidak valid. Temuan tersebut menunjukkan bahwa masih terdapat butir soal yang belum mampu mengukur kompetensi murid secara tepat sehingga perlu dilakukan revisi sebelum digunakan Kembali.

Hasil Uji Reliabilitas

Pengujian reliabilitas instrument adalah Langkah penting dalam penelitian untuk memastikan bahwa instrument pengukuran yang digunakan menghasilkan hasil yang konsisten dan dapat diandalkan. Hasil analisis yang dibantu dengan penggunaan aplikasi SPSS menunjukkan bahwa tingkat reliabilitas termasuk dalam kategori sangat tinggi dengan ditandai nilai Cronbach's Alpha sebesar 0,912.

Hasil ini menunjukkan bahwa instrument memiliki konsistensi internal yang sangat baik sehingga tes dapat dipercaya sebagai dasar pengambilan keputusan. Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh Putri et al. (2024) yang menyatakan bahwa instrument dengan koefisien reliabilitas tinggi menunjukkan konsistensi pengukuran yang baik dan

layak digunakan dalam proses evaluasi.

Sejalan dengan hal ini. Baihaqi et al. (2025) dalam penelitiannya juga menegaskan bahwa validitas dan reliabilitas merupakan dua syarat utama yang harus dipenuhi agar instrument sumatif dapat memberikan informasi yang akurat mengenai pencapaian belajar murid. Dapat ditarik kesimpulan, bahwa dalam penelitian ini tingginya nilai reliabilitas menjadi salah satu indikator bahwa instrument Sumatif Akhir Semester Genap mata Pelajaran Pendidikan Agama Islam memenuhi standar tes yang baik.

Daya Pembeda

Daya pembeda adalah kemampuan suatu butir soal dalam membedakan antara murid yang menjawab soal dengan tepat dan murid yang menjawab soal kurang tepat. Daya pembeda juga menjadi pembeda antara murid yang berkemampuan tinggi dengan murid yang berkemampuan rendah. Dengan adanya analisis daya pembeda pada butir soal, maka hasil yang didapatkan dapat mengkaji suatu butir soal dari segi kesanggupan tes tersebut dalam membedakan murid yang memiliki

kategori rendah dan kategori tinggi (Saputri, 2023).

Hasil penelitian mengkonfirmasi bahwa terdapat 14 butir soal berkategori sangat baik, 13 butir soal berkategori baik, 16 butir soal berkategori cukup, 6 butir soal berkategori jelek, dan 1 butir soal berkategori sangat jelek. Dengan kata lain. Sebagian besar butir soal telah mampu membedakan kemampuan murid dengan baik.

Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh Saputri et al. (2023) yang menemukan bahwa kualitas instrument yang baik ditunjukkan oleh dominasi butir soal dengan daya pembeda kategori cukup hingga baik. Dalam penelitian ini, butir soal yang memiliki daya pembeda sangat baik dan baik, sebaiknya dipertahankan, sedangkan butir soal dengan daya pembeda cukup dapat direvisi untuk meningkatkan kualitasnya. Butir soal yang memiliki daya pembeda jelek dan sangat jelek, sebaiknya diperbaiki secara menyeluruh atau diganti karena dianggap kurang efektif dalam membedakan kemampuan murid berdasarkan tingkat kemampuannya.

Tingkat Kesukaran

Tingkat kesukaran menunjukkan proporsi murid yang mampu menjawab butir soal dengan benar. Pada penelitian ini menunjukkan bahwa terdapat 40 butir soal dengan kategori mudah, 8 butir soal dengan kategori sedang, dan 2 butir soal dengan kategori sukar. Melihat hal ini, menunjukkan bahwa Sumatif Akhir Semester mata pelajaran Pendidikan Agama Islam didominasi oleh soal-soal berkategori mudah. Kondisi seperti ini dapat meningkatkan presentase keberhasilan murid dalam menjawab setiap butir soal, namun di sisi lain dapat mengurangi kemampuan tes dalam mengukur variasi kemampuan murid secara mendalam.

Dalam penelitian yang dilakukan Rahayu dan Djazari (2016), komposisi tingkat kesukaran yang proporsional, yaitu 25% soal mudah, 60% soal sedang, dan 15% soal sukar. Komposisi tersebut dianggap lebih ideal karena mampu mengukur kemampuan murid pada berbagai tingkat penguasaan materi.

Melihat hasil analisis dalam Sumatif Akhir Semester Genap ini, guru perlu meningkatkan jumlah soal kategori sedang dan sukar pada penyusunan

Sumatif Akhir Semester berikutnya. Komposisi soal yang lebih seimbang akan meningkatkan daya ukur instrument dan memberikan informasi yang lebih akurat mengenai pencapaian kompetensi murid.

D. Kesimpulan

Berdasarkan hasil analisis terhadap 50 butir soal Sumatif Akhir Semester Genap mata pelajaran Pendidikan Agama Islam kelas IX SMP Negeri 2 Ukui, dapat disimpulkan bahwa instrumen tes secara umum memiliki kualitas yang cukup baik dan layak digunakan sebagai alat evaluasi pembelajaran. Hal ini ditunjukkan oleh mayoritas butir soal yang valid, nilai reliabilitas yang sangat tinggi (Cronbach's Alpha = 0,912), serta daya pembeda yang sebagian besar berada pada kategori cukup hingga sangat baik. Namun demikian, masih terdapat beberapa butir soal yang tidak valid dan memiliki daya pembeda rendah sehingga perlu direvisi atau diganti. Selain itu, distribusi tingkat kesukaran soal belum proporsional karena didominasi oleh soal kategori mudah, sehingga perlu perbaikan agar kemampuan peserta didik dapat diukur secara lebih optimal.

Penelitian ini memiliki keterbatasan karena hanya melibatkan 31 murid dan menganalisis satu paket soal pada satu mata pelajaran, sehingga hasil penelitian belum dapat digeneralisasikan secara luas. Oleh karena itu, penelitian selanjutnya disarankan melibatkan jumlah responden yang lebih besar serta beberapa paket soal untuk memperoleh gambaran kualitas instrumen yang lebih komprehensif. Bagi guru, hasil penelitian ini dapat dijadikan bahan evaluasi dalam penyusunan soal dengan cara merevisi butir yang tidak valid, meningkatkan kemampuan soal dalam membedakan kemampuan peserta didik, serta menyusun komposisi tingkat kesukaran yang lebih seimbang antara soal mudah, sedang, dan sukar agar instrumen evaluasi menjadi lebih efektif dan akurat.

DAFTAR PUSTAKA

- Fariq, A. (2011). Perkembangan dunia konseling memasuki era globalisasi. *Pedagogi*, II Nov 2011(Universitas Negeri Padang), 255-262. Afrilianti, Dian, dkk. 2025. *Karakteristik Tes yang Baik. Journal of Innovative and Creativity*, 5 (2).

- <https://doi.org/10.31004/joecy.v5i2.274>
- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis validitas dan reliabilitas butir soal sumatif akhir semester ganjil mata pelajaran PAI. *Edukatif: Jurnal Ilmu Pendidikan*, 6(1), 964–975.
<https://doi.org/10.31004/edukatif.v6i1.5931>
- Arifin, Z. 2016. Evaluasi Pembelajaran. Bandung: PT Remaja Rosdakarya
- Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan* (3rd ed.). Bumi Aksara.
- Baihaqi, M., Mahmudah, M., Al Mubarak, A. J., Rabbaniyah, N., & Susanto, D. (2025). Studi validasi dan reliabilitas butir soal tes sumatif dalam pendidikan Bahasa Arab. *Cerdika: Jurnal Ilmiah Indonesia*, 5(4).
<https://doi.org/10.59141/cerdika.v5i4.2553>
- Bambang, Subali. 2010. Penilaian, Evaluasi, dan Remediasi Pembelajaran Biologi. Yogyakarta: Jurusan Biologi Fakultas MIPA
- Elizanoviyanti, F., & Saepurokhman, A. (2025). Analisis tingkat validitas dan reliabilitas butir soal penilaian akhir semester mata pelajaran Bahasa Indonesia kelas XI SMA Negeri Situraja. *Literat: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 4(2), 168–176.
- Kemendikbudristek. (2022). Panduan Pembelajaran dan Asesmen Kurikulum Merdeka. Jakarta: Kemendikbudristek
- Purba, Yeni Marchna Sari. 2021. Implementasi Program Diklat Berjenjang Tingkat Dasar dalam Jaringan Untuk Meningkatkan Kompetensi Pendidik PAUD. Yogyakarta: Universitas Pendidikan Indonesia
- Rahayu, R., Djazari, M., 2016. Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, Vol. XIV, No. 1., 85-94.
<https://doi.org/10.21831/jpai.v14i.11370>
- Saputra, H. D., dkk. 2022. Hasil Belajar Mahamurid: Analisis Butir Soal Tes, Edukasi: Jurnal Pendidikan, Volume 20 Nomor 1
- Saputri, H. A., Larasati, N. J., Zuhijrah, & Shaleh. (2023). Analisis instrumen asesmen: Validitas, reliabilitas, tingkat kesukaran dan daya beda butir soal. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(5), 2986–2995.
<https://doi.org/10.36989/didaktik.v9i5.2268>
- Sihombing, P. R., dkk. (2024) Analisis Butir Soal dalam Berbagai Software, Tangerang: Minhaj Pustaka