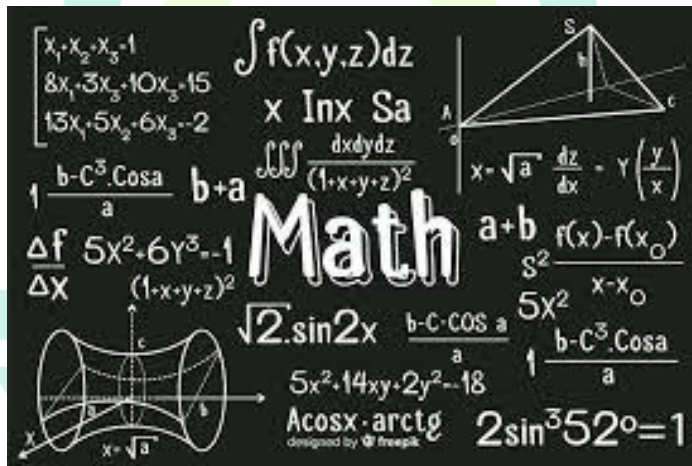




**ANALISIS KUALITAS BUTIR SOAL  
ASESMEN SUMATIF AKHIR  
SEMESTER (ASAS) MATEMATIKA  
KELAS VII SEMESTER GANJIL TAHUN  
AJARAN 2025/2026 SMPN 6 KOTA  
PEKALONGAN**



**HAEDAR BAQIR**

**NIM. 20622066**

**2026**

**ANALISIS KUALITAS BUTIR SOAL ASESMEN  
SUMATIF AKHIR SEMESTER (ASAS) MATEMATIKA  
KELAS VII SEMESTER GANJIL TAHUN AJARAN  
2025/2026 SMPN 6 KOTA PEKALONGAN**

**SKRIPSI**

**diajukan sebagai salah satu syarat untuk memperoleh  
gelar Sarjana Pendidikan (S.Pd.)**



Oleh :

**HAEDAR BAQIR**

**NIM. 20622066**

**PROGRAM STUDI TADRIS MATEMATIKA  
FAKULTAS TARBIYAH DAN ILMU KEGURUAN  
UNIVERSITAS ISLAM NEGERI  
K.H. ABDURRAHMAN WAHID PEKALONGAN**

**2026**

**ANALISIS KUALITAS BUTIR SOAL ASESMEN  
SUMATIF AKHIR SEMESTER (ASAS) MATEMATIKA  
KELAS VII SEMESTER GANJIL TAHUN AJARAN  
2025/2026 SMPN 6 KOTA PEKALONGAN**

**SKRIPSI**

**diajukan sebagai salah satu syarat untuk memperoleh  
gelar Sarjana Pendidikan (S.Pd.)**



Oleh :

**HAEDAR BAQIR**

**NIM. 20622066**

**PROGRAM STUDI TADRIS MATEMATIKA  
FAKULTAS TARBIYAH DAN ILMU KEGURUAN  
UNIVERSITAS ISLAM NEGERI  
K.H. ABDURRAHMAN WAHID PEKALONGAN**

**2026**

## SURAT PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya :

Nama : Haedar Baqir  
NIM : 20622066  
Program Studi : Tadris Matematika

Menyatakan bahwa yang tertulis dalam skripsi yang berjudul "**Analisis Kualitas Butir Soal Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII Semester Ganjil Tahun Ajaran 2025/2026 SMPN 6 Kota Pekalongan**" ini benar-benar karya saya sendiri, bukan jiplakan dari karya orang lain atau pengutipan yang melanggar etika keilmuan yang berlaku, baik Sebagian atau seluruhnya. Pendapat atau temuan orang lain yang terdapat dalam skripsi ini dikutip berdasarkan kode etik ilmiah. Apabila skripsi ini terbukti ditemukan pelanggaran terhadap etika keilmuan, maa saya bersedia menerima sanksi hukum yang dijatuhkan.

Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Pekalongan, 06 Maret 2026

Yang membuat pernyataan



Haedar Baqir

NIM. 20622066

## NOTA PEMBIMBING

Kepada  
Yth. Dekan Fakultas Tarbiyah dan Ilmu Keguruan  
UIN K.H. Abdurrahman Wahid Pekalongan  
c/q Ketua Program Studi Tadris Matematika  
di Pekalongan

*Assalamu'alaikum Wr. Wb.*

Setelah melakukan penelitian, bimbingan, dan koreksi naskah skripsi saudara:

Nama : Haedar Baqir  
NIM : 20622066  
Program Studi : Tadris Matematika  
Judul : Analisis Kualitas Butir Soal Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII Semester Ganjil Tahun Ajaran 2025/2026 SMPN 6 Kota Pekalongan

Saya menilai bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Tarbiyah dan Ilmu Keguruan UIN K.H. Abdurrahman Wahid Pekalongan untuk diajukan dalam sidang munaqasah.

Demikian nota pembimbing ini dibuat untuk digunakan sebagaimana mestinya. Atas perhatiannya, disampaikan terimakasih.

*Wassalamua'alaikum Wr. Wb.*

Pekalongan, 03 Maret 2026  
Pembimbing,



**Prof. Drs. Moh Muslih, M.Pd., Ph.D**  
NIP. 196707171999031001



### PENGESAHAN

Dekan Fakultas Tarbiyah dan Ilmu Keguruan Universitas Islam Negeri  
K.H. Abdurrahman Wahid Pekalongan mengesahkan Skripsi saudara/i:

Nama : HAEDAR BAQIR

NIM : 20622066

Program Studi: TADRIS MATEMATIKA

Judul Skripsi : ANALISIS KUALITAS BUTIR SOAL ASESMEN SUMATIF  
AKHIR SEMESTER (ASAS) MATEMATIKA KELAS VII  
SEMESTER GANJIL TAHUN AJARAN 2025/2026 SMPN 6  
KOTA PEKALONGAN

Telah diujikan pada hari Rabu tanggal 11 Maret 2026 dan dinyatakan **LULUS** serta  
diterima sebagai salah satu syarat guna memperoleh gelar Sarjana Pendidikan  
(S.Pd.).

Dewan Penguji

Penguji I

Umi Mahmudah, M.Sc., Ph.D.  
NIP. 198407102023214033

Penguji II

Heni Lilia Dewi, M.Pd.  
NIP. 199306222019032020



## MOTO DAN PERSEMBAHAN

"Setiap anak adalah jenius. Tetapi jika Anda mengevaluasi seekor ikan dari kemampuannya memanjat pohon, ia akan menjalani seluruh hidupnya dengan meyakini bahwa ia bodoh"

(Albert Einstein)

Puji syukur ke hadirat Allah Swt. atas segala rahmat dan karunia-Nya. Sebagai bentuk rasa syukur dan penghormatan, karya tulis ini penulis persembahkan kepada:

1. Kedua Orang Tua Saya, Bapak Arif Efendi dan Ibu Azizah. Terima kasih yang sebesar-besarnya atas segala doa, ketulusan, dan pengorbanan yang telah diberikan selama ini. Dukungan moral dan material dari Bapak dan Ibu merupakan landasan utama bagi penulis dalam menyelesaikan pendidikan ini. Semoga pencapaian ini dapat menjadi bentuk bakti dan kebanggaan bagi keluarga.
2. Kepada kedua kakak saya, Nurul Hanifah dan Halimatus Sa'diyah beserta keluarga, terima kasih atas dukungan, bantuan, serta motivasi yang senantiasa diberikan selama proses studi dan penyusunan skripsi ini berlangsung.
3. Kepada rekan-rekan mahasiswa Program Studi Tadris Matematika Angkatan 2022, terima kasih atas diskusi yang membangun, bantuan, dan solidaritas yang telah terjalin dengan baik selama masa perkuliahan hingga tahap penyelesaian tugas akhir ini.
4. Sebagai penutup, terima kasih kepada diri saya sendiri atas segala kerja keras, ketekunan, dan komitmen untuk tidak menyerah hingga berada di titik ini. Semoga pencapaian tugas akhir ini menjadi motivasi untuk terus bertumbuh dan memberikan manfaat di masa depan.

## ABSTRAK

Baqir, Haedar. 2026. Analisis Kualitas Butir Soal Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII Semester Ganjil Tahun Ajaran 2025/2026 SMPN 6 Kota Pekalongan. Skripsi. Program Studi Pendidikan Matematika. FTIK Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan. Pembimbing : Prof. Drs. Moh. Muslih, M.Pd., Ph.D.

Kata Kunci: Analisis Butir Soal, ASAS, Matematika, Teori Tes Klasik.

Evaluasi pembelajaran matematika yang berkualitas sangat esensial untuk mengukur penalaran logis dan ketercapaian kompetensi siswa secara akurat. Namun, dalam praktiknya, banyak instrumen tes buatan guru (*teacher-made test*) yang digunakan di sekolah belum melalui proses pengujian kualitas secara empiris. Hal ini sering kali disebabkan oleh keterbatasan waktu pendidik serta adanya anggapan bahwa proses analisis butir soal merupakan tahapan yang rumit dan membutuhkan perhitungan statistik yang kompleks.

Rumusan masalah dalam penelitian ini berfokus pada bagaimana kualitas butir soal Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII di SMPN 6 Kota Pekalongan jika ditinjau dari parameter psikometrik. Adapun tujuan utama dari penelitian ini adalah untuk mendeskripsikan karakteristik instrumen tersebut secara komprehensif berdasarkan tingkat validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh pada setiap butir soalnya.

Metode penelitian yang digunakan adalah kualitatif dengan pendekatan deskriptif analitis. Analisis data mengadopsi Model Interaktif dari Miles, Huberman, dan Saldaña yang meliputi pengumpulan data, kondensasi data, penyajian data, dan penarikan kesimpulan. Data primer berupa naskah soal dan lembar jawaban 156 siswa diolah terlebih dahulu menggunakan bantuan perangkat lunak ANATES Versi 4.0.9 sebagai pijakan awal (*entry point*) untuk kemudian

dibedah secara mendalam dari sisi pedagogis dan konstruksi materi.

Hasil penelitian menunjukkan bahwa: (1) Kualitas bentuk pilihan ganda sangat memuaskan dengan 100% soal valid, reliabilitas sangat tinggi (0,83), daya pembeda sangat baik, dan tingkat kesukaran proporsional, meskipun beberapa pengecoh perlu direvisi. (2) Kualitas bentuk benar-salah memiliki kelemahan fundamental; meskipun 100% valid, instrumen ini memiliki reliabilitas sangat rendah (0,09) akibat tingginya peluang tebakan buta (blind guessing), dengan satu butir soal memiliki daya pembeda jelek karena terlalu mudah. (3) Kualitas bentuk uraian sangat ideal, mencapai 100% valid, reliabilitas sangat tinggi (0,90), tingkat kesukaran sedang yang merata, dan daya pembeda sangat baik. Secara keseluruhan, kualitas instrumen ASAS tergolong sangat baik dengan rincian 19 butir soal (95%) layak masuk bank soal dan hanya satu butir soal (5%) yang memerlukan revisi.



## KATA PENGANTAR

Puji syukur senantiasa penulis panjatkan ke hadirat Allah Swt. serta selawat dan salam kepada Nabi Muhammad saw., sehingga penulis dapat menyelesaikan skripsi berjudul "Analisis Kualitas Butir Soal Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII Semester Ganjil Tahun Ajaran 2025/2026 SMPN 6 Kota Pekalongan". Penulis menyadari bahwa keberhasilan penyusunan skripsi ini tidak lepas dari doa, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan rasa terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Prof. Dr. H. Zaenal Mustakim, M.Ag., selaku Rektor UIN K.H. Abdurrahman Wahid Pekalongan, atas dedikasi dan kepemimpinannya dalam menciptakan iklim akademik yang kondusif, serta atas segala fasilitas pendidikan dan kebijakan kampus yang telah mendukung kelancaran proses studi penulis dari awal hingga penyelesaian tugas akhir ini..
2. Prof. Dr. H. Muhlisin, M.Ag., selaku Dekan Fakultas Tarbiyah dan Ilmu Keguruan (FTIK) UIN K.H. Abdurrahman Wahid Pekalongan, atas arahan strategis, motivasi akademis, serta kebijakannya dalam memfasilitasi perizinan dan dukungan administratif yang sangat membantu penulis selama proses penelitian di lapangan.
3. Santika Lya Diah Pramesti, M.Pd., selaku Ketua Program Studi Tadris Matematika, dan Heni Lilia Dewi, M.Pd., selaku Sekretaris Program Studi Tadris Matematika UIN K.H. Abdurrahman Wahid Pekalongan. Prof. Drs. Moh. Muslih, M.Pd., Ph.D., selaku Dosen Pembimbing yang telah berkenan meluangkan waktu, tenaga, dan pikiran untuk

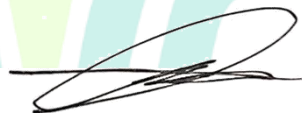
memberikan bimbingan, arahan, serta motivasi yang tak ternilai harganya selama proses penyusunan skripsi ini.

4. Segenap Bapak/Ibu Dosen dan Staf Akademik Fakultas Tarbiyah dan Ilmu Keguruan, khususnya Program Studi Tadris Matematika UIN K.H. Abdurrahman Wahid Pekalongan, yang telah memberikan bekal ilmu pengetahuan yang bermanfaat serta pelayanan administratif yang baik selama penulis menempuh pendidikan.
5. Qurratiani, S.Si., selaku Kepala SMP Negeri 6 Kota Pekalongan, dan Ibu Jumirah, S.Pd., selaku guru mata pelajaran Matematika Kelas VII, beserta segenap guru dan staf tata usaha yang telah memberikan izin, akses data, dan fasilitas selama penulis melaksanakan penelitian.

Semoga segala bantuan dan kebaikan yang telah diberikan mendapatkan balasan pahala yang berlipat ganda dari Allah Swt. Penulis berharap karya ilmiah ini dapat memberikan manfaat nyata bagi dunia pendidikan, khususnya dalam bidang evaluasi pembelajaran matematika.

Pekalongan, Februari 2026

Penulis,



Haedar Baqir

NIM. 20622066

## DAFTAR ISI

<b>JUDUL</b> .....	<b>i</b>
<b>SURAT PERNYATAAN KEASLIAN SKRIPSI</b> .....	<b>ii</b>
<b>NOTA PEMBIMBING</b> .....	<b>iii</b>
<b>PENGESAHAN</b> .....	<b>iv</b>
<b>MOTO DAN PERSEMBAHAN</b> .....	<b>v</b>
<b>ABSTRAK</b> .....	<b>vi</b>
<b>KATA PENGANTAR</b> .....	<b>viii</b>
<b>DAFTAR TABEL</b> .....	<b>xii</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiii</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xiv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang Masalah .....	1
1.2 Identifikasi Masalah.....	4
1.3 Pembatasan Masalah.....	5
1.4 Rumusan Masalah.....	6
1.5 Tujuan Penelitian .....	6
1.6 Manfaat Penelitian .....	7
<b>BAB II LANDASAN TEORI</b> .....	<b>9</b>
2.1 Deskripsi Teoritik.....	9
2.2 Kajian Penelitian Yang Relevan.....	39
2.3 Kerangka Berpikir.....	44
<b>BAB III METODE PENELITIAN</b> .....	<b>48</b>
3.1 Desain Penelitian .....	48
3.2 Fokus Penelitian.....	49
3.3 Data dan Sumber Data .....	50
3.4 Teknik Pengumpulan Data.....	51
3.5 Teknik Keabsahan Data .....	52
3.6 Teknik Analisis Data .....	55
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	<b>64</b>
4.1 Hasil Penelitian .....	64
4.2 Pembahasan .....	82

<b>BAB V PENUTUP .....</b>	<b>115</b>
5.1    Simpulan .....	115
5.2    Saran .....	117
<b>DAFTAR PUSTAKA .....</b>	<b>119</b>
<b>LAMPIRAN .....</b>	<b>123</b>



## DAFTAR TABEL

Tabel 3. 1 Kriteria Validitas .....	57
Tabel 3. 2 Kriteria Reliabilitas.....	59
Tabel 3. 3 Indeks Tingkat Kesukaran .....	59
Tabel 3. 4 Indeks Daya Pembeda.....	60
Tabel 3. 5 Kriteria Efektivitas Pengecoh .....	61
Tabel 4. 1 Hasil Uji Validitas Soal Pilihan Ganda .....	69
Tabel 4. 2 Hasil Uji Reliabilitas Soal Pilihan Ganda.....	70
Tabel 4. 3 Hasil Uji Tingkat Kesukaran Soal Pilihan Ganda.....	71
Tabel 4. 4 Hasil Uji Tingkat Daya Pembeda Soal Pilihan Ganda .....	72
Tabel 4. 5 Hasil Uji Efektivitas Pengecoh Soal Pilihan Ganda.....	73
Tabel 4. 6 Hasil Uji Validitas Soal Benar-Salah .....	75
Tabel 4. 7 Hasil Uji Reliabilitas Soal Benar-Salah.....	76
Tabel 4. 8 Hasil Uji Tingkat Kesukaran Soal Benar- Salah .....	77
Tabel 4. 9 Hasil Uji Tingkat Daya Pembeda Soal Benar- Salah .....	77
Tabel 4. 10 Hasil Uji Validitas Soal Uraian.....	79
Tabel 4. 11 Hasil Uji Reliabilitas Soal Uraian.....	79
Tabel 4. 12 Hasil Uji Tingkat Kesukaran Soal Uraian ....	80
Tabel 4. 13 Hasil Uji Tingkat Daya Pembeda Soal Uraian .....	81
Tabel 4. 14 Rekapitulasi Kualitas Instrumen.....	82
Tabel 4. 15 Perbandingan Standar Tingkat Kesukaran Pilihan Ganda .....	85
Tabel 4. 16 Perbandingan Standar Tingkat Kesukaran Benar-Salah .....	95
Tabel 4. 17 Perbandingan Standar Tingkat Kesukaran Uraian .....	103

## DAFTAR GAMBAR

Gambar 2. 1 Kerangka Berpikir Analisis Kualitas Soal Matematika .....	47
--	----



## DAFTAR LAMPIRAN

Lampiran 1. Surat Izin Penelitian .....	123
Lampiran 2. Surat Bukti Penelitian .....	124
Lampiran 3. Lembar Soal ASAS Matematika Kelas VII Semester Ganjil Tahun Ajaran 2025/2026 .....	125
Lampiran 4. Lembar Kunci Jawaban.....	127
Lampiran 5. Lembar Jawab Siswa.....	129
Lampiran 6. Hasil Rekap Nilai Pilihan Ganda .....	138
Lampiran 7. Hasil Rekap Nilai Soal Benar-Salah .....	143
Lampiran 8. Hasil Rekap Nilai Soal Uraian .....	148
Lampiran 9. Hasil Uji Validitas Soal Pilihan Ganda .....	153
Lampiran 10. Hasil Uji Reliabilitas Soal Pilihan Ganda .....	154
Lampiran 11. Hasil Uji Tingkat Kesukaran Soal Pilihan Ganda.....	155
Lampiran 12. Hasil Uji Tingkat Daya Pembeda.....	156
Lampiran 13. Hasil Uji Tingkat Efektivitas Pengecoh Soal Pilihan Ganda .....	157
Lampiran 14. Hasil Uji Validitas Soal Benar-Salah .....	158
Lampiran 15. Hasil Uji Reliabilitas Soal Benar-Salah ..	159
Lampiran 16. Hasil Uji Tingkat Kesukaran Soal Benar- Salah .....	160
Lampiran 17. Hasil Uji Tingkat Daya Pembeda.....	161
Lampiran 18. Hasil Uji Validitas Soal Uraian .....	162
Lampiran 19. Hasil Uji Reliabilitas Soal Uraian .....	163
Lampiran 20. Hasil Uji Tingkat Kesukaran Soal Uraian ..	164
Lampiran 21. Hasil Uji Tingkat Daya Pembeda Soal Uraian .....	165
Lampiran 22. Dokumentasi Kegiatan.....	166

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Dalam sistem pendidikan nasional, matematika memegang peranan vital sebagai disiplin ilmu yang melatih kemampuan berpikir logis, analitis, dan pemecahan masalah (Syaifudin, 2023). Untuk memastikan efektivitas pembelajaran matematika dan ketercapaian tujuan tersebut, diperlukan kegiatan evaluasi dan penilaian yang sistematis. Meskipun kedua istilah ini sering digunakan bergantian, terdapat perbedaan mendasar di antara keduanya. Penilaian (*assessment*), sebagaimana didefinisikan dalam Permendikbudristek Nomor 21 Tahun 2022, berfokus pada proses pengumpulan dan pengolahan informasi untuk mengetahui capaian hasil belajar peserta didik (Permendikbudristek, 2022). Sementara itu, evaluasi memiliki makna yang lebih luas sebagai proses pengambilan keputusan mengenai nilai dan keberhasilan suatu program pembelajaran berdasarkan data yang diperoleh dari penilaian (Ali & Khaeruddin, 2012). Sinergi keduanya menjadi sarana krusial untuk mengendalikan, menjamin, dan menetapkan mutu pendidikan secara berkelanjutan (Surbakti, 2025).

Keberhasilan sebuah evaluasi bergantung penuh pada kualitas instrumen yang digunakan. Dalam kegiatan penilaian sumatif, seperti Asesmen Sumatif Akhir Semester atau Ujian Sekolah, alat ukur utamanya adalah tes yang terdiri dari butir-butir soal (Nurlaila et al., 2024). Soal yang berkualitas akan mampu memberikan data yang akurat dan objektif mengenai pemahaman siswa. Sebaliknya, instrumen yang buruk akan menghasilkan data yang bias dan

menyesatkan. Jika soal yang digunakan tidak berkualitas, nilai yang diperoleh siswa tidak akan mencerminkan kemampuan mereka yang sebenarnya. Oleh karena itu, analisis kualitas butir soal merupakan langkah fundamental yang tidak dapat diabaikan untuk menjamin validitas dan reliabilitas instrumen penilaian (Hanna & Retnawati, 2022; Nisa & Hidayati, 2023).

Secara empiris, kualitas butir soal dapat dianalisis menggunakan landasan Teori Tes Klasik (*Classical Test Theory* atau CTT) (Sheptian et al., 2025). Teori ini menjelaskan bahwa skor yang tampak (X) pada lembar jawaban siswa merupakan gabungan dari skor murni (T) atau kemampuan asli siswa, ditambah dengan komponen kesalahan pengukuran (E) (Ayanwale et al., 2013; Syaifudin, 2023). Tujuan utama dari analisis butir soal adalah untuk memperkecil komponen kesalahan (E) tersebut. Menurut CTT, instrumen yang baik harus memenuhi empat kriteria utama: Validitas, yaitu ketepatan tes dalam mengukur apa yang seharusnya diukur; Reliabilitas, yakni tingkat konsistensi atau keandalan tes; Tingkat Kesukaran, yang mengidentifikasi apakah soal tergolong mudah, sedang, atau sukar; dan Daya Pembeda, yaitu kemampuan soal untuk membedakan antara siswa yang berkemampuan tinggi (kelompok atas) dan siswa yang berkemampuan rendah (kelompok bawah) (Ali & Khaeruddin, 2012; Karindi & Rufi'i, 2025).

Meskipun analisis butir soal merupakan pilar penting dalam evaluasi, kenyataan di lapangan menunjukkan adanya kesenjangan yang signifikan. Banyak perangkat soal yang digunakan di sekolah, baik yang disusun oleh guru secara individu maupun oleh tim Musyawarah Guru Mata Pelajaran (MGMP),

seringkali diterapkan tanpa melalui proses pengujian kualitas secara empiris (Surbakti, 2025). Kegiatan analisis butir soal ini cukup jarang dilakukan oleh guru (Nisa & Hidayati, 2023). Penyebab utamanya adalah karena proses analisis ini secara tradisional dianggap rumit, membutuhkan perhitungan statistik yang kompleks, dan sangat memakan waktu (Karindi & Rufi'i, 2025).

Berdasarkan studi pendahuluan dan observasi yang dilakukan di SMPN 6 Kota Pekalongan, sekolah tersebut telah secara rutin menyelenggarakan Asesmen Sumatif Akhir Semester (ASAS) sebagai tolak ukur ketercapaian kompetensi peserta didik, termasuk pada mata pelajaran Matematika Kelas VII. Instrumen soal ASAS yang digunakan merupakan instrumen buatan guru (teacher-made test) yang dirancang secara mandiri dan disajikan dalam beberapa bentuk variasi soal, yakni Pilihan Ganda, Benar-Salah, dan Uraian, guna menyesuaikan dengan tuntutan kurikulum yang berlaku.

Namun demikian, berdasarkan studi pendahuluan dan observasi, upaya penjaminan mutu instrumen tes matematika tersebut masih kurang konsisten dilakukan, terutama dalam tahapan analisis karakteristik butir soal secara kuantitatif-empiris. Pihak sekolah dan guru mata pelajaran pada dasarnya telah berupaya menyusun instrumen asesmen dengan baik sesuai capaian pembelajaran. Akan tetapi, keterbatasan waktu dan padatnya beban akademik sering kali menjadi kendala bagi guru untuk melakukan penelaahan statistik pasca-ujian secara mendalam dan berkesinambungan.

Sehingga dari keterbatasan tersebut, peta kualitas instrumen ASAS Matematika Kelas VII Semester Ganjil di SMPN 6 Kota Pekalongan belum terpetakan secara optimal dan komprehensif, baik dari segi validitas butir, tingkat keandalan instrumen (reliabilitas), proporsi tingkat kesukaran, daya pembeda, maupun efektivitas fungsi pengecoh pada soal objektifnya. Keterbatasan data empiris yang konsisten ini tentu berpotensi menimbulkan bias dalam pengambilan keputusan akademik terkait ketuntasan belajar siswa. Oleh karena itu, diperlukan suatu analisis empiris yang komprehensif untuk mengevaluasi kelayakan butir-butir soal tersebut.

Berangkat dari urgensi tersebut, peneliti tertarik untuk melakukan penelitian yang mendalam dengan judul "Analisis Kualitas Butir Soal Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII Semester Ganjil Tahun Ajaran 2025/2026 SMP Negeri 6 Kota Pekalongan".

## **1.2 Identifikasi Masalah**

Berdasarkan pemaparan latar belakang di atas, maka permasalahan pokok dalam penelitian ini dapat diidentifikasi sebagai berikut:

- a. Kegiatan analisis kualitas butir soal pasca-ujian secara kuantitatif-empiris masih kurang konsisten dilakukan oleh guru di lapangan, sering kali akibat kendala keterbatasan waktu dan padatnya beban akademik.
- b. Kualitas empiris dari instrumen buatan guru (teacher-made test) pada Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII Semester Ganjil di SMP Negeri 6 Kota

- Pekalongan belum pernah terpetakan secara pasti.
- c. Belum terdeskripsikannya parameter kualitas instrumen ASAS tersebut secara spesifik, yang meliputi tingkat validitas butir, tingkat reliabilitas tes, serta karakteristik butir soal (tingkat kesukaran, daya pembeda, dan efektivitas pengecoh).
  - d. Adanya potensi bias dalam pengambilan keputusan akademik terkait ketuntasan belajar peserta didik akibat penggunaan instrumen tes yang belum teruji kelayakannya secara empiris..

### **1.3 Pembatasan Masalah**

Mengingat luasnya permasalahan yang telah teridentifikasi, penelitian ini perlu dibatasi agar lebih fokus dan mendalam. Pembatasan masalah dalam penelitian ini adalah sebagai berikut:

- a. Analisis kualitas butir soal didasarkan pada landasan Teori Tes Klasik (CTT).
- b. Parameter kualitas butir soal yang dianalisis dibatasi pada tiga aspek utama, yaitu:
  - 1) Validitas butir soal
  - 2) Reliabilitas tes
  - 3) Karakteristik butir soal yang meliputi tingkat kesukaran, daya pembeda dan efektivitas pengecoh.
- c. Objek penelitian adalah naskah soal dan data empiris (jawaban siswa) dari Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII di SMP Negeri 6 Kota Pekalongan.

- d. Perangkat lunak (*software*) yang digunakan untuk membantu proses analisis data adalah ANATES 4.0.9.

#### **1.4 Rumusan Masalah**

Berdasarkan pembatasan masalah di atas, maka rumusan masalah dalam penelitian ini adalah:

- a. Bagaimana tingkat validitas butir soal pada Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII semester ganjil tahun ajaran 2025/2026 SMP Negeri 6 Kota Pekalongan?
- b. Bagaimana tingkat reliabilitas perangkat soal Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII semester ganjil tahun ajaran 2025/2026 SMP Negeri 6 Kota Pekalongan?
- c. Bagaimana karakteristik butir soal pada Asesmen Sumatif Akhir Semester (ASAS) ditinjau dari tingkat kesukaran, daya pembeda dan efektivitas pengecoh?

#### **1.5 Tujuan Penelitian**

- a. Mengetahui dan mendeskripsikan validitas butir soal pada tes Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII SMP Negeri 6 Kota Pekalongan.
- b. Mengetahui dan mendeskripsikan reliabilitas Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII SMP Negeri 6 Kota Pekalongan.
- c. Mengetahui dan mendeskripsikan karakteristik butir soal pada Asesmen Sumatif Akhir Semester (ASAS) ditinjau dari tingkat kesukaran, daya pembeda dan efektivitas pengecoh.

## 1.6 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- a. Manfaat Teoritis
  - 1) Menjadi landasan teoretis dan pijakan keilmuan bagi pendidik (guru) maupun pihak sekolah dalam memahami implementasi Teori Tes Klasik (CTT) untuk menyusun instrumen evaluasi asesmen yang berkualitas.
  - 2) Memberikan rujukan literatur yang bersifat empiris mengenai urgensi kalibrasi butir soal buatan guru (teacher-made test) sebagai upaya penjaminan mutu pembelajaran di tingkat satuan pendidikan..
- b. Manfaat Praktis
  - 1) Bagi Guru Matematika  
Memberikan umpan balik (umpan balik) yang objektif mengenai kualitas soal ASAS yang telah dibuat. Guru dapat menggunakan hasil analisis ini untuk memperbaiki, merevisi, atau membuang soal yang tidak layak dan menyusun "bank soal" yang berkualitas.
  - 2) Bagi Sekolah  
Memberikan data empiris mengenai kualitas alat evaluasi yang digunakan di sekolah, sebagai bahan masukan untuk proses penjaminan mutu (*quality assurance*) pembelajaran.
  - 3) Bagi Peneliti  
Memenuhi salah satu syarat untuk memperoleh gelar Sarjana Pendidikan, serta

menambah wawasan dan pengalaman praktis dalam melakukan analisis instrumen evaluasi.



## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Deskripsi Teoritik**

##### **2.1.1 Evaluasi Pendidikan**

Secara umum, literatur mengenai evaluasi pendidikan di Indonesia banyak merujuk pada pemikiran Suharsimi Arikunto yang memandang evaluasi sebagai proses pengambilan keputusan berdasarkan data hasil pengukuran. Namun, secara epistemologis, teori-teori tersebut merupakan sintesis dan adaptasi dari berbagai landasan teoretis primer yang dikembangkan oleh pakar evaluasi dunia.

Prinsip-prinsip evaluasi yang diterapkan dalam penelitian ini, khususnya mengenai dikotomi asesmen sumatif, sinkronisasi tujuan pembelajaran, hingga hierarki pengukuran, pada dasarnya berpijak pada fondasi pemikiran Ralph W. Tyler (1949) mengenai model orientasi tujuan (*goal-oriented model*), Michael Scriven (1967) yang mempopulerkan fungsi sumatif dalam evaluasi melalui karyanya *The Methodology of Evaluation*, serta Norman E. Gronlund (1985) yang meletakkan dasar-dasar pengukuran dan tes yang standar dalam karyanya *Measurement and Evaluation in Teaching*. Penelaahan terhadap sumber-sumber primer ini menjadi krusial untuk memastikan bahwa analisis kualitas butir soal ASAS Matematika yang dilakukan peneliti memiliki orisinalitas teoretis yang kuat dan akuntabel secara ilmiah.

Bagian ini menguraikan konsep-konsep fundamental yang menjadi landasan dalam penelitian, dimulai dari definisi pengukuran, penilaian, dan evaluasi, tujuan dan fungsinya, hingga tes dan asesmen sumatif yang menjadi objek penelitian.

a. Hakikat Pengukuran (*Measurement*)

Pengukuran merupakan fondasi kuantitatif dan langkah paling awal dalam keseluruhan rangkaian proses evaluasi pendidikan. Secara operasional, Arikunto (2018) mendefinisikan pengukuran sebagai suatu proses membandingkan sesuatu dengan satu ukuran standar tertentu yang sifatnya objektif. Sejalan dengan hal tersebut, dari perspektif psikometrik yang lebih mendalam, Magdalena et al. (2020) mengartikan pengukuran sebagai proses pemberian angka (kuantifikasi) kepada suatu atribut atau karakteristik spesifik yang dimiliki oleh peserta didik menurut aturan atau formulasi yang jelas dan terstruktur.

Dalam konteks ilmu pendidikan, pengukuran memiliki karakteristik filosofis yang khas dibandingkan dengan ilmu eksakta murni. Jika dalam ilmu fisika pengukuran bersifat langsung (*direct measurement*) seperti mengukur panjang meja dengan meteran maka dalam pendidikan, pengukuran bersifat tidak langsung (*indirect*

*measurement*). Objek yang diukur dalam pembelajaran matematika adalah konstruk laten atau tingkat kemampuan kognitif yang abstrak di dalam pikiran siswa. Kemampuan abstrak tersebut hanya dapat dikuantifikasi secara tidak langsung melalui indikator-indikator perilaku, yakni respons atau jawaban siswa terhadap pancingan berupa instrumen tes (Ali & Khaeruddin, 2012).

Lebih lanjut, esensi dari pengukuran adalah transformasi observasi menjadi data numerik (skor). Berdasarkan pandangan para ahli di atas, dapat disintesis bahwa proses pengukuran secara mutlak belum memberikan makna "baik" atau "buruk", "berhasil" atau "gagal" terhadap suatu capaian. Pengukuran murni hanya berfokus pada kuantifikasi matematis yang objektif tanpa melibatkan pertimbangan nilai (*value judgment*).

Sebagai ilustrasi konkret di sekolah, pengukuran terjadi tepat pada saat guru mengoreksi lembar jawaban ASAS peserta didik berdasarkan pedoman penskoran (*scoring rubric*) dan memperoleh angka mentah (*raw score*). Misalnya, seorang siswa berhasil menjawab benar 16 dari 20 soal matematika, maka hasil pengukurannya murni berupa angka 16 (Nursalam & Suardi, 2016). Skor tersebut bersifat

sangat netral dan belum bisa digunakan untuk justifikasi kelulusan sebelum ditafsirkan lebih lanjut. Oleh karena sifatnya yang menjadi penyuplai data kuantitatif pertama ini, alat ukur (instrumen tes) yang digunakan harus benar-benar presisi agar tidak menghasilkan bias data yang akan merusak proses evaluasi di tahap berikutnya.

b. Hakikat Penilaian (*Assessment*)

Berbeda dengan pengukuran yang berhenti pada penyajian data numerik murni, penilaian (*assessment*) bergerak lebih jauh ke ranah kualitatif yang selalu melibatkan pertimbangan nilai (*value judgment*). Dari sudut pandang kebijakan nasional, Permendikbudristek Nomor 21 Tahun 2022 mendefinisikan penilaian sebagai proses pengumpulan dan pengolahan informasi secara komprehensif untuk mengetahui kebutuhan belajar, profil kognitif, serta capaian perkembangan hasil belajar peserta didik (Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Nomor 21 Tahun 2022 Tentang Standar Penilaian Pendidikan., 2022). Di sisi lain, dari perspektif epistemologi evaluasi, Syaifudin (2023) mengonseptualisasikan penilaian sebagai serangkaian proses sistematis untuk menafsirkan angka-angka hasil

pengukuran (skor mentah) menjadi sebuah nilai atau predikat yang bermakna (*meaningful value*) melalui integrasi dengan kriteria atau acuan pembandingan tertentu.

Kompleksitas penilaian terletak pada proses pemaknaan data tersebut. Jika tahap pengukuran sekadar menjawab pertanyaan kuantitatif *How much?*"(berapa banyak skor atau item yang dijawab siswa dengan benar?), maka tahap penilaian memikul beban pedagogis untuk menjawab pertanyaan kualitatif *How good?*"(seberapa baik atau sejauh mana kualitas penguasaan kompetensi siswa tersebut?) (Ali & Khaeruddin, 2012). Untuk dapat menjawab pertanyaan esensial tersebut secara adil dan objektif, penilai (pendidik) mutlak membutuhkan tolok ukur yang terkalibrasi.

Dalam praksis pendidikan, penafsiran skor ini umumnya dieksekusi melalui dua pendekatan teoretis utama. Pertama, Penilaian Acuan Patokan (PAP) yang membandingkan skor aktual siswa dengan standar absolut yang telah ditetapkan secara institusional sebelumnya, seperti Kriteria Ketercapaian Tujuan Pembelajaran (KKTP). Kedua, Penilaian Acuan Norma (PAN) yang membandingkan skor seorang siswa dengan kurva

distribusi rata-rata kemampuan kelompok di kelasnya (Magdalena et al., 2020). Melalui mekanisme kognitif inilah, penilaian berfungsi sebagai "jembatan epistemologis" yang menerjemahkan data statistik yang kaku (hasil pengukuran) menjadi bahasa pedagogik yang deskriptif dan diagnostik, sebelum akhirnya digunakan sebagai landasan justifikasi pada tahap evaluasi.

c. Hakikat Evaluasi (*Evaluation*)

Evaluasi adalah tahapan tertinggi dan paling komprehensif yang bersifat pengambilan keputusan (*decision making*). Arikunto (2018) menegaskan bahwa evaluasi merupakan kegiatan yang mencakup mengukur dan menilai; kedua langkah tersebut mutlak harus dilalui sebelum suatu evaluasi dapat dilakukan secara utuh. Melengkapi definisi tersebut, Ali & Khaeruddin (2012) memandang evaluasi dari dimensi programatik, yakni suatu proses sistematis untuk menentukan sejauh mana tujuan instruksional telah dicapai oleh peserta didik, yang pada akhirnya digunakan untuk mengambil kebijakan strategis.

Lebih jauh lagi, evaluasi tidak hanya berpusat pada capaian akademis peserta didik, melainkan juga menilai efektivitas seluruh ekosistem

pembelajaran, mulai dari metode pengajaran, kesesuaian materi, hingga kualitas instrumen yang digunakan. Dengan kata lain, evaluasi memberikan umpan balik (*feedback*) yang komprehensif. Keputusan dari proses evaluasi ini memiliki dampak langsung, baik pada subjek didik, seperti penentuan ketuntasan belajar dan kelayakan naik kelas, maupun pada pendidik sebagai bahan refleksi untuk memperbaiki kualitas pedagogik pada masa mendatang (Magdalena et al., 2020).

d. Hubungan Antara Pengukuran, Penilaian, dan Evaluasi

Ketiga konsep di atas pengukuran, penilaian, dan evaluasi sering kali digunakan secara tumpang tindih oleh masyarakat awam maupun praktisi pendidikan. Padahal, secara teoretis, ketiganya memiliki hubungan hierarkis, sekuensial (berurutan), dan saling prasyarat. Pengukuran bertugas sebagai fondasi yang menyediakan data numerik yang objektif; penilaian bertugas menginterpretasikan data numerik tersebut ke dalam sebuah predikat kualitatif yang bermakna; sedangkan evaluasi berada di puncak hierarki, menggunakan informasi dari penilaian tersebut sebagai dasar justifikasi dan

pengambilan keputusan akhir (Ali & Khaeruddin, 2012; Syaifudin, 2023).

Dilihat dari ruang lingkupnya, evaluasi memiliki cakupan yang paling makro (luas), yang di dalamnya mewadahi proses penilaian dan pengukuran. Penilaian tidak dapat dilakukan tanpa adanya data hasil pengukuran yang presisi, dan evaluasi tidak dapat diputuskan tanpa adanya hasil penilaian yang akurat. Hal ini mengisyaratkan sebuah prinsip fundamental: kualitas suatu keputusan evaluasi sangat bergantung secara absolut pada kualitas alat ukurnya. Jika instrumen tes yang digunakan cacat (tidak valid/tidak reliabel), maka hasil pengukurannya akan bias, yang secara berantai akan menyebabkan penilaian menjadi keliru, dan akhirnya keputusan evaluasi menjadi tidak adil dan merugikan peserta didik (Magdalena et al., 2020).

Untuk memberikan gambaran yang lebih konkret dalam konteks pembelajaran, hubungan ketiganya dapat diilustrasikan sebagai berikut: Ketika guru memberikan instrumen ASAS Matematika dan mendapati seorang siswa berhasil menjawab 15 soal dengan benar dari total 20 soal, proses menghitung jawaban benar tersebut adalah tahap pengukuran

(menghasilkan angka murni 15). Selanjutnya, ketika guru mengonversi skor 15 tersebut menjadi nilai 75 dan membandingkannya dengan Kriteria Ketercapaian Tujuan Pembelajaran (KKTP) untuk menyimpulkan bahwa siswa tersebut "Tuntas" atau mendapat predikat "Baik", proses tersebut adalah penilaian. Terakhir, ketika dewan guru menggunakan kumpulan nilai tersebut (beserta aspek sikap dan keterampilan) untuk memutuskan bahwa siswa yang bersangkutan berhak "Naik ke Kelas VIII", proses pengambilan keputusan final itulah yang dinamakan evaluasi (Arikunto, 2018)..

e. Tujuan dan Fungsi Evaluasi Pembelajaran

Secara konseptual, tujuan utama dari evaluasi pembelajaran tidak sekadar menggugurkan kewajiban administratif guru dalam mengisi buku rapor. Tujuan esensialnya adalah untuk menghimpun informasi yang komprehensif dan objektif mengenai tingkat ketercapaian Tujuan Pembelajaran (TP) yang telah dirancang. Selain itu, evaluasi juga bertujuan untuk menelaah efektivitas strategi, model, dan media pembelajaran yang telah diimplementasikan oleh pendidik di dalam kelas (Pamulatsih & Zulfitriya, 2024).

Dalam konteks makro, evaluasi pendidikan berfungsi sebagai bentuk akuntabilitas publik dan transparansi sekolah kepada orang tua serta pemangku kepentingan (stakeholders) lainnya. Sementara itu, ditinjau dari fungsinya, para pakar (Arikunto, 2018; Pamulatsih & Zulfitria, 2023) mengklasifikasikan fungsi evaluasi pembelajaran ke dalam empat dimensi pokok yang saling terintegrasi:

1) Fungsi Pengukur Keberhasilan (Instruksional)

Fungsi ini berorientasi pada pencapaian kurikulum. Melalui evaluasi, guru dapat mengukur dan memetakan sejauh mana materi dalam hal ini konsep matematika telah diinternalisasi dan dikuasai oleh siswa.

2) Fungsi Diagnostik

Evaluasi yang baik mampu bertindak layaknya "sinar-X" (*X-ray*) yang mendiagnosis kelemahan kognitif atau kesulitan belajar spesifik siswa. Dalam pembelajaran matematika yang bersifat hierarkis, jika banyak siswa gagal pada satu butir soal analisis tertentu, guru dapat segera mendiagnosis adanya miskonsepsi

yang membutuhkan intervensi atau remediasi.

3) Fungsi Penempatan (*Placement*)

Data hasil evaluasi memungkinkan pendidik untuk memetakan profil kemampuan awal siswa. Hal ini sangat krusial dalam paradigma Kurikulum Merdeka, di mana guru dapat menggunakan data tersebut untuk merancang skenario pembelajaran berdiferensiasi (mengelompokkan siswa berdasarkan tingkat kesiapan belajarnya).

4) Fungsi Selektif dan Administratif

Pada dimensi administratif, evaluasi berfungsi sebagai dasar legitimasi kebijakan. Contoh konkret fungsinya adalah menyeleksi siswa yang berhak mendapat beasiswa, memutuskan ketuntasan belajar, menetapkan kelayakan naik jenjang ke kelas VIII, serta sebagai landasan penerbitan dokumen resmi Laporan Hasil Belajar (*rapor*) pada akhir semester.

f. Tes sebagai Instrumen Evaluasi

Dalam kajian evaluasi pendidikan, instrumen atau alat ukur secara umum diklasifikasikan menjadi dua kategori utama, yakni instrumen tes dan non-tes (seperti observasi, wawancara, dan

portofolio). Untuk mengukur domain kognitif atau tingkat penguasaan akademis secara presisi, khususnya pada mata pelajaran eksakta seperti matematika, instrumen yang paling mutlak dan representatif digunakan adalah tes. Arikunto (2018) mendefinisikan tes sebagai alat atau prosedur yang sistematis dan objektif untuk mengetahui atau mengukur sesuatu (baik berupa kemampuan, pemahaman, maupun bakat) dalam suasana, cara, dan aturan-aturan yang telah ditentukan secara ketat.

Sebagai instrumen formal pengumpul data, tes ibarat sebuah "timbangan analitik". Akurasi dari timbangan tersebut akan sangat menentukan keadilan dari setiap transaksinya. Dalam konteks pendidikan, kualitas dan objektivitas dari suatu keputusan evaluasi sangat bergantung secara absolut pada kualitas tes yang diselenggarakan. Jika instrumen tes tersebut disusun secara serampangan, cacat secara konstruk, atau tidak konsisten (tidak reliabel), maka data pengukuran yang dihasilkan akan sangat bias. Alih-alih memetakan kemampuan kognitif siswa secara adil, penggunaan tes yang buruk justru akan menghasilkan penilaian yang sangat menyesatkan dan pada akhirnya

merugikan pihak peserta didik (Masullah et al., 2024).(Sukma et al., 2025)

Lebih lanjut, agar sebuah tes layak disebut sebagai instrumen evaluasi yang standar, ia harus memiliki karakteristik psikometrik yang teruji secara empiris (Ali & Khaeruddin, 2012). Sebuah instrumen tes tidak boleh hanya mengandalkan intuisi atau asumsi subjektif guru pembuatnya (*face validity*), melainkan harus dikalibrasi dan dibuktikan melalui analisis butir soal untuk mendeteksi indeks kesukarannya, efektivitas daya bedanya, serta fungsi opsi pengecohnya. Hal inilah yang mendasari mengapa pengujian kualitas empiris instrumen tes menjadi prasyarat wajib sebelum hasil evaluasi tersebut disahkan sebagai nilai akhir.

g. Asesmen Sumatif Akhir Semester (ASAS)

Seiring dengan berlakunya Kurikulum Merdeka, terjadi pergeseran paradigma dalam pelaksanaan evaluasi pendidikan dasar dan menengah di Indonesia. Terminologi Penilaian Akhir Semester (PAS) kini bertransformasi menjadi Asesmen Sumatif Akhir Semester (ASAS). Perubahan nomenklatur ini mencerminkan pergeseran filosofis dari sekadar "memberikan nilai akhir"

menjadi proses "memetakan ketercapaian" kompetensi secara utuh berdasarkan Standar Capaian Pembelajaran (CP) yang telah ditetapkan secara institusional (Sukma et al., 2025).

Secara teoretis, terdapat perbedaan esensial antara asesmen formatif dan asesmen sumatif. Putri & Zakir (2023) menjelaskan bahwa asesmen formatif berkedudukan sebagai *assessment for learning* (penilaian untuk perbaikan proses belajar), sedangkan ASAS berkedudukan murni sebagai *assessment of learning* (penilaian atas hasil belajar). Artinya, ASAS tidak dirancang untuk memperbaiki metode di tengah jalan, melainkan dirancang secara khusus untuk merekam, menyimpulkan, dan mengkuantifikasi pencapaian kognitif peserta didik secara sistematis pada akhir periode pembelajaran (akhir semester).

Karena ASAS menghasilkan nilai akhir yang tertera pada dokumen rapor dan secara langsung menentukan kelayakan siswa untuk naik kelas atau lulus jenjang, asesmen ini diklasifikasikan sebagai ujian berisiko tinggi (*high-stakes testing*). Mengingat besarnya dampak dari hasil ASAS tersebut terhadap masa depan siswa, pemerintah melalui Permendikbudristek

Nomor 21 Tahun 2022 mewajibkan agar pelaksanaannya mematuhi lima prinsip fundamental penilaian, yakni: (1) Adil (tidak diskriminatif terhadap kondisi siswa), (2) Objektif (terbebas dari bias subjektivitas), (3) Komprehensif (menyeluruh), (4) Sistematis (terencana sesuai pedoman baku), dan (5) Akuntabel (dapat dipertanggungjawabkan metode dan hasilnya) (Nurhalifah et al., 2024; Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Nomor 21 Tahun 2022 Tentang Standar Penilaian Pendidikan., 2022).

Konsekuensi logis dari tuntutan prinsip objektif dan akuntabel tersebut adalah bahwa sekolah tidak boleh menggunakan instrumen soal yang disusun secara acak. Sebuah instrumen ASAS baru dapat dikatakan akuntabel dan berkeadilan jika telah terbukti secara empiris memiliki indeks kesukaran yang proporsional, daya pembeda yang tajam, serta reliabilitas yang tinggi. Hal inilah yang menjadi landasan rasional mengapa penelitian mengenai kualitas butir soal ASAS matematika menjadi sangat relevan dan mendesak untuk dilakukan.

Sintesis: Evaluasi dalam Konteks Penelitian

Berdasarkan paparan di atas, penelitian ini memosisikan ASAS Matematika

Kelas VII SMP Negeri 6 Pekalongan sebagai objek evaluasi. Karena ASAS berfungsi sebagai instrumen tes penentu keberhasilan siswa (Evaluasi), maka instrumen tersebut harus melalui proses verifikasi kualitas. Penelitian ini hadir untuk menjembatani kebutuhan tersebut dengan menganalisis kualitas butir soal ASAS menggunakan parameter yang akurat.

### 2.1.2 Kualitas Instrumen Tes (Teori Tes Klasik)

Setelah memahami hakikat evaluasi, bagian ini membahas metode untuk menguji kualitas instrumen tersebut menggunakan pendekatan Teori Tes Klasik dengan mengacu pada standar kriteria evaluasi pendidikan dari Suharsimi Arikunto.

#### a) Konsep Teori Tes Klasik (*Classical Test Theory / CTT*)

Analisis kualitas butir soal dalam penelitian ini didasarkan pada pendekatan Teori Tes Klasik. Inti dari teori ini adalah model skor murni, yang menyatakan bahwa skor tampak (*Observed Score*) yang diperoleh siswa adalah penjumlahan dari skor murni (*True Score*) dan skor kesalahan (*Error*), yang dirumuskan sebagai  $X = T + E$  (Sheptian et al., 2025); (Ayanwale et al., 2013). Tujuan utama analisis butir soal dalam kerangka CTT adalah untuk meminimalkan measurement error (*E*) sehingga skor tampak (*X*) sedekat mungkin

mencerminkan kemampuan asli siswa ( $T$ ) (Syarifudin, 2023).

Pemilihan pendekatan ini didasarkan pada pertimbangan bahwa CTT lebih aplikatif untuk analisis tes hasil belajar di tingkat kelas atau sekolah dengan ukuran sampel yang tidak terlalu besar (Naga, 1992 dalam Karindi & Ruffi'i, 2025). Berbeda dengan Teori Respon Butir (IRT) yang menuntut sampel besar dan asumsi matematis yang kompleks, CTT memberikan informasi yang langsung dapat diinterpretasikan guru untuk perbaikan pembelajaran, seperti tingkat kesukaran dan daya pembeda klasikal.

b) Validitas Tes

Validitas berasal dari kata *validity* yang secara terminologi bermakna sejauh mana ketepatan dan kecermatan suatu instrumen dalam menjalankan fungsi ukurnya. Menurut (Arikunto, 2018), sebuah tes dikatakan memiliki validitas yang tinggi apabila tes tersebut secara presisi mengukur apa yang memang hendak diukur. Dalam kajian evaluasi pendidikan yang komprehensif, parameter validitas tidak bersifat tunggal, melainkan dipecah menjadi dua dimensi utama, yaitu validitas logis (rasional) dan validitas empiris (statistik).

1) Validitas Logis (Rasional)

Validitas logis diperoleh melalui penalaran dan analisis teoretis yang mendalam oleh pakar (expert judgment) sebelum tes tersebut diujikan kepada siswa. Dimensi ini mencakup dua hal:

a) Validitas Isi (Content Validity): Parameter yang menuntut agar butir-butir soal di dalam tes ASAS Matematika

merepresentasikan secara proporsional materi yang telah diajarkan. Validitas isi dibuktikan melalui keselarasan antara butir soal dengan Kisi-Kisi dan Standar Capaian Pembelajaran (CP) kurikulum yang berlaku (Nisa & Hidayati, 2023).

b) Validitas Konstruk (Construct Validity): Berkaitan dengan aspek kognitif dan psikologis. Dalam matematika, validitas konstruk memastikan bahwa instrumen tersebut benar-benar menguji kompetensi logis, analitis, dan pemecahan masalah spasial/numerik siswa,

bukan malah menguji kemampuan linguistik (membaca) siswa akibat redaksi kalimat soal yang terlalu panjang, ambigu, atau membingungkan.

## 2) Validitas Empiris (Validitas Butir Soal)

Validitas empiris adalah validitas yang dibuktikan melalui data kuantitatif dari hasil uji coba di lapangan. Dalam Teori Tes Klasik (CTT) yang diterapkan pada algoritma perangkat lunak ANATES, konsep validitas yang digunakan adalah validitas butir (item validity). Validitas butir ditunjukkan secara statistik oleh indeks korelasi antara skor per butir soal dengan skor total tes secara keseluruhan secara komputasi.

a)) Dalam menganalisis validitas butir soal pada tes objektif (pilihan ganda) yang menerapkan skor 1 dan 0, peneliti menggunakan rumus korelasi Point Biserial.

b)) Untuk soal bentuk uraian/subjektif (polikotomi yang diskor secara bertingkat), analisis validitas butir dihitung menggunakan teknik korelasi Pearson Product Moment. Butir

soal yang terbukti valid secara empiris (nilai  $r_{hitung} > r_{tabel}$ ) mencerminkan bahwa soal tersebut konsisten dan selaras dengan arah pengukuran instrumen secara utuh. Sebaliknya, jika suatu butir soal memiliki korelasi yang rendah atau negatif, soal tersebut divonis cacat karena memberikan informasi yang bertolak belakang dengan kemampuan aktual peserta didik (Magdalena et al., 2020).

c) Reliabilitas Instrumen Tes

Jika validitas berbicara mengenai "ketepatan" arah sasaran pengukuran, maka reliabilitas berbicara mengenai "keajegan" (consistency) dan "keterandalan" (dependability) dari alat ukur tersebut. Arikunto (2018) mendefinisikan reliabilitas sebagai indikator krusial yang menunjukkan sejauh mana suatu instrumen dapat dipercaya sebagai alat pengumpul data yang stabil.

1) Konsep Filosofis Reliabilitas dalam CTT

Dari sudut pandang metodologis yang lebih kompleks, reliabilitas menuntut suatu kondisi di mana apabila tes yang sama diujikan kembali kepada subjek yang sama

pada waktu dan ruang yang berbeda (dengan asumsi kemampuan subjek statis), maka tes tersebut akan memproduksi skor yang konstan. Dalam kerangka Teori Tes Klasik ( $X = T + E$ ), reliabilitas secara matematis didefinisikan sebagai proporsi varians skor murni (*True Score Variance*) terhadap varians skor tampak (*Observed Score Variance*). Semakin kecil unsur Error (E) dalam suatu tes, maka koefisien reliabilitasnya akan semakin mendekati angka absolut 1,00 (Yustiandi & Saepuzaman, 2024).

2) Pendekatan Konsistensi Internal (*Internal Consistency*)

Mengingat instrumen ASAS di sekolah menengah umumnya hanya diujikan satu kali (*single administration*), pendekatan tes ulang (*test-retest*) sulit untuk dilakukan. Oleh karena itu, pendekatan yang digunakan untuk menguji keajegan tes adalah uji Konsistensi Internal. Pada pendekatan ini, analisis reliabilitas bertumpu pada tingkat kehomogenan antar-butir soal di dalam tes itu sendiri.

a)) ada instrumen soal pilihan ganda (dikotomi), keandalan tes umumnya diestimasi menggunakan formula Kuder-Richardson (K-R 20 atau K-R 21).

b)) Pada instrumen soal uraian (polikotomi), estimasi reliabilitas digunakan menggunakan formula Cronbach's Alpha (Surbakti, 2025).

### 3) Faktor Reduksi Reliabilitas

Tingkat reliabilitas sebuah instrumen sangat dipengaruhi oleh berbagai variabel dinamis, di antaranya: panjang tes (jumlah soal yang diberikan), tingkat penyebaran kemampuan (variance) peserta didik, serta tingkat objektivitas dalam pedoman penskoran. Khusus pada tes objektif seperti pilihan ganda, kejegan hasil tes sangat rentan tereduksi oleh tingginya faktor tebakan acak (guessing factor) dari kelompok siswa yang tidak menguasai materi.

Oleh karena itu, pembuktian reliabilitas secara statistik mutlak diperlukan untuk memastikan bahwa skor tinggi yang diraih siswa murni berasal dari

pemahaman logis matematika mereka, bukan karena kebetulan. Hal ini sejalan dengan dalil aksiomatik dalam evaluasi pendidikan yang menyatakan bahwa: "Sebuah instrumen tes yang valid secara logis dan empiris pada dasarnya pasti memiliki reliabilitas yang baik, namun instrumen tes yang memiliki reliabilitas tinggi belum tentu terjamin validitasnya" (Arikunto, 2018).

d) Karakteristik Butir Soal

Analisis butir soal dalam CTT juga mencakup parameter karakteristik soal, yaitu:

1) Tingkat Kesukaran (*Index of Difficulty*)

Dalam kerangka evaluasi empiris, tingkat kesukaran bermakna sejauh mana proporsi peserta didik yang mampu menjawab suatu butir soal dengan benar. Asumsi dasarnya adalah sebuah instrumen tes yang bermutu harus terdistribusi mengikuti kurva normal (Arikunto, 2018). Yustiandi & Saepuzaman (2024) mengemukakan rasionalisasi empiris bahwa indeks kesukaran butir yang paling ideal terletak pada rentang proporsional, yaitu antara 0,30 sampai dengan 0,70 (kategori Sedang).

Secara psikologis dan akademis, jika soal terlampau

mudah (indeks mendekati 1,00), maka tes tersebut gagal memberikan tantangan intelektual dan tidak akan menghasilkan varians skor (semua siswa mendapat nilai rata tinggi). Sebaliknya, jika tes terlampau sukar (indeks mendekati 0,00), hal tersebut dapat memicu keputusan akademis (frustrasi) pada siswa dan gagal memetakan capaian pembelajaran yang sebenarnya (Syaifudin, 2023).

Dalam pengembangan instrumen evaluasi yang berkualitas, distribusi tingkat kesukaran butir soal harus diperhatikan agar tes tersebut mampu memberikan gambaran kemampuan siswa yang objektif. Sebaran tingkat kesukaran yang ideal sering kali dikaitkan dengan prinsip Kurva Normal, di mana jumlah soal tidak boleh menumpuk pada satu kategori saja (terlalu mudah atau terlalu sukar).

Menurut Anas (2019), sebuah tes yang baik sebaiknya memiliki proporsi sebaran tingkat kesukaran yang seimbang. Standar yang sering digunakan dalam dunia pendidikan adalah perbandingan 3:5:2, dengan rincian sebagai berikut:

- a)) 30% soal kategori Mudah:  
Berfungsi untuk memberikan

motivasi kepada siswa agar tidak merasa putus asa dalam menghadapi tes.

b)) 50% soal kategori Sedang: Berfungsi sebagai inti dari pengukuran kompetensi dasar yang harus dicapai oleh mayoritas siswa.

c)) 20% soal kategori Sukar: Berfungsi untuk mendeteksi atau menyaring siswa yang memiliki kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skills*).

## 2) Daya Pembeda (*Discrimination Index*)

Daya pembeda merupakan parameter analitis yang sangat krusial dalam menentukan kualitas butir soal. Parameter ini menunjukkan tingkat efektivitas suatu soal dalam memisahkan (mendiskriminasi) demografi peserta didik yang telah mencapai ketuntasan belajar (kelompok atas / *upper group*) dari peserta didik yang belum tuntas (kelompok bawah / *lower group*) (Arikunto, 2018).

Suatu butir soal dikatakan memiliki daya diskriminasi yang sangat efektif (positif dan tinggi) apabila soal tersebut terbukti mampu dijawab secara tepat oleh

mayoritas peserta didik berkemampuan tinggi, namun mengecoh mayoritas peserta didik yang berkemampuan rendah (Elgadal & Mariod, 2021). Sebaliknya, butir soal yang memiliki daya pembeda negatif sangatlah berbahaya bagi validitas evaluasi karena bersifat menyesatkan (*misleading*). Soal semacam ini justru mengecoh siswa dari kelompok atas dan menguntungkan siswa dari kelompok bawah yang menebak secara acak (*blind guessing*).

### 3) Efektivitas Pengecoh (*Distractor Efficiency*)

Parameter ini berlaku secara eksklusif untuk tes bentuk objektif (seperti pilihan ganda). Konstruksi instrumen pilihan ganda yang baik tidak hanya bergantung pada rumusan jawaban yang benar (kunci), melainkan juga pada fungsi opsi jawaban yang salah (distraktor). Menurut (Ali & Khaeruddin, 2012), menyusun opsi pengecoh yang logis adalah aspek kognitif tersulit dalam pembuatan tes.

Pengecoh yang ideal secara psikologis harus terlihat masuk akal (*plausible*) dan memikat bagi siswa

yang tidak memahami materi (Radja et al., 2023). Berdasarkan pakem standar evaluasi dari Arikunto (2018), efektivitas distraktor terpenuhi jika ia mampu memikat sedikitnya 5% dari total peserta tes, dan secara dominan menyetatkan siswa dari kelompok bawah. Pengecoh yang tidak dipilih sama sekali (0%) merupakan pengecoh yang cacat konstruksi karena kesalahannya terlalu mencolok, sehingga memudahkan siswa menebak kunci jawaban.

### 2.1.3 Evaluasi Pembelajaran Matematika Berbantuan Teknologi

- a. Evaluasi dalam Pembelajaran Matematika Matematika merupakan disiplin ilmu eksakta yang berdiri di atas konstruksi konsep abstrak, pembuktian aksiomatik, dan struktur logis yang bersifat hierarkis. Mengingat karakteristik epistemologisnya yang kompleks, (Syarifudin, 2023) berpendapat bahwa evaluasi hasil belajar matematika tidak boleh direduksi sekadar menjadi alat uji hafalan rumus atau kemampuan operasi aritmetika dasar (berhitung prosedural). Evaluasi matematika tingkat menengah harus mampu menjangkau dimensi kognitif tingkat tinggi (*Higher Order Thinking Skills / HOTS*), yang mencakup kemampuan penalaran logis, representasi

spasial, koneksi antar-konsep, dan pemecahan masalah (*problem solving*).

Lebih jauh lagi, karena hierarki keilmuan matematika bersifat prasyarat dan akumulatif (penguasaan materi baru sangat bergantung pada penguasaan materi sebelumnya), maka evaluasi dalam matematika mengemban fungsi diagnostik yang sangat esensial. Sebuah instrumen asesmen matematika yang berkualitas harus mampu memetakan secara spesifik di mana letak miskonsepsi atau hambatan epistemologis siswa. Hasil evaluasi harus bisa membedakan apakah seorang siswa melakukan kesalahan pada tahap prosedural (kesalahan komputasi/hitung), atau justru mengalami kegagalan pada tahap konseptual (kegagalan menerjemahkan masalah nyata ke dalam model matematika yang tepat).

Kompleksitas tuntutan kognitif inilah yang menyebabkan instrumen evaluasi matematika sangat rentan terhadap bias pengukuran. Sebagai contoh, pada soal pemecahan masalah (soal cerita), redaksi kalimat yang ambigu, terlalu panjang, atau menggunakan diksi yang tidak familier dapat mengubah tes matematika tersebut menjadi sekadar "tes kemampuan membaca". Hal ini secara langsung mencederai validitas konstruk instrumen (Masullah et al., 2024). Oleh karena itu, instrumen asesmen matematika

yang disusun secara serampangan hanya akan memberikan vonis nilai yang tidak adil bagi siswa. Di sinilah analisis kualitas butir soal melalui Teori Tes Klasik menjadi langkah preventif mutlak untuk memvalidasi bahwa instrumen tersebut benar-benar murni mengukur kompetensi matematis..

b. Perangkat Lunak ANATES (Analisis Tes) Versi 4.0.9

Dalam melakukan analisis butir soal secara empiris menggunakan Teori Tes Klasik (CTT), proses perhitungan manual seringkali membutuhkan waktu yang lama dan rentan terhadap kesalahan (human error), terlebih jika jumlah sampel peserta tes dan butir soalnya banyak. Untuk mengatasi kendala teknis tersebut, pemanfaatan teknologi informasi menjadi solusi yang sangat relevan. Salah satu perangkat lunak (*software*) yang dirancang khusus untuk keperluan evaluasi pendidikan di Indonesia adalah ANATES (Analisis Tes).

ANATES merupakan program aplikasi komputer yang dikembangkan dan diprogramkan secara spesifik oleh Drs. Karno To, M.Pd. dan Prof. Dr. Yaya Sunarya, M.Pd. (Karno To, 1996). Perangkat lunak ini memiliki antarmuka yang berbahasa Indonesia sehingga sangat user-friendly (mudah digunakan) oleh para guru maupun praktisi pendidikan.

ANATES dikembangkan dalam dua versi utama, yaitu versi untuk soal bentuk Pilihan Ganda (*Multiple Choice*) dan versi untuk soal bentuk Uraian (*Essay*).

Penggunaan perangkat lunak ANATES memiliki beberapa keunggulan utama dalam kegiatan evaluasi pembelajaran, antara lain (Karindi & Rufi'i, 2025; Sheptian et al., 2025):

- 1) Otomatisasi Perhitungan: Mampu melakukan proses komputasi statistik yang rumit dalam hitungan detik setelah data jawaban mentah (*raw data*) diinputkan.
- 2) Analisis Komprehensif: Program ini mampu menghasilkan output analisis yang meliputi seluruh parameter Teori Tes Klasik secara serentak, yaitu: perhitungan skor (asli dan dibobot), reliabilitas tes, pengelompokan subjek (kelompok atas dan bawah), tingkat kesukaran, daya pembeda, korelasi skor butir dengan skor total (validitas), serta kualitas pengecoh (*distractor*).
- 3) Akurasi Tinggi: Mengurangi bias dan kesalahan hitung (*human error*) yang sering terjadi pada perhitungan manual.

Dalam penelitian ini, perangkat lunak yang digunakan adalah ANATES Versi 4.0.9. Pemilihan versi ini didasarkan pada tingkat kestabilan program dan relevansinya untuk menganalisis instrumen asesmen yang

mencakup soal pilihan ganda, benar-salah (dikotomis), maupun soal uraian (polikotomis).

## 2.2 Kajian Penelitian Yang Relevan

Untuk menghindari duplikasi dan memastikan keaslian penelitian, peneliti melakukan tinjauan terhadap beberapa karya ilmiah terdahulu yang relevan dengan topik penelitian. Berikut adalah beberapa penelitian yang dijadikan acuan:

a. Afifah (2020) dalam skripsinya yang berjudul "Analisis Kualitas Butir Soal Penilaian Akhir Semester Gasal Mata Pelajaran Matematika Kelas IX SMP Negeri 3 Ajibarang Kabupaten Banyumas". Dengan menggunakan ANATES V4, penelitian tersebut menelaah kualitas butir soal PAS Matematika berdasarkan validitas, reliabilitas, indeks kesukaran, daya diskriminasi, dan kualitas pengecoh. Kesimpulan yang diperoleh menunjukkan bahwa secara umum kualitas instrumen berada pada kategori cukup baik (62,5%), meskipun masih ditemukan anomali berupa butir soal tidak valid dan distraktor yang tidak efektif.

- 1) Persamaan: Penelitian ini memiliki korelasi metodologis yang erat dengan penelitian yang akan dilaksanakan, di mana keduanya menganalisis kualitas instrumen Matematika SMP kelas IX secara kuantitatif melalui bantuan aplikasi ANATES.
- 2) Perbedaan: Perbedaan terletak pada lokasi penelitian (SMP Negeri 3 Ajibarang vs SMP Negeri 6 Kota Pekalongan) dan istilah

asesmen yang digunakan saat itu (PAS vs ASAS).

- 3) Kebaruan (*Novelty*): Penelitian ini menggunakan instrumen yang dikembangkan berdasarkan Kurikulum Merdeka (ASAS), sedangkan penelitian Nur Afifah masih mengacu pada Kurikulum 2013 (PAS). Selain itu, penelitian ini menganalisis tiga bentuk soal (Pilihan Ganda, Benar-Salah, Uraian), sedangkan Nur Afifah hanya fokus pada Pilihan Ganda.
- b. Kurnia (2021) dalam skripsinya yang berjudul "Analisis Butir Soal UAS MGMP Matematika dengan Menggunakan Software Anates". Kajian ini berfokus pada evaluasi kualitas soal UAS Matematika susunan tim MGMP. Temuan riset mengindikasikan bahwa keterlibatan MGMP belum sepenuhnya menjamin kesempurnaan instrumen, dibuktikan dengan masih ditemukannya sejumlah soal yang tidak valid dan koefisien reliabilitas yang tergolong rendah.
- 1) Persamaan: Fokus utama penelitian adalah analisis butir soal matematika menggunakan bantuan software ANATES untuk mengetahui karakteristik empiris soal (validitas, reliabilitas, dll).
  - 2) Perbedaan: Objek penelitian Linda Kurnia adalah soal kolektif buatan MGMP yang digunakan bersama di beberapa sekolah, sedangkan penelitian ini berfokus pada soal ASAS yang digunakan secara spesifik di

satu satuan pendidikan, yaitu SMP Negeri 6 Kota Pekalongan.

- 3) Kebaruan (*Novelty*): Penelitian ini memberikan analisis mendalam terhadap kualitas soal buatan guru sekolah (*teacher-made test*) yang seringkali luput dari evaluasi kualitas dibandingkan soal standar MGMP.
- c. Utami (2025) dalam skripsinya yang berjudul "Analisis Butir Soal Asesmen Sumatif Akhir Semester (ASAS) Mata Pelajaran Matematika di SMP Al Fadllu Brangsong Kendal Pada Tahun Ajaran 2024/2025". Penelitian ini mengkaji kualitas soal ASAS Matematika di tingkat SMP. Hasil analisis menunjukkan variasi kualitas soal, di mana sebagian soal sudah memenuhi kriteria baik namun sebagian lain perlu direvisi, terutama pada aspek daya pembeda.
- 1) Persamaan: Memiliki kesamaan topik yang sangat spesifik yaitu analisis butir soal pada "Asesmen Sumatif Akhir Semester (ASAS)" mata pelajaran Matematika di jenjang SMP, serta tahun ajaran yang relatif baru.
  - 2) Perbedaan: Perbedaan utama terletak pada lokasi penelitian (SMP Al Fadllu Brangsong Kendal vs SMP Negeri 6 Kota Pekalongan) dan fokus penggunaan software (Penelitian Utami menggunakan pendekatan deskriptif umum, sedangkan penelitian ini secara spesifik menekankan penggunaan dan keunggulan fitur software ANATES 4.0.9).

- 3) Kebaruan (*Novelty*): Penelitian ini secara spesifik menguji efektivitas penggunaan ANATES versi terbaru (4.0.9) dalam menangani analisis soal campuran (objektif dan uraian) secara simultan, yang belum dibahas secara mendalam dalam penelitian Utami.
- d. Masullah et al. (2024) dalam jurnalnya "Analisis Butir Soal Ujian Akhir Semester Mata Pelajaran Matematika SMP Negeri 6 Praya Timur". Penelitian ini menemukan bahwa kualitas soal matematika SMP masih tergolong kurang baik, di mana terdapat 25 butir soal yang dinyatakan tidak valid dan perlu diperbaiki. Penelitian ini juga menyoroti pentingnya analisis butir soal untuk perbaikan pembelajaran.
  - a) Persamaan: Sama-sama meneliti kualitas soal matematika pada jenjang SMP (Kelas VII) yang mencakup aspek validitas, reliabilitas, dan tingkat kesukaran.
  - b) Perbedaan: Pendekatan manual dan wawancara yang diterapkan dalam pengumpulan data pada studi Masullah et al. menjadi garis pembeda utama. Sebagai bentuk inovasi teknis, penelitian ini mengintegrasikan penggunaan perangkat lunak ANATES 4.0.9 untuk meminimalisasi margin kesalahan (human error) sekaligus mempercepat proses analisis kuantitatifnya.
  - c) Kebaruan (*Novelty*): Penelitian ini menawarkan solusi metodologis yang lebih efisien melalui penggunaan teknologi

- (ANATES) dibandingkan metode manual yang digunakan Masullah, sehingga memberikan rekomendasi praktis bagi guru dalam melakukan analisis soal secara cepat.
- e. Nisa & Hidayati (2023) dalam jurnalnya "Analisis Butir Soal Penilaian Akhir Semester Matematika". Penelitian ini menganalisis 20 soal pilihan ganda dan 5 uraian PAS Matematika Kelas VII. Hasilnya menunjukkan reliabilitas tes yang tinggi (0,795), namun masih ditemukan soal dengan daya pembeda yang kurang baik.
- 1) Persamaan: Menganalisis instrumen PAS Matematika tingkat SMP, baik soal pilihan ganda maupun uraian, yang meliputi aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda.
  - 2) Perbedaan: Dari segi pengolahan data kuantitatif, penelitian Nisa dan Hidayati menggunakan software statistik umum berskala besar (SPSS 26). Sebaliknya, riset ini berfokus pada pemanfaatan ANATES Versi 4.0.9, sebuah aplikasi khusus (*dedicated software*) yang terbukti lebih praktis dan instan dalam membedah parameter psikometrik butir soal secara klasikal.
  - 3) Kebaruan (*Novelty*): Penelitian ini membandingkan penggunaan ANATES yang lebih spesifik untuk analisis butir soal (*item analysis*) dibandingkan SPSS yang merupakan software statistik umum, serta penerapannya pada konteks Kurikulum Merdeka.

f. Karindi & Rofi'i (2025) dalam jurnalnya "Penggunaan ANATES 4.0.9 dalam menguji kelayakan soal pilihan ganda sumatif matematika". Penelitian ini menunjukkan efektivitas penggunaan ANATES 4.0.9 dalam memetakan kualitas soal yang layak masuk ke dalam bank soal. Hasilnya merekomendasikan penggunaan software ini bagi guru untuk mempermudah evaluasi.

- 1) Persamaan: Penggunaan alat analisis spesifik (ANATES 4.0.9) dan fokus pada soal sumatif matematika.
- 2) Perbedaan: Penelitian Karindi dilakukan di jenjang SMA, sementara penelitian ini dilakukan di jenjang SMP, sehingga konteks materi dan karakteristik siswa berbeda.
- 3) Kebaruan (*Novelty*): Penelitian ini memperluas aplikasi ANATES 4.0.9 pada jenjang pendidikan dasar (SMP) dan pada jenis instrumen tes yang lebih variatif (tidak hanya pilihan ganda), memberikan wawasan baru tentang fleksibilitas software tersebut.

### **2.3 Kerangka Berpikir**

Kerangka berpikir dalam penelitian ini dirancang sebagai arsitektur logika untuk mengevaluasi kualitas instrumen Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII di SMP Negeri 6 Kota Pekalongan. Landasan utama dari kerangka ini adalah kesadaran bahwa instrumen evaluasi yang tidak terkalibrasi secara empiris berpotensi menghasilkan bias penilaian yang mencederai prinsip keadilan kognitif bagi peserta didik.

Alur berpikir penelitian ini secara sistematis dipetakan ke dalam empat fase integratif:

1. Fase Identifikasi Masalah dan Objek (Input)  
Proses dimulai dengan mengidentifikasi instrumen ASAS Matematika Kelas VII yang terdiri dari tiga modul variatif: Pilihan Ganda (10 butir), Benar-Salah (5 butir), dan Uraian (5 butir). Masalah fundamental yang diangkat adalah ketiadaan data empiris mengenai kelayakan psikometrik soal tersebut karena sifatnya yang merupakan teacher-made test yang langsung digunakan tanpa melalui proses uji coba (try-out).
2. Fase Analisis Psikometrik Kuantitatif (CTT)  
Pada tahap ini, peneliti menggunakan pisau bedah Teori Tes Klasik (CTT) dengan bantuan perangkat lunak ANATES V4.0.9. Analisis dilakukan secara terpisah berdasarkan tipe penskoran (dikotomi untuk PG/BS dan polikotomi untuk Uraian). Output yang dihasilkan berupa data numerik mengenai lima parameter kunci: validitas butir, koefisien reliabilitas, indeks kesukaran, daya pembeda, dan efektivitas pengecoh. Data ini memberikan profil statistik mengenai stabilitas dan fungsi mekanis instrumen.
3. Fase Pendalaman Pedagogis (Analisis Kualitatif)  
Sesuai dengan paradigma Pembelajaran Mendalam (Deep Learning) 2025, angka statistik tidak dipandang sebagai titik henti, melainkan sebagai "sinyal" untuk melakukan interpretasi makna. Menggunakan model interaktif Miles, Huberman, dan Saldaña, peneliti melakukan

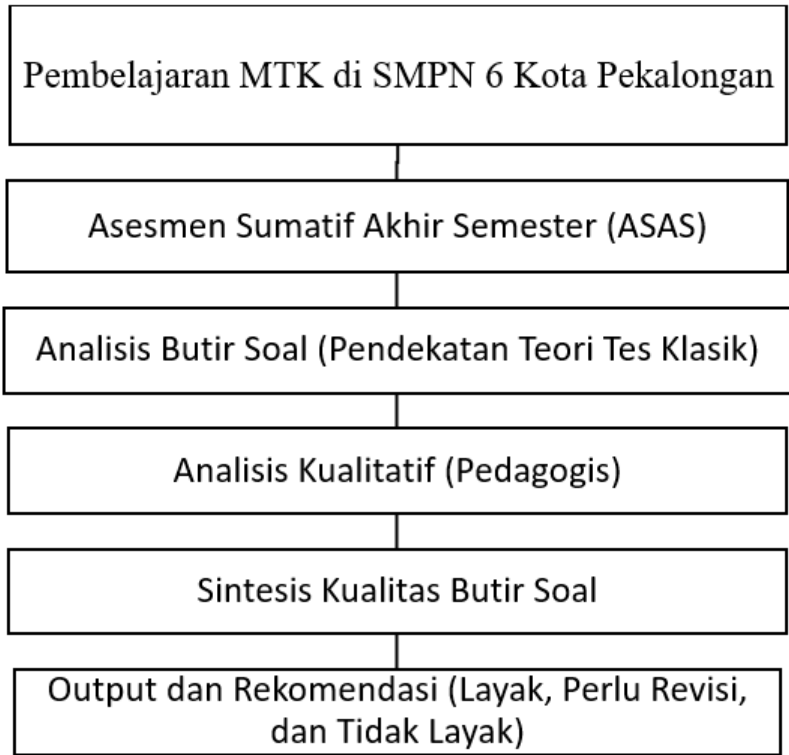
penelaahan mendalam terhadap naskah soal asli. Fokus analisis kualitatif diarahkan untuk mengungkap kompleksitas yang tersembunyi, meliputi:

- a. Analisis Hambatan Epistemologis: Menggali miskonsepsi aljabar yang dipicu oleh redaksi soal.
- b. Evaluasi Beban Kognitif (*Cognitive Load*): Menelaah kesesuaian tingkat kesulitan naskah dengan kapasitas kognitif siswa Kelas VII.
- c. Analisis Literasi Numerasi: Membedah efektivitas stimulus teks (seperti pada soal Benar-Salah) dalam memancing nalar kritis siswa.

#### 4. Fase Sintesis dan Rekomendasi (Output)

Tahap akhir adalah melakukan Triangulasi Teknik antara temuan statistik dan bedah dokumen naskah. Sintesis ini menghasilkan klasifikasi kualitas soal yang presisi: butir soal mana yang "Sangat Baik" untuk diarsipkan, mana yang "Bermasalah" (seperti fenomena kognitif redundan atau pengecoh menyesatkan) yang membutuhkan revisi redaksional, serta mana yang harus diganti.

Melalui alur kerangka berpikir ini, penelitian diharapkan mampu memberikan kontribusi nyata bagi SMP Negeri 6 Kota Pekalongan dalam membangun Bank Soal Sekolah yang tervalidasi, objektif, dan akuntabel, guna mendukung terciptanya ekosistem asesmen yang berkualitas dalam kerangka Kurikulum Merdeka. Hubungan antar variabel tersebut digambarkan dalam bagan berikut:



*Gambar 2. 1 Kerangka Berpikir Analisis Kualitas Soal Matematika*

## **BAB V**

### **PENUTUP**

#### **5.1 Simpulan**

Berdasarkan hasil analisis data dan pembahasan mendalam mengenai kualitas butir soal Asesmen Sumatif Akhir Semester (ASAS) Matematika Kelas VII di SMPN 6 Kota Pekalongan Tahun Ajaran 2025/2026, maka dapat ditarik simpulan sebagai berikut:

- a. **Tingkat Validitas:** Seluruh butir soal (100%) pada ketiga bentuk tes dinyatakan valid. Hal ini membuktikan bahwa instrumen secara substantif telah memiliki keselarasan yang kuat dengan Capaian Pembelajaran (CP) kurikulum yang berlaku dan mampu memotret indikator kompetensi matematika siswa secara presisi.
- b. **Tingkat Reliabilitas:**
  1. Instrumen soal Pilihan Ganda (0,83) dan Uraian (0,90) memiliki tingkat reliabilitas yang sangat tinggi. Hal ini menunjukkan stabilitas pengukuran yang sangat baik, terutama pada soal uraian yang didukung oleh rubrik penskoran analitik yang objektif.
  2. Instrumen soal Benar-Salah memiliki reliabilitas sangat rendah (0,09). Rendahnya angka ini dipicu oleh tingginya peluang tebakan (guessing factor) sebesar 50% dan adanya ketergantungan stimulus berantai antar nomor soal yang mengganggu stabilitas skor.

- c. Karakteristik Butir Soal secara Pedagogis:
1. Tingkat Kesukaran: Secara empiris sudah proporsional, namun pada soal uraian menunjukkan profil yang sangat ideal dengan 100% soal berkategori sedang. Hal ini efektif dalam menjaga motivasi akademik siswa tanpa mengurangi tantangan intelektual.
  2. Daya Pembeda: Mayoritas soal memiliki ketajaman diskriminasi yang sangat baik dalam memisahkan profil kognitif kelompok atas dan bawah. Namun, ditemukan fenomena "Kognitif Redundan" pada satu butir soal Benar-Salah (Nomor 14) yang menyebabkan daya pembedanya jatuh ke kategori jelek (0,19).
  3. Efektivitas Pengecoh: Pada soal Pilihan Ganda, pengecoh telah berfungsi efektif sebagai "jebakan miskonsepsi" yang logis. Keberhasilan pengecoh dalam menarik minat siswa kelompok bawah membuktikan bahwa desain opsi jawaban didasarkan pada analisis kesalahan prosedur yang umum dialami siswa.
- d. Kualitas Keseluruhan: Instrumen ASAS Matematika ini merupakan alat ukur yang sangat berkualitas dan akuntabel. Sebanyak 19 butir soal (95%) dinyatakan sangat layak untuk diarsipkan ke dalam Bank Soal Sekolah, sementara 1 butir soal (5%) memerlukan revisi pada aspek stimulus untuk menghindari redundansi kognitif.

## 5.2 Saran

Berdasarkan temuan penelitian tersebut, peneliti mengajukan beberapa saran sebagai berikut:

- a. Bagi Guru Mata Pelajaran:
  1. Guru disarankan untuk segera mengarsipkan 19 butir soal berkualitas tinggi hasil penelitian ini ke dalam Bank Soal sebagai aset evaluasi berkelanjutan.
  2. Mengingat rendahnya reliabilitas pada format Benar-Salah, guru hendaknya memodifikasi format tersebut menjadi "Benar-Salah Bersyarat" (dengan alasan) guna menekan faktor tebakan dan mendapatkan data diagnostik miskonsepsi yang lebih akurat.
  3. Perlu dilakukan perbaikan pada opsi pengecoh yang tidak dipilih sama sekali (0%) agar lebih homogen dan kompetitif.
- b. Bagi Pihak Sekolah:
  1. Sekolah hendaknya mendorong digitalisasi Bank Soal yang tervalidasi secara empiris untuk menjamin mutu asesmen di masa mendatang.
  2. Memfasilitasi kegiatan bedah soal secara berkala pasca-ujian sebagai bagian dari pengembangan profesionalisme guru dalam bidang evaluasi pendidikan.
- c. Bagi Peneliti Selanjutnya:
  1. Melakukan penelitian lanjutan dengan menggunakan pendekatan Teori Respon Butir (Item Response Theory) untuk mendapatkan karakteristik soal yang lebih

objektif dan tidak bergantung pada sampel (sample invariant).

2. Memperluas variabel penelitian dengan meninjau hubungan antara kualitas butir soal dengan tingkat kecemasan matematika (math anxiety) siswa saat menghadapi asesmen.



## DAFTAR PUSTAKA

- Afifah, N. (2020). *Analisis Kualitas Butir Soal Penilaian Akhir Semester Gasal Mata Pelajaran Matematika Kelas Ix Smp Negeri 3 Ajibarang Kabupaten Banyumas*. IAIN Purwokerto.
- Ali, S., & Khaeruddin. (2012). *Evaluasi Pembelajaran*. Badan Penerbit UNM.
- Anas, S. (2019). Pengantar Evaluasi Pendidikan (3 ed.). In *MoDuluS: Media Komunikasi Dunia Ilmu Sipil*.
- Arikunto, S. (2018). Dasar-dasar Evaluasi Pendidikan Edisi 3 - Suharsimi Arikunto - Google Buku. In *Pt. Bumi Aksara*.
- Ayanwale, M. A., Adeleke, J. O., & Mamadelo, T. I. (2013). *An Assessment of Item Statistics Estimates of Basic Education Certificate Examination through Classical Test Theory and Item Response Theory approach*. 55–67.
- Elgadal, A. H., & Mariod, A. A. (2021). *Item Analysis of Multiple-choice Questions ( MCQs ): Assessment Tool For Quality Assurance Measures*. 16(3), 334–346. <https://doi.org/10.18502/sjms.v16i3.9695>
- Gronlund, N. E. (1985). *Measurement and evaluation in teaching* (5th ed.). Macmillan.
- Hanna, W. F., & Retnawati, H. (2022). Analisis Kualitas Butir Soal Matematika Menggunakan Model Rasch Dengan Bantuan Software Quest. *AKSIOMA Journal of Mathematics Education*, 11(4), 3695–3704.
- Karindi, D. R., & Rofi'i. (2025). *ganda sumatif matematika Use of ANATES 4 . 0 . 9 in testing the feasibility of mathematics summative multiple choice questions*. 5(2).
- Kurnia, L. (2021). *Analisis Butir Soal Uas Mgmp Matematika Dengan Menggunakan Software Anates*. UIN Mataram.

- Magdalena, I., Fauzi, H. N., Putri, R., & Tangerang, U. M. (2020). Pentingnya Evaluasi Dalam Pembelajaran Dan Akibat Memanipulasinya. *Bintang : Jurnal Pendidikan Dan Sains*, 2, 244–257.
- Masullah, B. D., Zuhry, L. H., & Usman, L. H. (2024). *Analisis Butir Soal Ujian Akhir Semester Mata Pelajaran Matematika Smp Negeri 6 Praya Timur*. 5(September), 152–161.
- Milles, M. B., Huberman, M. A., & Saldana, J. (2014). *Qualitative Data Analysis A methods Sourcebook Edition 3 (Terjemahan Tjetjep Rohindi Rohidi)*. In *Sage Publications, Inc.*
- Nisa, Z. E. K., & Hidayati, K. (2023). *Analisis Butir Soal Penilaian Akhir Semester Matematika Universitas Negeri Yogyakarta , Yogyakarta , Indonesia E-mail : Abstrak Pendahuluan Pembelajaran merupakan kegiatan inti yang dilaksanakan dalam sebuah lembaga pendidikan , khususnya pada lembaga for. 12(3), 3516–3529.*
- Nurhalifah, S., Pangestika, R. R., & Ngazizah, N. (2024). Analisis Soal Asesmen Sumatif Akhir Semester Ganjil Mata Pelajaran Matematika Kelas IV SD Berdasarkan Taksonomi Bloom Revisi. *Jurnal Pendidikan Dasar*, 5(2), 29–35.
- Nurlaila, Remiswal, & Khadijah. (2024). *Analisis Kualitas Butir-Butir Soal Menggunakan Iteaman sebagai Evaluasi Hasil Belajar Pendidikan Agama Islam*. 5.
- Nursalam, & Suardi. (2016). *Evaluasi Pembelajaran Tes, Pengukuran dan Penilaian*. Badan Penerbit UNM.
- Pamulatsih, H., & Zulfitria. (2024). Evaluasi Pembelajaran Dalam Peningkatan Kompetensi Peserta Didik. *Al-Ansor: Jurnal Pendidikan*, 1(1), 65–78. <http://jurnal.ut.ac.id/index.php/jp/search/authors/view?givenName=Mery&familyName=&affiliation=UniversitasNoviyanti>

Terbuka&country=ID&authorName=Mery Noviyanti

- Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Nomor 21 Tahun 2022 Tentang Standar Penilaian Pendidikan. (2022).
- Putri, F., & Zakir, S. (2023). Mengukur Keberhasilan Evaluasi Pembelajaran: Telaah Evaluasi Formatif Dan Sumatif Dalam Kurikulum Merdeka. *Dewantara : Jurnal Pendidikan Sosial Humaniora*, 2(4), 172–180. <https://doi.org/10.30640/dewantara.v2i4.1783>
- Radja, S. P., Bano, V. O., & Ina, A. T. (2023). Analisis Kualitas Butir Soal Hasil Belajar Peserta Didik Berdasarkan Tingkat Kesukaran, Daya Beda Dan Efektivitas Pengecoh Di SMAN 1 Pandawai. *Jurnal Edusavana*, 1(1), 30–41.
- Scriven, M. (1967). *The methodology of evaluation*. Rand McNally.
- Sheptian, R., Sarifah, I., & Riyadi. (2025). *Classical Test Theory Analysis Using Anates: A Study Of Mathematics Readiness Test For Elementary School Students*. 5(1), 20–29.
- Sukma, N. L. M., Hani'ah, N., Dewi, E. E. K., & Inayati, N. L. (2025). Implementasi Asesmen Sumatif Pembelajaran Pendidikan Agama Islam Di SMPIT Insan Kamil Karanganyar. *Adiba: Journal Of Education*, 5(1), 150–159.
- Surbakti, I. W. (2025). *Analisis Kualitas Butir Soal Pada Uji Coba Evaluasi Pembelajaran Matematika Tingkat Sma*. 5(1).
- Syaifudin. (2023). *Analisis Butir Soal Penilaian Akhir Tahun Mata Pelajaran Matematika dengan Item And Test Analysis*. 1(2), 34–42.
- Tyler, R. W. (1949). *Basic Principles of Curriculum and*

Instruction. University of Chicago Press.

Utami, N. A. F. (2025). *Analisis Butir Soal Asesmen Sumatif Akhir Semester (Asas) Mata Pelajaran Matematika Di Smp Al Fadllu Brangsong Kendal Pada Tahun Ajaran 2024/2025*. Uin Salatiga.

Yustiandi, & Saepuzaman, D. (2024). Analisis Karakteristik Butir Soal Seleksi Penerimaan Siswa Baru Berdasarkan Pendekatan Teori Tes Klasik. *JURNAL BASICEDU*, 8(6), 4700–4706.

