

**Analisis Butir Soal Pendidikan Agama Islam dalam Ujian Akhir Semester Tingkat Sekolah Dasar: Studi Kasus SD Negeri Kaliwining 07 Rambipuji Jember**

**Rofiatus Surul**

Program Magister Pendidikan Guru Madrasah Ibtidaiyah Universitas Islam Negeri  
Sunan Kalijaga Yogyakarta  
rofiatus239@gmail.com

**Abstract**

The analysis of test items is a process of examining questions in a test with the aim of obtaining high-quality question items. Therefore, an analysis related to the quality of question items is necessary, considering difficulty level, discrimination power, and distractor effectiveness. The research conducted to analyze these question items used a quantitative research design, specifically descriptive quantitative research. The study's population consisted of fourth-grade students at SDN Kaliwining 07 Rambipuji Jember. The research sample included 21 students in one class, with Islamic Religious Education as the subject of the study. The instruments used included the final semester exam sheet, answer key, and student answer sheets, which were used as data for analysis. Data collection techniques involved observation and documentation. Quantitative data analysis was performed using computer software called *Iteman*, specialized for test item analysis. The results indicated that the difficulty level of the question items was good, as evidenced by difficulty values ( $P$ ) between 0.3-0.7, within the recommended range of 0.2-0.8. The items analyzed based on student answers showed good difficulty levels. However, the discrimination power results varied, indicating some items were poor, moderate, or good. Further improvement from the school or educators is needed, as some items failed to differentiate between students who mastered the competencies and those who did not. Distractor effectiveness had both good and very good criteria, ensuring that each question item was equally selected by students.

**Keywords:** PAI item analysis, difficulty level, discrimination power, distractor effectiveness.

**Abstrak**

Analisis butir soal merupakan proses pengkajian pertanyaan-pertanyaan dalam tes dengan tujuan untuk mendapatkan perangkat pertanyaan yang berkualitas. Oleh karena itu diperlukan analisis terkait dengan kualitas butir soal baik dari tingkat kesukaran, daya beda, dan daya pengecoh. Penelitian yang digunakan untuk menganalisis butir soal tersebut yakni menggunakan desain penelitian kuantitatif. Jenis penelitian deskriptif kuantitatif. Populasi dalam penelitian ini adalah siswa kelas IV SDN Kaliwining 07 Rambipuji Jember. Sampel penelitian terdiri dari 21 siswa dalam satu kelas, dengan mata pelajaran Pendidikan Agama Islam kelas IV sebagai objek penelitian. Instrumen yang digunakan dalam penelitian ini meliputi lembar soal ulangan akhir semester, kunci jawaban, serta lembar jawaban siswa yang digunakan sebagai data untuk analisis. Teknik pengumpulan data yang digunakan dalam penelitian ini adalah observasi dan dokumentasi. Analisis data kuantitatif dalam penelitian ini dilakukan menggunakan aplikasi komputer bernama *Iteman*, yang merupakan perangkat lunak khusus untuk analisis butir soal tes. Hasil dari analisis tersebut menghasilkan bahwa tingkat kesukaran dari butir soal mendapatkan hasil tiap butir soal memiliki tingkat kesukaran yang bagus, hal itu dapat dilihat dari butir soal

yang memiliki nilai kesukaran atau nilai P antara 0,3 -0,7 yang mana tingkat kesukaran itu dinyatakan bagus apabila memiliki rentan 0,2-0,8. Sehingga dari butirsoal yang dianalisis berdasarkan jawaban siswa memiliki tingkat kesukaran yang bagus. Sedangkan hasil daya beda dari butir soal mendapatkan beberapa ragam hasil, dari yang jelek, sedang, dan baik. Hal tersebut ditunjukkan oleh hasil Rpbis A. 0.496, B. 0.023, C.- 0.166, dan D. -0.166. Sehingga daya beda yang terdapat dalam butir soal tersebut sebagian belum mampu membedakan anatar peserta didik yang membedakan kemampuan siswa yang telah menguasai dan tidak menguasai kompetensi yang dicapai. Oleh karena itu diperlukan adanya perbaikan dari pihak sekolah ataupun pendidik. Daya pengecoh, yang memperoleh kriteria baik dan sangat baik baik, hal tersebut dapat dilihat dari hasil prop A. 0,095, B. 0,382, C. 0,429, D. 0,095. yang tergolong kriteria baik dan sangat baik, sehingga tiap butir soal dipilih dengan dipilih dengan proporsi yang sama oleh peserta didik.

**Kata Kunci:** Analisis butir soal PAI, tingkat kesukaran, daya beda, daya pengecoh.

## **PENDAHULUAN**

Pendidikan memiliki peran penting dalam kehidupan manusia untuk meningkatkan kualitas berpikir individu maupun kelompok, baik secara jasmani maupun rohani. Suatu negara dapat maju dan berhasil jika pendidikan diperoleh secara maksimal dan ditempuh hingga jenjang tertinggi. Penilaian adalah proses atau kegiatan sistematis dan berkesinambungan untuk mengumpulkan informasi tentang proses dan hasil belajar peserta didik dengan tujuan membuat keputusan berdasarkan kriteria dan pertimbangan tertentu (Arifin, 2014). Penilaian dilakukan untuk mengetahui sejauh mana kemampuan siswa yang telah dicapai dalam proses pembelajaran. Penilaian dalam bentuk tes digunakan oleh lembaga pendidikan untuk mengevaluasi kemampuan siswa. Penilaian diharapkan menjadi bukti nyata dari proses pembelajaran dan memberikan gambaran upaya siswa dalam mencapai hasil belajar yang baik. Penilaian juga diharapkan menjadi pedoman bagi orang tua dalam memantau belajar mengajar siswa untuk meningkatkan prestasinya.

Penilaian mencakup penilaian kognitif, psikomotor, dan afektif. Penilaian kognitif dilakukan untuk mengetahui pengetahuan siswa dalam penguasaan materi pembelajaran. Penilaian kognitif sering dilakukan oleh guru pada berbagai mata pelajaran di akhir semester dalam bentuk nilai. Penilaian kognitif berbentuk soal pilihan ganda dan bentuk tes lainnya. Tujuan dari penilaian kognitif adalah untuk mengetahui apakah siswa memahami materi pembelajaran dan apakah mereka berkembang. Dengan hasil penilaian kognitif, siswa dapat mengetahui tingkat pengetahuan masing-masing dan dapat menjadi dasar untuk memperbaiki hasil belajar.

Pelaksanaan penilaian tes kognitif dalam mata pelajaran Pendidikan Agama Islam di sekolah dasar di setiap daerah mungkin serupa, namun terdapat sedikit perbedaan dalam waktu pembelajarannya, terutama dalam hal jam praktik dan jam teori. Dalam pembelajaran Pendidikan

Agama Islam di SDN Kaliwining 07 Rambipuji. Ujian akhir semester Pendidikan Agama Islam di SDN Kaliwining 07 Rambipuji pada tahun ajaran 2023/2024 menggunakan soal dari pemerintah kabupaten dalam bentuk soal uraian yang mencakup pilihan ganda, isian, dan uraian. Tingkat kesulitan dan daya pembeda soal ujian akhir semester belum diketahui, serta belum diketahui apakah soal-soal tersebut memenuhi standar kelayakan untuk digunakan dalam tes.

Penilaian hasil belajar yang dilakukan di SDN Kalwining 07 Rambipuji belum sepenuhnya memperhatikan kualitas butir soal tes. Sehingga belum diketahui apakah butir soal tersebut memenuhi kriteria tes yang baik atau belum. Butir soal yang baik dapat dimasukkan ke dalam bank soal (item banking) dan dapat digunakan untuk tes berikutnya, sementara butir soal yang kurang baik atau berkualitas rendah dapat direvisi atau tidak digunakan kembali. Analisis butir soal berkaitan dengan apakah butir soal tes dapat mengukur hal yang seharusnya diukur dan apakah butir soal tersebut berkualitas.

Seperti halnya penelitian yang telah dilakukan oleh Sendi Dwi Ashari dengan judul Tingkat Kesukaran, Daya Pembeda dan Efektivitas Pengecoh Soal Ulangan Akhir Semester satu Kelas IV dan V SD Negeri Ngombol, Purworejo Tahun Ajaran 2020/2021 dengan hasil penelitian yang menunjukkan bahwa hasil analisis dari total keseluruhan 20 butir soal pada kelas IV yang ditinjau dari segi tingkat kesukaran butir soal terdapat sebanyak 15% butir soal termasuk dalam kategori sukar, 60% butir soal termasuk dalam kategori sedang dan 25% butir soal termasuk dalam kategori mudah. Ditinjau dari segi daya pembeda butir soal terdapat sebanyak 15% butir soal termasuk dalam kategori jelek, 45% butir soal termasuk dalam kategori sedang, 20% butir soal termasuk dalam kategori baik, 0% butir soal termasuk dalam kategori sangat baik dan 20% butir soal termasuk dalam kategori jelek sekali.

Selain itu penelitian yang sama juga telah dilakukan oleh Sedy Santosa dan Jami Ahmad Badawi dengan judul Analisis Butir Soal Pilihan Ganda Tema Pertumbuhan dan Perkembangan Makhluk Hidup Kelas III Madrasah Ibtidaiyah dengan hasil penelitian yang menunjukkan bahwa hasil penelitian menunjukkan 92% soal berkategori valid dan 8% soal berkategori tidak valid, tingkat reliabilitas tinggi dengan R1 sebesar 0.83, tingkat kesukaran soal dengan presentase sebesar 0%. soal berkategori sedang dengan presentase 12%, soal kategori mudah dengan presentase 88%. Pada tingkat daya beda terdapat 36% atau sebanyak 9 soal berkategori daya pembeda baik, 48% atau sebanyak 12 soal berkategori sedang dan 16% atau sebanyak 4 soal berkategori Buruk. Sedangkan analisis fungsi pengecoh menunjukkan 16% pengecoh soal berfungsi dengan baik yang terdapat pada 4 butir soal dan 84% pengecoh soal yang tidak berfungsi yang tersebar pada 21 butir soal.

Kemudian penelitian terkait dengan tema yang sama juga telah dilakukan oleh Weni Widayati, Bistari, dan Suparjan dengan judul Analisis Butir Soal Pilihan Ganda Penilaian Tengah Semester Pada Pembelajaran Tematik Kelas V Sekolah Dasar Negeri 39 Pontianak Kota dengan hasil penelitian tingkat kesukaran pada tema 6 sebanyak 13 butir (68,42%) kategori mudah, dan 2 butir (10,53%) kategori sukar, tema 7 sebanyak 11 butir (61,11%) kategori mudah, sehingga termasuk soal yang belum berkualitas baik dari segi tingkat kesukaran. (3) Daya pembeda tema 6 sebanyak 10 butir (52,63%) kategori jelek dan 1 butir (5,26%) kategori baik, tema 7 sebanyak 6 butir (33,33%) kategori jelek dan 3 butir (16,67%) kategori baik sehingga soal termasuk soal yang berkualitas sedang dari segi daya pembeda. Dengan adanya penelitian-penelitian tersebut menunjukkan letak pentingnya menganalisis butir soal yang akan diujikan, pasalnya dengan soal-soal tersebut pendidik akan dapat mengukur sejauh mana kemampuan peserta didik dalam menerima materi pembelajaran. Oleh karena itu, analisis butir soal ini dilakukan untuk mengetahui tingkat kesulitan soal, daya pembeda, dan efektivitas pilihan jawaban. Berdasarkan latar belakang tersebut di atas, peneliti tertarik untuk melakukan penelitian mengenai analisis butir soal tes hasil belajar siswa dengan judul “Analisis Daya Pembeda Pengecoh Dan Tingkat Kesukaran Soal Pendidikan Agama Islam dalam Ujian Ahir Semester Tingkat Sekolah Dasar: Studi Kasus SDN Kaliwinig 07 Rambipuji Jember”.

## **METODE**

Metode penelitian yang digunakan dalam penelitian ini adalah penelitian kuantitatif. Jenis penelitian deskriptif kuantitatif bertujuan untuk menggambarkan fenomena yang ada melalui deskripsi verbal, kalimat, dan data numerik, baik pada saat ini maupun masa lampau. (Penelitian ini dilakukan dengan tujuan untuk mendapatkan informasi dan data yang digunakan untuk menggambarkan Tingkat Kesukaran, Daya Pembeda, dan Efektivitas Pengecoh Soal Ulangan Akhir Semester Pendidikan Agama Islam kelas IV SDN Kaliwining 07 Rambipuji Jember.

Populasi dalam penelitian ini adalah siswa kelas IV SDN Kaliwining 07 Rambipuji Jember. Sampel penelitian terdiri dari 21 siswa dalam satu kelas, dengan mata pelajaran Pendidikan Agama Islam kelas IV sebagai objek penelitian. Instrumen yang digunakan dalam penelitian ini meliputi lembar soal ulangan akhir semester, kunci jawaban, serta lembar jawaban siswa yang digunakan sebagai data untuk analisis. Teknik pengumpulan data yang digunakan dalam penelitian ini adalah observasi dan dokumentasi. Analisis data kuantitatif dalam penelitian ini dilakukan menggunakan aplikasi komputer bernama *Iteman*, yang merupakan perangkat lunak khusus untuk analisis butir soal tes.

## HASIL DAN PEMBAHASAN

Berikut ini hasil analisis dari tingkat kesukaran, daya pembeda, dan tingkat pengecoh dengan menggunakan aplikasi *iteman* dari butir soal ujian akhir semester Pendidikan Agama Islam kelas IV di SDN Kaliwining 07 Rambipuji pada tahun ajaran 2023/2024 yaitu:

Tabel 1. Hasil Analisis Tingkat Kesukaran

Item ID	P / ItemMean	R	Flag(s)
item 4	0.476	0.271	K, LR
item 5	0.333	0.089	K
item 8	0.762	0.064	K
item 10	0.381	0.319	K, LR
item 11	0.476	0.146	K
item 15	0.333	0.009	K
item 20	0.333	0.089	K

Dalam pemilihan soal jika dilihat dari aspek kesukaran, sesuai dengan teori yang dikatakan oleh Arifin yang menyebutkan bahwa perhitungan tingkat kesukaran soal adalah pengukuran sejauh mana suatu soal dianggap sulit (Arifin, 2014). Jika tingkat kesukaran suatu soal seimbang atau proporsional, maka dapat disimpulkan bahwa soal tersebut baik. Sebaiknya, soal tes tidak terlalu sulit atau terlalu mudah. Sehingga disarankan untuk menggunakan butir soal dengan tingkat kesukaran yang seimbang, yaitu 25% sulit, 50% sedang, dan 25% mudah.

Berdasarkan hasil analisis yang dilakukan di SDN Kaliwining 07 Rambipuji dalam mata pelajaran Pendidikan Agama Islam mendapatkan hasil bahwasanya tiap butir soal memiliki tingkat kesukaran yang bagus, hal itu dapat dilihat dari butir soal yang memiliki nilai kesukaran atau nilai P antara 0,3-0,7 yang mana tingkat kesukaran itu dinyatakan bagus apabila memiliki rentang 0,2-0,8. Sehingga dari butir soal yang dianalisis berdasarkan jawaban siswa memiliki tingkat kesukaran yang bagus. Kemudian hasil analisis dari daya pembeda butir soal ujian akhir semester Pendidikan Agama Islam kelas IV di SDN Kaliwining 07 Rambipuji pada tahun ajaran 2023/2024 yaitu:

Tabel 2. Hasil Analisis Daya Pembeda

Option	N	Prop.	Rpbis	Rbis	Mean	SD	Color	
--------	---	-------	-------	------	------	----	-------	--

A	3	0.143	0.496	0.770	17.333	5.686	Maroon	
B	6	0.286	0.023	0.030	12.667	2.338	Green	
C	2	0.095	-0.166	-0.288	10.500	6.364	Blue	
D	10	0.476	-0.271	-0.340	12.400	3.836	Olive	<b>**KEY*</b> *
Omit	0							
Not Admin	0							

Berikut ini indek deskriminasi item, untuk melihat dan menganalisis daya beda dari masing-masing butir soal.

Tabel 3. Tolak Ukur Daya Beda

<b>Indeks deskriminasi item</b>	<b>Klasifikasi</b>
Kurang dari 0,20	Jelek
0,20-0,40	Sedang
0,41-0,70	Baik
0,71-1,00	Sangat baik
Bertanda negative	Jelek sekali

Berdasarkan data yang diperoleh dari hasil analisis butir soal yang dilakukan di SDN Kaliwining 07 Rambipuji dalam mata pelajaran Pendidikan Agama Islam mendapatkan beberapa ragam hasil dari yang jelek, sedang, dan baik. Hal tersebut ditunjukkan oleh hasil Rpbis A. 0.496, B. 0.023, C. -0.166, dan D. -0.166. Sehingga daya beda yang terdapat dalam butir soal tersebut sebagian belum mampu membedakan anatar peserta didik yang membedakan kemampuan siswa yang telah menguasai dan tidak menguasai kompetensi yang dicapai. Oleh karena itu, diperlukan adanya perbaikan dari pihak sekolah ataupun pendidik. Hal tersebut tentunya sejalan dengan teori yang dikatakan oleh Arifin yang menjelaskan bahwa perhitungan daya pembeda adalah pengukuran sejauh mana suatu butir soal mampu membedakan antara peserta didik yang telah menguasai kompetensi dengan peserta didik yang belum atau kurang menguasai kompetensi, berdasarkan kriteria tertentu (Arifin, 2014). Semakin tinggi koefisien daya pembeda suatu butir

soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang kurang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi.

Selanjutnya hasil dari analisis tingkat pengecoh, berdasarkan hasil analisis dengan *itemen* mendapatkan hasil sebagai berikut.

Tabel 4. Hasil Tingkat Pengecoh

Option	N	Prop.	Rpbis	Rbis	Mean	SD	Color	
A	2	0.095	0.037	0.065	13.000	2.828	Maroon	
B	8	0.381	-0.278	-0.354	11.250	3.196	Green	
C	9	0.429	0.203	0.255	14.444	5.114	Blue	<b>**KEY*</b> *
D	2	0.095	0.081	0.140	13.500	2.121	Olive	
Omit	0							
Not Admi n	0							

Berikut ini kriteria untuk menilai penggunaan pengecoh menurut Arifin yaitu sebagai berikut (Arifin, 2014):

Sangat Baik IP=76%-125%

Baik IP=51%-75% atau 126%-150%

Kurang Baik IP=26%-50% atau 151%-175%

Jelek IP=0%-25% atau 176%-200%

Sangat jelek IP= lebih dari 200%

Berdasarkan hasil analisis dan nilai yang diperoleh kriteria baik dan sangat baik, hal tersebut dapat dilihat dari hasil prop A. 0,095, B. 0,382, C. 0,429, D. 0,095. yang tergolong kriteria baik dan sangat baik, sehingga tiap butir soal dipilih dengan dipilih dengan proporsi yang sama oleh peserta didik. Hal tersebut sesuai dengan pernyataan Arifin yang menyatakan bahwa dalam soal pilihan ganda terdapat alternatif jawaban yang berfungsi sebagai distractort (Arifin, 2014). Dalam butir soal yang baik, distractort akan dipilih dengan proporsi yang sama oleh peserta didik yang menjawab salah. Sebaliknya, pada butir soal yang kurang baik, distractort akan dipilih secara tidak merata. Kualitas distractort dianggap baik jika jumlah peserta didik yang memilih distractort tersebut sama atau mendekati jumlah ideal.

## **PEMBAHASAN**

Evaluasi adalah kegiatan sistematis, berkelanjutan, dan menyeluruh dalam mengendalikan, menjamin, dan menetapkan kualitas komponen pembelajaran berdasarkan pertimbangan dan kriteria tertentu (Farida, 2017). Evaluasi merupakan proses sistematis dan berkelanjutan untuk menentukan kualitas suatu hal berdasarkan pertimbangan dan kriteria tertentu dalam pengambilan keputusan (Arifin, 2014). Evaluasi adalah pernyataan tentang kualitas, kebaikan, kelebihan, nilai, atau kelayakan suatu hal yang telah diukur dan berdampak pada pengambilan keputusan. Melakukan evaluasi berarti meliputi dua kegiatan, yaitu penilaian dan pengukuran. Penilaian dalam pendidikan merupakan upaya untuk mengetahui tingkat keberhasilan kegiatan pendidikan, serta proses untuk mengetahui pengetahuan dan keterampilan peserta didik yang digunakan sebagai evaluasi (Arikunto, 2013). Maka dapat disimpulkan evaluasi melibatkan pengumpulan data objektif dan koreksi terhadap hal yang telah terjadi, dengan tujuan mendapatkan informasi akurat yang menjadi acuan dalam pengambilan keputusan untuk mencapai tujuan yang sebenarnya.

Hasil belajar mencerminkan kemampuan siswa setelah mereka mendapatkan pengalaman dan pembelajaran. Proses belajar melibatkan transformasi dalam diri individu, termasuk perkembangan berpikir, perilaku, dan kemampuan dalam membuat keputusan. Perbedaan hasil belajar antara manusia dan hewan terletak pada kemampuan manusia untuk terus berkembang dan mengalami perubahan, sedangkan hewan tidak mengalami perkembangan yang sama seperti manusia. Hasil belajar melibatkan perubahan dalam tiga aspek, yaitu kognitif, afektif, dan psikomotorik. Hasil belajar merupakan tujuan utama dari kegiatan pembelajaran di sekolah dan mencerminkan pencapaian siswa selama proses pembelajaran.

Evaluasi pembelajaran memiliki tujuan untuk mengukur keefektifan sistem pembelajaran dalam hal tujuan, materi, metode, sumber belajar, lingkungan, dan sistem penilaian. Secara umum, evaluasi pendidikan memiliki dua tujuan, yaitu mengumpulkan bukti mengenai perkembangan peserta didik setelah mengikuti proses pembelajaran dalam jangka waktu tertentu, dan mengevaluasi keefektifan metode pengajaran yang digunakan selama periode pembelajaran. Tujuan khusus evaluasi pembelajaran mencakup memotivasi peserta didik dalam mengikuti program pendidikan dan mengidentifikasi faktor-faktor yang mempengaruhi keberhasilan atau kegagalan peserta didik, sehingga dapat menemukan solusi perbaikan.

Menurut Subali evaluasi dalam proses dan hasil belajar memiliki beberapa fungsi, yaitu: 1). Memberikan arah dan petunjuk dalam pelaksanaan pembelajaran baik bagi gurumaupun subjek belajar. 2). Memberikan gambaran tentang perkembangan subjek belajar dalam hal kemampuan dan personalitasnya, sehingga mereka dapat mengenali diri sendiri dan meningkatkan



produktivitas mereka. 3). Mendorong dan memotivasi subjek belajar untuk meraih prestasi yang lebih baik. 4). Memberikan masukan untuk perbaikan dan pelaksanaan program pembelajaran di masa yang akan datang (Subali, 2012).

Arifin juga menyampaikan pandangan lain mengenai fungsi evaluasi secara menyeluruh, yang meliputi: 1). Secara psikologis, evaluasi membantu peserta didik mengetahui sejauh mana kegiatan yang telah dilakukan sesuai dengan tujuan yang ingin dicapai. Evaluasi juga memberikan rasa puas dan ketenangan kepada peserta didik. 2). Secara sosiologis, evaluasi berfungsi untuk menilai kemampuan peserta didik dalam berinteraksi dan beradaptasi dengan masyarakat yang memiliki berbagai karakteristik. 3 (Arifin, 2014)). Secara didaktis- metodis, evaluasi membantu guru dalam menempatkan peserta didik pada kelompok yang sesuai dengan kemampuan masing-masing, serta membantu perbaikan proses pembelajaran. 4). Evaluasi digunakan untuk mengetahui posisi peserta didik dalam kelompoknya, apakah termasuk anak pandai, sedang, atau kurang pandai. 5). Evaluasi membantu menentukan tingkat kesiapan peserta didik dalam mengikuti program pendidikan. 6). Evaluasi mendukung guru dalam memberikan bimbingan dan seleksi, seperti menentukan jenis pendidikan, jurusan, atau kenaikan kelas. 7). Secara administratif, evaluasi memberikan laporan tentang kemajuan peserta didik kepada orang tua, pejabat pemerintah, kepala sekolah, guru, dan peserta didik itu sendiri.

Menurut Arifin terdapat beberapa prinsip evaluasi yang perlu diperhatikan, yaitu: 1). Kontinuitas, prinsip ini menekankan bahwa evaluasi harus dilakukan secara terus- menerus dan tidak hanya secara kebetulan. Evaluasi harus mencakup seluruh proses pembelajaran, termasuk aspek input dan prosesnya, serta tidak hanya terfokus pada produk atau hasil akhir belajar peserta didik. 2). Komprehensif, prinsip ini mengarahkan guru untuk mengevaluasi seluruh aspek objek yang dievaluasi. Misalnya, jika objek evaluasi adalah peserta didik, maka seluruh dimensi kepribadian peserta didik, baik kognitif, afektif, maupun psikomotorik, harus dievaluasi. 3) (Arifin, 2014). Adil dan objektif, prinsip ini menekankan pentingnya keadilan dalam melaksanakan evaluasi bahwa guru harus bersikap adil dan tidak memihak. Semua peserta didik harus diperlakukan sama tanpa memihak. Selain itu, guru juga harus bertindak secara objektif, menghindari preferensi pribadi, perasaan, keinginan, dan prasangka negatif. Evaluasi harus didasarkan pada fakta dan data yang sebenarnya, bukan hasil manipulasi atau rekayasa. 4). Kooperatif, prinsip ini menunjukkan pentingnya kerjasama dalam kegiatan evaluasi antara guru dengan semua pihak terkait, seperti orang tua peserta didik, rekan guru, kepala sekolah, dan peserta didik sendiri. Hal ini bertujuan agar semua pihak merasa puas dengan hasil evaluasi dan merasa dihargai. 5). Praktis, prinsip ini menekankan bahwa alat evaluasi yang digunakan harus mudah digunakan, baik oleh guru yang

menyusunnya maupun oleh pihak lain yang akan menggunakannya. Oleh karena itu, perlu memperhatikan bahasa dan petunjuk dalam mengerjakan soal evaluasi agar praktis dan mudah dipahami.

Tes adalah sebuah instrumen yang digunakan untuk mengumpulkan informasi mengenai individu atau objek. Instrumen ini dapat berbentuk pertanyaan dalam bentuk kertas atau wawancara, observasi terhadap penampilan fisik dan perilaku, checklist, atau catatan anekdot. Tes didefinisikan sebagai alat pengukuran yang digunakan untuk mendapatkan data tentang karakteristik khusus individu atau kelompok. Sudijono menjelaskan bahwates merupakan cara atau prosedur yang digunakan dalam pengukuran dan penilaian pendidikan (Sudijono, 2015). Tes melibatkan pemberian tugas atau serangkaian tugas kepada peserta tes, baik berupa pertanyaan yang harus dijawab maupun perintah yang harus dilakukan, untuk menghasilkan nilai yang mencerminkan tingkah laku atau prestasi peserta tes. Nilai tersebut dapat dibandingkan dengan nilai peserta tes lainnya atau dengan nilai standar yang telah ditentukan.

Tes merupakan rangsangan yang diberikan kepada seseorang dengan tujuan untuk mendapatkan jawaban yang digunakan sebagai dasar penentuan skor atau angka (Susetyo, 2015). Arifin mengemukakan bahwa tes adalah teknik atau metode yang digunakan dalam kegiatan pengukuran. Tes ini berisi berbagai pertanyaan, pernyataan, atau serangkaian tugas yang harus dikerjakan atau dijawab oleh peserta didik untuk mengukur aspek perilaku mereka (Arifin, 2014). Dalam mengingat pentingnya sebuah tes, terutama jika digunakan sebagai alat pengambil keputusan, diperlukan tes yang baik yang memenuhi karakteristik tes yang baik.

Karakteristik tes yang baik meliputi validitas, reliabilitas, objektivitas, praktikabilitas, dan ekonomis (Sudijono, 2015). 1). Validitas, mengacu pada sejauh mana tes tersebut benar-benar mengukur apa yang ingin diukur. Misalnya, jika tes ditujukan untuk mengukur kemampuan berbicara, maka tes harus berbentuk tes lisan, bukan tes tertulis. Validitas dapat dibedakan menjadi validitas logis, yaitu validitas yang dianalisis secara logis berdasarkan teori-teori ahli, dan validitas empiris, yaitu validitas yang dianalisis berdasarkan data empiris. 2). Reliabilitas, reliabilitas mengacu pada konsistensi hasil tes. Jika tes diberikan beberapa kali kepada individu yang sama, hasilnya harus tetap konsisten. Meskipun tidak harus sama, namun secara keseluruhan jika hasil tes menurun, maka hasil semua peserta tes juga cenderung menurun. 3). Reliabilitas menggambarkan konsistensi yang dapat dipercaya, seperti seseorang yang berbicara dengan konsisten. 4). Objektivitas, objektivitas mengacu pada ketetapan dan konsistensi dalam sistem penilaian. Objektivitas menunjukkan bahwa hasil tes tidak dipengaruhi oleh faktor pribadi. Hasil tes seharusnya mampu menunjukkan kemampuan peserta tes seobjektif mungkin. 5).

Praktikabilitas, praktikabilitas mengacu pada kemudahan dan kepraktisan dalam penggunaan dan administrasi tes. Tes yang praktis mudah dilaksanakan, diperiksa, dan dilengkapi dengan petunjuk yang jelas. Tes yang baik harus sederhana dan lengkap. 6). Ekonomis, ekonomis menunjukkan bahwa tes tidak memerlukan biaya yang mahal, waktu yang lama, dan banyak. Tes harus dapat diselenggarakan dengan baik tanpa membebani sumber daya secara berlebihan.

Tes yang baik perlu direncanakan dengan hati-hati dan teliti. Beberapa petunjuk yang diberikan meliputi kesesuaian tes dengan tujuan kurikulum, proporsi isi materi yang penting, perhatian terhadap tujuan tes, dan penyesuaian tes dengan tingkat kemampuan siswa. Kemudian berkaitan dengan analisis soal, yang mana analisis soal merupakan suatu kegiatan untuk mengevaluasi setiap item atau butir soal guna menilai kualitasnya. Analisis butir soal merupakan proses pengkajian pertanyaan-pertanyaan dalam tes dengan tujuan untuk mendapatkan perangkat pertanyaan yang berkualitas (Sudjana, 2014). Arikunto juga mengemukakan bahwa analisis butir soal adalah prosedur sistematis yang memberikan informasi spesifik mengenai butir tes yang disusun (Arikunto, 2013). Analisis soal tes melibatkan identifikasi soal yang baik dan buruk, serta pengevaluasian alasan mengapa suatu soal dianggap baik atau buruk. Dengan melakukan analisis soal, kita dapat memperoleh tiga hal penting dari setiap soal, yaitu: 1). Tingkat kesukaran soal (*difficulty level of an item*), mencerminkan sejauh mana tingkat kesulitan suatu soal. 2). Daya pembeda (*discriminating power*), menunjukkan kemampuan soal untuk membedakan antara kelompok peserta didik yang pandai dan yang kurang pandai. 3). Alternatif jawaban (*options*), mengevaluasi ke-menarik-an jawaban-jawaban yang disajikan dalam soal, termasuk keputusan apakah semua alternatif jawaban menarik atau ada yang tidak menarik dan tidak perlu disertakan dalam soal.

Selanjutnya teknik analisis soal berdasarkan tingkat kesukaran, menurut Arifin perhitungan tingkat kesukaran soal adalah pengukuran sejauh mana suatu soal dianggap sulit (Arifin, 2014). Jika tingkat kesukaran suatu soal seimbang atau proporsional, dapat disimpulkan bahwa soal tersebut baik. Sebaiknya, soal tes tidak terlalu sulit atau terlalu mudah. Dalam menyusun soal tes, disarankan untuk menggunakan butir soal dengan tingkat kesukaran yang seimbang, yaitu 25% sulit, 50% sedang, dan 25% mudah. Rumus untuk menghitung tingkat kesukaran (P) adalah sebagai berikut:

$$\sum B P = N$$

Keterangan :

P = tingkat kesukaran

$\sum B$  = jumlah peserta didik yang menjawab benar

N = jumlah peserta didik.

Selain itu, Arifin juga menyatakan bahwa semakin tinggi indeks tingkat kesukaran (P), maka soal dianggap semakin mudah. Sebaliknya, semakin rendah tingkat kesukaran, maka soal dianggap semakin sulit. Untuk menafsirkan tingkat kesukaran tersebut, dapat menggunakan kriteria berikut:

Tabel 5. Tabel Kriteria Kesukaran

Indeks deskriminasi item	Klasifikasi
Kurang dari 0,20	Jelek
0,20-0,40	Sedang
0,41-0,70	Baik
0,71-1,00	Sangat baik
Bertanda negative	Jelek sekali

Kemudian berkaitan dengan daya beda, Kunandar menyebutkan bahwa tingkat diskriminasi soal merujuk pada kemampuan suatu soal untuk membedakan antara peserta didik yang telah menguasai materi dan peserta didik yang belum menguasai materi (kompetensi) (Kunandar, 2014). Hal ini sejalan dengan pandangan Sudijono yang menyatakan bahwa daya pembeda suatu item adalah kemampuan suatu butir item tes hasil belajar untuk membedakan antara testee yang memiliki kemampuan tinggi dengan testee yang memiliki kemampuan rendah (Sudijono, 2015). Arifin juga menjelaskan bahwa perhitungan daya pembeda adalah pengukuran sejauh mana suatu butir soal mampu membedakan antara peserta didik yang telah menguasai kompetensi dengan peserta didik yang belum atau kurang menguasai kompetensi, berdasarkan kriteria tertentu (Arifin, 2014). Semakin tinggi koefisien daya pembeda suatu butir soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang kurang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi. Tes dalam bentuk objektif digunakan untuk menghitung indeks diskriminasi dengan rumus:

$$D = PA - PB$$

Keterangan:

D : Daya pembeda yang dicari

BA : Batas atas

BB : Batas bawah

JA : Jumlah batas atas

JB : Jumlah batas bawah

PA= BA : Proporsi atas yang benar (p, sebagai indeks kesukaran) JA

PB= BB : Proporsikelompok bawah yang menjawab benar JB JB

Setelah mendapatkan hasil diskriminasi maka hasil tersebut diklarifikasi berdasarkan kualitas soal. Ini berguna untuk mempermudah dalam menentukan kualitas soal yang telah dibuat sesuai dengan hasil perhitungan tersebut. Kemudian Klasifikasi Daya Pembeda butir soal disajikan sebagai berikut:

Tabel 6. Klasifikasi Daya Pembeda

Indeks deskriminasi item	Klasifikasi
Kurang dari 0,20	Jelek
0,20-0,40	Sedang
0,41-0,70	Baik
0,71-1,00	Sangat baik
Bertanda negative	Jelek sekali

Sudijono menjelaskan bahwa distractor telah berfungsi dengan baik ketika setidaknya 5% dari seluruh peserta tes memilih distractortersebut. Arifin menyatakan bahwa dalam soal pilihan ganda terdapat alternatif jawaban yang berfungsi sebagai distractort (Arifin, 2014). Dalam butir soal yang baik, distractort akan dipilih dengan proporsi yang sama oleh peserta didik yang menjawab salah. Sebaliknya, pada butir soal yang kurang baik, distractort akan dipilih secara tidak merata. Kualitas distractort dianggap baik jika jumlah pesertadidik yang memilih distractort tersebut sama atau mendekati jumlah ideal.

$$IP = \frac{P}{(N-B)/(n-1)} \times 100\%$$

Indeks distractort dihitung menggunakan rumus:

Keterangan:

IP : indeks pengecoh

P : Jumlah peserta didik yang memilih pengecoh

N : Jumlah peserta didik yang ikut tes

B : Jumlah peserta didik yang menjawab benar pada setiap soal

n : Jumlah alternatif jawaban (opsi) l : Bilangan tetap

Kriteria untuk menilai penggunaan pengecoh yaitu sebagaiberikut (Arifin, 2014):

Sangat Baik IP=76%-125%

Baik IP=51%-75% atau 126%-150%

Kurang Baik IP=26%-50% atau 151%-175% Jelek IP=0%-25% atau 176%-200%

Sangat jelek IP= lebih dari 200%

## **KESIMPULAN**

Berdasarkan hasil yang diperoleh dari analisis butir soal yang dilakukan, dapat disimpulkan bahwa tingkat kesukaran dari butir soal Pendidikan Agama Islam kelas IV di SDN Kaliwining 07 Rambipuji Jember, mendapatkan hasil bahwasanya tiap butir soal memiliki tingkat kesukaran yang bagus, hal itu dapat dilihat dari butir soal yang memiliki nilai kesukaran atau nilai P. Sehingga dari butir soal yang dianalisis berdasarkan jawaban siswa memiliki tingkat kesukaran yang bagus. Sedangkan hasil daya beda dari butir soal mendapatkan beberapa ragam hasil, dari yang jelek, sedang, dan baik. Sehingga daya beda yang terdapat dalam butir soal tersebut sebagian belum mampu membedakan antar peserta didik yang membedakan kemampuan siswa yang telah menguasai dan tidak menguasai kompetensi yang dicapai. Oleh karena itu diperlukan adanya perbaikan dari pihak sekolah ataupun pendidik. Dan yang terakhir daya pengecoh, yang memperoleh kriteria baik dan sangat baik, hal tersebut dapat dilihat dari hasil prop A yang tergolong kriteria baik dan sangat baik, sehingga tiap butir soal dipilih dengan proporsi yang sama oleh peserta didik.

## **REFERENSI**

- Arifin, Z. (2014). *Evaluasi Pembelajaran*. Remaja Rosdakarya.
- Arikunto, S. (2013). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Rineka cipta.
- Farida, I. (2017). *Evaluasi Pembelajaran Berdasarkan Kurikulum Nasional*. Remaja Rosdakarya.
- Kunandar. (2014). *Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013)*. Rajawali Press.
- Subali, B. (2012). *Prinsip Asesmen & Evaluasi Pembelajaran*. UNY Press.
- Sudijono, A. (2015). *Pengantar Evaluasi pendidikan*. Raja Grafindo Persada.
- Sudjana, N. (2014). *Penilaian Hasil Proses Belajar Mengajar*. Remaja Rosdakarya.
- Susetyo, B. (2015). *Prosedur penyusunan dan analisis tes untuk penilaian hasil belajar bidang kognitif*. Refika Aditama.