

Analisis Butir Soal Penilaian Akhir Semester (PAS) pada Mata Pelajaran Matematika Kelas X di SMK Mahardika Ditinjau dari Teori Klasik (Classical Theory)

Syifa Amelia Nuraini¹, Nurhayati², Sumiah Nasution³
Program Studi Pendidikan Matematika, Universitas Indraprasta PGRI

INFO ARTICLES

Key Words:

Item Analysis; Classical Test Theory; Mathematics Assessment; Validity; Reliability



This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Abstract: This study aims to determine the quality of test items in the Odd Semester Final Assessment (PAS) for Mathematics subject in Grade X at SMK Mahardika for the 2025/2026 academic year, reviewed through classical theory, encompassing validity, reliability, difficulty level, discriminating power, and distractor effectiveness. This study employs a descriptive quantitative method. The analysis was conducted on 25 multiple-choice items with 50 students using Microsoft Excel. The results indicate that the overall quality of the PAS test instrument is fairly good, although there are still a number of items that require revision. Out of the 25 items analyzed, 16 items (64%) were categorized as very good, 1 item (4%) as good, 3 items (12%) as moderate, 5 items (20%) as very poor, and no items fell under the poor category.

Abstrak: Penelitian ini bertujuan mengetahui kualitas butir soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Matematika kelas X di SMK Mahardika tahun ajaran 2025/2026 ditinjau dari teori klasik, meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda serta efektivitas distraktor. Penelitian ini menggunakan metode kuantitatif deskriptif. Analisis dilakukan terhadap 25 butir soal pilihan ganda dengan 50 peserta didik menggunakan program *Microsoft Excel*. Hasil penelitian menunjukkan bahwa kualitas instrumen soal PAS secara umum sudah cukup baik, meskipun masih terdapat sejumlah butir soal yang memerlukan perbaikan. Dari 25 butir soal yang dianalisis, sebanyak 16 butir soal (64%) berkategori sangat baik, 1 butir soal (4%) berkategori baik, 3 butir soal (12%) berkategori sedang, 5 butir soal (20%) berkategori sangat tidak baik, dan tidak terdapat butir soal dengan kategori tidak baik.

Correspondence Address: Jl. Raya Tengah Kelurahan Gedong, Pasar Rebo - Jakarta Timur 13760, Indonesia; e-mail: syifaamelia16@gmail.com

How to Cite (APA 6th Style): Nuraini, S. A., Nurhayati., & Nasution, S. (2026). Analisis Butir Soal Penilaian Akhir Semester (PAS) pada Mata Pelajaran Matematika Kelas X di SMK Mahardika ditinjau dari Teori Klasik (*Classical Theory*). *Prosiding Diskusi Panel Nasional Pendidikan Matematika*, 193-204.

Copyright: Syifa Amelia Nuraini, Nurhayati, & Sumiah Nasution, (2026)

PENDAHULUAN

Pendidikan merupakan fondasi penting bagi pembangunan bangsa. Pendidikan yang berkualitas tinggi dapat menentukan terbentuknya sumber daya manusia yang handal, kreatif, dan siap bersaing secara internasional (Satria et al., 2025). Untuk menciptakan pendidikan yang berkualitas, diperlukan pembelajaran yang baik, implementasi yang baik, serta penilaian yang sistematis dan berkelanjutan agar tujuan pembelajaran dapat tercapai. Salah satu aspek penting dalam mendukung proses pendidikan adalah evaluasi. Berdasarkan Pasal 10 ayat 1 Undang-Undang Nomor 2 Tahun 2005 tentang Guru dan Dosen, seorang guru diharapkan menguasai empat kompetensi, salah satunya adalah kompetensi pedagogik (Amalia et al., 2021). Dalam kompetensi ini, guru diwajibkan untuk mengevaluasi hasil belajar siswa, yang bertujuan untuk menilai kemajuan belajar, memperbaiki program pengajaran, dan menentukan apakah siswa telah menguasai materi yang diajarkan.

Salah satu evaluasi yang dilakukan untuk mengukur kemampuan siswa adalah Penilaian Akhir Semester (PAS). Sebuah tes seperti PAS tidak hanya berfungsi sebagai alat ukur pencapaian belajar, tetapi juga harus memiliki kualitas yang baik agar hasil yang diperoleh benar-benar mencerminkan kemampuan siswa. Tes yang baik haruslah terdiri dari soal-soal yang ditulis dengan baik agar dapat mengukur kemampuan dan keterampilan peserta didik sesuai yang diharapkan oleh tujuan pembelajaran. Soal yang berkualitas yaitu soal yang dapat memberikan informasi setepat-tepatnya, sehingga dapat diketahui siswa yang telah menguasai materi dan yang belum. Maka, dalam membuat soal sebagai instrument penilaian tidak bisa dibuat secara sembarangan, melainkan harus melalui proses analisis butir soal yang sistematis agar setiap butir soal yang dibuat dan digunakan memenuhi kriteria yang baik dan mampu mengukur kemampuan siswa secara akurat serta sesuai dengan tujuan pembelajaran. Analisis soal dapat dihitung berdasarkan beberapa hal seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan fungsi pengecoh (Arikunto, 2012 dalam Ulfahyana, 2022).

Meskipun analisis butir soal memiliki peran yang sangat penting dalam memastikan kualitas instrumen penilaian, pada kenyataannya belum banyak guru yang melakukan analisis butir soal secara kuantitatif (Basri et al., 2025). Selain itu karena keterbatasan waktu, banyaknya beban administratif yang harus dikerjakan, tidak adanya tuntutan resmi dari sekolah, kurangnya pelatihan serta pengetahuan teknis dalam menganalisis butir soal membuat guru tidak melakukan analisis butir soal ketika membuat soal yang akan diberikan kepada siswa. Berdasarkan hasil wawancara dengan guru mata pelajaran Matematika di SMK Mahardika, diketahui bahwa soal-soal pada evaluasi yang dilakukan sampai saat ini belum pernah dianalisis terhadap setiap butir soalnya. Hal ini berdampak pada ketidakmampuan soal tersebut untuk mengukur capaian kompetensi siswa. Dari 50 peserta didik yang mengikuti PAS semester ganjil tahun ajaran 2025/2026, hanya 13 peserta didik (26%) yang dinyatakan tuntas, sementara 37 peserta didik (74%) belum tuntas dengan rata-rata nilai sebesar 64,08. Kondisi ini mendorong dilakukannya penelitian analisis butir soal PAS secara lebih mendalam.

Penelitian ini bertujuan untuk mengetahui kualitas butir soal Penilaian Akhir Semester (PAS) ganjil pada mata pelajaran matematika kelas X di SMK Mahardika tahun ajaran 2025/2026, ditinjau dari teori klasik (*Classical Theory*) yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor..

METODE

Penelitian ini menggunakan metode kuantitatif deskriptif. Penelitian kuantitatif deskriptif adalah jenis penelitian yang mendeskripsikan serta menjawab persoalan berdasarkan pada

pengamatan dan data nyata, dengan hasil berupa angka-angka yang bisa dihitung dan dianalisis menggunakan statistik melalui teknik pengukuran terhadap variabel yang akan diteliti untuk menarik kesimpulan. Penelitian dilaksanakan di SMK Mahardika yang berlokasi di Jl. SKB No.03, Karadenan, Kec. Cibinong, Kabupaten Bogor, Jawa Barat, pada bulan Maret hingga Juli 2026.

Subjek penelitian adalah peserta didik kelas X di SMK Mahardika tahun ajaran 2025/2026, terdiri dari 2 kelas yaitu kelas X RTL (37 siswa) dan X AKL (13 siswa), dengan total 50 siswa. Objek penelitian adalah soal PAS semester ganjil mata pelajaran matematika kelas X yang terdiri dari 25 butir soal pilihan ganda dengan 5 pilihan jawaban. Teknik pengumpulan data yang digunakan meliputi wawancara semi terstruktur dengan guru matematika, observasi non sistematis terhadap soal dan lembar jawaban, serta dokumentasi berupa soal PAS, kunci jawaban, dan lembar jawaban siswa. Data dianalisis menggunakan Microsoft Excel 2021. Teknik analisis data meliputi: (1) Validitas butir, menggunakan rumus korelasi Product Moment Pearson; butir valid jika $r_{hitung} > r_{tabel}$ (0,2787). (2) Reliabilitas, menggunakan rumus KR-20; kategori: sangat tinggi (0,80-1,00), tinggi (0,60-0,79), sedang (0,40-0,59), rendah (0,20-0,39), sangat rendah (0,00-0,19). (3) Tingkat kesukaran (P); kategori: mudah ($P > 0,70$), sedang ($0,30 \leq P \leq 0,70$), sukar ($P < 0,30$). (4) Daya pembeda (D); kategori: jelek (0,00-0,20), cukup (0,21-0,40), baik (0,41-0,70), baik sekali (0,71-1,00). (5) Efektivitas distraktor; berfungsi baik jika dipilih $> 5\%$ peserta tes, dikategorikan dengan adaptasi Skala Likert.

HASIL

Penelitian ini menganalisis 25 butir soal PAS pilihan ganda yang diikuti 50 peserta didik kelas X di SMK Mahardika tahun ajaran 2025/2026. Hasil analisis butir soal ditinjau dari teori klasik sebagai berikut :

Validitas

Pengujian validitas dilakukan menggunakan korelasi *product moment* dengan cara mengorelasikan skor pada setiap butir soal dengan skor total. Setelah diperoleh nilai koefisien korelasi, selanjutnya nilai tersebut dibandingkan dengan nilai r_{tabel} . Apabila nilai $r_{hitung} > r_{tabel}$, maka butir soal dinyatakan valid secara empiris (Kurniawan et al, 2022). Subjek berjumlah 50 peserta didik dan variabel dikorelasikan sebanyak 2 buah maka df atau db adalah 48 ($50 - 2$) sehingga jika ditarik pada taraf signifikansi 5% maka akan mendapatkan harga $r_{tabel} = 0,2787$. Berdasarkan hasil analisis menunjukkan bahwa 19 butir soal (76%) dinyatakan valid dan 6 butir soal (24%) dinyatakan tidak valid. Butir soal yang tidak valid adalah nomor 1, 3, 4, 15, 16, dan 17. Berikut rekapitulasi hasil analisis validitas :

Tabel 1. Rekapitulasi Butir Soal Berdasarkan Validitas

Klasifikasi	Butir Soal	Jumlah	Persentase
Valid	2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 18, 19, 20, 21, 22, 23, 24, 25	19	76%
Tidak Valid	1, 3, 4, 15, 16, 17	6	24%

Sumber: Data Primer Diolah

Reliabilitas

Analisis reliabilitas terhadap butir soal PAS ganjil matematika kelas X dilakukan menggunakan rumus KR-20 dengan bantuan *Microsoft excel*. Hasil dari uji reliabilitas kemudian diinterpretasikan berdasarkan kategori koefisien korelasi dengan rentang 0,80 – 1,00 termasuk kategori sangat tinggi, rentang 0,60 – 0,79 termasuk kategori tinggi, rentang 0,40 – 0,59 termasuk

kategori sedang, rentang 0,20 – 0,39 termasuk kategori rendah, dan rentang 0,00 – 0,19 termasuk kategori sangat rendah (Guilford, 1965 dalam Tanta, 2025). Hasil menunjukkan nilai koefisien reliabilitas sebesar 0,83, yang termasuk dalam kategori sangat tinggi (rentang 0,80 – 1,00). Hasil ini menunjukkan instrumen soal PAS memiliki tingkat konsistensi yang sangat baik.

Tingkat Kesukaran

Analisis tingkat kesukaran terhadap butir soal PAS ganjil matematika kelas X dihitung menggunakan rumus indeks kesukaran (P) dengan bantuan *Microsoft excel*. Tingkat kesukaran diperoleh dengan membandingkan jumlah siswa yang menjawab benar pada setiap butir soal dengan jumlah seluruh siswa yang mengikuti tes. Berdasarkan hasil dari analisis, tingkat kesukaran menunjukkan distribusi yang belum proporsional karena tidak ditemukan satupun butir soal berkategori sukar. Berikut rekapitulasi hasil analisis tingkat kesukaran :

Tabel 2. Rekapitulasi Butir Soal Berdasarkan Tingkat Kesukaran

Indeks Kesukaran	Kategori	Butir Soal	Jumlah	Persentase
$P > 0,70$	Mudah	1, 3, 10, 12, 13, 15, 16, 17	8	32%
$0,30 \leq P \leq 0,70$	Sedang	2, 4, 5, 6, 7, 8, 9, 11, 14, 18, 19, 20, 21, 22, 23, 24, 25	17	68%
$P < 0,30$	Sukar	-	0	0%

Sumber: Data Primer Diolah

Daya Pembeda

Analisis daya pembeda terhadap butir soal PAS ganjil matematika kelas X dihitung menggunakan rumus indeks diskriminasi (D) dengan bantuan *Microsoft excel*. Daya pembeda soal dihitung dengan menghitung selisih proporsi antara siswa yang menjawab benar dari kelompok atas dan kelompok bawah. Sebelum menghitung daya pembeda, jumlah peserta apabila kurang dari 100 orang, seluruh peserta dibagi sama banyak, yaitu 50% sebagai kelompok atas dan 50% sebagai kelompok bawah (Ramadhan et al., 2023). Dalam penelitian terdapat 50 peserta didik maka terlebih dahulu diurutkan berdasarkan skor total dari yang tertinggi hingga terendah, lalu dibagi menjadi dua kelompok yang sama besar, yaitu 25 peserta didik kelompok atas (skor tertinggi) dan 25 peserta didik kelompok bawah (skor terendah). Analisis daya pembeda menunjukkan bahwa tidak ditemukan butir soal dengan daya pembeda negatif, yang menandakan tidak ada butir yang fungsinya terbalik. Berikut rekapitulasi hasil analisis daya pembeda :

Tabel 3. Rekapitulasi Butir Soal Berdasarkan Daya Pembeda

Indeks Daya Pembeda	Klasifikasi	No. Butir Soal	Jumlah	Persentase
0,00 – 0,20	Jelek (Poor)	1, 3, 15, 16, 17	5	20%
0,21 – 0,40	Cukup (Satisfactory)	2, 4, 6, 7, 8, 10, 12, 18, 20	9	36%
0,41 – 0,70	Baik (Good)	5, 9, 13, 14, 19, 21, 22, 23, 24, 25	10	40%
0,71 – 1,00	Baik Sekali (Excellent)	11	1	4%
< 0,00 (Negatif)	Sangat Jelek	-	0	0%

Sumber: Data Primer Diolah

Efektivitas Distraktor

Pengecoh (*distractor*) dalam tes biasanya terdiri atas tiga sampai lima pilihan jawaban. Dari beberapa pilihan yang tersedia, terdapat satu jawaban benar, sedangkan pilihan lainnya berperan sebagai pengecoh. Distraktor dikatakan berfungsi dengan baik apabila dipilih oleh lebih dari 5% peserta tes ($p > 5\%$), sedangkan apabila dipilih oleh kurang atau sama dengan 5% peserta tes ($p \leq 5\%$), maka distraktor dinyatakan tidak berfungsi dengan baik (Sudidjono, 2011 dalam Sohilait, 2021), lalu hasil tersebut dikategorikan berdasarkan adaptasi dari *Skala Likert*, distraktor yang berfungsi sebanyak 4 termasuk kategori sangat baik, 3 termasuk kategori baik, 2 termasuk kategori cukup baik, 1 termasuk kategori kurang baik, dan 0 termasuk kategori tidak baik (Limbong, 2022). Kriteria ini diadaptasi dari pendekatan penilaian yang berbasis jumlah pengecoh aktif, sebagaimana digunakan dalam berbagai penelitian analisis butir soal (Muchlizani et al., 2023). Berikut hasil rekapitulasi hasil analisis efektivitas distraktor :

Tabel 4. Rekapitulasi Butir Soal Berdasarkan Efektivitas Distraktor

Distraktor Berfungsi	Kriteria	No. Butir Soal	Jumlah	Persentase
4 distraktor	Sangat Baik	7, 8, 18, 19, 23	5	20%
3 distraktor	Baik	2, 5, 6, 9, 11, 14, 19, 20, 21, 22, 24	11	44%
2 distraktor	Cukup Baik	1, 3, 4, 10, 25	5	20%
1 distraktor	Kurang Baik	13	1	4%
0 distraktor	Tidak Baik	15, 16, 17	3	12%

Sumber: Data Primer Diolah

Rekapitulasi Kualitas Butir Soal

Berdasarkan hasil pengujian yang telah dilakukan, kualitas butir soal selanjutnya ditafsirkan menggunakan beberapa kriteria penilaian yang diadaptasi dari Skala Likert, sebagai berikut (Oktanin, 2015 dalam Muchlizani et al., 2023) :

Tabel 4. Kriteria Kualitas Butir Soal

Jumlah Kriteria Terpenuhi	Kualitas Butir Soal	Tindak Lanjut	Masuk Bank Soal
4	Sangat Baik	Tidak Perlu Revisi	Ya
3	Baik	Revisi Minor	Belum
2	Sedang	Revisi	Belum
1	Tidak Baik	Dibuang	Tidak
0	Sangat Tidak Baik	Dibuang	Tidak

Sumber : Oktanin, 2015, dalam Muchlizani et al., 2023

Selain memenuhi syarat-syarat pada setiap butir soal berdasarkan kriteria di atas, tes secara keseluruhan harus reliabel atau memiliki reliabilitas yang baik. Berdasarkan seluruh aspek analisis (validitas, reliabilitas, tingkat kesukaran, daya pembeda, efektivitas distraktor), maka kualitas butir soal dapat dilihat sebagai berikut:

Tabel 5. Rekapitulasi Kualitas Butir Soal PAS Semester Ganjil Matematika Kelas X

Kategori Kualitas	No. Butir Soal	Jumlah Butir	Persentase	Tindak Lanjut
Sangat Baik	2, 5, 6, 7, 8, 9, 11, 14, 18, 19, 20, 21, 22, 23, 24, 25	16	64%	Masuk Bank Soal
Baik	13	1	4%	Masuk Bank Soal, Revisi Minor
Sedang	4, 10, 12	3	12%	Perlu Direvisi
Tidak Baik	-	0	0%	Dibuang/Ganti
Sangat Tidak Baik	1, 3, 15, 16, 17	5	20%	Dibuang/Ganti

Sumber: Data Primer Diolah

PEMBAHASAN

Validitas

Validitas butir dilakukan dengan teknik korelasi Product Moment yang dikemukakan oleh Karl Pearson, dengan ketentuan butir soal dinyatakan valid apabila nilai $r_{hitung} > r_{tabel}$ (Tanta, 2025). Hasil analisis validitas menunjukkan bahwa dari 25 butir soal PAS ganjil Matematika kelas X, sebanyak 19 butir soal (76%) dinyatakan valid dan 6 butir soal (24%) dinyatakan tidak valid. Tingginya persentase soal yang valid (76%) menunjukkan bahwa sebagian besar butir soal PAS Matematika kelas X sudah memenuhi kriteria validitas yang diharapkan dan mampu mengukur kemampuan peserta didik sesuai dengan tujuan pembelajaran yang ditetapkan. Keenam butir yang tidak valid menunjukkan nilai r_{hitung} di bawah r_{tabel} 0,2787, dan lima di antaranya memiliki indeks kesukaran di atas 0,70 (kategori mudah), bahkan nomor 15, 16, dan 17 memiliki indeks kesukaran di atas 0,90. Ketika hampir semua peserta menjawab benar, variasi skor antar peserta pada butir tersebut menjadi sangat kecil sehingga korelasi tidak dapat bekerja dengan baik secara statistik. Hal ini sejalan dengan pandangan bahwa terdapat hubungan antara validitas butir soal dengan daya pembeda dan tingkat kesukaran, di mana soal yang terlalu mudah cenderung menghasilkan daya pembeda yang jelek sekaligus validitas yang rendah (Nurhalimah et al., 2022). Kondisi serupa juga ditemukan pada analisis soal matematika di SMK, yang menyimpulkan bahwa butir soal dengan indeks kesukaran terlalu tinggi secara konsisten menghasilkan koefisien korelasi yang tidak signifikan (Ramadhan et al., 2023).

Berbeda pada butir soal nomor 4 yang memiliki indeks kesukaran sedang ($P = 0,70$) namun tetap tidak valid ($r_{hitung} = 0,148$), yang mengindikasikan adanya masalah pada konstruksi soal itu sendiri, kemungkinan besar berkaitan dengan kalimat yang ambigu, pengecoh yang tidak berfungsi sebagaimana mestinya, atau materi yang tidak selaras dengan indikator pembelajaran. Apabila penyusunan pilihan pengecoh lebih banyak mengandalkan intuisi berdasarkan pengalaman mengajar dan bukan berdasarkan analisis sistematis tentang pola kesalahan peserta didik, ada kemungkinan pengecoh pada butir soal nomor 4 kurang mencerminkan miskonsepsi nyata yang dialami peserta sehingga pola responnya tidak memiliki korelasi yang bermakna dengan skor total. Sebaliknya, butir soal dengan nilai r_{hitung} tertinggi adalah nomor 11 ($r_{hitung} = 0,752$), nomor 22 ($r_{hitung} = 0,711$), dan nomor 21 ($r_{hitung} = 0,623$) yang tidak hanya valid tetapi memiliki korelasi kuat dengan skor total. Butir soal yang tidak valid sebaiknya tidak langsung dibuang, melainkan dilakukan telaah ulang terhadap konstruksi soal, kesesuaiannya

dengan kisi-kisi, serta penyederhanaan kalimat yang ambigu dan jika setelah revisi masih tidak memenuhi kriteria validitas, barulah soal tersebut layak diganti atau dihilangkan dari instrumen penilaian (Kurniawan et al., 2022).

Reliabilitas

Dalam penelitian ini, reliabilitas soal PAS dihitung menggunakan rumus *Kuder-Richardson 20* (KR-20) yang diperuntukkan bagi soal dengan penilaian dikotomi, yaitu benar (skor 1) dan salah (skor 0). Hasil perhitungan menggunakan rumus *Kuder-Richardson 20* (KR-20) menunjukkan bahwa nilai koefisien reliabilitas soal PAS Matematika kelas X adalah sebesar 0,83, yang berada pada rentang 0,80 – 1,00 dan termasuk kategori sangat tinggi. Reliabilitas sebesar 0,83 ini dapat tercapai meskipun terdapat enam butir soal yang tidak valid, karena KR-20 menghitung reliabilitas berdasarkan varians total skor dan jumlah varians setiap butir. Dari data perhitungan, nilai $\sum pq = 5,01$ dan varians total $st^2 = 24,57$, di mana varians total yang relatif besar ini mencerminkan bahwa ada cukup banyak variasi kemampuan antar peserta didik yang berhasil ditangkap oleh instrumen secara keseluruhan, terutama oleh 19 butir yang valid. Enam butir yang tidak valid dengan nilai pq yang kecil justru memberikan kontribusi sangat kecil terhadap $\sum pq$, sehingga dampak negatifnya terhadap reliabilitas keseluruhan relatif terbatas.

Nilai reliabilitas sebesar 0,83 menunjukkan bahwa instrumen soal PAS ini memiliki tingkat konsistensi yang sangat baik, artinya jika tes ini diberikan kembali kepada peserta didik yang sama pada kondisi yang relatif serupa, hasil yang diperoleh akan cenderung sama. Capaian reliabilitas yang tinggi juga didukung oleh jumlah butir soal yang digunakan, yakni 25 butir. Jumlah butir soal yang lebih banyak umumnya dapat meningkatkan reliabilitas dengan cara memperluas cakupan materi yang diukur (Bhakti et al., 2022). Selain itu, format soal pilihan ganda yang digunakan menghasilkan penilaian yang objektif dan bebas dari unsur subjektivitas penilai, sehingga konsistensi skor antar pengerjaan dapat terjaga dengan baik. Hal ini sejalan dengan temuan bahwa tes pilihan ganda yang diskor secara dikotomi menggunakan KR-20 secara konsisten menghasilkan koefisien reliabilitas yang lebih tinggi dibandingkan tes uraian, karena tidak ada faktor variasi penilai yang masuk ke dalam pengukuran (Astuti et al., 2024).

Reliabilitas yang sangat tinggi ini menjadi modal penting dalam pengambilan keputusan penilaian, karena tes yang sangat reliabel memberikan kepercayaan bahwa hasil yang diperoleh mencerminkan kemampuan sesungguhnya peserta didik, bukan sekadar keberuntungan atau faktor kondisional semata. Namun perlu digarisbawahi bahwa reliabilitas yang tinggi tidak serta-merta menjamin validitas. Suatu soal bisa sangat reliabel namun tidak valid, jika soal tersebut secara konsisten mengukur sesuatu yang bukan tujuan penilaian (Ramadhan et al., 2023).

Tingkat Kesukaran

Dalam penelitian ini, tingkat kesukaran soal PAS dihitung menggunakan indeks kesukaran (P). Butir soal yang baik adalah butir soal yang tingkat kesulitannya dapat diidentifikasi, tidak terlalu sulit maupun terlalu mudah (Supriyadi, 2022). Distribusi tingkat kesukaran pada instrumen ini menunjukkan hasil yang tidak merata, yaitu 17 butir soal (68%) berkategori sedang dan 8 butir soal (32%) berkategori mudah, sementara tidak ada satu pun butir yang berkategori sukar. Proporsi ini jauh dari ideal karena komposisi yang disarankan dalam analisis butir soal umumnya mengacu pada sekitar 25% soal sukar, 50% sedang, dan 25% mudah agar dapat diketahui siswa yang benar-benar memahami materi dan yang belum (Magdalena, 2022). Ketidakhadiran soal berkategori sukar ini dapat dipahami dalam konteks bahwa penyusunan soal PAS bersumber dari LKS yang digunakan peserta didik sebagai bahan latihan sehari-hari, sehingga soal yang sudah akrab bagi peserta didik cenderung terasa lebih mudah dan indeks kesukaran empirisnya pun menjadi tinggi. Kondisi ini sejalan dengan temuan bahwa instrumen tes yang tidak menyertakan soal berkategori sukar cenderung menghasilkan distribusi skor yang miring ke kanan, sehingga kurang mampu merepresentasikan kemampuan peserta didik pada level tinggi secara akurat (Sohilait, 2021). Butir

soal nomor 15, 16, dan 17 memiliki indeks kesukaran tertinggi yaitu 0,94, 0,96, dan 0,96, yang menunjukkan bahwa hampir seluruh peserta didik menjawab soal tersebut dengan benar sehingga soal tersebut terlalu mudah dan kurang mampu membedakan kemampuan peserta didik.

Sebanyak 17 butir berkategori sedang merupakan butir-butir yang paling informatif dalam instrumen ini, karena soal dengan kesukaran sedang memiliki variasi p_q yang paling besar mendekati 0,25, sehingga memberikan kontribusi paling optimal terhadap reliabilitas dan paling mampu memunculkan perbedaan kemampuan antar peserta. Sejalan dengan pendapat yang mengatakan soal dengan tingkat kesukaran sedang merupakan soal yang paling informatif karena dapat membedakan kemampuan peserta didik secara optimal (Anggraini, 2021). Tidak adanya soal berkategori sukar perlu menjadi perhatian serius dalam penyusunan tes, karena soal semacam itu kurang memberikan informasi diagnostik yang bermakna tentang penguasaan materi pada level yang lebih tinggi dan membuat instrumen kehilangan kemampuan untuk membedakan peserta didik. Perbaikan butir soal berkategori mudah dapat dilakukan dengan meningkatkan kompleksitas pertanyaan, memperketat konteks soal, atau merujuk pada indikator kompetensi yang menuntut kemampuan analisis dan evaluasi, sementara 17 butir soal berkategori sedang yang sudah ada dapat langsung disimpan di bank soal karena sudah memenuhi kriteria tingkat kesukaran yang diharapkan.

Daya Pembeda

Dari hasil analisis, distribusi daya pembeda butir soal terbagi ke dalam empat kategori: 5 butir jelek (20%), 9 butir cukup (36%), 10 butir baik (40%), dan 1 butir baik sekali (4%). Tidak ditemukan satu pun butir dengan daya pembeda negatif, yang menunjukkan bahwa tidak ada butir yang justru terbalik fungsinya. Butir soal nomor 11 yang memiliki daya pembeda baik sekali ($D = 0,76$) menunjukkan kemampuan diskriminasi yang sangat tinggi, di mana dari 25 peserta kelompok atas sebanyak 23 menjawab benar sementara dari 25 peserta kelompok bawah hanya 4 yang berhasil menjawab benar. Sebaliknya, butir soal nomor 16 memperoleh nilai $D = 0,00$ karena jumlah peserta pada kelompok atas dan kelompok bawah yang sama-sama menjawab benar masing-masing sebanyak 24 dari 25 peserta, menandakan soal terlalu mudah sehingga kurang berfungsi sebagai alat ukur untuk membedakan kemampuan peserta didik.

Kelima butir soal yang berkategori jelek semuanya memiliki pola kegagalan yang sama yaitu tidak valid, tingkat kesukaran mudah, dan hampir seluruh peserta didik dari kedua kelompok menjawab benar. Dalam teori tes klasik, ketiga aspek validitas, tingkat kesukaran, dan daya pembeda memang saling terkait secara matematis, di mana soal yang dijawab benar oleh hampir semua peserta tidak akan menghasilkan korelasi bermakna dengan skor total, tidak akan membedakan kelompok atas dan bawah secara efektif, dan kontribusinya terhadap variasi skor total pun sangat kecil. Temuan ini mengonfirmasi bahwa soal yang terlalu mudah secara konsisten akan menghasilkan daya pembeda yang jelek dan validitas yang rendah secara bersamaan (Nurhalimah et al., 2022). Tindak lanjut yang direkomendasikan ialah butir soal dengan daya pembeda baik sekali dan baik (11 butir) layak langsung masuk bank soal, butir soal cukup (9 butir) masih dapat digunakan namun perlu direvisi, sementara butir soal dengan daya pembeda jelek (5 butir) sebaiknya direvisi secara menyeluruh atau dibuang karena tidak memberikan informasi yang berguna tentang perbedaan kemampuan peserta didik.

Efektivitas Distraktor

Hasil analisis menunjukkan bahwa 5 butir (20%) masuk kategori sangat baik, 11 butir (44%) baik, 5 butir (20%) cukup baik, 1 butir (4%) kurang baik, dan 3 butir (12%) tidak baik. Butir soal nomor 15, 16, dan 17 yang berkategori tidak baik memiliki pengecoh yang hampir seluruhnya tidak berfungsi; pada soal nomor 16 misalnya, opsi C dan D sama sekali tidak dipilih (0%), sementara B dan E hanya dipilih oleh masing-masing 2% peserta, yang mengindikasikan pengecoh yang terlalu jelas sehingga soal kehilangan fungsi diskriminasinya. Pengecoh yang efektif harus

dibuat sedemikian rupa sehingga terasa masuk akal bagi peserta yang belum menguasai materi, namun jelas salah bagi yang sudah menguasainya, dan idealnya didasarkan pada kesalahan-kesalahan konsep yang umum dilakukan peserta didik secara sistematis, bukan hanya berdasarkan intuisi. Pengecoh yang dirancang berdasarkan analisis kesalahan konseptual peserta didik terbukti meningkatkan fungsi distraktor secara signifikan dibandingkan pengecoh yang dibuat berdasarkan intuisi penyusun soal semata (Muchlizani et al., 2023). Selain itu, penelitian lain menambahkan perspektif penting bahwa kegagalan distraktor umumnya bukan semata-mata karena pilihan jawaban yang salah dibuat terlalu jelas, tetapi juga karena peserta didik sudah sangat terbiasa dengan tipe soal sehingga mampu mengeliminasi pengecoh secara otomatis tanpa proses berpikir mendalam (Bhakti et al., 2022).

Butir soal nomor 13 yang berkategori kurang baik hanya memiliki satu pengecoh yang berfungsi efektif yaitu opsi D (14%), padahal dari sisi validitas dan daya pembeda soal nomor 13 sudah cukup baik ($D = 0,44$), yang menunjukkan bahwa meskipun soalnya sudah mampu membedakan peserta berkemampuan tinggi dan rendah, pilihan jawaban salah yang tersedia kurang mampu mendiagnosis kesalahan konsep peserta. Perbaikan pada butir ini relatif mudah dilakukan dengan mengganti pilihan jawaban yang tidak diminati dengan alternatif yang lebih mencerminkan miskonsepsi umum pada topik tersebut. Secara keseluruhan, 64% butir soal sudah memiliki distraktor yang berfungsi efektif, namun 16% butir soal yang berkategori kurang baik dan tidak baik perlu segera dilakukan revisi, di mana perbaikan efektivitas distraktor pada butir-butir bermasalah harus dimulai dari akar masalahnya yaitu meningkatkan kesukaran soal itu sendiri, karena tanpa perbaikan konten soal, merevisi pengecoh saja tidak akan cukup.

Kualitas Butir Soal

Penilaian kualitas butir soal dilakukan dengan mempertimbangkan seluruh hasil analisis yang telah dilakukan, meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Penentuan kualitas mengacu pada kriteria adaptasi Skala Likert di mana semakin banyak kriteria yang terpenuhi mencerminkan kualitas butir yang semakin baik, sehingga penilaian bersifat komprehensif dengan mempertimbangkan seluruh dimensi analisis secara bersamaan (Oktanin, 2015 dalam Muchlizani et al., 2023). Pendekatan Teori Tes Klasik yang digunakan dalam penelitian ini terbukti tetap efektif meskipun jumlah peserta tes terbatas, karena pendekatannya yang sederhana, praktis, dan dapat dilakukan dengan bantuan komputer (Kurniawan et al., 2022).

Secara keseluruhan, hasil analisis kualitas butir soal PAS ganjil Matematika kelas X di SMK Mahardika Tahun Ajaran 2025/2026 menunjukkan sebanyak 16 butir soal (64%) berkualitas sangat baik karena memenuhi seluruh empat kriteria penilaian yaitu valid, tingkat kesukaran sedang, daya pembeda baik atau lebih, serta efektivitas distraktor yang baik atau sangat baik, sehingga layak langsung disimpan dalam bank soal dan dapat digunakan kembali tanpa revisi. Satu butir soal (4%) berkualitas baik yaitu nomor 13 yang memenuhi tiga dari empat kriteria namun memerlukan revisi minor khususnya pada distraktor, sedangkan tiga butir soal (12%) berkualitas sedang memerlukan revisi yang cukup signifikan terutama dengan meningkatkan kesukaran soal agar tidak terlalu mudah dijawab. Penentuan kualitas ini mengacu pada kriteria adaptasi Skala Likert di mana semakin banyak kriteria yang terpenuhi mencerminkan kualitas butir yang semakin baik, sehingga penilaian bersifat komprehensif dengan mempertimbangkan seluruh dimensi analisis secara bersamaan.

Lima butir soal (20%) berkualitas sangat tidak baik karena tidak memenuhi satupun dari empat kriteria penilaian, semuanya tidak valid, bertingkat kesukaran mudah, berdaya pembeda jelek, dan memiliki efektivitas distraktor yang buruk. Kegagalan menyeluruh pada kelima butir ini tidak terlepas dari absennya mekanisme evaluasi soal yang formal dan terstruktur di sekolah, di mana selama ini tinjauan soal hanya dilakukan secara informal dengan mengidentifikasi soal yang

paling banyak dijawab salah, sementara soal yang hampir semua peserta jawab benar justru tidak menjadi perhatian karena secara sekilas tampak seolah berhasil. Padahal soal yang terlalu mudah tidak dapat membedakan antara peserta didik yang berkemampuan tinggi dengan yang berkemampuan rendah, sehingga tidak mampu memberikan informasi yang bermakna tentang penguasaan materi peserta didik (Ramadhan et al., 2023). Dengan nilai reliabilitas tes yang sangat tinggi ($KR-20 = 0,83$), instrumen ini secara keseluruhan masih layak digunakan sebagai alat penilaian dengan syarat butir-butir berkualitas sangat tidak baik segera diganti dan butir berkualitas sedang direvisi sebelum digunakan kembali.

SIMPULAN

Berdasarkan hasil analisis butir soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Matematika kelas X di SMK Mahardika tahun ajaran 2025/2026 yang ditinjau dari teori klasik, meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor, dapat disimpulkan bahwa kualitas instrumen soal PAS tersebut secara umum sudah cukup baik, meskipun masih terdapat sejumlah butir soal yang memerlukan perbaikan. Dari hasil analisis terdapat 16 butir soal (64%) memiliki kualitas sangat baik, 1 butir soal (4%) memiliki kualitas baik, 3 butir soal (12%) memiliki kualitas sedang, 5 butir soal (20%) memiliki kualitas sangat tidak baik dan tidak ada butir soal dengan kualitas tidak baik.

UCAPAN TERIMA KASIH

Penulis mengucapkan banyak terima kasih kepada dosen pembimbing materi, dosen pembimbing teknik, kepala sekolah SMK Mahardika, guru mata pelajaran Matematika SMK Mahardika yang telah memberikan data dan informasi yang diperlukan dalam penelitian ini, serta kepada seluruh peserta didik kelas X yang menjadi subjek penelitian.

DAFTAR RUJUKAN

- Amalia, N.R., Halik, A., & Mukhlisa, N. (2021). Analisis Butir Soal Matematika Pada Siswa Sekolah Dasar. *Pinisi Journal of Education*, 1(1), 219–230. <https://ojs.unm.ac.id/PJE/article/view/25840>
- Anggraini, D. (2021). Analisis butir soal pilihan ganda ujian akhir semester pelajaran matematika di SMA Negeri 1 Padang Bolak. *JURNAL MathEdu (Mathematic Education Journal)*, 4(1), 92–100. <https://doi.org/10.37081/mathedu.v4i1.2465>
- Astuti, S., Hapsan, A., Herainto, H., Mutmainna, M., Warsyidah, A.A., Riskawati, R., Mahmud, N., Febriana, B.W., & Toron, V.B. (2024). Prinsip-Prinsip Pengukuran Dan Evaluasi Pendidikan: Disertai dengan Contoh Kasus. CV. Ruang Tentor.
- Basri, M., Sultan, S., Rapi, M., Baharman, B., & Sakaria, S. (2025). Meningkatkan Kompetensi Evaluasi Pembelajaran melalui Pelatihan Analisis Butir Soal bagi Mahasiswa Calon Guru. *MALLOMO: Journal of Community Service*, 5(2), 633–639. <https://doi.org/10.55678/mallomo.v5i2.2031>
- Bhakti, Y. B., Awaludin, A.A., Chusni, M.M., Astuti, I.A.D., Mariasih, M., Dewi, S., Fitriani, A., Okyranida, I.Y., Haryadi, D., & Gumilar, E.B. (2022). Evaluasi Pembelajaran dalam Bidang Pendidikan. CV. Bintang Semesta Media.
- Kurniawan, A., Febrianti, A.N., Risan, R., Merris, D., & Sari, M. (2022). Evaluasi pembelajaran. PT Global Eksekutif Teknologi.
- Limbong, S.S.P. (2022). Model Pembelajaran CTL dan RME Terhadap Hasil Belajar Matematika Siswa. Guepedia.

- Magdalena, I. (2022). *Teori dan Praktik Evaluasi Pembelajaran SD*. CV Jejak (Jejak Publisher).
- Masullah, B.D., Zuhry, L., Usman, L., & Maulana, L. (2024). Analisis Butir Soal Ujian Akhir Semester Mata Pelajaran Matematika SMP Negeri 6 Praya Timur. *LIPS: Jurnal Pendidikan Matematika*, 5(2), 152–161. <https://doi.org/10.47650/elips.v5i2.1219>
- Muchlizani, A.N., Mania, S., & Rasyid, M.N.A. (2023). Analisis kualitas butir soal ujian akhir semester mata pelajaran akidah akhlak kelas V MI Radhiatul Adawiyah Makassar. *Inspiratif Pendidikan*, 12(1), 224–240. <https://doi.org/10.24252/ip.v12i1.32040>
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W.P. (2022). Hubungan antara Validitas Item dengan Daya Pembeda dan Tingkat Kesukaran Soal Pilihan Ganda PAS. *NSER: Natural Science Education Research*, 4(3), 249–257. <https://doi.org/10.21107/nser.v4i3.8682>
- 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. <https://doi.org/10.21093/twt.v10i2.6155>
- Satria, D., Kusasih, I.H., & Gusmaneli, G. (2025). Analisis Rendahnya Kualitas Pendidikan di Indonesia Saat Ini: Suatu Kajian Literatur. *Jurnal Bintang Pendidikan Indonesia*, 3(2), 292–309. <https://doi.org/10.55606/jubpi.v3i2.3838>
- Sohilait, E. (2021). *Buku Ajar: Evaluasi Pembelajaran Matematika*. PT. Rajagrafindo Persada.
- Supriyadi, S. (2022). *Evaluasi Pembelajaran*. PT Nasya Expanding Management (NEM).
- Tanta, T. (2025). *Penilaian Hasil Pembelajaran*. PT. Sonpedia Publishing Indonesia.
- Ulfahyana, H. (2022). Analisis Kualitas Butir Soal Ujian Sekolah Kelas IX Mata Pelajaran Matematika. *PRISMA (Jurnal Penalaran dan Riset Matematika)*, 1(2), 70–78. <https://doi.org/10.62388/prisma.v1i2.212>

