

Analisis Butir Soal Asesmen Sumatif Akhir Semester Fiqih di Madrasah Tsanawiyah

Febrilia Khoiri Dewi Azizah^{1✉}, Joko Subando²

(1) Pendidikan Agama Islam, Institut Islam Mamba'ul 'Ulum Surakarta, Indonesia

(2) Institut Islam Mamba'ul 'Ulum, Indonesia

✉ Corresponding author
[febriliakdewi@gmail.com]

Abstrak

Penelitian ini bertujuan untuk mengetahui seberapa sukar butir soal, mengetahui daya beda soal dan mengetahui keberfungsian distraktor butir soal fiqih pada siswa Madrasah Tsanawiyah Muhammadiyah Blimbing, Wonorejo, Polokarto, Sukoharjo. Penelitian ini adalah penelitian kuantitatif. Sampel penelitian ini adalah 26 siswa kelas VIII. Teknik pengumpulan data menggunakan dokumentasi. Teknik analisis data menggunakan teori tes klasik dengan bantuan Anates V4. Hasil penelitian menunjukkan: tingkat kesukaran soal asesmen masih belum ideal karena lebih didominasi kategori soal mudah, dari daya pembedanya, soal asesmen ini tidak baik dan perlu untuk direvisi/tidak digunakan pada tes selanjutnya, dan efektivitas pengecoh soal ini lebih didominasi dengan pengecoh yang tidak berfungsi. Soal ini tergolong soal yang mudah dan tidak menantang bagi peserta didik dan perlu dilakukan perbaikan pada kualitas soal, khususnya dalam hal daya pembeda dan efektivitas pengecoh untuk meningkatkan kualitas penilaian. Analisis butir soal membantu meningkatkan efektivitas dan keadilan evaluasi pendidikan serta mendukung pengembangan profesional guru.

Kata Kunci: *analisis soal, asesmen sumatif*

Abstract

This study aims to assess the difficulty level of test items, determine the discriminating power of the questions, and evaluate the effectiveness of distractors in Fiqh test items for students at Madrasah Tsanawiyah Muhammadiyah Blimbing, Wonorejo, Polokarto, Sukoharjo. This is a quantitative study with a sample of 26 eighth-grade students. Data collection was performed using documentation, and data analysis was conducted using classical test theory with the assistance of Anates V4 software. The findings indicate that (1) the difficulty level of the assessment items is still not ideal, as they are predominantly accessible, (2) the discriminating power of the assessment items is poor and needs revision or should not be used in future tests, and (3) non-functional distractors primarily dominate the effectiveness of the distractors. These questions are considered uncomplicated and not challenging for students, and there is a need for improvements in question quality, particularly in discriminating power and distractor effectiveness, to enhance the quality of assessment. Item analysis helps improve the effectiveness and fairness of educational evaluations and supports the professional development of teachers.

Keywords: *question analysis, summative assessment*

PENDAHULUAN

Pada akhir setiap semester, Madrasah Tsanawiyah (MTs) melaksanakan asesmen sumatif sebagai bagian dari evaluasi hasil belajar siswa tak terkecuali di Madrasah Tsanawiyah Muhammadiyah Blimbing. Dengan adanya evaluasi, guru dapat terus meningkatkan kualitas dari pembelajaran yang dilakukan agar kemampuan peserta didik terus meningkat. Peningkatan

kemampuan peserta didik akan mempengaruhi kualitas dari suatu pendidikan. Kualitas pendidikan itu sebanding dengan kemampuan peserta didik. Jika kemampuan peserta didik tinggi, maka kualitas pendidikan pun akan baik. Sedangkan apabila kemampuan peserta didik rendah, maka kualitas pendidikan pun akan buruk (Fiska et al., 2021). Asesmen sumatif ini bertujuan untuk mengukur sejauh mana siswa telah memahami materi yang diajarkan selama satu semester penuh.

Dalam konteks mata pelajaran Fikih, yang merupakan salah satu mata pelajaran inti di MTs, asesmen sumatif akhir semester gasal memiliki peran penting dalam mengevaluasi kemampuan siswa dalam memahami dan menerapkan hukum-hukum syariah Islam dalam kehidupan sehari-hari. Untuk memastikan bahwa asesmen tersebut benar-benar efektif dan objektif dalam mengukur kompetensi siswa, perlu dilakukan analisis butir soal terhadap soal-soal yang digunakan dalam ujian tersebut.

Analisis butir soal secara klasik *Classical Test Theory* (CTT) adalah proses penelaahan butir soal melalui informasi dari jawaban peserta didik guna meningkatkan mutu butir soal yang bersangkutan dengan menggunakan teori tes klasik. Sedangkan Analisis butir soal secara modern yaitu penelaahan butir soal dengan menggunakan *Item Response Theory* (IRT) atau teori jawaban butir soal. Teori ini merupakan suatu teori yang menggunakan fungsi matematika untuk menghubungkan antara peluang menjawab benar suatu soal dengan kemampuan siswa. Nama lain IRT adalah *latent trait theory* (LTT) atau *characteristics curve theory* (ICC) (Ahmad Asyasyahru, 2020). Aspek yang perlu diperhatikan dalam analisis butir soal secara klasik adalah setiap butir soal ditelaah dari segi: tingkat kesukaran butir, daya pembeda butir, dan penyebaran pilihan jawaban (untuk soal bentuk obyektif) atau frekuensi jawaban pada setiap pilihan jawaban.

Berdasarkan penelusuran literatur, sudah banyak penelitian mengenai analisis butir soal dengan teori klasik. Namun demikian, di Madrasah Tsanawiyah Muhammadiyah Blimbing belum pernah dilakukan analisis butir soal Fikih. Sehingga perlu dilakukan penelitian mengenai analisis butir soal asesmen sumatif akhir semester gasal mata pelajaran Fikih di Madrasah Tsanawiyah Muhammadiyah Blimbing agar bisa diketahui tingkat kesukaran soal, daya pembeda dan keberfungsian distraktor soal mata pelajaran Fikih.

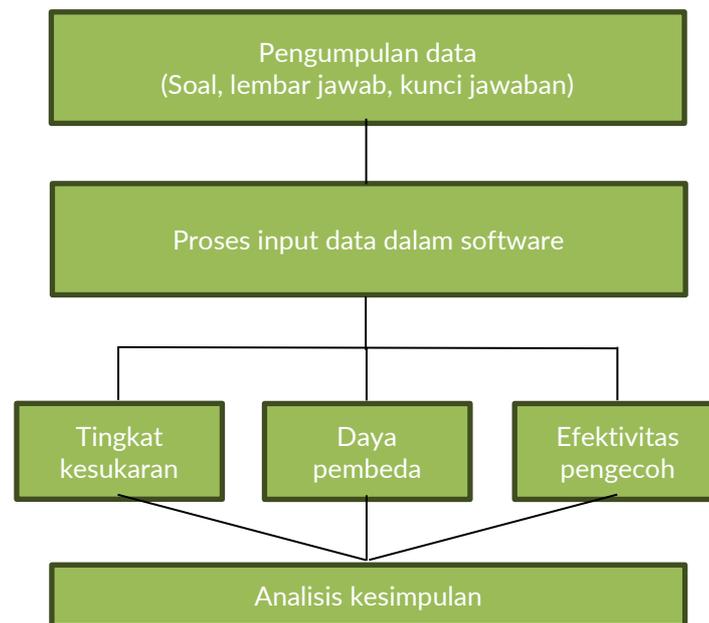
Atas dasar pemaparan diatas, penelitian ini bertujuan untuk (1) mengetahui tingkat kesukaran butir soal fikih kelas VIII pada asesmen sumatif akhir semester di Madrasah Tsanawiyah, (2) mengetahui daya beda soal fikih kelas VIII pada asesmen sumatif akhir semester di Madrasah Tsanawiyah, (3) mengetahui keberfungsian distraktor soal fikih kelas VIII pada asesmen sumatif akhir semester di Madrasah Tsanawiyah. Penelitian ini penting untuk sebagai bahan evaluasi guru dalam menyusun soal. Penelitian ini penting untuk sebagai bahan evaluasi guru dalam menyusun soal dan dijadikan informasi untuk mengetahui kualitas dari soal yang dibuat sehingga guru dapat terus meningkatkan soal sesuai dengan penyebab permasalahannya.

METODE PENELITIAN

Penelitian ini merupakan jenis penelitian kuantitatif. Penelitian ini menggunakan statistik deskriptif, yaitu mendeskripsikan atau menggambarkan data sesuai dengan kenyataan yang sebenarnya, analisis ini tidak untuk menolak/menerima hipotesis atau mencari hubungan antar variabel, namun hanya sekedar mendeskripsikan (Subando, 2021). Penelitian berusaha melaporkan keadaan objek yang diteliti sesuai dengan apa adanya, yaitu untuk mengetahui kualitas soal. Dengan melakukan pengumpulan data berupa butir soal dan lembar jawaban siswa. Setelah mendapatkan data, peneliti melakukan analisis soal dan kemudian mendeskripsikan hasil penelitian secara kuantitatif.

Penelitian ini dilaksanakan di Madrasah Tsanawiyah Muhammadiyah Blimbing, Wonorejo, Polokarto, Sukoharjo. Populasi pada penelitian ini adalah seluruh siswa kelas VIII Madrasah Tsanawiyah Muhammadiyah Blimbing unit non pondok tahun ajaran 2023/2024 sebanyak 104 siswa. Arikunto (2006) menyatakan bahwa jika jumlah subjek kurang dari 100, lebih baik mengambil seluruh subjek tersebut sehingga penelitian tersebut menjadi penelitian populasi. Namun, jika jumlah subjek melebihi 100, dapat diambil sampel sebesar 10-15% atau 20-25% dari populasi untuk memenuhi teknik pengambilan sampel (Juafsan Liadia Cici, Tuti Sarwita, dan Irfandi, 2022). Maka sampel dari penelitian ini adalah 26 siswa. Teknik pengumpulan data yang digunakan pada penelitian ini adalah dokumentasi. Teknik dokumentasi ini dilakukan untuk memperoleh dokumen

terkait pelaksanaan asesmen sumatif akhir semester gasal mata pelajaran Fiqih kelas VIII di Madrasah Tsanawiyah Muhammadiyah Blimbing tahun ajaran 2023/2024 meliputi kisi-kisi soal, naskah soal, kunci jawaban soal, dan lembar jawaban siswa.



Gambar 1. Bagan Langkah Penelitian

Data yang sudah diperoleh akan dianalisis untuk mengetahui tingkat kesukaran, daya pembeda dan efektivitas pengecoh soal menggunakan software Anates V4. Kategori kriteria identifikasi 3752istrac kesukaran butir soal, daya pembeda butir soal, dan keberfungsian 3752istractor adalah sebagai berikut:

Tabel 1. Kriteria Tingkat Kesukaran Butir Soal

Nilai P	Interpretasi
$P = 0,00 < P \leq 0,10$	Sangat Sukar
$0,11 < P \leq 0,30$	Sukar
$0,31 < P \leq 0,70$	Sedang
$0,71 < P < 0,90$	Mudah
$P = 1,00$	Sangat Mudah

Tabel 2. Kriteria Daya Pembeda Butir Soal

Indeks Daya Pembeda	Interpretasi
$DP \leq 0,00$	Sangat Jelek
$0,00 < DP \leq 0,20$	Jelek
$0,20 < DP \leq 0,40$	Cukup
$0,40 < DP \leq 0,70$	Baik
$0,70 < DP \leq 1,00$	Sangat Baik

Tabel 3. Kriteria Indeks Pengecoh Butir Soal

Nilai IP	Interpretasi
$>200\%$	Sangat Jelek
$0\% - 25\%$ atau $176\% - 200\%$	Jelek
$26\% - 50\%$ atau $151\% - 175\%$	Kurang Baik
$51\% - 75\%$ atau $126\% - 150\%$	Baik
$76\% - 125\%$	Sangat Baik

HASIL DAN PEMBAHASAN

Arikunto (2016) menjelaskan bahwa tes merupakan salah satu metode untuk mengevaluasi hasil belajar melalui kegiatan pengukuran yang menuntut peserta didik menjawab berbagai pertanyaan dan pernyataan untuk mengukur aspek-aspek perilaku peserta (Rishan & Sulaiman, 2023). Tes menurut Ainin (2016) adalah prosedur atau alat yang digunakan untuk pengukuran dan penilaian, testing berarti saat dilaksanakannya atau peristiwa berlangsungnya pengukuran dan penilaian (Ardiansyah, 2022). Tes adalah serangkaian pertanyaan dengan jawaban benar atau salah. Tes tertulis adalah tes yang menuntut peserta tes memberi jawaban secara tertulis berupa pilihan ganda dan uraian, tes pilihan ganda terdiri atas sebuah pernyataan atau kalimat yang belum lengkap, kemudian diikuti oleh sejumlah pernyataan atau bentuk yang dapat digunakan untuk melengkapinya. Dari sejumlah "pelengkap" tersebut, hanya satu yang tepat, yang lain merupakan pengecoh (distractors). Soal pilihan ganda dapat mengukur pengetahuan mulai dari tingkat pengetahuan, pemahaman dan aplikasi (Sunarti & Rahmawati, 2014). Tes yang digunakan dalam asesmen sumatif akhir semester gasal kelas VIII mata pelajaran Fikih di Madrasah Tsanawiyah Muhammadiyah Blimbing adalah tes tertulis yang terdiri dari soal pilihan ganda sebanyak 40 soal dan 5 soal uraian. Penelitian ini hanya menganalisis soal yang berbentuk pilihan ganda menggunakan program *Anates Versi 4*.

Soal yang baik adalah soal yang memiliki tingkat kesulitan yang seimbang, tidak terlalu mudah dan tidak terlalu sulit. Soal yang mudah tidak menantang peserta tes dalam memecahkan suatu masalah. Sebaliknya, soal yang terlalu sulit dapat membuat peserta tes merasa putus ada karena jarak antara soal tersebut dan pemahaman mereka terlalu besar. Soal yang baik memiliki proporsi yang seimbang antara tingkat kesukaran rendah, sedang, dan tinggi. Biasanya dengan perbandingan 3:5:2 atau 3:4:3. Oleh karena itu, dalam membuat soal penting untuk menjaga keseimbangan antara soal yang mudah, sedang, dan sulit (Fiska et al., 2021). Perhitungan tingkat kesukaran soal Fikih menggunakan program *Anates Versi 4*.

Tingkat kesukaran soal menurut Arifin (2003) adalah pengukuran seberapa besar derajat kesukaran soal. Suatu soal dikatakan baik, apabila memiliki tingkat kesukaran soal yang seimbang (proporsional) dalam artian soal tersebut tidak terlalu mudah atau terlalu sukar (Rahmaini & Taufiq, 2018). Misalnya, untuk keperluan ujian semester digunakan butir soal dengan tingkat kesukaran sedang, untuk seleksi dengan butir soal tingkat kesukaran tinggi, dan untuk keperluan diagnosis digunakan butir soal dengan tingkat kesukaran mudah.

Tabel 4. Analisis Tingkat Kesukaran Soal Program Anates V4

No.	Jumlah Betul	Tingkat Kesukaran (%)	Tafsiran
1	25	96,15	Sangat Mudah
2	25	96,15	Sangat Mudah
3	28	69,23	Sedang
4	22	84,62	Mudah
5	21	80,77	Mudah
6	18	69,23	Sedang
7	20	76,92	Mudah
8	24	92,31	Sangat Mudah
9	23	88,46	Sangat Mudah
10	16	61,54	Sedang
11	21	80,77	Mudah
12	6	23,08	Sukar
13	26	100,00	Sangat Mudah
14	26	100,00	Sangat Mudah
15	23	88,46	Sangat Mudah
16	19	73,08	Mudah
17	23	88,46	Sangat Mudah
18	24	92,31	Sangat Mudah
19	23	88,46	Sangat Mudah
20	24	92,31	Sangat Mudah
21	25	96,15	Sangat Mudah

22	23	88,46	Sangat Mudah
23	9	34,62	Sedang
24	12	46,15	Sedang
25	25	96,15	Sangat Mudah
26	18	69,23	Sedang
27	23	88,46	Sangat Mudah
28	14	53,85	Sedang
29	23	88,46	Sangat Mudah
30	23	88,46	Sangat Mudah
31	24	92,31	Sangat Mudah
32	16	61,54	Sedang
33	17	65,38	Sedang
34	18	69,23	Sedang
35	10	38,46	Sedang
36	14	53,85	Sedang
37	11	42,31	Sedang
38	21	80,77	Mudah
39	24	92,31	Sangat Mudah
40	23	88,46	Sangat Mudah

Berdasarkan hasil perhitungan diatas, tingkat kesukaran butir soal asesmen sumatif akhir semester gasal mata pelajaran Fikih adalah sebagai berikut:

Tabel 5 Distribusi Tingkat Kesukaran Soal

Kategori	Nomor Soal	Jumlah	Presentase
Sukar (0,11-0,30)	12	1	2,5%
Sedang (0,31-0,70)	3, 6, 10, 23, 24, 26, 28, 32, 33, 34, 35, 36, 37	13	32,5%
Mudah (0,71-0,90)	4, 5, 7, 11, 16, 38	6	15%
Sangat Mudah (0,91-1,00)	1, 2, 8, 9, 13, 14, 15, 17, 18, 19, 20, 21, 22, 25, 27, 29, 30, 31, 39, 40	20	50%

Berdasarkan tabel hasil analisis 40 soal pilihan ganda asesmen sumatif akhir semester gasal kelas VIII mata pelajaran Fikih di MTs Muhammadiyah Blimbing termasuk soal yang tergolong kurang baik. Hal ini disebabkan karena 32,5% soal yang baik, sementara 67,5% sisanya tidak memenuhi kriteria, yaitu 50% sangat mudah, 15% mudah, dan 2,5% sukar. Kriteria kesimpulan kualitas butir soal yang baik berdasarkan tingkat kesukaran adalah soal dengan indeks kesukaran antara 0,31 hingga 0,70, atau tingkat kesukaran sedang. Soal yang baik adalah yang tidak terlalu mudah maupun terlalu sukar, terdapat 1 butir soal yang berkategori sedang yaitu pada soal nomor 12. Indeks soal ini terdapat pada rentang 0,11 - 0,30 oleh karena itu, soal-soal ini sebaiknya tidak digunakan kembali pada tes-tes berikutnya. Soal yang terlalu sulit dapat mengurangi semangat dan motivasi siswa atau peserta tes dalam menyelesaikan soal tersebut karena berada di luar kemampuan mereka. Sementara itu, soal yang terlalu mudah tidak akan merangsang siswa untuk berpikir. Soal yang baik adalah yang sesuai dengan kemampuan siswa, tidak terlalu sulit ataupun terlalu mudah, dan termasuk dalam kategori sedang (Halik et al., 2019). Oleh karena itu, butir soal nomor 3, 6, 10, 23, 24, 26, 28, 32, 33, 34, 35, 36, dan 37 yang masuk dalam kategori sedang dapat digunakan kembali untuk ujian mendatang. Sementara itu, untuk butir soal yang berkategori mudah, sebaiknya diteliti lebih lanjut untuk mengetahui alasan soal tersebut dapat dengan mudah dikerjakan oleh peserta tes, serta mengapa peserta dapat dengan mudah menemukan kunci jawaban dan membedakannya dari pengecoh. Dari uraian di atas, dapat disimpulkan bahwa kualitas butir soal berdasarkan tingkat kesukaran termasuk dalam kategori kurang baik.

Analisis daya beda bertujuan untuk menilai kemampuan butir-butir soal dalam membedakan antara siswa yang menguasai materi dengan baik, dan siswa yang tidak menguasai materi dengan baik. Butir soal yang baik seharusnya dapat membedakan peserta tes yang kompeten, kurang

kompeten, yaitu dengan cara soal tersebut dijawab dengan benar oleh peserta yang mampu, dan dijawab salah oleh peserta yang kurang mampu (Rusmayani, 2020). Analisis daya pembeda dimaksudkan untuk mengkaji kemampuan soal untuk membedakan antara siswa yang mempunyai prestasi tinggi dan prestasi rendah (Uno & Koni, 2012). Besarnya daya pembeda ditunjukkan dengan indeks diskriminasi atau daya pembeda dengan menggunakan simbol D. Kisaran indeks daya pembeda sama dengan indeks kesukaran soal yaitu 0,00 sampai 1,00. Semakin tinggi indeks pembeda soal, maka soal tersebut mampu membedakan antara siswa yang pintar dengan siswa yang kurang pintar. Tanda negatif yang dijumpai pada perhitungan indeks diskriminasi soal menunjukkan bahwa soal menggambarkan kualitas peserta tes secara terbalik, dimana siswa pandai disebut bodoh dan siswa bodoh disebut pandai (Arikunto, 2016:226). Berdasarkan hasil analisis butir soal berdasarkan daya pembedanya pada mata pelajaran Fikih menggunakan program Anates V4, maka diperoleh hasil perhitungan sebagai berikut :

Tabel 6. Analisis Daya Pembeda Soal Program Anates Versi 4

No.	Kelompok Atas	Kelompok Bawah	Daya Beda
1	7	6	14,29
2	7	6	14,29
3	7	2	71,43
4	7	5	28,57
5	7	5	28,57
6	7	3	57,14
7	7	4	42,86
8	7	7	0,00
9	6	7	-14,29
10	7	2	71,43
11	7	4	42,86
12	5	0	71,43
13	7	7	0,00
14	7	7	0,00
15	7	5	28,57
16	6	5	14,29
17	7	6	14,29
18	7	6	14,29
19	7	5	28,57
20	7	7	0,00
21	7	6	14,29
22	7	4	42,86
23	3	2	14,29
24	5	1	57,14
25	7	6	14,29
26	7	3	57,14
27	7	5	28,57
28	7	0	100,00
29	7	4	42,86
30	7	5	28,57
31	7	6	14,29
32	7	0	100,00
33	5	4	14,29
34	7	4	42,86
35	6	1	71,43
36	5	2	42,86
37	5	1	57,14
38	7	3	57,14
39	7	6	14,29
40	6	6	0,00

Berdasarkan perhitungan diatas, daya pembeda butir soal asesmen sumatif akhir semester gasal mata pelajaran fikih adalah sebagai berikut:

Tabel 7. Distribusi Daya Pembeda Soal

Kategori	Nomor Soal	Jumlah	Presentase
Sangat Jelek (<0,00)	9	1	2,5%
Jelek (0,00-0,20)	1, 2, 8, 13, 14, 16, 17, 18, 20, 21, 23, 25, 31, 33, 39, 40	16	40%
Cukup (0,21-0,40)	4, 5, 15, 19, 27, 30	6	15%
Baik (0,41-0,70)	6, 7, 11, 12, 24, 26, 29, 34, 36, 37, 38	11	27,5%
Baik Sekali (0,71-1,00)	3, 10, 12, 28, 32, 35	6	15

Berdasarkan hasil analisis diatas, diperoleh bahwa dari 40 soal yang diujikan terdapat 1 soal atau sebesar 2,5% kategori sangat jelek karena hasilnya negative, 16 soal atau sebesar 40% kategori jelek, 6 soal atau sebesar 15% dalam kategori cukup, 11 soal atau sebesar 27,5% memiliki kategori baik, dan 6 soal atau sebesar 15% dengan kategori baik sekali. Berdasarkan kesimpulan kualitas soal dilihat dari daya pembeda, terdapat beberapa soal yang masuk dalam kategori baik atau sangat baik. Ada 11 butir soal yang dikategorikan baik dengan indeks daya pembeda berkisar 0,41-0,70, yaitu nomor 6, 7, 11, 12, 24, 26, 29, 34, 36, 37, dan 38, serta 6 butir soal yang sangat baik dengan indeks daya pembeda berkisar antara 0,71-1,00, yaitu nomor 3, 10, 12, 28, 32, dan 35. Soal-soal ini dapat digunakan kembali dan dimasukkan ke dalam bank soal.

Butir soal dengan kategori cukup berdasarkan daya pembeda adalah soal yang masih mampu membedakan siswa yang paham dengan materi dan yang tidak. Namun, soal dalam kategori cukup dapat menyebabkan siswa yang belum mengerti materi bisa secara kebetulan menebak jawaban yang benar. Soal-soal dalam kategori cukup dapat digunakan kembali jika telah direvisi atau diperbaiki dan kemudian disimpan dalam bank soal (Halik et al., 2019). Butir soal yang kurang baik berjumlah 6 soal, yakni pada nomor 4, 5, 15, 19, 27, dan 30.

Soal yang tidak baik (jelek) berdasarkan daya pembeda adalah soal dengan daya pembeda jelek atau sangat jelek. Soal yang jelek adalah soal yang tidak dapat membedakan antara siswa yang menguasai materi dan yang tidak. Soal yang jelek bisa disebabkan oleh banyaknya siswa yang tidak menguasai materi tetapi bisa menjawab soal dengan benar. Terdapat 17 butir soal yang tidak baik (16 butir kategori jelek dan 1 butir kategori sangat jelek), yaitu nomor 1, 2, 8, 13, 14, 16, 17, 18, 20, 21, 23, 25, 31, 33, 39, 40 dan 9. Soal-soal dalam kategori jelek ini sebaiknya dibuang dan diganti dengan soal yang baru.

Arifin (2014) menjelaskan pengecoh adalah alternatif jawaban (opsi) yang ada pada bentuk soal pilihan ganda. Butir soal yang baik, pengecohnya akan dipilih secara merata oleh peserta didik yang menjawab salah. Sebaliknya, butir soal yang kurang baik, pengecohnya akan dipilih secara tidak merata dalam (Al-Fathiyah, 2019). Sukiman (2012) analisis yang berhubungan dengan distraktor dilakukan khusus untuk bentuk tes pilihan ganda (multiple choice item) yang mempunyai opsi jawaban antara 3 sampai dengan 5 buah. Di dalam soal pilihan ganda terdapat satu jawaban yang benar dan beberapa jawaban salah atau distraktor. Keberfungsian distraktor atau efektivitas pengecoh bertujuan untuk menentukan apakah pilihan jawaban yang dibuat dapat mengecoh peserta tes yang tidak memahami materi pelajaran.

Peserta pada penelitian ini berjumlah 26 peserta didik. 5% dari 26 adalah 1,3 dibulatkan menjadi 1, Dengan demikian, efektivitas pengecoh berfungsi apabila jawaban A, B, C, dan D dipilih minimal oleh 1 peserta didik. Analisis efektivitas pengecoh soal Fikih dilakukan dengan bantuan program Anates Versi 4 sebagai berikut:

Tabel 8. Analisis Keberfungsian Distraktor Soal

Nomor Soal	Kualitas Pengecoh			
	A	B	C	D
1	0--	25**	0--	1---
2	25**	0--	1---	0--
3	18**	7---	1-	0--
4	1+	1+	2+	22**
5	1+	4---	21**	0--
6	2+	3++	18**	3++
7	0--	1-	5---	20**
8	0--	24**	2---	0--
9	1++	1++	23**	1++
10	10---	0--	0--	16**
11	0--	21**	2++	3--
12	4++	9+	6**	7++
13	0	0	26**	0
14	0	26**	0	0
15	1++	23**	2--	0--
16	1-	0--	19**	6---
17	23**	3---	0--	0--
18	1+	0--	1+	24**
19	3--	23**	0--	1++
20	0--	2---	0--	24**
21	1---	0--	25**	0--
22	0--	2--	1++	23**
23	12---	9**	5++	0--
24	3+	12**	7+	4++
25	0--	25**	0--	1---
26	18**	2+	2+	4+
27	0--	23**	2--	1++
28	4++	14**	3+	4++
29	1++	1++	23**	1++
30	0--	1++	2--	23**
31	1+	24**	1+	0--
32	16**	3++	6--	1-
33	17**	3++	4+	2+
34	1-	18**	5--	2+
35	10**	6++	7+	3+
36	1--	4++	7-	14**
37	10--	11**	4++	1--
38	0--	21**	4---	1+
39	24**	1+	1+	0--
40	0--	23**	1++	2--

Jumlah pengecoh yang berfungsi kemudian didistribusikan kedalam skala Likert yaitu: apabila keempat jawaban pengecoh berfungsi maka soal mempunyai efektivitas pengecoh yang sangat baik, apabila ada tiga jawaban pengecoh yang berfungsi maka soal mempunyai efektivitas pengecoh yang baik, jika terdapat dua jawaban pengecoh yang berfungsi maka soal mempunyai efektivitas pengecoh yang cukup baik sedangkan jika hanya satu jawaban pengecoh yang berfungsi maka efektivitas pengecoh soal tersebut kurang baik. Hasil analisis soal ujian Fikih jika disesuaikan dengan skala Likert yaitu 2 soal (5%) masuk dalam kategori tidak baik, 8 soal (20%) masuk dalam

kategori kurang baik, 16 soal (40%) masuk dalam kategori cukup baik, dan 14 soal (35%) masuk kategori soal memiliki efektivitas pengecoh yang baik. Apabila didistribusikan dalam bentuk tabel, hasilnya sebagai berikut:

Tabel 9. Distribusi Efektivitas Pengecoh Soal

Kategori	Nomor Soal	Jumlah	Presentase
Baik	4, 6, 9, 12, 24, 26, 28, 29, 32, 33, 34, 35, 36, 37	14	35%
Kurang Baik	3, 5, 7, 11, 15, 16, 18, 19, 22, 23, 27, 30,31, 38, 39, 40	16	40%
Jelek	1, 2, 8, 10, 17, 20, 21, 25	8	20%
Sangat Jelek	13, 14	2	5%

Berdasarkan hasil analisis jika dilihat dari efektivitas pengecohnya, menunjukkan bahwa dari 40 butir soal terdapat sebanyak 14 butir soal (35%) baik, 16 butir soal (40%) kurang baik, 8 butir soal (20%) jelek, dan 2 butir soal (5%) sangat jelek. Kualitas soal yang baik berdasarkan efektivitas pengecoh adalah soal yang termasuk dalam kategori baik atau sangat baik, di mana maksimal hanya ada satu opsi yang tidak berfungsi. Jika semua pengecoh pada suatu soal dipilih oleh setidaknya 5% peserta tes, atau terdapat satu pengecoh yang tidak dipilih oleh minimal 5% peserta, maka soal tersebut dianggap baik. Terdapat 14 butir soal yang baik, yaitu nomor 24, 6, 9, 12, 24, 26, 28, 29, 32, 33, 34, 35, 36, dan 37. Hal ini menunjukkan bahwa opsi pengecoh pada butir-butir soal tersebut memiliki kemungkinan besar untuk dipilih oleh peserta tes. Oleh karena itu, butir-butir soal tersebut dapat dimasukkan ke dalam bank soal untuk digunakan dalam tes atau ujian lainnya.

Kualitas soal yang cukup baik berdasarkan efektivitas pengecoh adalah soal yang termasuk dalam kategori buruk, di mana terdapat dua opsi yang tidak berfungsi. Jika ada dua pengecoh yang tidak dipilih oleh minimal 5% peserta tes, maka soal tersebut dianggap cukup baik. Terdapat 16 butir soal yang cukup baik, yaitu nomor 3, 5, 7, 11, 15, 16, 18, 19, 22, 23, 27, 30,31, 38, 39, dan 40. Soal-soal dalam kategori cukup baik dapat dimasukkan ke dalam bank soal dengan syarat sudah direvisi. Butir soal nomor 1, 2, 8, 10, 13, 14, 17, 20, 21, dan 25 termasuk dalam kategori soal yang jelek, sehingga soal-soal ini sebaiknya tidak digunakan lagi.

Berdasarkan penjelasan di atas, dapat disimpulkan bahwa salah satu syarat agar suatu soal dikatakan baik adalah jika soal tersebut memiliki kualitas pengecoh yang baik. Butir soal dianggap baik jika opsi pengecohnya memiliki peluang yang setara untuk dipilih oleh peserta tes (Sulistiawan, 2016). Sebaliknya, jika opsi pengecohnya tidak memiliki peluang yang sama, maka butir soal pilihan ganda tersebut dapat dianggap kurang baik.

Implikasi dari analisis butir soal Fikih di MTs dapat mencakup beberapa aspek yang penting untuk meningkatkan kualitas pembelajaran dan evaluasi antara lain: meningkatkan kualitas evaluasi baik dari segi akurasi pengukuran maupun penyesuaian soal, pengembangan strategi pembelajaran yang focus pada kelemahan serta kelebihan siswa, meningkatkan kemampuan dan keterampilan siswa, pengembangan dan evaluasi kurikulum, meningkatkan kredibilitas penilaian dan meningkatkan keterampilan evaluasi pendidik yang pada akhirnya mendukung pencapaian tujuan pendidikan yang lebih baik

Penelitian mengenai analisis butir soal ini telah dilakukan sesuai dengan prosedur ilmiah, namun masih memiliki beberapa keterbatasan antara lain sebagai berikut: peneliti hanya menganalisis soal pilihan ganda, sehingga hasil penelitiannya bergantung pada sampel yang digunakan. Jika soal diujikan pada sampel yang berbeda, maka hasilnya kemungkinan akan berbeda pula. Program Anates Versi 4 yang digunakan untuk menganalisis soal dalam penelitian ini, hanya menganalisis butir soal secara umum. Program ini tidak dapat digunakan untuk menganalisis secara detail, misalnya untuk 64 menentukan apakah siswa benar-benar memahami materi atau hanya menjawab dengan benar secara kebetulan.

SIMPULAN

Berdasarkan hasil penelitian analisis butir soal asesmen sumatif akhir semester gasal kelas VIII mata pelajaran Fikih di MTs Muhammadiyah Blimbing menggunakan program Anates Versi 4 dapat disimpulkan tingkat kesukaran soal dalam asesmen sumatif akhir semester ini masih belum memenuhi standar ideal, karena sebagian besar soal yang diberikan tergolong dalam kategori mudah. Hal ini mengakibatkan hasil evaluasi kurang efektif dalam mengukur variasi kemampuan siswa secara menyeluruh. Selain itu, dari segi daya pembeda, soal-soal tersebut tidak berfungsi dengan baik untuk membedakan antara siswa yang memiliki pemahaman mendalam dan mereka yang hanya memiliki pemahaman dasar. Oleh karena itu, soal-soal ini perlu direvisi atau tidak digunakan dalam tes selanjutnya untuk memastikan bahwa asesmen mampu menilai secara lebih akurat dan adil tingkat pemahaman dan kemampuan siswa. Efektivitas pengecoh soal asesmen ini banyak yang tidak berfungsi, sehingga perlu dilakukan perbaikan dengan menciptakan pilihan ganda yang memiliki distraktor yang setara.

DAFTAR PUSTAKA

- Ahmad Asyasyahru, W. (2020). *Analisis Soal Penilaian Akhir Tahun (PAT) Mata Pelajaran Fikih Kelas IX di Madrasah Aliyah NU Salafiyah Kenduren Tahun Pelajaran 2019/2020 [Skripsi Sarjana, Dipublikasikan]*.
- Al-Fathiyah, Fathimah, S. (2019). Analisis Butir Soal Pelajaran Bahasa Arab. *Tarbiyatuna: Jurnal Pendidikan Ilmiah*, 4(1), 76–100.
- Ardiansyah. (2022). Analisis Butir Soal Tes Uraian pada Mata Pelajaran SKI Kelas IX MTsS TI Candung Tahun Ajaran 2020/2021. *Didaktika: Jurnal Kependidikan*, 11(1), 1–8. <https://doi.org/10.58230/27454312.113>
- Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2021). Analisis Butir Soal Ulangan Harian Ipa Menggunakan Software Anates Pada Pendekatan Teori Tes Klasik. *Natural Science Education Research*, 4(1), 65–76. <https://doi.org/10.21107/nser.v4i1.8133>
- Halik, A. S., Mania, S., & Nur, F. (2019). Analisis Butir Soal Ujian Akhir Sekolah (Uas) Mata Pelajaran Matematika Pada Tahun Ajaran 2015/2016 Smp Negeri 36 Makassar. *Al Asma : Journal of Islamic Education*, 1(1), 11. <https://doi.org/10.24252/asma.v1i1.11249>
- Juafsan Liadia Cici, Tuti Sarwita, dan Irfandi, U. B. B. G. (2022). Survei Tingkat Presepsi Guru dan Orang Tua Terhadap Pembelajaran Daring (Terhadap Jaringan) Selama Covid 19 Pada SMA 1 Baitusalam Aceh Besar. *Jurnal Ilmiah Mahasiswa*, 3(2).
- Nilmasari, N. (2014). Memahami Studi Dokumen Dalam Penelitian Kualitatif. *Wacana*, 13(2), 1–5.
- Rahmaini, A., & Taufiq, A. N. (2018). Analisis Butir Soal Pendidikan Agama Islam Di SMK N 1 Sedayu Tahun Ajaran 2017/2018 (Analisis Tingkat Kesukaran, Daya Pembeda dan Fungsi Distraktor pada Soal Pilihan Ganda Kelas XI). *Jurnal Mudarrisuna*, 8(1), 1–24. <https://www.jurnal.ar-raniry.ac.id/index.php/mudarrisuna/article/view/2787>
- Rishan, M., & Sulaiman, S. (2023). Analisis Butir Soal Tipe Multiple Choice Ujian Akhir Semester Ganjil Kelas X Tahun Ajaran 2022/2023 Mata Pelajaran Pendidikan Agama Islam di SMA Negeri 1 Padang. *Islamika*, 5(3), 981–997. <https://doi.org/10.36088/islamika.v5i3.3517>
- Rusmayani. (2020). Analisis Butir Soal Penilaian Akhir Semester Genap Mata Pelajaran Pendidikan Agama Islam di SMP Bintang Persada Tabanan-Bali. *Widya Balina*, 5(1), 41–49. <https://doi.org/10.53958/wb.v5i1.50>
- Subando, J. (2021). *Teknik Analisis Data Kuantitatif Teori dan Aplikasi dengan SPSS*. Lakeisha.
- Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Insan Madani.
- Sulistiwawan, C. H. (2016). Kualitas Soal Ujian Sekolah Matematika Program Ipa Dan Kontribusinya Terhadap Hasil Ujian Nasional. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(1), 1–10. <https://doi.org/10.21831/pep.v20i1.7516>
- Sunarti, & Rahmawati, S. (2014). *Penilaian dalam Kurikulum 2013: Membantu Guru dan Calon Guru Mengetahui Langkah-Langkah Penilaian Pembelajaran*. CV. Andi Offset.
- Uno, B. H., & Koni, S. (2012). *Asesment Pembelajaran*. Bumi Aksara.