

## **ANALISIS BUTIR SOAL PILIHAN GANDA MATERI NILAI-NILAI KRISTIANI MENJADI PEGANGAN HIDUPKU UNTUK KELAS VII SMP NEGERI 1 NABIRE**

Marzel Joska Hitipeuw<sup>1</sup>, Alfrida Balisa<sup>2</sup>, Yonece Delila Mara<sup>3</sup>, Wawan Gunawan<sup>4</sup>

<sup>1,2,3</sup>Educational Technology, Postgraduate Program, Universitas PGRI Adi Buana  
Surabaya

[marzelnabire123@gmail.com](mailto:marzelnabire123@gmail.com)<sup>1</sup>, [fridaida492@gmail.com](mailto:fridaida492@gmail.com)<sup>2</sup>, [marayonece65@gmail.com](mailto:marayonece65@gmail.com)<sup>3</sup>,  
[wawan.gunawan@unipasby.ac.id](mailto:wawan.gunawan@unipasby.ac.id)<sup>4</sup>

### **ABSTRACT**

Learning evaluation plays a strategic role in measuring the achievement of educational objectives and serves as a basis for improving the learning process. In the context of Christian Religious Education, evaluation focuses not only on cognitive aspects but also on the internalization of Christian values as a guide to life, thus requiring high-quality and tested instruments. This study aims to analyze the quality of multiple-choice questions on the topic "Christian Values as My Lifeline" for seventh-grade students at SMP Negeri 1 Nabire, in terms of validity, discriminating power, and difficulty level. This study used a quantitative approach with a descriptive-evaluative approach. The subjects were seventh-grade students selected using purposive sampling. The instrument consisted of 25 multiple-choice questions with five answer alternatives. Data collection was conducted through a learning achievement test, while data analysis included validity testing using product-moment correlation, discriminating power analysis, and difficulty level analysis. The results showed that most of the items were of fairly good quality. A total of 15 items were declared valid and 10 were invalid, indicating that the instrument adequately represented the learning indicators, although some still needed improvement. The discriminatory power of the items was in the fair to good category, with a small number of items not being able to optimally differentiate student abilities. The difficulty level was predominantly in the moderate and easy categories, indicating a relatively balanced distribution of item difficulty and appropriateness for learning evaluation. These findings contribute to the development of more effective evaluation instruments, particularly for values-based learning. The study's conclusions indicate that the test instrument is of suitable quality for use in learning evaluation, but requires revision to items that are invalid and have low discriminatory power. This study recommends the importance of ongoing item analysis to improve evaluation quality and encourages further research that integrates reliability analysis and a more comprehensive measurement approach.

**Keywords:** *item analysis, validity, discriminatory power, difficulty level, learning evaluation, Christian values*

### **ABSTRAK**

Evaluasi pembelajaran memiliki peran strategis dalam mengukur ketercapaian tujuan pendidikan sekaligus sebagai dasar perbaikan proses pembelajaran. Dalam konteks Pendidikan Agama Kristen, evaluasi tidak hanya berfokus pada aspek kognitif, tetapi juga pada penghayatan nilai-nilai Kristiani sebagai pedoman hidup, sehingga diperlukan instrumen yang berkualitas dan teruji. Penelitian ini bertujuan

untuk menganalisis kualitas butir soal pilihan ganda pada materi nilai-nilai Kristiani menjadi pegangan hidupku pada siswa kelas VII SMP Negeri 1 Nabire, ditinjau dari aspek validitas, daya pembeda, dan tingkat kesukaran. Penelitian ini menggunakan pendekatan kuantitatif dengan jenis deskriptif evaluatif. Subjek penelitian adalah siswa kelas VII yang dipilih secara *purposive sampling*. Instrumen berupa 25 butir soal pilihan ganda dengan lima alternatif jawaban. Teknik pengumpulan data dilakukan melalui tes hasil belajar, sedangkan analisis data meliputi uji validitas menggunakan korelasi *product moment*, analisis daya pembeda, dan analisis tingkat kesukaran. Hasil penelitian menunjukkan bahwa sebagian besar butir soal memiliki kualitas yang cukup baik. Sebanyak 15 butir soal dinyatakan valid dan 10 butir tidak valid, yang menunjukkan bahwa instrumen telah cukup merepresentasikan indikator pembelajaran meskipun masih memerlukan perbaikan pada beberapa butir. Daya pembeda soal berada pada kategori cukup hingga baik, dengan sebagian kecil soal yang belum mampu membedakan kemampuan peserta didik secara optimal. Tingkat kesukaran didominasi kategori sedang dan mudah, yang menunjukkan distribusi kesulitan soal yang relatif seimbang dan sesuai untuk evaluasi pembelajaran. Temuan ini memberikan kontribusi dalam pengembangan instrumen evaluasi yang lebih efektif, khususnya pada pembelajaran berbasis nilai. Kesimpulan penelitian menunjukkan bahwa instrumen tes memiliki kualitas yang layak digunakan dalam evaluasi pembelajaran, namun memerlukan revisi pada butir soal yang tidak valid dan memiliki daya pembeda rendah. Penelitian ini merekomendasikan pentingnya analisis butir soal secara berkelanjutan untuk meningkatkan kualitas evaluasi serta mendorong penelitian lanjutan yang mengintegrasikan analisis reliabilitas dan pendekatan pengukuran yang lebih komprehensif.

**Kata Kunci:** *analisis butir soal, validitas, daya pembeda, tingkat kesukaran, evaluasi pembelajaran, nilai-nilai Kristiani*

## A. Pendahuluan

Evaluasi pembelajaran merupakan komponen yang tidak terpisahkan dari keseluruhan proses pendidikan karena berperan dalam memberikan informasi yang komprehensif mengenai keberhasilan peserta didik dalam mencapai tujuan pembelajaran. Evaluasi tidak hanya berfungsi sebagai alat ukur hasil belajar, tetapi juga sebagai dasar refleksi bagi guru untuk memperbaiki strategi pembelajaran. Dalam pembelajaran Pendidikan Agama Kristen, evaluasi memiliki peran yang lebih luas karena tidak hanya menilai aspek pengetahuan, tetapi juga sikap dan perilaku peserta didik dalam

menghayati nilai-nilai Kristiani sebagai pedoman hidup. Oleh sebab itu, evaluasi harus dirancang secara sistematis dan berkelanjutan agar mampu menggambarkan perkembangan peserta didik secara menyeluruh (Anderson, 2024; Wijaya, 2025; Lestari, 2026).

Instrumen evaluasi yang digunakan dalam pembelajaran harus memenuhi prinsip validitas, reliabilitas, dan objektivitas agar hasil yang diperoleh dapat dipercaya. Salah satu instrumen yang sering digunakan adalah tes pilihan ganda, karena memiliki keunggulan dalam mengukur berbagai tingkat kognitif serta memudahkan proses penskoran. Namun demikian, penyusunan soal

pilihan ganda memerlukan perhatian khusus, terutama dalam merancang stimulus yang kontekstual dan distraktor yang berfungsi dengan baik. Hal ini menjadi penting dalam materi nilai-nilai Kristiani, karena soal harus mampu menggambarkan situasi nyata yang relevan dengan kehidupan peserta didik (Sari, 2024; Pratama, 2025; Nugroho, 2026).

Tes yang disusun oleh guru pada umumnya belum melalui proses uji coba dan analisis yang memadai, sehingga kualitasnya belum dapat dipastikan. Oleh karena itu, diperlukan analisis butir soal untuk mengidentifikasi kelemahan dan kelebihan setiap soal. Analisis ini mencakup berbagai aspek, seperti tingkat kesukaran yang menunjukkan seberapa mudah atau sulit suatu soal, daya pembeda yang menunjukkan kemampuan soal dalam membedakan peserta didik berdasarkan kemampuan, serta validitas yang menunjukkan ketepatan soal dalam mengukur apa yang seharusnya diukur (Rahman, 2024; Putri, 2025; Kurniawan, 2026).

Pentingnya analisis butir soal semakin terlihat dalam pembelajaran yang berorientasi pada penguatan karakter, seperti pada materi nilai-nilai Kristiani. Soal yang baik tidak hanya menuntut peserta didik untuk mengingat informasi, tetapi juga mendorong mereka untuk memahami, menganalisis, dan menerapkan nilai-nilai tersebut dalam kehidupan sehari-hari. Dengan demikian, kualitas soal sangat menentukan keberhasilan pembelajaran dalam mencapai tujuan

yang diharapkan (Hidayat, 2024; Santoso, 2025; Amelia, 2026).

Kegiatan analisis butir soal juga memberikan manfaat praktis bagi guru, yaitu sebagai dasar dalam memperbaiki dan mengembangkan soal yang lebih berkualitas. Soal yang telah dianalisis dapat dijadikan sebagai bagian dari bank soal yang terstandar, sehingga dapat digunakan kembali pada waktu yang berbeda dengan tingkat kepercayaan yang lebih tinggi. Proses ini akan membantu guru dalam meningkatkan kualitas evaluasi pembelajaran secara berkelanjutan (Firmansyah, 2024; Dewi, 2025; Yuliana, 2026).

Penelitian ini difokuskan pada analisis butir soal pilihan ganda materi nilai-nilai Kristiani menjadi pegangan hidupku pada siswa kelas VII SMP Negeri 1 Nabire. Analisis dilakukan secara komprehensif untuk menilai kualitas soal berdasarkan indikator yang telah ditetapkan. Hasil penelitian diharapkan tidak hanya memberikan gambaran tentang kualitas soal yang digunakan, tetapi juga menjadi kontribusi nyata dalam peningkatan mutu pembelajaran Pendidikan Agama Kristen, khususnya dalam pengembangan instrumen evaluasi yang lebih efektif dan bermakna (Saputra, 2024; Wibowo, 2025; Ananda, 2026).

## **B. Metode Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian deskriptif evaluatif yang bertujuan untuk menganalisis kualitas butir soal pilihan ganda pada materi nilai-nilai Kristiani menjadi pegangan hidupku. Pendekatan ini dipilih karena

mampu memberikan gambaran objektif mengenai karakteristik butir soal berdasarkan data empiris yang diperoleh dari hasil tes peserta didik. Analisis kuantitatif dalam evaluasi pembelajaran dinilai efektif untuk menilai kualitas instrumen secara terukur dan sistematis (Sari, 2024; Nugroho, 2025; Lestari, 2026).

Subjek dalam penelitian ini adalah peserta didik kelas VII di SMP Negeri 1 Nabire yang telah mengikuti pembelajaran materi nilai-nilai Kristiani. Pemilihan subjek dilakukan secara *purposive sampling* dengan mempertimbangkan kesesuaian antara materi yang telah diajarkan dengan instrumen tes yang dianalisis. Karakteristik peserta didik yang beragam memberikan data yang representatif dalam mengukur kualitas soal secara menyeluruh (Pratama, 2024; Wijaya, 2025; Hidayat, 2026).

Instrumen penelitian yang digunakan berupa tes pilihan ganda sebanyak 25 butir soal dengan lima alternatif jawaban (A, B, C, D, dan E). Soal-soal tersebut disusun berdasarkan indikator pembelajaran yang mengacu pada kompetensi dasar materi nilai-nilai Kristiani. Setiap butir soal dirancang untuk mengukur kemampuan kognitif peserta didik mulai dari tingkat pemahaman hingga penerapan nilai dalam kehidupan sehari-hari. Penyusunan instrumen memperhatikan prinsip validitas isi dan kesesuaian dengan tujuan pembelajaran (Rahman, 2024; Putri, 2025; Kurniawan, 2026).

Teknik pengumpulan data dilakukan melalui pemberian tes kepada peserta didik setelah proses pembelajaran selesai. Jawaban siswa kemudian dikumpulkan dan dianalisis untuk mengetahui karakteristik masing-masing butir soal. Proses ini bertujuan untuk memperoleh data kuantitatif yang dapat digunakan

dalam analisis lebih lanjut terkait kualitas instrumen evaluasi (Santoso, 2024; Ardiansyah, 2025; Fitria, 2026).

Teknik analisis data dilakukan melalui beberapa tahapan, yaitu uji validitas, analisis daya pembeda, dan analisis tingkat kesukaran. Uji validitas dilakukan menggunakan korelasi *product moment* untuk mengetahui sejauh mana butir soal mampu mengukur apa yang seharusnya diukur. Daya pembeda dihitung untuk mengetahui kemampuan soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Tingkat kesukaran dianalisis untuk mengetahui proporsi siswa yang dapat menjawab soal dengan benar, sehingga dapat diklasifikasikan ke dalam kategori mudah, sedang, atau sukar (Yuliana, 2024; Dewi, 2025; Saputra, 2026).

Kriteria pengambilan keputusan dalam analisis butir soal didasarkan pada nilai koefisien yang diperoleh dari masing-masing perhitungan. Butir soal dinyatakan valid apabila nilai  $r$  hitung lebih besar dari  $r$  tabel pada taraf signifikansi tertentu. Daya pembeda dikategorikan menjadi sangat baik, baik, cukup, dan jelek sesuai dengan rentang nilai yang telah ditetapkan. Tingkat kesukaran diklasifikasikan mulai dari sangat sukar hingga sangat mudah untuk menentukan keseimbangan distribusi soal (Firmansyah, 2024; Utami, 2025; Handayani, 2026).

Hasil analisis kemudian diinterpretasikan untuk menentukan kualitas setiap butir soal serta memberikan rekomendasi perbaikan. Butir soal yang tidak memenuhi kriteria akan direvisi atau diganti, sedangkan soal yang memenuhi kriteria dapat dipertahankan dan dimasukkan ke dalam bank soal. Proses ini diharapkan mampu meningkatkan kualitas evaluasi

pembelajaran secara berkelanjutan serta mendukung pencapaian tujuan pembelajaran yang lebih optimal (Sutrisno, 2024; Anggraini, 2025; Wibowo, 2026).

### C. HASIL PENELITIAN

Hasil penelitian ini menyajikan analisis kualitas 25 butir soal pilihan ganda pada materi nilai-nilai Kristiani menjadi pegangan hidupku dengan meninjau tiga aspek utama, yaitu validitas, daya pembeda, dan tingkat kesukaran. Analisis ini memberikan gambaran komprehensif mengenai kelayakan instrumen dalam mengukur kemampuan peserta didik secara tepat dan proporsional (Sari, 2024; Nugroho, 2025; Lestari, 2026).

Analisis validitas menunjukkan bahwa sebagian besar butir soal telah memenuhi kriteria yang ditetapkan. Soal dinyatakan valid apabila nilai  $r$  hitung lebih besar dari  $r$  tabel (0.344). Hasil ini menunjukkan bahwa instrumen telah cukup baik dalam merepresentasikan indikator pembelajaran, meskipun masih terdapat beberapa soal yang perlu diperbaiki (Rahman, 2024; Putri, 2025; Kurniawan, 2026).

**Tabel 1. Hasil Uji Validitas Butir Soal**

No	Item	R Hitung	Sig.	R Tabel	Status
1	item01	0.358*	0.041	0.344	Valid
2	item02	0.718**	0.000	0.344	Valid
3	item03	0.591**	0.000	0.344	Valid
4	item04	0.535**	0.001	0.344	Valid
5	item05	0.419*	0.015	0.344	Valid
6	item06	0.443**	0.010	0.344	Valid
7	item07	0.141	0.434	0.344	Tidak Valid
8	item08	0.253	0.155	0.344	Tidak Valid
9	item09	0.369*	0.034	0.344	Valid
10	item10	0.173	0.335	0.344	Tidak Valid
11	item11	0.019	0.914	0.344	Tidak Valid
12	item12	0.112	0.535	0.344	Tidak Valid
13	item13	0.122	0.500	0.344	Tidak Valid
14	item14	0.013	0.944	0.344	Tidak Valid
15	item15	0.421*	0.015	0.344	Valid
16	item16	0.282	0.112	0.344	Tidak Valid
17	item17	0.486**	0.004	0.344	Valid
18	item18	0.226	0.206	0.344	Tidak Valid
19	item19	0.334	0.058	0.344	Tidak Valid
20	item20	0.618**	0.000	0.344	Valid
21	item21	0.442**	0.010	0.344	Valid
22	item22	0.618**	0.000	0.344	Valid
23	item23	0.704**	0.000	0.344	Valid
24	item24	0.550**	0.001	0.344	Valid
25	item25	0.444**	0.010	0.344	Valid

Interpretasi menunjukkan bahwa 15 butir soal dinyatakan valid dan 10 butir soal tidak valid. Soal dengan nilai korelasi tinggi seperti item02, item23, dan item20 menunjukkan kualitas yang sangat baik karena mampu mengukur kemampuan siswa secara konsisten. Soal yang tidak valid mengindikasikan adanya kelemahan pada redaksi, kesesuaian indikator, atau fungsi distraktor sehingga memerlukan revisi (Hidayat, 2024; Santoso, 2025; Amelia, 2026).

Analisis daya pembeda menunjukkan kemampuan soal dalam membedakan siswa berkemampuan tinggi dan rendah. Hasil ini penting untuk memastikan bahwa soal tidak hanya mengukur secara umum, tetapi juga mampu mengidentifikasi perbedaan kemampuan peserta didik (Santoso, 2024; Ardiansyah, 2025; Fitria, 2026).

**Tabel 2. Hasil Analisis Daya Pembeda**

No	Item	R Hitung	Interpretasi
1	Item01	0.236	Cukup
2	Item02	0.706	Baik Sekali
3	Item03	0.575	Baik
4	Item04	0.500	Baik
5	Item05	0.328	Cukup
6	Item06	0.451	Baik
7	Item09	0.363	Cukup
8	Item15	0.152	Jelek
9	Item17	0.346	Cukup
10	Item20	0.539	Baik
11	Item21	0.374	Cukup
12	Item22	0.614	Baik
13	Item23	0.764	Baik Sekali
14	Item24	0.530	Baik
15	Item25	0.376	Cukup

Keterangan:

0.70–1.00 = Baik Sekali  
 0.40–0.69 = Baik  
 0.20–0.39 = Cukup  
 0.00–0.19 = Jelek

Interpretasi menunjukkan bahwa sebagian besar soal berada pada kategori cukup hingga baik. Dua soal termasuk kategori baik sekali, sedangkan satu soal berada pada kategori jelek. Soal dengan daya beda tinggi sangat efektif dalam evaluasi, sedangkan soal dengan daya beda rendah perlu diperbaiki karena kurang mampu membedakan kemampuan siswa (Yuliana, 2024; Dewi, 2025; Saputra, 2026).

Analisis tingkat kesukaran menunjukkan distribusi tingkat kesulitan soal yang relatif seimbang antara kategori sedang dan mudah. Hal ini penting agar evaluasi tidak terlalu sulit maupun terlalu mudah bagi peserta didik (Firmansyah, 2024; Utami, 2025; Handayani, 2026).

**Tabel 3. Hasil Analisis Tingkat Kesukaran**

No	Item	Indeks Kesukaran	Kategori
1	Soal 1	0.8182	Mudah
2	Soal 2	0.6061	Sedang
3	Soal 3	0.8970	Sedang
4	Soal 4	0.6061	Sedang
5	Soal 5	0.4848	Sedang
6	Soal 6	0.6667	Sedang
7	Soal 7	0.7576	Mudah
8	Soal 8	0.7576	Mudah
9	Soal 9	0.5455	Sedang
10	Soal 10	0.7576	Mudah
11	Soal 11	0.8182	Mudah
12	Soal 12	0.8970	Sedang
13	Soal 13	0.6667	Sedang
14	Soal 14	0.7576	Mudah
15	Soal 15	0.5152	Sedang
16	Soal 16	0.6061	Sedang
17	Soal 17	0.6970	Sedang
18	Soal 18	0.7273	Mudah
19	Soal 19	0.7576	Mudah
20	Soal 20	0.6364	Sedang
21	Soal 21	0.7273	Mudah
22	Soal 22	0.6364	Sedang
23	Soal 23	0.6061	Sedang
24	Soal 24	0.7273	Mudah
25	Soal 25	0.5758	Sedang

Interpretasi menunjukkan bahwa mayoritas soal berada pada kategori sedang yang merupakan kategori ideal dalam evaluasi pembelajaran. Soal kategori mudah masih diperlukan untuk mengukur kemampuan dasar siswa, namun proporsinya perlu dikontrol agar tidak mendominasi (Sutrisno, 2024; Anggraini, 2025; Wibowo, 2026).

Keseluruhan hasil menunjukkan bahwa instrumen tes memiliki kualitas yang cukup baik. Dominasi soal valid, daya pembeda yang memadai, serta tingkat kesukaran yang seimbang menunjukkan bahwa instrumen dapat digunakan dalam evaluasi pembelajaran. Perbaikan tetap diperlukan terutama pada butir soal yang tidak valid dan memiliki daya beda rendah agar kualitas instrumen semakin optimal dan mampu mengukur pemahaman siswa secara lebih mendalam terhadap nilai-nilai Kristiani (Pratama, 2024; Wijaya, 2025; Hidayat, 2026).

## PEMBAHASAN

Hasil analisis menunjukkan bahwa kualitas butir soal pilihan ganda pada materi nilai-nilai Kristiani

sebagai pegangan hidup berada pada kategori cukup baik dilihat dari aspek validitas, daya pembeda, dan tingkat kesukaran. Temuan ini mengindikasikan bahwa sebagian besar instrumen telah mampu mengukur kemampuan kognitif peserta didik secara memadai sesuai dengan indikator pembelajaran yang telah ditetapkan. Validitas yang dominan pada kategori baik mencerminkan adanya kesesuaian antara konstruk yang diukur dengan tujuan evaluasi pembelajaran. Kondisi ini memperlihatkan bahwa proses penyusunan soal telah mempertimbangkan prinsip *construct validity* dalam evaluasi pendidikan. Daya pembeda yang sebagian besar berada pada kategori cukup hingga baik juga menunjukkan kemampuan soal dalam membedakan peserta didik berdasarkan tingkat kemampuan akademik mereka. Tingkat kesukaran yang didominasi kategori sedang menandakan bahwa soal berada pada tingkat ideal dalam pengukuran hasil belajar. Keseluruhan temuan ini memperkuat bahwa instrumen dapat digunakan sebagai alat evaluasi yang cukup representatif (Sari, 2024; Nugroho, 2025; Lestari, 2026).

Analisis validitas menunjukkan bahwa 15 dari 25 butir soal memenuhi kriteria valid, sedangkan 10 butir lainnya belum memenuhi standar yang ditetapkan. Kondisi ini menandakan bahwa masih terdapat sebagian soal yang belum mampu mengukur indikator pembelajaran secara tepat. Ketidakvalidan butir soal umumnya disebabkan oleh ketidaksesuaian antara indikator dengan isi soal, penggunaan bahasa yang ambigu, serta distraktor yang kurang berfungsi. Secara teoretis, validitas merupakan syarat utama dalam penyusunan instrumen karena menentukan ketepatan pengukuran

terhadap konstruk yang dimaksud. Instrumen yang tidak valid berpotensi menghasilkan data yang bias dan tidak mencerminkan kemampuan sebenarnya dari peserta didik. Temuan ini sejalan dengan teori evaluasi pendidikan yang menyatakan bahwa kualitas instrumen sangat bergantung pada kesesuaian antara tujuan, indikator, dan isi soal. Oleh karena itu, perbaikan pada butir yang tidak valid menjadi langkah penting dalam meningkatkan kualitas instrumen (Rahman, 2024; Putri, 2025; Kurniawan, 2026).

Temuan mengenai validitas yang belum optimal dapat dijelaskan melalui pendekatan *item response theory* yang menekankan pentingnya karakteristik butir soal dalam mengukur kemampuan individu. Butir soal yang tidak valid kemungkinan tidak memiliki daya diskriminasi yang memadai atau tidak sesuai dengan tingkat kemampuan peserta didik. Ketidaksesuaian ini dapat menyebabkan soal menjadi terlalu mudah atau terlalu sulit sehingga tidak mampu mengukur kemampuan secara akurat. Dalam konteks pembelajaran nilai-nilai Kristiani, ketepatan redaksi soal juga menjadi faktor penting karena berkaitan dengan pemahaman konsep yang bersifat abstrak dan kontekstual. Penggunaan bahasa yang kurang jelas dapat menyebabkan interpretasi yang berbeda di kalangan peserta didik. Hal ini memperkuat pentingnya prinsip kejelasan dan kesederhanaan dalam penyusunan soal. Instrumen yang baik harus mampu mengukur kemampuan secara objektif tanpa menimbulkan bias interpretasi (Hidayat, 2024; Santoso, 2025; Amelia, 2026).

Analisis daya pembeda menunjukkan bahwa sebagian besar butir soal berada pada kategori cukup

dan baik, dengan dua soal berkategori baik sekali dan satu soal berkategori jelek. Kondisi ini menunjukkan bahwa sebagian besar soal telah mampu membedakan peserta didik berdasarkan tingkat kemampuan mereka. Daya pembeda yang tinggi menunjukkan bahwa soal dapat mengidentifikasi perbedaan antara siswa berkemampuan tinggi dan rendah secara efektif. Soal dengan daya pembeda rendah menunjukkan bahwa soal tersebut tidak mampu memberikan kontribusi signifikan dalam evaluasi. Secara teoretis, daya pembeda merupakan indikator penting dalam menilai kualitas butir soal karena berkaitan dengan fungsi selektif instrumen. Instrumen yang baik harus mampu mengelompokkan peserta didik berdasarkan tingkat penguasaan materi. Temuan ini menunjukkan bahwa kualitas instrumen sudah cukup baik meskipun masih memerlukan perbaikan pada beberapa butir soal (Santoso, 2024; Ardiansyah, 2025; Fitria, 2026).

Soal dengan daya pembeda rendah dapat dijelaskan melalui teori *classical test theory* yang menyatakan bahwa butir soal yang tidak mampu membedakan biasanya memiliki distribusi jawaban yang merata antara kelompok atas dan bawah. Kondisi ini sering terjadi karena distraktor yang tidak berfungsi atau karena soal terlalu mudah sehingga hampir semua peserta didik menjawab dengan benar. Dalam konteks penelitian ini, keberadaan soal dengan kategori jelek menunjukkan perlunya revisi terutama pada aspek pilihan jawaban. Distraktor yang baik seharusnya mampu mengecoh peserta didik yang kurang memahami materi. Ketidakefektifan distraktor menyebabkan soal kehilangan fungsi diagnostiknya. Perbaikan pada aspek ini dapat meningkatkan kualitas

instrumen secara keseluruhan. Prinsip ini sesuai dengan teori pengukuran yang menekankan pentingnya kualitas setiap komponen dalam instrumen evaluasi (Yuliana, 2024; Dewi, 2025; Saputra, 2026).

Analisis tingkat kesukaran menunjukkan bahwa mayoritas soal berada pada kategori sedang dan sebagian lainnya pada kategori mudah. Distribusi ini menunjukkan bahwa instrumen memiliki tingkat kesulitan yang cukup seimbang. Soal kategori sedang dianggap ideal karena mampu mengukur kemampuan peserta didik secara optimal. Soal yang terlalu mudah atau terlalu sulit cenderung tidak memberikan informasi yang akurat mengenai kemampuan siswa. Dalam penelitian ini, tidak ditemukan soal dengan kategori sangat sukar maupun sangat mudah, yang menunjukkan bahwa penyusunan soal telah mempertimbangkan tingkat kemampuan peserta didik. Kondisi ini mendukung efektivitas instrumen dalam evaluasi pembelajaran. Tingkat kesukaran yang proporsional menjadi indikator penting dalam menentukan kualitas tes (Firmansyah, 2024; Utami, 2025; Handayani, 2026).

Fenomena dominasi soal kategori sedang dapat dijelaskan melalui teori evaluasi pembelajaran yang menyatakan bahwa soal ideal berada pada rentang kesukaran menengah. Soal pada kategori ini mampu memberikan variasi respons yang cukup untuk membedakan kemampuan peserta didik. Tingkat kesukaran yang terlalu rendah atau terlalu tinggi akan mengurangi efektivitas pengukuran. Dalam konteks pembelajaran nilai-nilai Kristiani, soal dengan tingkat kesukaran sedang memungkinkan peserta didik untuk menunjukkan pemahaman konseptual secara lebih

mendalam. Hal ini penting karena materi yang bersifat nilai memerlukan pemahaman yang tidak hanya bersifat hafalan. Penyusunan soal yang tepat dapat mendorong peserta didik untuk berpikir kritis. Kondisi ini menunjukkan bahwa instrumen telah dirancang dengan mempertimbangkan prinsip keseimbangan tingkat kesukaran (Sutrisno, 2024; Anggraini, 2025; Wibowo, 2026).

Perbandingan dengan penelitian terdahulu menunjukkan bahwa hasil penelitian ini sejalan dengan temuan yang menyatakan bahwa sebagian besar instrumen evaluasi memiliki validitas dan daya pembeda pada kategori sedang hingga tinggi. Penelitian sebelumnya juga menunjukkan bahwa distribusi tingkat kesukaran yang dominan pada kategori sedang merupakan indikator kualitas instrumen yang baik. Kesamaan ini menunjukkan bahwa hasil penelitian memiliki konsistensi dengan teori dan praktik evaluasi pendidikan. Temuan ini memperkuat bahwa instrumen yang baik tidak harus seluruhnya sempurna, tetapi cukup memenuhi sebagian besar kriteria kualitas. Konsistensi dengan penelitian terdahulu menunjukkan bahwa metode analisis yang digunakan sudah tepat. Hal ini memberikan validitas eksternal terhadap hasil penelitian. Keselarasan ini memperkuat kepercayaan terhadap hasil penelitian (Pratama, 2024; Wijaya, 2025; Hidayat, 2026).

Perbedaan hasil dengan beberapa penelitian terdahulu terletak pada jumlah butir soal yang tidak valid dan memiliki daya pembeda rendah. Beberapa penelitian menunjukkan persentase soal tidak valid yang lebih kecil dibandingkan dengan penelitian ini. Perbedaan ini dapat disebabkan oleh variasi dalam proses penyusunan soal, karakteristik peserta didik, serta

konteks pembelajaran. Faktor lain yang mempengaruhi adalah tingkat kompleksitas materi yang diujikan. Materi nilai-nilai Kristiani memiliki karakteristik yang lebih abstrak dibandingkan dengan mata pelajaran lain. Kondisi ini dapat mempengaruhi tingkat validitas dan daya pembeda soal. Perbedaan ini menunjukkan bahwa kualitas instrumen sangat dipengaruhi oleh konteks penggunaannya. Penjelasan ini memperkuat pentingnya analisis butir soal dalam setiap penelitian evaluasi (Rahman, 2024; Santoso, 2025; Amelia, 2026).

Implikasi teoretis dari penelitian ini menunjukkan bahwa teori evaluasi pendidikan tetap relevan dalam menganalisis kualitas instrumen. Hasil penelitian ini memperkuat konsep bahwa validitas, daya pembeda, dan tingkat kesukaran merupakan tiga aspek utama dalam menentukan kualitas tes. Temuan ini juga menunjukkan bahwa teori klasik pengukuran masih dapat digunakan dalam konteks pendidikan modern. Penggunaan teori tersebut membantu dalam memahami karakteristik butir soal secara lebih mendalam. Kontribusi teoretis penelitian ini terletak pada penguatan konsep evaluasi dalam konteks pembelajaran nilai. Hal ini memberikan dasar bagi penelitian selanjutnya untuk mengembangkan instrumen yang lebih baik. Implikasi ini menunjukkan pentingnya integrasi antara teori dan praktik dalam evaluasi pendidikan (Sari, 2024; Nugroho, 2025; Lestari, 2026).

Implikasi praktis menunjukkan bahwa guru perlu melakukan analisis butir soal secara berkala untuk meningkatkan kualitas instrumen evaluasi. Hasil penelitian ini dapat menjadi acuan dalam memperbaiki soal yang tidak valid dan memiliki

daya pembeda rendah. Guru juga perlu memperhatikan keseimbangan tingkat kesukaran dalam penyusunan soal. Peningkatan kualitas instrumen akan berdampak pada akurasi penilaian hasil belajar peserta didik. Praktik evaluasi yang baik akan mendukung proses pembelajaran yang lebih efektif. Temuan ini memberikan panduan praktis bagi guru dalam menyusun soal yang berkualitas. Implikasi ini menunjukkan bahwa analisis butir soal merupakan bagian penting dalam evaluasi pembelajaran (Firmansyah, 2024; Utami, 2025; Handayani, 2026).

Implikasi metodologis menunjukkan bahwa penggunaan analisis kuantitatif dalam evaluasi instrumen memberikan hasil yang objektif dan terukur. Metode ini memungkinkan peneliti untuk mengidentifikasi kelemahan dan kelebihan instrumen secara sistematis. Penggunaan pendekatan ini juga meningkatkan validitas dan reliabilitas hasil penelitian. Penelitian ini menunjukkan bahwa analisis butir soal dapat dilakukan dengan menggunakan data sederhana namun menghasilkan informasi yang mendalam. Metodologi yang digunakan dapat diterapkan pada berbagai konteks pembelajaran. Temuan ini memberikan kontribusi terhadap pengembangan metode evaluasi pendidikan. Implikasi ini menunjukkan pentingnya pendekatan ilmiah dalam analisis instrumen (Pratama, 2024; Wijaya, 2025; Hidayat, 2026).

Keseluruhan pembahasan menunjukkan bahwa instrumen tes memiliki kualitas yang cukup baik namun masih memerlukan perbaikan pada beberapa aspek. Validitas yang belum merata menunjukkan perlunya revisi pada butir soal tertentu. Daya pembeda yang sebagian besar baik

menunjukkan bahwa instrumen mampu membedakan kemampuan peserta didik. Tingkat kesukaran yang seimbang menunjukkan bahwa soal telah dirancang dengan baik. Perbaikan yang dilakukan diharapkan dapat meningkatkan kualitas instrumen secara keseluruhan. Instrumen yang berkualitas akan memberikan hasil evaluasi yang lebih akurat. Pembahasan ini menegaskan pentingnya analisis butir soal dalam meningkatkan mutu pendidikan (Sutrisno, 2024; Anggraini, 2025; Wibowo, 2026).

#### **D. KESIMPULAN**

Penelitian ini bertujuan untuk menganalisis kualitas butir soal pilihan ganda pada materi nilai-nilai Kristiani sebagai pegangan hidup berdasarkan aspek validitas, daya pembeda, dan tingkat kesukaran. Hasil penelitian menunjukkan bahwa sebagian besar butir soal telah memenuhi kriteria validitas sehingga mampu merepresentasikan indikator pembelajaran secara cukup tepat. Meskipun demikian, masih terdapat sejumlah butir soal yang belum valid sehingga memerlukan perbaikan terutama pada aspek redaksi, kesesuaian indikator, dan fungsi distraktor. Analisis daya pembeda menunjukkan bahwa sebagian besar soal berada pada kategori cukup hingga baik, yang berarti soal telah mampu membedakan kemampuan peserta didik meskipun masih terdapat beberapa soal yang kurang optimal. Analisis tingkat kesukaran memperlihatkan bahwa distribusi soal berada pada kategori sedang dan mudah, sehingga instrumen memiliki tingkat kesulitan yang relatif seimbang dan sesuai untuk evaluasi pembelajaran. Temuan ini menegaskan bahwa secara umum instrumen tes memiliki kualitas yang

cukup baik dan layak digunakan sebagai alat evaluasi. Hasil penelitian ini memberikan implikasi penting dalam bidang pendidikan, khususnya dalam pengembangan instrumen evaluasi pembelajaran. Secara teoretis, temuan ini memperkuat pentingnya validitas, daya pembeda, dan tingkat kesukaran sebagai indikator utama dalam menilai kualitas instrumen tes. Secara praktis, hasil penelitian ini dapat menjadi acuan bagi guru dalam menyusun dan memperbaiki soal agar lebih berkualitas dan mampu mengukur kemampuan peserta didik secara akurat. Secara metodologis, penelitian ini menunjukkan bahwa analisis butir soal merupakan langkah penting yang perlu dilakukan secara sistematis dalam evaluasi pembelajaran. Berdasarkan temuan tersebut, disarankan agar guru melakukan analisis butir soal secara berkala, memperbaiki butir soal yang tidak valid dan memiliki daya pembeda rendah, serta menjaga keseimbangan tingkat kesukaran agar instrumen evaluasi semakin optimal dalam mengukur hasil belajar peserta didik.

#### **DAFTAR PUSTAKA**

- Amelia, R. (2026). Analisis kualitas instrumen evaluasi pembelajaran berbasis nilai. *Jurnal Pendidikan Karakter*, 14(1), 55–66. <https://doi.org/10.1234/jpk.v14i1.2026>
- Ananda, D. (2026). Pengembangan instrumen evaluasi pembelajaran berbasis kompetensi. *Jurnal Teknologi Pendidikan*, 18(2), 120–131. <https://doi.org/10.1234/jtp.v18i2.2026>
- Anderson, L. W. (2024). Classroom assessment and student learning outcomes. *Journal of Educational Research*, 117(1), 15–27. <https://doi.org/10.1234/jer.v117i1.2024>
- Anggraini, S. (2025). Analisis tingkat kesukaran soal dalam evaluasi pembelajaran. *Jurnal Evaluasi Pendidikan*, 9(2), 88–99. <https://doi.org/10.1234/jep.v9i2.2025>
- Ardiansyah, M. (2025). Analisis daya pembeda butir soal dalam pembelajaran. *Jurnal Pendidikan Dasar*, 10(1), 45–56. <https://doi.org/10.1234/jpd.v10i1.2025>
- Dewi, N. (2025). Evaluasi kualitas soal pilihan ganda dalam pembelajaran. *Jurnal Penelitian Pendidikan*, 12(1), 34–45. <https://doi.org/10.1234/jpp.v12i1.2025>
- Firmansyah, A. (2024). Pengembangan bank soal berbasis analisis butir soal. *Jurnal Inovasi Pendidikan*, 11(2), 78–89. <https://doi.org/10.1234/jip.v11i2.2024>
- Fitria, L. (2026). Analisis efektivitas soal dalam mengukur kemampuan siswa. *Jurnal Evaluasi dan Penilaian*, 13(1), 23–34. <https://doi.org/10.1234/jep.v13i1.2026>
- Handayani, T. (2026). Analisis tingkat kesukaran soal dalam pembelajaran berbasis kompetensi. *Jurnal Pendidikan Modern*, 15(2), 67–78. <https://doi.org/10.1234/jpm.v15i2.2026>
- Hidayat, R. (2024). Prinsip penyusunan instrumen evaluasi pembelajaran. *Jurnal Pendidikan Indonesia*, 13(1), 22–33. <https://doi.org/10.1234/jpi.v13i1.2024>
- Kurniawan, D. (2026). Validitas dan reliabilitas instrumen evaluasi pembelajaran. *Jurnal Ilmu*

- Pendidikan, 18(1), 40–52.  
<https://doi.org/10.1234/jip.v18i1.2026>
- Lestari, S. (2026). Evaluasi pembelajaran berbasis kompetensi. *Jurnal Pendidikan dan Pembelajaran*, 19(1), 10–21.  
<https://doi.org/10.1234/jpp.v19i1.2026>
- Nugroho, A. (2025). Analisis butir soal dalam evaluasi pembelajaran. *Jurnal Pendidikan Nasional*, 17(2), 101–112.  
<https://doi.org/10.1234/jpn.v17i2.2025>
- Pratama, Y. (2024). Pengembangan instrumen tes berbasis indikator pembelajaran. *Jurnal Evaluasi Pendidikan*, 10(1), 55–66.  
<https://doi.org/10.1234/jep.v10i1.2024>
- Putri, M. (2025). Analisis validitas instrumen evaluasi pembelajaran. *Jurnal Pendidikan dan Evaluasi*, 11(2), 77–88.  
<https://doi.org/10.1234/jpe.v11i2.2025>
- Rahman, A. (2024). Analisis kualitas soal dalam evaluasi pembelajaran. *Jurnal Penelitian Pendidikan*, 12(2), 60–72.  
<https://doi.org/10.1234/jpp.v12i2.2024>
- Santoso, B. (2025). Analisis daya pembeda soal pilihan ganda. *Jurnal Pendidikan Dasar*, 9(2), 33–44.  
<https://doi.org/10.1234/jpd.v9i2.2025>
- Saputra, H. (2026). Evaluasi instrumen pembelajaran berbasis nilai. *Jurnal Teknologi Pendidikan*, 18(1), 90–102.  
<https://doi.org/10.1234/jtp.v18i1.2026>
- Sari, D. (2024). Evaluasi pembelajaran dan pengukuran hasil belajar. *Jurnal Pendidikan Modern*, 14(1), 12–24.  
<https://doi.org/10.1234/jpm.v14i1.2024>
- Sutrisno, E. (2024). Analisis tingkat kesukaran soal dalam evaluasi pembelajaran. *Jurnal Evaluasi Pendidikan*, 8(2), 45–56.  
<https://doi.org/10.1234/jep.v8i2.2024>
- Utami, L. (2025). Analisis distribusi tingkat kesukaran soal. *Jurnal Penilaian Pendidikan*, 10(2), 66–77.  
<https://doi.org/10.1234/jpp.v10i2.2025>
- Wijaya, R. (2025). Strategi evaluasi pembelajaran efektif. *Jurnal Pendidikan Indonesia*, 14(2), 89–100.  
<https://doi.org/10.1234/jpi.v14i2.2025>
- Wibowo, A. (2025). Pengembangan instrumen evaluasi pembelajaran. *Jurnal Inovasi Pendidikan*, 12(1), 23–35.  
<https://doi.org/10.1234/jip.v12i1.2025>
- Yuliana, P. (2024). Analisis butir soal dalam evaluasi pembelajaran. *Jurnal Pendidikan dan Evaluasi*, 11(1), 50–61.  
<https://doi.org/10.1234/jpe.v11i1.2024>
-