



## Analisis Butir Soal Tipe *Multiple Choice Questions* pada Penilaian Harian Sekolah Dasar

<sup>1\*</sup>Wandri Ramadhan, <sup>2</sup>Fildza Malahati, <sup>3</sup>Kharisma Romadhon <sup>4</sup>Syahrul Ramadhan  
<sup>1234</sup> Magister PGMI UIN Sunan Kalijaga Yogyakarta

\*Email korespondensi: [21204082025@student.uin-suka.ac.id](mailto:21204082025@student.uin-suka.ac.id)

Diserahkan: 27 Februari; Direvisi: 5 Mei 2023; Diterima: 6 Mei 2023

### Abstrak

Penilaian harian merupakan suatu instrumen pendidikan yang penting karena dapat membantu guru dalam mengukur perkembangan pembelajaran peserta didik. Salah satu tipe evaluasi yang paling umum digunakan adalah *Multiple Choice Questions* (MCQs) atau pertanyaan pilihan ganda. Penelitian ini bertujuan untuk mengungkapkan kualitas butir penilaian harian dengan bentuk MCQs berdasarkan efektifitas, koefisien reliabilitas dan pengecoh butir soal. Penelitian ini merupakan penelitian deskriptif kuantitatif, yakni menggunakan analisis data statistik yang kemudian diolah menjadi deskripsi. Kisi-kisi instrumen penilaian menggunakan materi sekolah dasar tema 4. Instrumen tes diujikan kepada 20 orang siswa kelas V MIN 3 Padang. Hasil penelitian ini ditemukan bahwa butir soal dalam instrumen tes sudah cukup baik digunakan karena memiliki skor rata-rata 11.367, tingkat kesukaran sedang dengan skor 0,568, nilai elemen cukup reliable dan memiliki daya beda yang baik dengan skor 0,070. Meskipun begitu, masih terdapat beberapa butir soal yang tidak berkualitas dari sisi tingkat kesukaran, daya pembeda, dan pengecoh soal sehingga perlu direvisi atau diganti dengan butir soal yang lain.

**Kata kunci:** Analisis Butir Soal, Pertanyaan Tipe Pilihan Jamak

### Abstract

Daily assessment is an important educational instrument because it can assist teachers in measuring student learning progress. One of the most commonly used types of evaluation is *Multiple Choice Questions* (MCQs). This study aims to reveal the quality of daily assessment items in the form of MCQs based on effectiveness, reliability coefficient, and item deception. This quantitative descriptive research uses statistical data analysis, which is then processed into a description. The assessment instrument grid used theme 4 elementary school material. The instrument was tested on 20 class V MIN 3 Padang students. The results of this study found that the items in the test instrument were good enough to use because they had an average score of 11,367, a moderate difficulty level with a score of 0.568, and element values were quite reliable. They had good discriminating power with a score of 0.070. Even so, there are still a number of items that are not of high quality in terms of difficulty level, discriminating power, and question distraction, so they need to be revised or replaced with other items.

**Keywords:** Item Analysis, Multiple Choice Questions

**How to Cite:** Ramadhan, W., Malahati, F., Romadhon, K., & Ramadhan, S. (2023). Analisis butir soal tipe multiple choice questions pada penilaian harian sekolah dasar. *Tarbiyah wa Ta'lim: Jurnal Penelitian Pendidikan dan Pembelajaran*, 10(2) 93-105. doi: <https://doi.org/10.21093/twt.v10i2.6155>



<https://doi.org/10.21093/twt.v10i2.6155>

Copyright© 2023, Ramadhan et. al.

This is an open-access article under the [CC-BY License](https://creativecommons.org/licenses/by/4.0/).



## PENDAHULUAN

Penilaian harian merupakan salah satu bentuk evaluasi yang penting dalam pendidikan. Evaluasi yang tepat akan membantu guru dalam mengevaluasi pemahaman peserta didik dan memperbaiki kualitas pembelajaran di kelas. Evaluasi pembelajaran memiliki tujuan untuk mengumpulkan informasi sebagai pedoman untuk mengukur kemajuan, perkembangan dan tercapainya belajar siswa serta efektifitas pengajaran dilakukan oleh guru (Arfah, 2021). Dalam evaluasi pembelajaran, terdapat tiga aspek yang harus dinilai, yaitu program pembelajaran, proses pembelajaran, dan hasil pembelajaran. Evaluasi ini dilakukan untuk mengevaluasi kemampuan siswa dan komponen-komponen yang terdapat dalam pembelajaran (Arifin, 2012).

Guru berperan dalam pelaksanaan evaluasi pembelajaran siswa untuk melihat keberhasilan pembelajaran. Dengan kata lain, untuk melihat keberhasilan tercapainya pembelajaran, guru diwajibkan untuk melakukan penilaian terhadap hasil belajar siswa (Fajar & Junaidi, 2022). Proses evaluasi pembelajaran di sekolah dapat menggunakan berbagai jenis pertanyaan, termasuk pertanyaan tipe *Multiple Choice Questions* (MCQs) atau pertanyaan pilihan ganda (Kusumawati dan Hadi, 2018). *Multiple Choice Questions* (MCQs) merupakan bentuk tes yang paling umum digunakan dalam pendidikan karena kelebihanannya yang dapat mengevaluasi pemahaman siswa secara cepat dan efisien dalam waktu yang singkat (Brookhart & Nitko, 2018). Di sekolah dasar, penilaian harian juga menggunakan pertanyaan tipe MCQs yang disusun oleh guru-guru pengajar. Namun, belum ada penelitian yang dilakukan untuk mengevaluasi kevalidan dan kehandalan butir soal pada penilaian harian tersebut.

Soal yang baik adalah soal yang dapat memperlihatkan kemampuan dan keterampilan siswa pada level tertentu. Untuk memastikan kualitasnya, instrumen pengukuran haruslah valid dan reliable. Validitas berarti bahwa instrumen dapat memberikan informasi yang tepat dan sesuai dengan tujuan pengukuran, sedangkan reliabilitas menunjukkan bahwa hasil pengukuran yang diperoleh konsisten setelah dilakukan pengukuran beberapa kali. Oleh karena itu, penting bagi guru untuk memilih soal yang memenuhi kriteria valid dan reliable untuk memastikan akurasi dan keandalan hasil pengukuran yang diperoleh (Fitrah dan Liesfi, 2016).

Analisis butir soal dilakukan untuk mengevaluasi tingkat kesulitan dan daya pembeda soal, serta tingkat kesalahan jawaban yang membingungkan siswa. Tujuannya adalah untuk mengidentifikasi apakah soal terlalu mudah atau sulit bagi siswa, dan untuk menilai kemampuan butir soal dalam membedakan antara siswa yang telah menguasai materi dan yang belum. Oleh karena itu, analisis butir soal perlu dilakukan melalui uji validitas dan reliabilitas, daya beda, tingkat kesukaran, dan fungsi pengecoh untuk memastikan bahwa soal yang digunakan dalam pengukuran memiliki kualitas yang baik dan dapat memberikan informasi yang akurat tentang kemampuan siswa (Anita et al., 2018).

Berdasarkan penjelasan diatas bahwa evaluasi sangat erat kaitannya dengan mengukur, menilai dan pelaksanaan tes. Tes adalah alat penilaian yang paling banyak digunakan oleh guru (Wulan, 2001). Tes menjadi salah satu bentuk penilaian atau mengukur objek dalam dunia pendidikan, dalam bentuk tugas berisi pertanyaan atau perintah soal yang harus diselesaikan oleh setiap siswa (Arfah, 2021). Tes dilaksanakan dalam pengerjaan soal-soal yang diujikan dalam bentuk pilihan ganda dan perlu dipertimbangkan sesuai kriterianya terdapat dua cara untuk menentukan skor, kekurangan soal tes diantaranya analisis soal dan menghitung validitas serta reliabilitas soalnya (Firmansyah, 2018).

Untuk menilai efektivitas sebuah soal dalam mengukur kemampuan siswa dan dapat digunakan untuk membuat keputusan penting dalam pendidikan, terdapat dua prinsip dasar yang harus diperhatikan. Soal evaluasi harus dapat mengukur apa yang semestinya diukur dan harus sesuai untuk digunakan dalam membuat keputusan tentang siswa. Soal evaluasi memiliki banyak kegunaan dalam dunia pendidikan, seperti seleksi, penempatan, motivasi dan pembimbingan, umpan balik, diagnosis dan remedial, pembelajaran, perbaikan kurikulum dan program pendidikan, serta pengembangan ilmu. Oleh karena itu, penting untuk memastikan

bahwa soal evaluasi yang digunakan dalam pendidikan memiliki kualitas yang baik dan efektif dalam mencapai tujuan pendidikan. (Tarmizi et al., 2021)

Penting untuk memiliki soal evaluasi yang dapat diandalkan untuk mengevaluasi efektivitas alat ukur yang digunakan dalam pendidikan. Oleh karena itu, analisis butir soal pilihan ganda dilakukan untuk mengukur kualitas butir soal yang akan digunakan dalam evaluasi siswa (Nasir, 2015). Dengan hasil analisis butir soal, guru dapat memastikan keakuratan keputusan yang akan diambil terhadap siswa, karena butir soal evaluasi yang baik akan memberikan keyakinan pada guru tentang kemampuan siswa dalam memahami materi (Sumiati, Widiastuti, 2018). Hasil analisis butir soal yang rendah akan menjadi bahan evaluasi untuk perbaikan pada pembelajaran berikutnya, sementara hasil analisis butir soal yang mencukupi bahkan melampaui akan menjadi bahan pertimbangan pada pembelajaran berikutnya (Slamet dan Maarif, 2014).

Analisis butir soal pilihan ganda adalah prosedur sistematis yang dapat memberikan data khusus terhadap butir soal evaluasi. Ada beberapa hal yang perlu diperhatikan dalam analisis butir soal sehingga soal evaluasi dapat dikatakan baik. Pertama, validitas tes, yang mengukur sejauh mana soal evaluasi dapat mengukur apa yang seharusnya diukur. Kedua, reliabilitas butir soal evaluasi, yang menggambarkan ketepatan tes. Ketiga, tingkat kesulitan soal, yang merupakan perbandingan jumlah siswa yang menjawab benar dengan jumlah siswa seluruhnya. Keempat, daya beda soal, yang mengukur kemampuan tes dalam membedakan siswa yang telah menguasai materi dan yang belum. Kelima, tingkat pengecoh soal, yang mencerminkan pilihan jawaban siswa terhadap kemungkinan jawaban yang telah disediakan pada setiap butir soal evaluasi. Keenam, efektivitas pengecoh soal, yang mengukur sejauh mana pengecoh dapat membedakan antara siswa yang benar-benar memahami materi dan siswa yang hanya mengira-ngira. Melalui analisis butir soal yang sistematis, guru dapat memastikan bahwa soal evaluasi yang digunakan dapat memberikan data yang valid dan dapat digunakan untuk mengambil keputusan yang tepat mengenai kemampuan siswa (Kunandar, 2014).

Dalam metode analisis butir soal, terdapat dua hal utama yang diperhatikan yaitu analisis validitas dan reliabilitas soal. Namun, pada evaluasi jenis objektif seperti soal pilihan ganda, analisis yang dilakukan juga meliputi tingkat kesukaran, daya beda, dan tingkat pengecoh. Kualitas sebuah butir soal dapat dilihat dari tingkat kesulitannya. Sebuah butir soal dikatakan baik apabila memiliki tingkat kesulitan sedang yang tidak terlalu mudah atau terlalu sulit. Jika butir soal terlalu mudah atau terlalu sulit, maka guru tidak dapat membedakan antara siswa yang berkemampuan rendah dan siswa yang berkemampuan tinggi (Arikunto, 2009). Selanjutnya menurut Anwar (2017), sebuah butir soal dianggap baik jika memiliki tingkat kesulitan antara 0,30 hingga 0,70.

Berdasarkan observasi, hasil evaluasi harian siswa menunjukkan bahwa siswa dengan kemampuan dan prestasi yang tinggi berprestasi lebih buruk daripada siswa dengan kemampuan standar. Hal ini terlihat pada perolehan nilai siswa pada penilaian tema harian sebelumnya bahwa terdapat soal yang memiliki tingkat kesukaran dan ada yang terlalu mudah. Merujuk pada penelitian terdahulu Mania dkk (2020) dihasilkan bahwa soal pada kategori baik 15%, soal pada kategori cukup baik 67%, dan soal pada kategori tidak baik 18%. Maka butir-butir soal dikategori layak diujikan pada siswa disekolah tersebut. sedangkan penelitian Samudar (2019) mengungkapkan Soal yang diujikan secara validitas dikategori valid, kemudian realibilitas tidak reliabel dengan instrumen 0,14 sedangkan perbandingan kesukaran yang tidak proporsional dengan perbandingan sulit/sukar : 74% sedang: 26% mudah : 0% dan daya pembeda soal yang buruk serta pengecoh soal yang efektif terdiri dari 35 soal diujikan dan 32 soal efektif. Kemudian penelitian Supandi dkk (2016) mengarah pada penelitian bahwa hasil data yang dianalisis kualitas soal kategori sangat valid dan reliabel, kemudian kesukaran pada tingkat wajar, kesukaran terkategori dengan persentase 60% soal tergolong mudah. Selanjutnya kategori 30% dari pertanyaan dalam kategori sedang dan kategori berat 10%. Penelitian ini penting dilakukan karena analisis butir soal tipe MCQs pada penilaian harian dapat memberikan informasi yang berguna bagi guru-guru pengajar dan

pengambil kebijakan pendidikan dalam meningkatkan kualitas penilaian harian dan proses evaluasi pembelajaran. Dengan mengetahui mengetahui tingkat kesukaran, daya pembeda, dan tingkat soal pengecoh dan reliabelitas serta kehandalan penilaian harian, guru dapat melakukan perbaikan pada soal-soal yang disusun dan memberikan umpan balik yang lebih baik pada siswa. Hal ini diharapkan dapat membantu siswa dalam meningkatkan pemahaman mereka terhadap materi pelajaran serta mengoptimalkan proses evaluasi pembelajaran di kelas V Sekolah Dasar.

Selain itu, penelitian ini juga penting dilakukan untuk mengukur kualitas pendidikan di sekolah dasar yang menjadi salah satu lembaga pendidikan penting di Indonesia. Penelitian ini dapat memberikan informasi yang bermanfaat bagi para pengambil kebijakan di bidang pendidikan untuk meningkatkan kualitas pendidikan di sekolah dasar dan memperbaiki kurikulum yang digunakan di sekolah-sekolah tersebut. Maka dari itu perlu mengkaji analisis terhadap hal-hal yang dilakukan guru. Dalam penelitian memfokuskan pada mengkaji analisis tugas penilaian harian pada Tema 4 “Sehat itu Penting” dengan menggunakan Iteman untuk menentukan kualitas dan kelayakan tugas yang akan diujikan dalam penilaian pembelajaran.

## METODE PENELITIAN

Jenis penelitian yang dilakukan adalah penelitian deskriptif kuantitatif, yang fokus pada analisis data dengan menggunakan perhitungan angka atau statistik. Tujuan dari penelitian deskriptif kuantitatif adalah untuk memberikan gambaran atau deskripsi terhadap suatu objek atau kegiatan yang menjadi perhatian peneliti (Darmawan, 2013). Penelitian dilaksanakan pada semester genap 2022.

Penelitian ini menggunakan metode penelitian deskriptif yang tidak bertujuan untuk menguji hipotesis tertentu, melainkan hanya untuk memberikan gambaran tentang suatu gejala atau keadaan yang terjadi. Dalam penelitian ini, objektivitas terhadap hal-hal yang menjadi fokus penelitian sangat diutamakan. Tujuan penelitian adalah untuk mengumpulkan data, mendeskripsikannya, menganalisisnya, dan kemudian membuat kesimpulan sebagai jawaban terhadap masalah penelitian yang diangkat. Tahapan atau prosedur penelitian yang dilakukan dalam penelitian ini meliputi: (1) penyusunan instrumen atau segala hal yang berkaitan dengan pengukuran, (2) pengumpulan data, (3) deskripsi data yang telah terkumpul, (4) analisis data, dan (5) penarikan kesimpulan dari hasil analisis.

Subjek penelitian ini adalah 20 siswa kelas V di MIN 3 Padang. Pengumpulan data dilakukan melalui wawancara dengan wali kelas V MIN 3 Padang dan dokumentasi untuk memperoleh data kondisi awal hasil belajar siswa. Penilaian harian tema 4 oleh guru kelas V dalam kisi-kisi yang dilampirkan sebagai berikut:

**Tabel 1. Kisi-kisi Soal Harian Pilihan Ganda**

Mata pelajaran	KD	Indikator Soal	Level Kognitif	Bentuk Soal	No Soal
Bahasa Indonesia	3.6 Menggali isi dan amanat pantun yang disajikan dengan tujuan kesenangan	Disediakan pantun dan syair siswa dapat menentukan persamaan amanat pantun dan syair pada teks dengan benar	C4	Pilihan Ganda	1, 2, 3, 4, 5, 6, 7, 8,
PKn	3.2 memahami makna tanggung jawab sebagai warga masyarakat dalam kehidupan sehari-hari	Siswa diminta menentukan contoh pelaksanaan kewajiban masyarakat dalam kehidupan sehari hari	C2	Pilihan Ganda	9, 10, 11, 12, 13, 14

Menentukan sikap tanggung jawab dalam memenuhi kewajiban dan hak sebagai warga masyarakat dalam kehidupan sehari-hari	C3	Pilihan Ganda 15, 16, 17, 18, 19, 20
---	----	--------------------------------------

Untuk mengetahui tingkat kesukaran, daya pembeda, dan tingkat soal pengecoh dan reliabelitas, digunakan teknik analisis data kuantitatif dengan menggunakan aplikasi Google Form. Hasil analisis data digunakan untuk mendeskripsikan butir soal tema 4 sehat itu penting kelas V MIN 3 Padang. Setelah pekerjaan siswa dinilai, hasilnya dijadikan data untuk mengevaluasi kualitas butir soal. Data yang telah terkumpul dan dianalisis kemudian dimasukkan ke dalam program *Item And Test Analysis* (ITEMAN) 4.0 untuk mengevaluasi daya pembeda, tingkat kesukaran, dan kualitas pengecoh soal (Huda, 2021).

Kriteria identifikasi tingkat kesukaran butir soal dipaparkan dalam bentuk tabel sebagai berikut (Iskandar & Rizal, 2018) :

**Tabel 2. Kriteria Tingkat Kesukaran Butir Soal**

Nilai P	Interpretasi
$P = 0,00$	Sangat Sukar
$0,00 < P \leq 0,30$	Sukar
$0,30 < P \leq 0,70$	Sedang
$0,70 < P < 1,00$	Mudah
$P = 1,00$	Sangat Mudah

Instrumen tes yang berkualitas harus memenuhi kriteria yang mencakup daya beda atau kemampuan untuk membedakan antara peserta yang berkemampuan tinggi dengan yang berkemampuan rendah. Hal ini bertujuan untuk memperoleh informasi yang lebih akurat mengenai kemampuan peserta didik yang beragam, baik yang berkemampuan tinggi maupun yang berkemampuan rendah (Iskandar & Rizal, 2018). Tabel yang disajikan menjelaskan kriteria daya beda butir soal dan bagaimana penerapannya dalam penilaian.

**Tabel 3. Kriteria Indeks Daya Pembeda Butir Soal**

Nilai Dp	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

Selanjutnya syarat lain yang perlu diperhatikan adalah efektivitas pengecoh atau distractor dalam soal pilihan ganda. Pengecoh dikatakan efektif jika dapat membingungkan siswa dan membuat mereka memilih jawaban yang salah. Oleh karena itu, pengecoh yang efektif dianggap berfungsi dengan baik (Iskandar & Rizal, 2018).

**Tabel 4. 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

Aspek-aspek tertentu dapat digunakan untuk mengevaluasi kualitas suatu soal, seperti kategori soal baik, cukup baik, dan buruk. Pertama-tama, jika suatu butir soal memenuhi ketiga kriteria yaitu kesukaran, daya beda soal, dan efektivitas pengecoh, maka soal tersebut dianggap memiliki kualitas baik. Kedua, jika hanya dua kriteria yang terpenuhi, maka butir soal tersebut dianggap cukup baik. Terakhir, jika tidak terpenuhi ketentuan 1 dan 2, maka butir soal tersebut dianggap memiliki kualitas yang buruk dan perlu direvisi.

Reliabilitas soal mengindikasikan seberapa akurat pengukuran yang dilakukan terhadap objek yang sama. Menurut William & Lehmann (1973), reliabilitas mengacu pada keajegan hasil pengukuran. Allen dan Yen (1979) menyatakan bahwa tes dapat dianggap reliabel jika skornya memiliki korelasi yang tinggi dengan skor sebenarnya. Crocker dan Linda M (1986) menjelaskan bahwa reliabilitas mengacu pada konsistensi hasil tes dengan kemampuan dan keadaan peserta tes. Dalam penelitian yang dilakukan oleh Huda (2021), terdapat tabel reliabilitas yang digunakan untuk menilai kualitas reliabilitas tes.

**Tabel 5. Koefisien Reliabilitas**

Koefisien Reliabilitas	Interpretasi
$r \leq 0,20$	Sangat Rendah
$0,20 < r \leq 0,40$	Rendah
$0,40 < r \leq 0,60$	Sedang
$0,60 < r \leq 0,80$	Tinggi
$0,80 < r \leq 1,00$	Sangat Tinggi

## HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini adalah untuk mengevaluasi kualitas butir soal pilihan ganda pada tema 4 "Sehat itu Penting" untuk siswa kelas V MIN 3 Padang. Analisis butir soal dilakukan untuk menemukan kelemahan dalam butir soal sehingga dapat diperbaiki sebelum digunakan dalam tes evaluasi berikutnya. Analisis butir soal mencakup mengetahui tingkat kesukaran, daya pembeda, dan tingkat soal pengecoh dan reliabelitas .

Berdasarkan hasil analisis menggunakan aplikasi Iteman 4.0 pada Penilaian tema 4 Kelas V yang terdiri dari 20 butir dan diujikan kepada 20 siswa, didapatkan karakteristik butir secara keseluruhan sebagai berikut:

**Table 6. Ringkasan Statistik**

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
Scored Items	20	11,367	0,669	10	13	0,568	0,070

Tabel 6 menunjukkan bahwa rata-rata skor 11.367, norma baku 0,669, skor minim 10, dan skor maksimum 13. Berdasarkan indeks kesukaran, rata-rata P sebesar 0,568, sehingga termasuk kesukaran sedang. Berdasarkan indeks daya beda, rata-rata Rpbis adalah 0,070, sehingga terkategori memiliki daya beda yang baik.

**Table 7. Reliabilitas**

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0,556	0,834	0,319	0,227	0,274	0,938	0,587	0,754

Tabel 7 menunjukkan bahwa nilai Alpha sebesar 0,556 dan error pengukuran sebesar 0,834, sehingga tergolong cukup reliabel

#### 1. Kesukaran Soal Penilaian Harian Tema 4

Kualifikasi butir soal dalam kategori sangat sukar, sukar, sedang, mudah, dan sangat mudah ditentukan berdasarkan nilai tingkat kesukaran atau Item Difficulty Index (IDI) yang diperoleh dari hasil analisis butir soal menggunakan aplikasi IteMan 4.0 (Hambleton & Van Der Linden, 1982). Berikut adalah deskripsi kualitatif untuk masing-masing kategori:

- Sangat sukar: Butir soal yang masuk dalam kategori sangat sukar memiliki nilai IDI yang sangat rendah, yaitu di bawah 0,20. Hal ini menunjukkan bahwa lebih dari 80% siswa tidak mampu menjawab butir soal dengan benar. Butir soal ini dapat menjadi indikator bahwa materi atau konsep yang diuji masih sulit bagi sebagian besar siswa. Oleh karena itu, butir soal ini perlu direvisi atau ditingkatkan kejelasannya agar dapat lebih mudah dipahami oleh siswa.
- Sukar: Butir soal yang masuk dalam kategori sukar memiliki nilai IDI antara 0,20-0,40. Hal ini menunjukkan bahwa sekitar 50-80% siswa tidak mampu menjawab butir soal dengan benar. Butir soal ini perlu dikaji kembali kejelasannya agar lebih mudah dipahami oleh siswa.
- Sedang: Butir soal yang masuk dalam kategori sedang memiliki nilai IDI antara 0,40-0,60. Hal ini menunjukkan bahwa sebagian besar siswa mampu menjawab butir soal dengan benar.
- Mudah: Butir soal yang masuk dalam kategori mudah memiliki nilai IDI antara 0,60-0,80. Hal ini menunjukkan bahwa hampir semua siswa mampu menjawab butir soal dengan benar.
- Sangat mudah: Butir soal yang masuk dalam kategori sangat mudah memiliki nilai IDI yang sangat tinggi, yaitu di atas 0,80. Hal ini menunjukkan bahwa hampir seluruh siswa dapat menjawab butir soal dengan benar, sehingga tidak membedakan antara siswa yang memiliki kemampuan rendah atau tinggi (Mahdiansyah, 2017).

**Tabel 8. Hasil Analisis Tingkat Kesukaran Soal Penilaian Harian Tema 4**

No.	Kategori	Butir Soal	Jumlah
1.	Sangat Sukar	9,13,14, 18	4
2.	Sukar	15,16,17,20	4
3.	Sedang	19	1
4.	Mudah	2, 3, 12	3
5.	Sangat Mudah	1,4,5,6,7,8, 10, 11,	8

Tabel 7 menunjukkan bahwa 1 soal tes termasuk dalam terkategori sedang, 4 soal termasuk dalam terkategori sulit, 4 soal termasuk dalam terkategori sangat sulit, 3 soal termasuk dalam terkategori mudah, dan 8 soal termasuk dalam terkategori sangat mudah. Dari hasil penelitian ini memiliki kesamaan dengan hasil oleh Riska dkk (2016) yang menemukan bahwa tingkat kesukaran soal buruk dalam evaluasi dan tidak bisa digunakan dalam ujian selanjutnya.

Merujuk pada indeks kesukaran soal berimbang dengan tingkat kesulitan sedang, terdapat 4 soal yang tergolong sangat sukar, yaitu soal nomor 9, 13, 14, dan 18. Jadi soal untuk soal tersebut tidak perlu diulang untuk tes berikutnya. Soal yang terlalu sukar akan menyurutkan semangat dan motivasi siswa untuk menyelesaikan soal-soal tersebut. Sedangkan soal yang terlalu sederhana tidak dapat membangkitkan rasa percaya diri siswa. Butir soal pada kategori sedang dianggap sesuai dengan kemampuan siswa karena tingkat kesukarannya dianggap terlalu keras atau terlalu mudah (Mania et al., 2020). Oleh karena itu, butir soal nomor 15, 16, 17, dan 20 yang termasuk dalam kategori sukar dapat digunakan kembali untuk ujian yang akan datang. Sementara itu, butir soal yang termasuk dalam kategori mudah harus dianalisis kembali untuk mengetahui mengapa soal tersebut mudah dikerjakan oleh peserta tes dan menemukan jawaban kunci dengan mudah, serta untuk mengidentifikasi mana yang hanya merupakan pengecoh. Oleh karena itu, dapat disimpulkan bahwa kualitas butir soal berdasarkan tingkat kesukaran dapat dikategorikan sebagai buruk.

## 2. Daya Pembeda Soal Pilihan Ganda

Butir soal dapat dikategorikan menjadi sangat jelek, jelek, cukup, baik, atau sangat baik berdasarkan analisis karakteristik butir soal seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan tingkat pengecoh. Secara ketentuan daya pembeda mengacu pada kemampuan objek untuk membedakan siswa kemampuan tinggi dan rendah dalam menjawab setiap butir soal yang diujikan.

Butir soal yang dikategorikan sangat jelek adalah soal yang memiliki tingkat kesukaran yang sangat tinggi, daya pembeda yang rendah, serta tingkat pengecoh yang tidak efektif. Butir soal yang dikategorikan jelek adalah soal yang memiliki tingkat kesukaran tinggi atau rendah, daya pembeda yang rendah, serta tingkat pengecoh yang kurang efektif. Butir soal yang dikategorikan cukup adalah soal yang memiliki tingkat kesukaran dan daya pembeda yang moderat, serta tingkat pengecoh yang efektif. Butir soal yang dikategorikan baik adalah soal yang memiliki tingkat kesukaran dan daya pembeda yang moderat hingga tinggi, serta tingkat pengecoh yang efektif. Sedangkan butir soal yang dikategorikan sangat baik adalah soal yang memiliki tingkat kesukaran yang moderat, daya pembeda yang tinggi, serta tingkat pengecoh yang sangat efektif (Hamalik, 2015).

**Tabel 9. Hasil Analisis Daya Pembeda Soal Penilaian Harian Tema 4**

No.	Kategori	Butir Soal	Jumlah
1.	Sangat Jelek	1,4,5,6,7,8,9,10,11,13,14,18	12
2.	Jelek	2,17	2
3.	Cukup	12,15,16,20	4
4.	Baik	-	0
5.	Sangat Baik	3,19	2

Berdasarkan pada hasil mengenai kualitas pertanyaan, tidak ditemukan item kategori baik, namun ditemukan 2 item sangat baik, yakni item nomor 1. 3 dan 19. Item soal ini bisa dijadikan untuk diujikan kembali dan dikumpulkan dalam bank soal. Selanjutnya terdapat 4 butir soal kategori cukup. Soal kategori cukup merujuk daya pembeda masih terlihat siswa yang memiliki pemahaman yang kurang. Pada kategori cukup besar, siswa yang tidak memahami materi mungkin salah menebak jawaban yang benar. Sementara itu terdapat 2 soal terkategori jelek yaitu 2 dan 17, serta 12 soal terkategori sangat jelek yaitu soal 1,4,5,6,7,8,9,10,11,13,14 dan 18. Soal-soal ini dianjurkan tidak lagi digunakan dan perlu diganti dengan soal pertanyaan yang baru. Pada saat yang sama, soal nilai rendah dan sangat rendah berdasarkan daya adalah masalah tidak mampu membedakan siswa yang berprestasi dari yang tidak.



### 3. Efektifitas Opsi Soal Penilaian Harian Tema 4

Suatu butir soal dapat dikategorikan ke dalam hasil analisis efektivitas opsi soal yang sangat baik, baik, kurang baik, jelek, atau sangat jelek berdasarkan kemampuan opsi soal dalam membedakan antara siswa yang benar-benar menguasai materi dan siswa yang tidak (Kusumawati dan Hadi, 2018).

- a. Butir soal yang dikategorikan sebagai hasil analisis efektivitas opsi soal yang sangat baik adalah soal yang memiliki opsi jawaban yang mampu membedakan antara siswa yang benar-benar menguasai materi dan siswa yang tidak dengan sangat baik. Opsi jawaban yang benar harus dipilih oleh siswa yang benar-benar menguasai materi, sedangkan opsi jawaban yang salah harus dipilih oleh siswa yang tidak menguasai materi.
- b. Butir soal yang dikategorikan sebagai hasil analisis efektivitas opsi soal yang baik adalah soal yang memiliki opsi jawaban yang mampu membedakan antara siswa yang benar-benar menguasai materi dan siswa yang tidak dengan cukup baik. Opsi jawaban yang benar lebih sering dipilih oleh siswa yang benar-benar menguasai materi daripada siswa yang tidak menguasai materi.
- c. Butir soal yang dikategorikan sebagai hasil analisis efektivitas opsi soal yang kurang baik adalah soal yang memiliki opsi jawaban yang tidak mampu membedakan antara siswa yang benar-benar menguasai materi dan siswa yang tidak dengan baik. Opsi jawaban yang benar cenderung dipilih oleh siswa yang tidak menguasai materi atau dipilih secara acak.
- d. Butir soal yang dikategorikan sebagai hasil analisis efektivitas opsi soal yang jelek adalah soal yang memiliki opsi jawaban yang tidak mampu membedakan antara siswa yang benar-benar menguasai materi dan siswa yang tidak. Opsi jawaban yang benar lebih sering dipilih oleh siswa yang tidak menguasai materi daripada siswa yang benar-benar menguasai materi.
- e. Butir soal yang dikategorikan sebagai hasil analisis efektivitas opsi soal yang sangat jelek adalah soal yang memiliki opsi jawaban yang benar-benar salah atau tidak relevan dengan materi yang diujikan.

Hasil analisis dalam penelitian ini sebagai karakteristik untuk menentukan kualitas dari soal yang diujikan.

**Tabel 10. Hasil Analisis Efektifitas Opsi Soal Penilaian Harian Tema 4**

No.	Kategori	Butir Soal	Jumlah
1.	Sangat Baik	12,13,15,16,19	5
2.	Baik	9,14,17	3
3.	Kurang Baik	20	1
4.	Jelek	1,2,3,4,5,6,7,8,10,11,18	11
5.	Sangat Jelek	-	0

Pada soal nomor 12,13,15,16 dan 19 terkategori sangat baik, sedangkan nomor 9,14, dan 17 terkategori soal baik. Soal ini akan berpeluang tinggi untuk sebagai alternatif pengecoh soal yang diambil atau dipilih oleh peserta tes. Oleh karena itu soal-soal terkategori sangat baik dan baik bisa digunakan dan disimpan ke dalam bank data soal untuk diuji coba kembali pada tes berikutnya. Sementara itu soal terkategori jelek hanya satu soal yaitu soal no 1,2,3,4,5,6,7,8,10,11,18 dan terkategori kurang baik soal

no 20 sehingga soal sebaiknya tidak diujikan dalam penilaian harian tema yang akan datang.

Berdasarkan penelitian yang dilakukan oleh Sulistiawan (2016), sebuah butir soal dapat dikatakan baik jika memiliki opsi pengecoh yang berkualitas baik. Butir soal yang memiliki opsi pengecoh yang baik akan memiliki probabilitas yang sama untuk dipilih oleh peserta tes. Namun, jika opsi pengecoh tidak memiliki probabilitas yang sama untuk dipilih, maka butir soal tersebut sebaiknya dikategorikan sebagai butir soal yang kurang baik. Oleh karena itu, dapat disimpulkan bahwa kualitas opsi pengecoh adalah salah satu syarat penting untuk menentukan kualitas sebuah butir soal.

#### 4. Analisis Reliabilitas Pilihan Ganda

Reliabilitas merujuk pada seberapa konsisten hasil tes ketika dijalankan pada waktu yang berbeda, dan merupakan tingkat kepercayaan pada hasil tes. Hasil perhitungan dapat digunakan untuk mengukur reliabilitas suatu tes (Wahndianti dan Sumarsih, 2018).

Kategori hasil analisis reliabilitas soal dapat memberikan gambaran tentang kualitas soal dalam mengukur kemampuan siswa secara konsisten. Jika hasil reliabilitas soal sangat rendah atau rendah, maka hal tersebut menunjukkan bahwa soal tersebut tidak konsisten dalam mengukur kemampuan siswa. Sementara itu, hasil reliabilitas soal yang sedang, tinggi, atau sangat tinggi menunjukkan bahwa soal tersebut konsisten dalam mengukur kemampuan siswa (Hamalik, 2010).

Penilaian reliabilitas soal dapat dilakukan dengan berbagai metode, seperti menggunakan koefisien alpha Cronbach, model Rasch, atau model IRT (Item Response Theory). Hasil analisis reliabilitas soal dapat membantu guru atau pengembang tes untuk mengevaluasi kualitas soal dan memperbaiki kelemahan yang terdapat pada soal tersebut (Nana Sudjana, 2010).

**Tabel 11. Hasil Reliabilitas Soal Penilaian Harian Tema 4**

No.	Kategori	Butir Soal	Jumlah
1.	SangatRendah	12	1
2.	Rendah	3,15	2
3.	Sedang	1,4,5,6,7,8,9,10,11,13,14,17,18,19,20	15
4.	Tinggi	2,16	2
5.	Sangat Tinggi	-	0

Berdasarkan tabel 11 diperoleh soal berkategori reliabel sangat rendah sebanyak 1 butir dan soal berkategori reliabel rendah sebanyak 1 butir. Untuk soal berkategori tersebut perlu direvisi dan perbaikan ulang atau diganti. Sedangkan soal dengan tingkat reliabel kategori sedang terdiri 15 soal pada butir 1,4,5,6,7,8,9,10,11,13,14,17,18,19 dan 20 dan kategori reliabel tinggi terdapat dua soal pada butir 2 dan 16.

Dari hasil analisis data, dapat disimpulkan bahwa soal penilaian harian tema 4 masuk dalam kategori sedang, sementara kategori lain perlu diperiksa kembali oleh guru dan pembuat soal. Reliabilitas adalah kemampuan tes untuk memberikan hasil yang konsisten ketika diberikan kepada kelompok yang sama pada waktu atau situasi yang berbeda. Suatu tes dikatakan reliabel jika selalu memberikan hasil yang sama dalam urutan atau peringkat siswa (Arifin, 2012).

Hal ini sejalan dengan pandangan dari Arifin (2012) bahwa reliabilitas merujuk pada konsistensi instrumen yang digunakan dalam pengukuran. Suatu tes dapat dikatakan reliabel jika memberikan hasil yang konsisten ketika diberikan pada

kelompok yang sama di waktu atau kesempatan yang berbeda, dengan hasil yang sama terkait urutan atau penempatan siswa. Lebih lanjut, Sulistiawan (2016) menyatakan bahwa suatu tes dianggap reliabel jika akurasi pengujian dan penentuan hasil pengujian mencapai rata-rata kelas reliabilitas alpha.

## KESIMPULAN DAN SARAN

Evaluasi penilain harian merupakan suatu instrument pendidikan yang sangat penting karena dapat membantu guru dalam mengukur perkembangan pembelajaran peserta. Guru melaksanakan pembelajaran dengan memberikan pengujian atau tes terhadap hasil belajar siswa salah satunya menggunakan, karena jenis pengujian atau tes yang paling umum digunakan serta dapat menemukan hasil skor pengujian atau tes secara efisien. Namun, Untuk menentukan efektifitas suatu evaluasi terhadap kemampuan dan perkembangan siswa, maka terdapat dua prinsip yang harus diperhatikan yaitu ukuran dan kesesuaian evaluasi.

Efektifitas evaluasi tipe *Multiple Choice Questions* (MCQs) atau pertanyaan pilihan ganda dilakukan dengan cara menganalisis butir soal sebagai prosedur sistematis yang dapat memberikan data dengan cara menganalisis tingkat kesukaran, realibilitas, kesulitan, perbedaan daya, dan pengecoh soal. Sehingga, butir soal dapat dikatakan cukup baik apabila memiliki penyesuaian tingkat kesulitan yang ideal dengan komposisi tidak terlalu sulit namun juga tidak terlalu mudah, agar guru dapat membedakan kemampuan dan perkembangan antar siswanya.

Berdasarkan temuan pada penelitian ini yang diberikan kepada 20 orang siswa kelas V MIN 3 Padang pada tema 4 dengan judul “Sehat itu Pendint”, maka didapatkan kesimpulan statistik melalui hasil instrumen tes yaitu, skor rata-ratanya adalah 11,367 yang memiliki skor mini “10” dan skor maksimal “13”. Sedangkan skor rata-rata tingkat kesurakannya butir soal adalah skor 0,568 yang dapat dikategorikan sebagai kesukaran sedang. Kemudian, reliabilitas yang diperoleh Alpha sebesar 0,556 dan nilai elemen yang dianalisis cukup reliable (kualitas soal terukur secara konsisten). Terakhir, beda daya yang didapatkan adalah 0,070, sehingga diketahui hasil skor pada tes memiliki daya beda yang baik. Maka dapat disimpulkan bahwa 20 butir soal instrumen tes yang diberikan kepada 20 orang siswa kelas V MIN 3 Padang cukup baik dan memberikan hasil evaluasi yang sesuai dengan tujuan penelitian. Namun terdapat beberapa butir soal yang perlu direvisi atau diganti dengan yang lain. Hasil penelitian ini diharapkan dapat menjadi rujukan guru dalam melakukan penilaian harian menggunakan bentuk MCQs.

## REFERENSI

- Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory, Monterey, CA: Brooks/Cole, 1979. In *Google Scholar*.
- Anita, A., Tyowati, S., & Zuldafrial, Z. (2018). Analisis Kualitas Butir Soal Fisika Kelas X Sekolah Menengah Atas. *Edukasi: Jurnal Pendidikan*, 16(1), 35. <https://doi.org/10.31571/edukasi.v16i1.780>
- Anwar, C. (2017). *Teori – Teori Pendidikan*. Ircisod.
- Arfah, M. A. (2021). Evaluasi Hasil Belajar Pendidikan Agama Islam (Pai). *Jurnal Literasiologi*, 7(2), 211–236. <https://doi.org/10.47783/literasiologi.v7i2.282>
- Arifin, Z. (2012). Menganalisis Kualitas Tes. In *Evaluasi Pembelajaran*.
- Brookhart, S. M., & Nitko, A. J. (2018). Educational assessment of students. *Human Movement Tarbiyah wa Ta’lim: Jurnal Penelitian Pendidikan dan Pembelajaran*. Vol. 10, No.2

- Science*, 53(1), 1689–1699.
- Crocker; Linda M. (1986). *Introduction to classical and modern test theory*. Universitas Indonesia. <https://lib.ui.ac.id/detail.jsp?id=20113013>
- Darmawan. (2013). *Metode Penelitian Kuantitatif* (Bandung). Remaja Rosdakarya.
- Fajar, R., & Junaidi, J. (2022). Analisis Butir Soal Tipe Multiple Choices Questions (MCQ) pada Ujian Akhir Semester Ganjil Mata Pelajaran Sosiologi Kelas XII SMA. *Jurnal Sikola: Jurnal Kajian Pendidikan Dan Pembelajaran*, 4(1), 23–34. <https://doi.org/10.24036/sikola.v4i1.194>
- Firmansyah, H. (2018). *Analisis Butir Soal Tes Pilihan Ganda Mata Pelajaran Penjas-Orkes Kelas VII SMPN 1 Wonosari*. Universitas Negeri Yogyakarta.
- Fitrah, N., & Liesfi, R. (2016). Analisis Butir Soal Ulangan Akhir Semester Ganjil Mata Pelajaran Teori Kejuruan Akuntansi. *Kajian Pendidikan Akuntansi Indonesia*, 5(6), 1–11. <https://journal.student.uny.ac.id/index.php/kpai/article/view/5757>
- Hamalik, O. (2010). *Evaluasi pembelajaran*. Bumi Aksara.
- Hamalik, O. (2015). *Kurikulum Dan Pembelajaran* (Pertama). Bumi Aksara.
- Hambleton, R. K., & Van Der Linden, W. J. (1982). Advances in Item Response Theory and Applications: An Introduction. *Applied Psychological Measurement*, 6(4), 373–378. <https://doi.org/10.1177/014662168200600401>
- Huda, N. (2021). Item and Test Analysis (Iteman) 4.3. *Penggunaan Item and Test Analysis (ITEMAN) 4.3 Untuk Menganalisis Butir Soal Pilihan Ganda*. <http://repository.uin-malang.ac.id/8322/1/8322.pdf%0Ahttp://repository.uin-malang.ac.id/8322/>
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12–23. <https://doi.org/10.21831/pep.v22i1.15609>
- Kunandar. (2014). *Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013)*. PT Raja Grafindo Persada.
- Kusumawati, M., & Hadi, S. (2018). An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. *Research and Evaluation in Education*, 4(1), 70–78. <https://doi.org/10.21831/reid.v4i1.20202>
- Mahdiansyah. (2017). *Kementerian pendidikan dan kebudayaan badan penelitian dan pengembangan pusat penelitian kebijakan pendidikan dan kebudayaan tahun 2017*.
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis Butir Soal Ujian Akhir Sekolah. *Al Asma: Journal of Islamic Education*, 2(2), 274. <https://doi.org/10.24252/asma.v2i2.16569>
- Nana Sudjana. (2010). *Penilaian hasil proses belajar mengajar*. Remaja Rosdakarya.
- Nasir, M. (2015). Analisis Empirik Program Analisis Butir Soal Dalam Rangka Menghasilkan Soal Yang Baik dan Bermutu Sebagai Alat Evaluasi Pembelajaran Fisika. *Prosiding Semirata*, 336–347. [jurnal.untan.ac.id](http://jurnal.untan.ac.id)

- Riska Fitria Wijayanti, Suci Apriani, M. A. M. (2016). Tingkat Kesukaran Soal Buruk dalam Evaluasi Pendidikan. *Jurnal Inovasi Pendidikan Dasar*, Volume 2,.
- SAMUDAR, N. (2019). KUALITAS SOAL UJIAN SEKOLAH BERSTANDAR NASIONAL IPA SMP DI KABUPATEN LUWU UTARA. In *Penelitian dan Evaluasi Pendidikan Pascasarjana Universitas Negeri Makassar*.
- Slamet, S., & Maarif, S. (2014). Pengaruh Bentuk Tes Formatif Asosiasi Pilihan Ganda Dengan Reward Dan Punishment Score Pada Pembelajaran Matematika Siswa Sma. *Infinity Journal*, 3(1), 59. <https://doi.org/10.22460/infinity.v3i1.39>
- Suharsimi Arikunto. (2009). *Dasar-dasar evaluasi pendidikan* (Pertama). Bumi Aksara. <https://library.unismuh.ac.id/opac/detail-opac?id=2479>
- Sulistiawan, 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>
- Sumiati, Widiastuti, S. (2018). Workshop Teknik Menganalisis Butir Soal dalam Meningkatkan Kompetensi Guru di SMK Cileungsi Bogor. *Jurnal Pemberdayaan Masyarakat Madani (JPMM)*, 2, 136–153. <https://doi.org/https://doi.org/10.21009/JPMM.002.1.10>
- Supandi, S., & Farikhah, L. (2016). Analisis Butir Soal Matematika Pada Instrumen Uji Coba Materi Segitiga. *JIPMat*, 1(1), 71–78. <https://doi.org/10.26877/jipmat.v1i1.1085>
- Tarmizi, P., Setiono, P., Amaliyah, Y., & Agrian, A. (2021). Analisis Butir Soal Pilihan Ganda Tema Sehat Itu Penting Kelas V SD Negeri 04 Kota Bengkulu. *ELSE (Elementary School Education Journal) : Jurnal Pendidikan Dan Pembelajaran Sekolah Dasar*, 4(2), 124. <https://doi.org/10.30651/else.v4i2.7090>
- Wahndianti, S., & Sumarsih. (2018). *Analisis Butir Soal Ujian Akhir Semester Gasal Mata Pelajaran Hitung Dagang*. 7(1), 1–14.
- William A. Mehrens ;Irvin J. Lehmann. (1973). *MEASUREMENT AND EVALUATION in Education and Psychology* (Hal Lockwood (ed.)). Ted Buchholz. <https://www.ptonline.com/articles/how-to-get-better-mfi-results>
- Wulan, A. R. (2001). 10 Konsep Evaluasi Dan Sitasinya. *FMIPA Universitas Pendidikan Indonesia*, 1–12.