

## **JURNAL MERAH PUTIH SEKOLAH DASAR**

Volume 02 No. 06 Bulan Juli Tahun 2025

*Jurnal Merah Putih Sekolah Dasar (JMPSD) memuat artikel yang berkaitan tentang hasil penelitian, pendidikan, pembelajaran dan pengabdian kepada masyarakat di sekolah dasar.*

<https://jurnal.unimed.ac.id/2012/index.php/jmpsd>

### **ANALISIS KUALITAS BUTIR SOAL PILIHAN GANDA MATA PELAJARAN AGAMA DAN BUDI PEKERTI KELAS III SD ADVENT NABIRE MENGGUNAKAN ITEMEN**

Efa Nurlinda Purba<sup>1</sup>, Hermenius Sujati<sup>2</sup>

<sup>1</sup>Universitas PGRI Yogyakarta, Yogyakarta, Indoensia

<sup>2</sup>Universitas Negeri Yogyakarta, Yogyakarta, Indoensia

Surel: efapurbastap@gmail.com

#### **ABSTRACT**

*This study aimed to analyze the quality of 20 multiple-choice test items used in the second-semester Religion and Character Education subject for Grade III students at SD Advent Nabire. The analysis, conducted using the ITEMEN program, focused on evaluating the difficulty level, discrimination index, and effectiveness of distractors. The findings revealed that 65% of the items were of medium difficulty and 35% were easy, with a complete absence of difficult items. This distribution indicates a lack of varied challenges within the test. Furthermore, a significant majority of the items (65%) exhibited a low to very low discrimination index, suggesting their inadequacy in effectively differentiating students based on their understanding levels. Despite these concerns, 85% of the distractors were found to be effective, demonstrating a well-designed set of incorrect options. In conclusion, although the distractors performed well, the overall quality of the test items necessitates substantial improvement in both their discrimination power and the range of difficulty levels to enhance the validity of the evaluation instrument. Recommendations stemming from this study include revising problematic items, strengthening teachers' capacity in test item construction, and developing a comprehensive, high-quality item bank.*

**Keywords:** Quality of Questions, Multiple Choice, Religion and Ethics.

#### **ABSTRAK**

Tujuan penelitian ini menganalisis kualitas 20 butir soal ulangan semester II pilihan ganda mata pelajaran Agama dan Budi Pekerti Kelas III semester II SD Advent Nabire menggunakan program ITEMEN. Fokus analisis mencakup tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasilnya menunjukkan bahwa 65% soal berkategori sedang dan 35% mudah, tanpa adanya soal sukar, yang mengindikasikan kurangnya variasi tantangan. Mayoritas butir soal (65%) memiliki daya pembeda rendah hingga sangat rendah, menunjukkan ketidakmampuannya dalam membedakan siswa berdasarkan tingkat pemahaman. Kendati demikian, 85% pengecoh berfungsi efektif, menunjukkan desain distraktor yang baik. Disimpulkan bahwa, meski pengecoh sudah baik, kualitas soal secara keseluruhan perlu perbaikan signifikan pada daya pembeda dan distribusi tingkat kesukaran agar instrumen evaluasi lebih valid. Rekomendasi mencakup revisi soal bermasalah, peningkatan kapasitas guru dalam penyusunan soal, dan pengembangan bank soal berkualitas.

**Kata Kunci:** Kualitas Butir Soal, Pilihan Ganda, Agama dan Budi Pekerti

Copyright (c) 2025 Efa Nurlinda Purba<sup>1</sup>, Hermenius Sujati<sup>2</sup>

✉ Corresponding author (Perwakilan Tim) :

Email : sujati@uny.ac.id

HP : -

Received 08 Mei 2025, Accepted 19 Juni 2025, Published 01 Juli 2025

## PENDAHULUAN

Penilaian hasil belajar merupakan aspek penting dalam proses pendidikan, di mana instrumen evaluasi yang valid dan reliabel menjadi kunci utama dalam mengukur kompetensi siswa secara objektif Arikunto, 2019 Dalam konteks pembelajaran di SD Advent Nabire, mata pelajaran Agama dan Budi Pekerti memegang peranan strategis , sehingga membutuhkan alat ukur yang akurat dan efektif Instrumen berupa soal pilihan ganda dipilih karena kemudahannya dalam analisis statistik dan kemampuannya dalam mengukur berbagai aspek kompetensi secara efisien. Tujuan Umum Pendidikan Agama dan Budi Pekerti pada Pendidikan Dasar untuk membentuk peserta didik menjadi individu yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, serta berakhhlak mulia. Ini selaras dengan visi Profil Pelajar Pancasila, yang mencakup enam dimensi utama(1) Beriman, Bertakwa kepada Tuhan Yang Maha Esa, dan Berakhhlak Mulia (2) Berkebinekaan Global (3) Bergotong Royong (4) Mandiri (5) Bernalar Kritis (6) Kreatif (Kemendikbudristek 2022/2023).

Menurut para pakar, kualitas soal sangat dipengaruhi oleh aspek tingkat kesukaran, daya pembeda, serta efektivitas pilihan pengecoh (distractors) Spesifiknya, soal yang baik harus mampu membedakan siswa yang menguasai materi dengan yang belum, serta memiliki tingkat kesulitan yang seimbang sesuai kebutuhan pengukuran kompetensi secara komprehensif ( Sudijono, 2017). Pentingnya aspek ini bukan hanya untuk mendapatkan gambaran hasil belajar yang akurat, tetapi juga untuk memastikan proses pembelajaran yang berlangsung mampu mendorong siswa mencapai kompetensi yang kompetitif dan optimal.

Selain itu, pentingnya analisis kualitas butir soal dilakukan agar dapat mengidentifikasi kelemahan dan kekurangan dalam konstruksi soal, seperti ambiguitas, pilihan jawaban yang homogen, atau tingkat kesulitan yang tidak sesuai. Penggunaan perangkat lunak analisis seperti ITEMEN telah

menjadi salah satu solusi modern yang dapat membantu para pendidik dan peneliti dalam melakukan penilaian secara objektif dan sistematis terhadap karakteristik butir soal . Dengan demikian, keberadaan analisis ini dapat mendukung upaya peningkatan mutu instrumen evaluasi yang dapat digunakan secara berkelanjutan.

Berdasarkan uraian tersebut, penting dilakukan penelitian untuk menganalisis kualitas butir soal pilihan ganda pada mata pelajaran Agama dan Budi Pekerti kelas III di SD Advent Nabire. Penelitian ini bertujuan untuk mengidentifikasi aspek-aspek kritis dalam kualitas soal, seperti tingkat kesukaran, daya pembeda, dan efektivitas pengecoh, serta memberikan rekomendasi perbaikan agar instrumen evaluasi yang digunakan dapat lebih valid dan reliabel sesuai dengan standar yang dianjurkan (Arikunto, 2019). Hasil dari penelitian ini diharapkan dapat menjadi acuan dalam penyusunan soal yang lebih berkualitas dan mampu memberikan gambaran yang akurat terhadap hasil belajar siswa.

Selain itu, penelitian ini sangat relevan untuk meningkatkan kompetensi guru dalam penyusunan instrumen assessment yang sesuai standar psikometris dan pedagogis. Dengan begitu, proses evaluasi tidak hanya sekadar mengukur pencapaian siswa, tetapi juga menjadi alat pembelajaran yang mendukung peningkatan kualitas pendidikan di SD Advent Nabire secara berkelanjutan. Keberhasilan dalam meningkatkan kualitas soal akan berdampak positif terhadap peningkatan hasil belajar siswa dan efektivitas proses pembelajaran secara keseluruhan.

## METODOLOGI PENELITIAN

### Jenis Penelitian

Penelitian ini menggunakan pendekatan deskriptif kuantitatif. Data yang dianalisis adalah hasil penggeraan soal ujian akhir semester mata pelajaran Agama dan Budi Pekerti Kelas III SD Advent Nabire pada Tahun Ajaran 2024/2025

### Waktu dan Tempat Penelitian

Penelitian ini telah dilaksanakan pada Ulangan Semester II kelas III tahun ajaran 2024/2025 di SD Advent

### Populasi dan Sampel

Populasi penelitian adalah seluruh siswa kelas III SD Advent Nabire. Sampel penelitian adalah seluruh siswa Kelas III yang mengikuti ujian akhir semester II mata pelajaran Agama dan Budi Pekerti. Jumlah soal yang dianalisis adalah 20 butir soal pilihan ganda.

### Instrumen dan Teknik Pengumpulan Data

Instrumen pengumpulan data yakni mengambil lembar jawaban siswa pada panitia kemudian menganalisis data yang meliputi tingkat kesukaran soal, daya pembeda soal, pengecoh soal reliabilitas soal tes dan kesalahan baku pengukuran soal tes dan Teknik yang digunakan dalam pengumpulan data penelitian adalah teknik dokumentasi, yaitu mengambil lembar jawaban siswa pada Ulangan Akhir Sekolah kelas III semester II bidang studi Agama dan Budi Pekerti SD Advent Nabire

### Tingkat Kesukaran (P)

Analisis data yang diajukan analisis kuantitatif yang digunakan untuk mengolah data hasil penelitian, dalam hal ini menganalisis data, meliputi tingkat kesukaran soal, daya pembeda soal, pengecoh soal, reliabilitas soal tes, dan kesalahan baku pengukuran soal tes. Tingkat Kesukaran Yang dimaksud dengan tingkat kesukaran soal adalah seberapa mudah atau seberapa sulitnya suatu butir soal bagi sekelompok siswa. Makin besar siswa yang menjawab benar, makin mudah butir soal itu. Rumus yang digunakan untuk mencari tingkat kesukaran adalah:

Mengukur seberapa sulit atau mudah suatu butir soal bagi siswa. Rumus yang digunakan adalah

$$P = \frac{B}{N} \quad (\text{Arikunto 2019})$$

Di mana:

**P** = Indeks kesukaran (tingkat kesukaran)

**B** = Banyaknya siswa yang menjawab soal dengan benar

**N** = Jumlah seluruh siswa yang mengikuti tes Untuk penafsiran tingkat kesukaran soal diklasifikasikan sebagai berikut:

### Nilai P Kategori

0,00 – 0,30 Soal Sukar

0,31 – 0,70 Soal Sedang

0,71 – 1,00 Soal Mudah

### Daya Beda (D)

Yang dimaksud dengan daya pembeda adalah kemampuan suatu soal untuk membedakan siswa yang pandai dengan siswa yang kurang pandai. Analisis daya pembeda.

Mengukur kemampuan suatu butir soal untuk membedakan siswa kelompok atas (pandai) dan kelompok bawah (kurang pandai). Rumus yang digunakan adalah

$$r_{pbis} = \frac{Mp - Mt}{S} \cdot \sqrt{\frac{p}{q}}$$

Dimana :

$r_{pbis}$  : Koefisien korelasi biserial

$Mp$  : rata skor dari subyek yang menjawab benar bagi item yang cari                    daya pembedanya

$Mt$  : rata-rata skor total

$S$  : standar deviasi dari skor total

$p$  : proporsi peserta yang menjawab benar

$P$  :  $\frac{\text{Banyaknya peserta yang Benar}}{\text{Jumlah seluruh Peserta}}$

$q$  : proporsi peserta yang menjawab salah

$q$  :  $(1-p)$

Untuk mengadakan interpretasi mengenai besarnya koefisien korelasi adalah sebagai berikut:

$0,80 \leq y_{pb} \leq 1,00$  Sangat Tinggi

$0,80 \leq y_{pb} < 0,60$  Tinggi

$0,40 \leq y_{pb} < 0,60$  Cukup

$0,20 \leq y_{pb} < 0,40$  Rendah

$0,00 \leq y_{pb} < 0,20$  Sangat Rendah

### Efektivitas Pengecoh

Menganalisis proporsi siswa yang memilih setiap pilihan jawaban, termasuk pilihan pengecoh. Pengecoh yang baik adalah pengecoh yang dipilih oleh setidaknya 5% dari siswa kelompok bawah, dan lebih sedikit dipilih oleh siswa kelompok atas.

Tes dikatakan valid karena peneliti menggunakan tes yang dibuat oleh guru mata pelajaran kelas 3 SD Advent Nabire pada ulangan semester. Reliabilitas suatu tes adalah untuk menetapkan taraf ketelitian soal tes, bila digunakan untuk mengukur seseorang. Reliabilitas berhubungan dengan masalah kepercayaan suatu tes. Suatu tes dikatakan dipercaya jika memberikan hasil yang tetap meskipun diteskan berkali-kali. Ketelitian itu berlaku untuk setiap orang yang diukur dengan alat ukur yang sama.

Untuk mencari reliabilitas digunakan rumus:

$$R_{11} = \left( \frac{K}{K-1} \right) \left( 1 - \frac{\sum p_i q_i}{\sigma_t^2} \right)$$

(Arikunto 2019)

Dimana :

**R<sub>11</sub>** : Koefisien reliabilitas instrumen (reliabilitas yang dicari).

k : Jumlah butir soal (item) dalam instrumen.

$\sum p_i q_i$  : Jumlah dari hasil perkalian proporsi jawaban benar (pi) dan proporsi jawaban salah (qi) untuk setiap butir soal.

pi : Proporsi subjek yang menjawab benar pada butir ke-i.

qi : Proporsi subjek yang menjawab salah pada butir ke-i ( $qi=1-pi$ ).

$\sigma_t^2$  : Varians total dari skor seluruh subjek.

Untuk menafsirkan Interpretasi Koefisien Reliabilitas ( $r_{11}$ ): soal diklasifikasikan sebagai berikut :

0.80–1.00: Sangat tinggi (excellent)

0.60–0.79: Tinggi (good)

0.40–0.59: Cukup (fair)

0.20–0.39: Rendah (poor)

0.00–0.19: Sangat rendah (very poor) atau tidak reliabel

"Reliabilitas instrumen soal tes menunjukkan

nilai Alpha Cronbach sebesar 0,799, berada dalam kategori tinggi ( $0.70 \leq r_{11} < 0.90$ ), sehingga instrumen tersebut dapat dikatakan handal (reliable). Untuk mengetahui selisih antara hasil pengukuran sebenarnya dengan hasil pengukuran yang diperoleh, digunakan Standar Kesalahan Pengukuran (SEM).

Menurut teori tes klasik, SEM dihitung menggunakan rumus:

$$SEM = S_x \sqrt{1 - \alpha}$$

di mana SEM adalah standar kesalahan pengukuran, SX adalah standar deviasi skor, dan  $\alpha$  adalah koefisien reliabilitas instrumen (Nitko & Brookhart, Azwar 2022).

Berdasarkan perhitungan, nilai SEM sebesar 1,967. Hal ini menunjukkan bahwa skor tes yang diperoleh masing-masing siswa memiliki kesalahan pengukuran sebesar 1,967. Dengan kata lain, jika tes yang sama diberikan kembali, skor siswa cenderung akan berada dalam rentang  $\pm 1,967$  dari skor yang diperoleh."

## HASIL PENELITIAN DAN PEMBAHASAN

### Hasil Penelitian

Berdasarkan analisis menggunakan program ITEMEN terhadap 20 butir soal pilihan ganda mata pelajaran Agama dan Budi Pekerti Kelas III SD Advent Nabire, diperoleh temuan sebagai berikut:

#### Analisis Tingkat Kesukaran

**Tabel 1. Distribusi Tingkat Kesukaran Butir Soal**

| Kateg<br>ori<br>soal | Juml<br>ah | Perse<br>ntase | Nomor Butir Soal               |
|----------------------|------------|----------------|--------------------------------|
| Sukar                | -          | -              | -                              |
| Sedang               | 13         | 65             | 2,4,6,7,8,10,14,16,17,18,19,20 |
| Mudah                | 7          | 35             | 1,3,5,11,12,13,15              |
| Total                | 20         | 100            |                                |

Berdasarkan tabel 1 Tingkat kesukaran soal

Analisis 20 butir soal pilihan ganda Agama dan Budi Pekerti Kelas III SD Advent Nabire menunjukkan dominasi soal berkategori sedang (65%), yang merupakan kondisi ideal karena efektif membedakan tingkat penguasaan materi siswa. Meskipun demikian, sepertiga soal (35%) tergolong mudah. Soal-soal ini, meski dapat memotivasi siswa, berpotensi kurang efektif mengukur pemahaman mendalam atau

kemampuan berpikir tingkat tinggi, mungkin karena konsep yang terlalu sederhana atau pengecoh yang lemah. Tidak adanya soal sangat sukar (0%) adalah indikasi positif bahwa materi telah diajarkan dengan baik dan soal tidak terlalu kompleks. Secara keseluruhan, distribusi tingkat kesukaran ini cukup baik, namun disarankan agar guru meninjau dan merevisi soal-soal yang mudah. Tujuannya adalah untuk meningkatkan tantangan atau memastikan soal tersebut memang menguji konsep dasar esensial, dengan fokus pada perbaikan redaksi, kualitas pengecoh, dan relevansi konteks soal.

### Analisis Daya Beda

Tabel 2. Distribusi Daya Beda Butir Soal

| Kategori soal | Jumlah    | Percentase |
|---------------|-----------|------------|
| Sangat Tinggi | 1         | 5          |
| Tinggi        | 2         | 10         |
| Cukup         | 4         | 20         |
| Rendah        | 8         | 40         |
| Sangat Rendah | 5         | 25         |
|               | <b>20</b> | <b>100</b> |

Berdasarkan Tabel 2, Daya Pembeda Soal

Analisis daya pembeda pada 20 butir soal pilihan ganda Agama dan Budi Pekerti Kelas III SD Advent Nabire menunjukkan bahwa sebagian besar soal belum efektif dalam membedakan kemampuan siswa. Hanya 1 butir soal (5%) memiliki daya pembeda sangat tinggi dan 2 butir soal (10%) tergolong tinggi. Sementara itu, 4 butir soal (20%) cukup membedakan, sisanya justru didominasi oleh masalah serius: 8 butir (40%) berdaya pembeda rendah dan 5 butir (25%) sangat rendah. Proporsi mayoritas soal dengan daya pembeda rendah hingga sangat rendah ini (total 65%) menandakan bahwa sebagian besar butir soal gagal memisahkan antara siswa yang menguasai materi dengan baik dan yang belum. Kondisi ini mengindikasikan perlunya revisi menyeluruh pada butir-butir soal bermasalah untuk meningkatkan akurasi pengukuran dan validitas tes.

### Analisis Efektivitas Pengecoh

Tabel 3. Analisis Efektivitas Pengecoh

| Kateg<br>ori<br>soal | Jum<br>lah<br>soal | Persen<br>tase | Nomor<br>Butir soal                         |
|----------------------|--------------------|----------------|---|
| Baik                 | 17                 | 85             | 1,2,3,4,5,7,8,9,10,11,12,<br>13,14,17,18,19 |
| Tida<br>k<br>Baik    | 3                  | 15             | 6,15,20                                     |

| Total | 20 | 100 |
|-------|----|-----|
|-------|----|-----|

Berdasarkan Tabel 3 Efektivitas pengecoh

Pada butir soal pilihan ganda Agama dan Budi Pekerti Kelas III semester II SD Advent Nabire menunjukkan hasil yang sangat positif. Sebanyak 17 butir soal (85%) memiliki pengecoh yang berfungsi dengan baik, artinya pilihan jawaban yang salah mampu menarik minat siswa yang tidak memahami materi. Ini mengindikasikan bahwa pengecoh dirancang secara relevan dan plausibel, sehingga soal benar-benar menguji pemahaman dan bukan sekadar keberuntungan. Meskipun demikian, 3 butir soal (15%) masih memiliki pengecoh yang tidak efektif. Pada soal-soal ini, pengecoh mungkin terlalu jelas salah atau tidak ~~menarik minat~~ tidak berhasil menarik perhatian siswa bahkan dari kelompok bawah. Secara keseluruhan, kualitas pengecoh butir soal ujian ini sangat baik, dengan hanya sedikit butir soal yang memerlukan peninjauan ulang untuk memperbaiki desain pengecohnya agar lebih menantang dan berfungsi optimal.

### PEMBAHASAN

Berdasarkan analisis butir instrumen pilihan ganda mata pelajaran Agama dan Budi Pekerti Kelas III semester II SD Advent Nabire, dapat disimpulkan bahwa kualitas butir soal secara umum cukup baik, namun masih memerlukan beberapa penyesuaian untuk mencapai kondisi ideal sesuai pedoman Depdikbud (1985, p.37) serta prinsip-prinsip psikometris.

Dari segi tingkat kesukaran, hasil analisis menunjukkan dominasi soal sedang (65%) Nomor (2,4,6,7,8,8,10,14,16,17,18,19,20) yang sedikit melebihi proporsi ideal 50%. Sementara itu, 35% soal tergolong mudah Nomor (1,3,5,11,12,13,15) lebih tinggi dari rekomendasi 30%. Yang paling menonjol adalah tidak adanya soal sukar (0%), padahal pedoman Depdikbud menganjurkan 20%. Kondisi ini mengindikasikan bahwa tes mungkin kurang mampu membedakan secara tajam antara siswa yang sangat unggul dengan yang unggul, serta kurang memberikan tantangan yang cukup.

Pada aspek daya pembeda, ditemukan

bahwa sebagian besar soal (65%) Nomor memiliki daya pembeda rendah nomor (2,7,10,11,15,16,17,25) hingga sangat rendah. Nomor ( 6,8,13,14,18) hanya 5% daya pembeda sangat tinggi nomor (1) Ini adalah temuan krusial yang menunjukkan bahwa butir-butir soal tersebut belum efektif dalam membedakan siswa yang menguasai materi dengan yang tidak. Meskipun terdapat beberapa soal dengan daya pembeda tinggi dan cukup, proporsi yang dominan ini mengisyaratkan adanya masalah serius dalam konstruksi soal, seperti ambiguitas, pilihan jawaban yang terlalu homogen, atau materi yang belum sepenuhnya dipahami oleh kelompok atas.

Di sisi lain, analisis efektivitas pengecoh menunjukkan hasil yang sangat positif, di mana 85% pengecoh Nomor (1,2,3,4,5,7,8,9,10,11,12,13,14,17,18,19) berfungsi dengan baik. Ini menandakan bahwa sebagian besar pilihan jawaban pengecoh dirancang secara relevan dan mampu menarik perhatian siswa yang tidak mengetahui jawaban benar. Hanya sebagian kecil (15%) Nomor (6,15,20) pengecoh yang perlu diperbaiki .

Secara keseluruhan, meskipun efektivitas pengecoh sudah sangat baik, masalah utama terletak pada daya pembeda soal yang mayoritas rendah, ditambah dengan ketidadaan soal sukar dan proporsi soal mudah yang sedikit berlebihan. Untuk mencapai kualitas butir soal yang optimal sesuai standar Depdikbud dan prinsip evaluasi, revisi menyeluruh sangat diperlukan, khususnya pada butir soal dengan daya pembeda rendah dan sangat rendah. Guru perlu meninjau ulang konstruksi butir soal, memastikan kesesuaian tingkat kesukaran dengan kompetensi yang diukur, serta meningkatkan kompleksitas soal untuk menciptakan rentang kemampuan yang lebih luas.

## KESIMPULAN

Berdasarkan analisis komprehensif terhadap 20 butir soal pilihan ganda mata pelajaran Agama dan Budi Pekerti Kelas III SD Advent

Nabire, dapat disimpulkan bahwa kualitas butir soal secara umum berada pada kategori cukup baik, namun masih memiliki celah signifikan yang memerlukan perbaikan. Distribusi tingkat kesukaran menunjukkan dominasi soal sedang (65%) dan proporsi soal mudah (35%) yang melebihi rekomendasi, sementara tidak ada soal sukar (0%). Hal ini mengindikasikan bahwa tes mungkin belum cukup menantang untuk mengukur rentang kemampuan siswa secara optimal. Masalah paling krusial ditemukan pada daya pembeda, di mana mayoritas butir soal (65%) memiliki daya pembeda rendah hingga sangat rendah, menandakan ketidakmampuan soal dalam membedakan siswa yang menguasai materi dengan yang tidak. Kendati demikian, aspek efektivitas pengecoh menunjukkan kualitas yang sangat baik (85%), dengan sebagian besar pilihan pengecoh berfungsi sebagaimana mestinya. Secara keseluruhan, butir soal memerlukan revisi substansial pada aspek daya pembeda dan distribusi tingkat kesukaran agar dapat menjadi instrumen evaluasi yang lebih valid dan reliabel.

## SARAN

Mengacu pada temuan penelitian ini, disarankan beberapa langkah strategis untuk meningkatkan kualitas instrumen evaluasi di SD Advent Nabire:1)Revisi Butir Soal: Guru diharapkan untuk segera merevisi butir-butir soal, khususnya yang memiliki daya pembeda rendah hingga sangat rendah. Fokus perbaikan harus mencakup kejelasan redaksi, relevansi pilihan jawaban, dan peningkatan kompleksitas untuk mengubah soal mudah menjadi sedang atau sukar. 2)Peningkatan Kapasitas Guru: Penting untuk memberikan pelatihan atau lokakarya berkelanjutan bagi guru dalam penyusunan butir soal yang memenuhi standar psikometris, terutama dalam merumuskan indikator soal, membuat soal dengan tingkat kesukaran bervariasi, dan menciptakan pengecoh yang efektif.3)Pengembangan Bank Soal: Sekolah dapat membangun bank soal yang terstandardisasi, di mana setiap butir soal telah

melalui proses analisis dan validasi. Ini akan memudahkan guru dalam memilih soal berkualitas untuk evaluasi selanjutnya.4) Analisis Berkala: Lakukan analisis butir soal secara rutin setelah setiap ujian untuk memantau kualitas soal yang digunakan dan mengidentifikasi area yang memerlukan perbaikan berkelanjutan.

#### DAFTAR PUSTAKA

Arikunto, S. (2019). Dasar-Dasar Evaluasi Pendidikan. Bumi Aksara.

Azwar, Saifuddin. (2016). Dasar-Dasar Psikometri. Yogyakarta: Pustaka Pelajar.

Anastasi, A., & Urbina, S. (1997). Psychological Testing (7th ed.). Prentice Hall.

Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. Holt, Rinehart and Winston.

Depdikbud. (1985). Petunjuk Penulisan Butir Soal Tes. Jakarta: Departemen Pendidikan dan Kebudayaan

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theory. Sage Publications.

Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2022). Keputusan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia

Nomor 56/M/2022 tentang Pedoman Penerapan Kurikulum dalam Rangka Pemulihan Pembelajaran. Jakarta: Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

Mardapi, D. (2012). Pengukuran, Penilaian, dan Evaluasi Pendidikan. Nuha Medika.

Nitko, A. J., & Brookhart, S. M. (2022). Educational Assessment of Students (8th ed.). Pearson.

Popham, W. J. (2005). *Classroom Assessment: What Teachers Need to Know* (4th ed.). Pearson Education.

Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian*. Parama Publishing.

Sudijono, Anas. (2011). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.

Suharsimi. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

Sudijono, A. (2017). *Pengantar Evaluasi Pendidikan*. PT Raja Grafindo Persada.