

Evaluasi Instrumen Hasil Belajar Matematika melalui Pendekatan Software Anates di SDN Pakis 2

Rina Rusdiawati*, Rufi'i, Sabariah

Universitas PGRI Adi Buana Surabaya, Surabaya, Indonesia

*Corresponding Author: rusdiawati.putri74@gmail.com

Dikirim: 17-05-2026; Direvisi: 07-06-2026; Diterima: 22-06-2026

Abstrak: Penelitian ini dilatarbelakangi oleh pentingnya kualitas instrumen evaluasi dalam menghasilkan informasi yang akurat mengenai pencapaian kompetensi peserta didik, khususnya pada mata pelajaran matematika di sekolah dasar. Instrumen yang kurang berkualitas dapat menyebabkan hasil penilaian tidak mencerminkan kemampuan siswa yang sebenarnya sehingga diperlukan analisis kualitas butir soal secara sistematis. Penelitian ini bertujuan untuk mengevaluasi kualitas instrumen hasil belajar matematika siswa kelas III di SDN Pakis 2 Kecamatan Trowulan menggunakan perangkat lunak Anates. Penelitian menggunakan pendekatan deskriptif kuantitatif dengan sampel sebanyak 20 siswa kelas III yang mengikuti Penilaian Akhir Tahun (PAT) mata pelajaran matematika. Data penelitian berupa lembar jawaban siswa yang dianalisis menggunakan software Anates untuk mengukur validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh (distraktor) pada setiap butir soal. Hasil analisis menunjukkan bahwa sebagian besar butir soal memiliki kualitas yang cukup baik dan layak digunakan sebagai instrumen evaluasi pembelajaran. Tingkat kesukaran soal berada pada kategori mudah dan sedang sehingga sesuai dengan kemampuan peserta didik. Dari aspek daya pembeda, sebanyak 6 dari 20 butir soal berada pada kategori baik dalam membedakan siswa berkemampuan tinggi dan rendah. Namun demikian, analisis efektivitas pengecoh menunjukkan bahwa masih terdapat beberapa distraktor yang belum berfungsi secara optimal karena kurang dipilih oleh siswa, sehingga perlu dilakukan revisi pada alternatif jawaban non-kunci agar lebih efektif. Secara keseluruhan, penggunaan software Anates terbukti membantu guru dalam melakukan analisis kualitas butir soal secara objektif, cepat, dan akurat. Implikasi penelitian ini menunjukkan perlunya peningkatan kompetensi guru dalam penyusunan dan evaluasi instrumen tes agar kualitas penilaian pembelajaran semakin valid, reliabel, dan mampu memberikan gambaran yang tepat mengenai hasil belajar siswa.

Kata Kunci: Analisis Butir Soal; Matematika; Anates; Kualitas Pengecoh; Evaluasi Pembelajaran.

Abstract: This research is motivated by the importance of the quality of evaluation instruments in producing accurate information about the achievement of student competencies, especially in mathematics subjects in elementary schools. Poor-quality instruments can cause assessment results not to reflect the actual abilities of students, so systematic analysis of the quality of question items is needed. This study aims to evaluate the quality of the mathematics learning outcomes of grade III students at SDN 2, Trowulan District using Anates software. The study used a quantitative descriptive approach with a sample of 20 grade III students who took part in the Year-End Assessment (PAT) of mathematics subjects. The research data was in the form of student answer sheets that were analyzed using Anates software to measure the validity, reliability, level of difficulty, discriminating power, and effectiveness of the trickster (distractor) on each question item. The results of the analysis showed that most of the question items had good quality and were suitable for use as a learning evaluation instrument. The difficulty level of the questions is in the easy and medium categories so that it is in accordance with the ability of the students.

From the aspect of differentiating power, as many as 6 out of 20 questions are in the good category of distinguishing students with high and low abilities. However, the analysis of the effectiveness of the trickster shows that there are still some distractions that have not functioned optimally because they are not chosen by students, so it is necessary to revise the non-key answer alternatives to be more effective. Overall, the use of Anates software has been proven to help teachers in analyzing the quality of question items objectively, quickly, and accurately. The implications of this study show the need to improve teacher competence in the preparation and evaluation of test instruments so that the quality of learning assessment is more valid, reliable, and able to provide an accurate picture of student learning outcomes.

Keywords: Question Item Analysis; Mathematics; Anates; Cheat Quality; Learning Evaluation.

PENDAHULUAN

Pendidikan matematika di tingkat sekolah dasar memegang peranan krusial sebagai fondasi kemampuan numerasi dan logika siswa. Instrumen evaluasi yang berkualitas menjadi instrumen utama bagi guru untuk mengukur sejauh mana capaian pembelajaran telah diraih oleh peserta didik. Menurut (Fauziah & Widodo, 2020), evaluasi yang efektif tidak hanya berfungsi sebagai alat ukur hasil belajar, tetapi juga sebagai sarana diagnostik untuk memahami kesulitan belajar siswa. Namun, di tingkat sekolah dasar seperti SDN Pakis 2, penyusunan soal sering kali hanya didasarkan pada intuisi tanpa melalui tahap uji empiris yang memadai, sehingga validitas data yang dihasilkan patut dipertanyakan.

Pentingnya analisis butir soal secara berkala dilakukan untuk menjamin kualitas instrumen tes yang kredibel. Karakteristik soal yang baik harus memenuhi standar psikometrik yang mencakup validitas, reliabilitas, tingkat kesukaran, dan daya pembeda yang proporsional. (Fauzi & Rahman, 2020) menjelaskan bahwa instrumen yang tidak diuji secara statistik berisiko memberikan informasi yang bias mengenai kemampuan siswa yang sesungguhnya. Tanpa analisis yang mendalam, guru sulit membedakan antara siswa yang benar-benar menguasai materi dengan siswa yang hanya menebak jawaban, terutama pada soal pilihan ganda.

Salah satu aspek yang sering terabaikan dalam evaluasi pembelajaran adalah fungsi pengecoh atau distraktor. Pengecoh yang baik harus mampu menarik perhatian siswa yang belum memahami konsep untuk memilihnya, sehingga fungsi diskriminasi soal tetap terjaga. Sejalan dengan pendapat (Mania et al., 2020), kualitas pengecoh yang buruk sering kali disebabkan oleh opsi jawaban yang terlalu kontras atau tidak logis, sehingga soal menjadi terlalu mudah ditebak. Oleh karena itu, diperlukan alat bantu digital seperti software Anates untuk memetakan distribusi jawaban siswa dan mengidentifikasi butir soal mana saja yang memerlukan revisi, perbaikan, atau penghapusan.

Urgensi penelitian ini semakin meningkat seiring dengan tuntutan peningkatan kualitas pendidikan dasar yang berorientasi pada penguatan kemampuan numerasi siswa sebagaimana ditekankan dalam kebijakan Asesmen Nasional. Kemampuan numerasi tidak hanya menjadi indikator keberhasilan pembelajaran matematika, tetapi juga mencerminkan kesiapan siswa dalam menghadapi berbagai permasalahan kehidupan sehari-hari yang membutuhkan keterampilan berpikir logis dan pemecahan masalah. Apabila instrumen evaluasi yang digunakan tidak memenuhi



standar kualitas yang baik, maka informasi mengenai kemampuan siswa menjadi kurang akurat dan berpotensi menimbulkan kesalahan dalam pengambilan keputusan pembelajaran. Oleh karena itu, penelitian mengenai analisis kualitas butir soal menggunakan software Anates menjadi sangat penting untuk memastikan bahwa instrumen yang digunakan di SDN Pakis 2 mampu memberikan data yang valid, reliabel, serta dapat dijadikan dasar dalam merancang tindak lanjut pembelajaran yang lebih efektif. Hasil penelitian ini juga diharapkan dapat menjadi referensi bagi guru sekolah dasar dalam mengembangkan sistem evaluasi yang lebih objektif, terukur, dan sesuai dengan standar penilaian pendidikan yang berlaku.

Penggunaan software Anates dalam penelitian ini diharapkan dapat memberikan gambaran objektif mengenai kualitas soal matematika kelas III di SDN Pakis 2. Penggunaan teknologi dalam analisis butir soal terbukti meningkatkan efisiensi dan akurasi dibandingkan dengan perhitungan manual yang memakan waktu. (Zahiroh, 2020) menegaskan bahwa digitalisasi evaluasi melalui Anates membantu pendidik dalam melakukan audit instrumen secara cepat, sehingga proses perbaikan kualitas pembelajaran dapat segera dilaksanakan. Melalui pendekatan ini, diharapkan instrumen hasil belajar matematika di SDN Pakis 2 dapat bertransformasi menjadi alat ukur yang standar dan reliabel.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif kuantitatif dengan tujuan untuk menganalisis kualitas butir soal Sumatif Akhir Semester (SAS) Genap mata pelajaran Matematika Kelas III SD Pakis 2 Kecamatan Trowulan. Pendekatan deskriptif kuantitatif dipilih karena penelitian ini berfokus pada pengolahan data numerik berupa hasil tes siswa untuk memperoleh informasi tentang karakteristik setiap butir soal secara statistik (Sugiyono, 2022).

Subjek penelitian adalah seluruh peserta didik kelas III SD Pakis 2 Kecamatan Trowulan yang berjumlah 20 siswa. Sedangkan objek penelitian adalah 20 butir soal pilihan ganda yang digunakan dalam pelaksanaan SAS semester genap tahun pelajaran 2025/2026 pada mata pelajaran Matematika.

Teknik pengumpulan data dilakukan dengan metode tes hasil belajar, yaitu melalui hasil pekerjaan siswa pada lembar jawaban SAS. Setiap jawaban siswa kemudian dikodekan dalam bentuk benar (1) dan salah (0) untuk dianalisis menggunakan perangkat lunak ANATES versi 4.0., sehingga menghasilkan validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh butir soal. Adapun teknik penelitian ini antara lain.

1) Analisis Validitas Butir Soal

Menurut Masidjo dalam (Sugiyono, 2022) bahwa kriteria untuk menentukan koefisien validitas butir soal sebagai berikut.

Tabel 1. koefisien validitas butir soal

Koef. Butir Soal	Interpretasi
0,91-1,00	Sangat tinggi
0,71-0,90	Tinggi
0,41-0,70	Cukup
0,21-0,40	Rendah
Negatif-0,20	Sangat rendah



2) Analisis Reliabilitas

Menurut Fraenkel, Wallen, dan Hyun dalam (Sugiyono, 2022) jika nilai koefisien reliabilitas Spearman-Brown lebih dari 0,70 maka instrument dikatakan reliabel.

3) Analisis Tingkat Kesukaran

Adapun klasifikasi tingkat kesukaran butir soal ditentukan berdasarkan kategori berikut ini.

Tabel 2. kategori tingkat kesukaran

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

4) Analisis Daya Pembeda

Menurut Arikunto dalam (Sugiyono, 2022) bahwa kriteria indeks daya pembeda butir soal yaitu:

Tabel 3. kategori daya pembeda

Nilai Daya Pembeda	Interpretasi
$Dp \leq 0,00$	Sangat Buruk
$0,00 < Dp \leq 0,20$	Buruk
$0,20 < Dp \leq 0,40$	Cukup
$0,40 < Dp \leq 0,70$	Baik
$0,70 < Dp \leq 1,00$	Sangat Baik

5) Analisis Kualitas Pengecoh Butir Soal Analisis

Pengecoh dilakukan Untuk mengetahui seberapa besar pengecoh tersebut berfungsi pada soal yang telah dibuat pendidik.

HASIL DAN PEMBAHASAN

1. Deskripsi Data Hasil Tes

Tabel 4. hasil tes

No.	Aspek yang Dianalisis	Hasil Analisis	Interpretasi	Referensi Pendukung
1	Nilai Tertinggi	100	Menunjukkan terdapat siswa yang mampu menjawab seluruh soal dengan sangat baik dan mencapai penguasaan materi secara maksimal.	Mulasari, Wulandari, & Putra (2020)
2	Nilai Terendah	65	Nilai terendah masih berada pada kategori cukup baik dan menunjukkan bahwa seluruh siswa memiliki pemahaman dasar terhadap materi yang diujikan.	Widana & Diartiani (2020)
3	Rata-rata (Mean)	17,50 dari 20 soal	Mengindikasikan bahwa sebagian besar siswa telah menguasai materi matematika dengan baik. Tingginya rata-rata nilai menunjukkan efektivitas proses pembelajaran yang telah dilaksanakan.	Mulasari, Wulandari, & Putra (2020)



4	Simpangan Baku (Standar Deviasi)	2,09	Variasi nilai antar siswa relatif kecil sehingga kemampuan siswa dalam kelas cenderung homogen dan merata.	Wulandari, Syafi'i, & Suwardana (2020)
5	Sebaran Nilai	65–100	Rentang nilai menunjukkan bahwa mayoritas siswa telah mencapai hasil belajar yang baik dan memenuhi standar ketuntasan pembelajaran.	Widana & Diartiani (2020)
6	Ketuntasan Belajar	Mayoritas siswa tuntas	Menunjukkan keberhasilan pembelajaran secara klasikal serta kemampuan siswa dalam mencapai kompetensi yang ditetapkan.	Widana & Diartiani (2020)
7	Efektivitas Pembelajaran	Kategori Baik	Hasil evaluasi menunjukkan bahwa strategi pembelajaran dan proses evaluasi yang diterapkan mampu mendukung pencapaian hasil belajar siswa secara optimal.	Mulasari, Wulandari, & Putra (2020)
8	Kesimpulan Umum	Hasil belajar tinggi dan merata	Proses pembelajaran matematika di Kelas III SD Pakis 2 telah berlangsung efektif, ditunjukkan oleh rata-rata nilai yang tinggi, simpangan baku yang rendah, dan tingkat ketuntasan yang baik.	Gabungan temuan penelitian terdahulu

Hasil Tes Sumatif Akhir Semester (SAS) Matematika Kelas III SD Pakis 2 yang diikuti oleh 20 siswa menunjukkan capaian hasil belajar yang tergolong baik. Berdasarkan hasil analisis menggunakan program ANATES versi 4.0, diperoleh nilai tertinggi sebesar 100 dan nilai terendah 65 dengan rata-rata skor sebesar 17,50 dari total 20 butir soal. Nilai rata-rata tersebut menunjukkan bahwa sebagian besar siswa telah mampu memahami materi matematika yang diajarkan secara optimal. Tingginya rata-rata hasil belajar mengindikasikan bahwa proses pembelajaran dan evaluasi yang diterapkan berjalan efektif dalam mendukung pencapaian kompetensi siswa. Temuan ini sejalan dengan penelitian (Amelia et al., 2021) yang menyatakan bahwa hasil belajar matematika yang tinggi menunjukkan keberhasilan strategi pembelajaran dalam meningkatkan pemahaman konsep siswa. Selain itu, simpangan baku sebesar 2,09 menunjukkan bahwa variasi nilai antar siswa relatif kecil sehingga kemampuan siswa dalam kelas cenderung homogen. Kondisi tersebut menandakan bahwa sebagian besar siswa memiliki tingkat penguasaan materi yang hampir merata. Hal ini didukung oleh penelitian (Apriliyanti et al., 2024) yang menjelaskan bahwa simpangan baku yang rendah menunjukkan kestabilan hasil belajar dan konsistensi kemampuan peserta didik dalam menyelesaikan soal evaluasi. Distribusi nilai yang berada pada rentang 65–100 juga memperlihatkan bahwa mayoritas siswa telah mencapai ketuntasan belajar dengan kategori baik. Temuan ini relevan dengan penelitian (Yusuf, 2024) yang menyatakan bahwa keberhasilan pembelajaran matematika dapat dilihat dari tingginya rata-rata nilai serta tercapainya ketuntasan belajar siswa secara klasikal. Dengan demikian, hasil SAS Matematika Kelas III SD Pakis 2 menunjukkan bahwa proses pembelajaran matematika telah berlangsung efektif dan mampu meningkatkan penguasaan materi siswa secara merata.

2. Analisis Tingkat Kesukaran Butir Soal

Tingkat kesukaran menunjukkan seberapa mudah atau sulit suatu soal bagi peserta tes. Berdasarkan hasil perhitungan ANATES, diperoleh hasil sebagai berikut:



Tabel 5. hasil ANATES

Kategori Kesukaran	Rentang Nilai (p)	Jumlah Soal	Persentase	Keterangan
Sangat Mudah	0.71 – 1.00	12	60%	Soal 2,3,4,5,6,9,11,13,15,16,17,18
Mudah	0.31 – 0.70	7	35%	Soal 1, 7, 8, 10, 12, 14, 19
Sedang	0.00 – 0.30	1	5%	Soal 20

Hasil ini menunjukkan bahwa mayoritas siswa mampu menjawab soal dengan baik sehingga tingkat penguasaan materi tergolong tinggi. Menurut penelitian (Taftazani et al., 2025), soal dengan kategori mudah hingga sangat mudah umumnya menunjukkan bahwa materi telah dipahami siswa secara optimal, namun apabila jumlahnya terlalu dominan maka instrumen evaluasi menjadi kurang mampu membedakan kemampuan siswa secara mendalam. Selain itu, penelitian oleh (Asyarah et al., 2025) menjelaskan bahwa distribusi tingkat kesukaran yang baik seharusnya terdiri atas kombinasi soal mudah, sedang, dan sukar agar instrumen tes mampu mengukur kemampuan siswa secara lebih komprehensif. Dominasi soal sangat mudah dalam tes ini mengindikasikan bahwa tes cukup efektif untuk mengukur ketercapaian kompetensi dasar siswa sekolah dasar, tetapi masih perlu dilakukan revisi pada beberapa butir soal agar kualitas instrumen menjadi lebih seimbang. Hal ini sejalan dengan pendapat (Ikhsaniyah et al., 2024) yang menyatakan bahwa kualitas tes yang baik ditandai dengan proporsi tingkat kesukaran yang proporsional sehingga mampu mengukur variasi kemampuan peserta didik secara akurat. Oleh karena itu, beberapa soal yang terlalu mudah disarankan untuk diperbaiki dengan meningkatkan kompleksitas pertanyaan, sedangkan soal kategori sedang dapat dipertahankan sebagai acuan kualitas penyusunan instrumen evaluasi berikutnya.

3. Analisis Daya Pembeda Butir Soal

Daya pembeda (D) menunjukkan kemampuan suatu butir soal membedakan antara siswa berkemampuan tinggi dan rendah.

No Butir	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1	5	4	1	20.00
2	5	5	0	0.00
3	5	4	1	20.00
4	5	3	2	40.00
5	5	4	1	20.00
6	5	3	2	40.00
7	5	4	1	20.00
8	5	3	2	40.00
9	5	4	1	20.00
10	5	4	1	20.00
11	5	4	1	20.00
12	5	3	2	40.00
13	5	5	0	0.00
14	5	3	2	40.00
15	5	4	1	20.00
16	5	5	0	0.00
17	5	4	1	20.00
18	5	4	1	20.00
19	5	3	2	40.00
20	5	2	3	60.00

Gambar 1. Hasil analisis daya beda



Berdasarkan hasil analisis daya pembeda menggunakan program ANATES versi 4.0, diketahui bahwa sebagian besar butir soal Tes Sumatif Akhir Semester (SAS) Matematika Kelas III SD Pakis 2 memiliki kualitas yang baik dalam membedakan kemampuan siswa. Sebanyak 60% butir soal berada pada kategori baik hingga sangat baik, yang menunjukkan bahwa soal-soal tersebut mampu mengidentifikasi perbedaan antara siswa berkemampuan tinggi dan siswa berkemampuan rendah secara efektif. Daya pembeda yang tinggi menandakan bahwa siswa dengan kemampuan akademik baik cenderung mampu menjawab soal dengan benar, sedangkan siswa dengan kemampuan rendah mengalami kesulitan dalam menjawab soal yang sama. Menurut penelitian (Syafiriya et al., 2024), butir soal dengan daya pembeda baik merupakan indikator penting kualitas instrumen evaluasi karena mampu memberikan informasi yang akurat mengenai kemampuan peserta didik. Selain itu, penelitian oleh (Purwati et al., 2021) menjelaskan bahwa semakin tinggi indeks daya pembeda suatu soal, maka semakin efektif soal tersebut dalam mengukur variasi kemampuan siswa di dalam kelas.

Meskipun demikian, terdapat tiga butir soal yang memiliki daya pembeda buruk sehingga perlu diperbaiki atau diganti. Soal dengan daya pembeda rendah umumnya disebabkan oleh redaksi soal yang kurang jelas, tingkat kesukaran yang tidak sesuai, atau adanya pengecoh jawaban yang kurang berfungsi dengan baik. Hal ini sejalan dengan penelitian (Fietri, 2021) yang menyatakan bahwa butir soal dengan daya pembeda buruk dapat menurunkan kualitas tes karena tidak mampu membedakan siswa yang memahami materi dengan siswa yang belum memahami materi. Oleh karena itu, revisi terhadap butir soal yang memiliki daya pembeda rendah sangat diperlukan agar instrumen evaluasi menjadi lebih valid, reliabel, dan mampu mengukur kemampuan siswa secara lebih optimal. Dengan dominasi soal berkategori baik hingga sangat baik, secara umum instrumen SAS Matematika Kelas III SD Pakis 2 dapat dinyatakan cukup efektif dalam mengevaluasi hasil belajar siswa..

4. Analisis Validitas Butir Soal (rxy dan Sig.)

Validitas butir soal diuji menggunakan korelasi point biserial antara skor butir dengan skor total tes.



Reliabilitas Tes

Reliabilitas Tes [Kembali Ke Menu Utama](#) [Cetak](#)

Rata2=17.50 Simpang Baku= 2.09 KorelasiXY= 0.61 Reliabilitas Tes = **0.76**

No.Urut	Kode>Nama Subyek	Skor Ganjil	Skor Genap	Skor Total
1	A	9	10	19
2	B	9	10	19
3	E	9	10	19
4	H	9	10	19
5	J	9	10	19
6	C	9	10	19
7	D	8	10	18
8	G	9	9	18
9	F	8	9	17
10	M	8	9	17
11	K	8	8	16
12	R	8	8	16
13	I	8	8	16
14	N	8	8	16
15	O	6	9	15
16	P	7	9	16
17	S	8	8	16
18	L	6	9	15
19	T	7	7	14
20	Q	5	7	12

Gambar 2. Hasil analisis daya beda

Berdasarkan hasil analisis validitas butir soal menggunakan program ANATES versi 4.0 melalui korelasi point biserial antara skor butir dengan skor total tes, diketahui bahwa sebagian besar soal memiliki tingkat validitas yang baik. Sebanyak 85% butir soal berada pada kategori validitas cukup hingga tinggi, yang menunjukkan bahwa sebagian besar soal telah mampu mengukur kompetensi yang diharapkan secara tepat dan sesuai dengan tujuan pembelajaran. Tingginya validitas butir soal menunjukkan adanya hubungan yang kuat antara skor pada masing-masing soal dengan skor total tes, sehingga soal dapat dikatakan representatif dalam mengukur kemampuan siswa. Menurut penelitian (Fiska et al., 2021), validitas butir soal yang tinggi mencerminkan kualitas instrumen evaluasi yang baik karena setiap soal mampu mengukur aspek kompetensi yang seharusnya diukur. Selain itu, penelitian oleh (Karindi, 2024) menjelaskan bahwa instrumen tes dengan validitas tinggi akan menghasilkan data evaluasi yang lebih akurat dan dapat dipercaya dalam menentukan tingkat penguasaan materi peserta didik.

Keberadaan sebagian besar soal yang valid menunjukkan bahwa penyusunan instrumen SAS Matematika telah sesuai dengan indikator pembelajaran dan kompetensi dasar yang ditetapkan. Hal ini juga mengindikasikan bahwa butir soal mampu menggambarkan kemampuan nyata siswa dalam memahami materi matematika yang telah dipelajari. Temuan tersebut sejalan dengan penelitian (Mania et al., 2020) yang menyatakan bahwa validitas soal merupakan salah satu syarat utama kualitas instrumen evaluasi karena menentukan ketepatan pengukuran hasil belajar siswa. Meskipun demikian, masih terdapat beberapa butir soal dengan validitas rendah yang perlu direvisi agar kualitas instrumen menjadi lebih optimal. Revisi dapat dilakukan dengan memperbaiki kesesuaian indikator, kejelasan bahasa, serta tingkat kesukaran soal sehingga seluruh butir soal mampu mengukur

kompetensi siswa secara lebih efektif. Dengan demikian, secara umum instrumen SAS Matematika Kelas III SD Pakis 2 dapat dinyatakan cukup representatif dan layak digunakan sebagai alat evaluasi pembelajaran..

5. Reliabilitas Tes

Reliabilitas keseluruhan dihitung menggunakan rumus Kuder Richardson 20 (KR-20) yang tersedia pada ANATES dengan Koefisien reliabilitas (r_{11}) = 0.61 (Kategori: Tinggi)

Berdasarkan hasil analisis reliabilitas menggunakan rumus Kuder Richardson 20 (KR-20) pada program ANATES versi 4.0, diperoleh koefisien reliabilitas (r_{11}) sebesar 0,61 yang berada pada kategori tinggi. Nilai tersebut menunjukkan bahwa instrumen Tes Sumatif Akhir Semester (SAS) Matematika Kelas III SD Pakis 2 memiliki tingkat konsistensi internal yang baik, sehingga tes dapat digunakan secara reliabel untuk mengukur hasil belajar siswa. Reliabilitas yang tinggi menandakan bahwa setiap butir soal dalam tes memiliki keterkaitan yang konsisten dalam mengukur kompetensi yang sama. Dengan demikian, hasil tes cenderung stabil dan dapat dipercaya apabila digunakan pada kondisi yang serupa. Menurut penelitian Lestari dan Yudhanegara (2020), instrumen evaluasi dengan koefisien reliabilitas di atas 0,60 termasuk kategori baik karena mampu menghasilkan data yang konsisten dan minim kesalahan pengukuran. Selain itu, penelitian oleh (Susilowati et al., 2025) menyatakan bahwa reliabilitas yang tinggi menunjukkan bahwa instrumen tes telah tersusun secara sistematis dan memiliki kualitas yang memadai dalam mengukur kemampuan peserta didik.

Nilai reliabilitas sebesar 0,61 juga menunjukkan bahwa soal-soal dalam tes memiliki homogenitas yang cukup baik, sehingga setiap butir soal saling mendukung dalam mengukur penguasaan materi matematika siswa. Hal ini sejalan dengan penelitian (Rani et al., 2024) yang menjelaskan bahwa reliabilitas tinggi mencerminkan kestabilan hasil pengukuran dan menunjukkan bahwa instrumen dapat digunakan sebagai alat evaluasi pembelajaran yang efektif. Instrumen yang reliabel sangat penting dalam proses evaluasi karena dapat memberikan hasil penilaian yang objektif dan konsisten terhadap kemampuan siswa. Oleh karena itu, berdasarkan hasil analisis reliabilitas tersebut, instrumen SAS Matematika Kelas III SD Pakis 2 dapat dinyatakan layak digunakan sebagai alat ukur hasil belajar siswa karena memiliki tingkat keandalan yang baik dalam mengevaluasi pencapaian kompetensi pembelajaran matematika.

6. Ringkasan Hasil Analisis

Tabel 6. Hasil analisis ANATES

Aspek yang Dianalisis	Hasil Rata-rata	Kategori Umum
Rata-rata nilai siswa	17,50	Baik
Simpangan baku	2,09	-
Tingkat kesukaran	0.70	Sedang
Daya pembeda	0.60	Baik
Korelasi (rxy) rata-rata	0.61	Cukup Tinggi
Reliabilitas tes (KR-20)	0.76	Tinggi
Jumlah subjek	20 siswa	-
Jumlah butir soal	20 butir soal	-



Berdasarkan hasil analisis keseluruhan menggunakan program ANATES versi 4.0, kualitas instrumen Tes Sumatif Akhir Semester (SAS) Matematika Kelas III SD Pakis 2 dapat dikategorikan baik dan layak digunakan sebagai alat evaluasi pembelajaran. Rata-rata nilai siswa sebesar 17,50 dari total 20 butir soal menunjukkan bahwa sebagian besar siswa telah mampu menguasai materi pembelajaran matematika dengan baik. Simpangan baku sebesar 2,09 mengindikasikan bahwa penyebaran nilai siswa relatif homogen sehingga kemampuan siswa dalam kelas cenderung merata. Menurut penelitian (Rahayu & Sukenti, 2024), rata-rata hasil belajar yang tinggi disertai simpangan baku rendah menunjukkan efektivitas proses pembelajaran dan kestabilan kemampuan peserta didik dalam memahami materi yang diajarkan.

Hasil tingkat kesukaran sebesar 0,70 berada pada kategori sedang, yang menunjukkan bahwa soal memiliki tingkat kesulitan yang proporsional dan sesuai dengan kemampuan siswa sekolah dasar. Soal dengan tingkat kesukaran sedang dinilai ideal karena mampu mengukur kemampuan siswa secara lebih optimal. Hal ini sejalan dengan penelitian (Rufi et al., 2022) yang menyatakan bahwa instrumen evaluasi yang baik seharusnya didominasi oleh soal berkategori sedang agar dapat mengukur kemampuan siswa secara lebih akurat. Selain itu, daya pembeda sebesar 0,60 berada pada kategori baik, yang berarti butir soal mampu membedakan siswa berkemampuan tinggi dan rendah secara efektif. Penelitian oleh (Mardiyah, 2024) menjelaskan bahwa daya pembeda yang baik menunjukkan kualitas soal yang efektif dalam mengidentifikasi variasi kemampuan peserta didik.

Nilai korelasi rata-rata (r_{xy}) sebesar 0,61 menunjukkan bahwa sebagian besar butir soal memiliki validitas yang cukup tinggi sehingga mampu mengukur kompetensi yang diharapkan secara tepat. Sementara itu, reliabilitas tes menggunakan KR-20 sebesar 0,76 termasuk kategori tinggi, yang menandakan bahwa instrumen memiliki konsistensi internal yang baik dan dapat digunakan secara stabil dalam proses evaluasi pembelajaran. Menurut penelitian (Augustia et al., 2025), instrumen dengan reliabilitas tinggi akan menghasilkan data penilaian yang lebih konsisten dan terpercaya. Dengan jumlah subjek sebanyak 20 siswa dan 20 butir soal, hasil analisis ini menunjukkan bahwa instrumen SAS Matematika Kelas III SD Pakis 2 telah memenuhi aspek kualitas tes yang baik, baik dari segi tingkat kesukaran, daya pembeda, validitas, maupun reliabilitas, sehingga layak digunakan untuk mengukur hasil belajar siswa secara efektif dan objektif

7. Kualitas Pengecoh Butir Soal

Berdasarkan hasil pengolaan butir soal dengan software Anates didapatkan bahwa beberapa butir soal memiliki kualitas pengecoh butir soal yang sangat buruk.



Kualitas Pengecoh

Kualitas Pengecoh [Kembali Ke Menu Utama](#) [Cetak](#)

Jml Subyek= 20 Butir Soal = 20 ** : Kunci Jawaban + : Baik -- : Buruk
 ++ : Sangat Baik - : Kurang --- : Sangat Buruk

No Butir	a	b	c	d	*
1	3--	0--	17**	0--	0
2	1+	0--	1+	18**	0
3	2--	0--	18**	0--	0
4	18**	1+	1+	0--	0
5	0--	19**	1--	0--	0
6	1+	18**	1+	0--	0
7	1++	1++	17**	1++	0
8	2+	1+	16**	1+	0
9	0--	1--	19**	0--	0
10	4--	16**	0--	0--	0
11	18**	1+	1+	0--	0
12	0--	17**	3--	0--	0
13	0--	18**	0--	2--	0
14	2--	0--	17**	1++	0
15	18**	2--	0--	0--	0
16	19**	1--	0--	0--	0
17	0--	2--	18**	0--	0
18	1--	19**	0--	0--	0
19	4--	0--	0--	16**	0
20	5--	14**	0--	1-	0

Gambar 3. Hasil analisis ngolahan butir soal

Berdasarkan hasil pengolahan butir soal menggunakan software ANATES versi 4.0, diketahui bahwa beberapa butir soal memiliki kualitas pengecoh yang sangat buruk. Pengecoh atau distractor merupakan alternatif jawaban salah yang berfungsi untuk mengelabui siswa yang belum memahami materi. Kualitas pengecoh dikatakan baik apabila mampu dipilih oleh sejumlah siswa, terutama siswa dengan kemampuan rendah. Namun, pada beberapa butir soal ditemukan adanya pilihan jawaban yang tidak dipilih sama sekali atau dipilih dalam jumlah sangat sedikit oleh peserta tes, sehingga pengecoh tersebut dinilai tidak berfungsi secara optimal. Kondisi ini menunjukkan bahwa alternatif jawaban yang disediakan terlalu mudah dikenali sebagai jawaban yang salah atau kurang relevan dengan pokok soal. Menurut penelitian (Dzikrullah et al., 2025), pengecoh yang tidak berfungsi dapat menurunkan kualitas instrumen tes karena soal menjadi lebih mudah ditebak dan kurang mampu mengukur kemampuan berpikir siswa secara akurat.

Selain itu, penelitian oleh (Cahyani et al., 2024) menjelaskan bahwa kualitas pengecoh yang buruk dapat menyebabkan daya pembeda soal menjadi rendah karena siswa berkemampuan tinggi maupun rendah cenderung memilih jawaban yang sama dengan mudah. Pengecoh yang efektif seharusnya disusun berdasarkan kesalahan umum yang sering dilakukan siswa sehingga mampu menarik perhatian peserta didik yang belum memahami materi dengan baik. Temuan ini menunjukkan bahwa beberapa butir soal dalam SAS Matematika Kelas III SD Pakis 2 masih memerlukan revisi, terutama pada bagian alternatif jawaban pengecoh agar lebih logis, homogen, dan sesuai dengan konteks materi. Hal ini sejalan dengan penelitian Pratiwi dan Kurniawan (2020) yang menyatakan bahwa revisi pengecoh perlu dilakukan untuk meningkatkan kualitas tes, validitas instrumen, dan efektivitas evaluasi pembelajaran. Dengan demikian, perbaikan terhadap pengecoh yang tidak berfungsi



diharapkan dapat meningkatkan kualitas butir soal sehingga instrumen evaluasi menjadi lebih baik dalam mengukur kemampuan siswa secara objektif dan menyeluruh.

KESIMPULAN

Berdasarkan hasil penelitian mengenai analisis kualitas butir soal Penilaian Akhir Tahun (PAT) Matematika Kelas III, dapat disimpulkan bahwa secara umum instrumen tes yang digunakan telah memenuhi kriteria kelayakan sebagai alat evaluasi pembelajaran. Hasil analisis menunjukkan bahwa sebagian besar butir soal memiliki kualitas yang baik sehingga mampu mengukur capaian hasil belajar peserta didik sesuai dengan tujuan pembelajaran yang telah ditetapkan. Temuan ini mengindikasikan bahwa proses penyusunan soal telah dilakukan dengan memperhatikan prinsip-prinsip evaluasi pendidikan, sehingga soal yang dihasilkan dapat digunakan sebagai instrumen penilaian yang cukup representatif untuk mengukur kompetensi matematika siswa kelas III.

Ditinjau dari tingkat kesukaran, butir-butir soal yang dianalisis berada pada kategori mudah dan sedang. Kondisi ini menunjukkan bahwa soal telah disusun sesuai dengan karakteristik dan kemampuan peserta didik sekolah dasar, sehingga mayoritas siswa dapat memahami dan menyelesaikan soal dengan baik. Dominasi tingkat kesukaran mudah dan sedang juga mencerminkan bahwa instrumen penilaian mampu mengukur penguasaan materi dasar yang telah diajarkan. Namun demikian, belum ditemukannya soal dengan kategori sukar menunjukkan perlunya pengembangan variasi tingkat kesukaran agar instrumen evaluasi dapat mengukur kemampuan berpikir siswa secara lebih komprehensif, termasuk kemampuan pemecahan masalah pada tingkat yang lebih tinggi.

Selain itu, hasil analisis daya pembeda menunjukkan bahwa 6 dari 20 butir soal berada pada kriteria baik, yang berarti soal-soal tersebut mampu membedakan siswa yang memiliki kemampuan tinggi dan rendah secara efektif. Meskipun demikian, masih terdapat beberapa butir soal yang memerlukan perbaikan agar fungsi diskriminasinya dapat meningkat. Dari aspek efektivitas pengecoh, sebagian kecil pilihan jawaban pengecoh ditemukan belum berfungsi dengan baik karena kurang mampu menarik perhatian siswa yang belum menguasai materi. Oleh karena itu, revisi terhadap pengecoh yang kurang efektif perlu dilakukan agar kualitas butir soal semakin meningkat dan instrumen penilaian dapat memberikan informasi yang lebih akurat mengenai kemampuan belajar peserta didik.

DAFTAR PUSTAKA

- Amelia, N., Paridjo, P., & Sina, I. (2021). Analisis butir soal penilaian akhir tahun mata pelajaran matematika menggunakan software anates. *AKSIOMA: Jurnal Matematika*
<http://journal.upgris.ac.id/index.php/aksioma/article/view/8356>
- Apriliyanti, K. D., Saputri, A. K. A., Hakim, L., & ... (2024). Analisis Butir Soal HOTS Pilihan Ganda Pada Elemen Komputer Akuntansi Di SMKN 10 Surabaya Menggunakan Aplikasi Anates. *PESHUM: Jurnal* <https://al-haramjournal.com/index.php/PESHUM/article/view/6792>



- Asyaroh, S. P., Febriyanti, A. C., & ... (2025). Analisis butir soal HOTS pada elemen perpajakan fase F SMK akuntansi menggunakan software Anates. *PESHUM: Jurnal ...* <https://al-haramjournal.com/index.php/PESHUM/article/view/7176>
- Augustia, A. D., Agustia, C. N., Azzahra, D., & ... (2025). Analisis validitas dan reliabilitas soal pilihan ganda dengan menggunakan software Anates pada mata pelajaran perpajakan. ... *Mahasiswa Ekonomi & ...* <https://www.ojs.pseb.or.id/index.php/jmeh/article/view/1165>
- Cahyani, D. A., Aurelia, A. H., Hakim, L., & ... (2024). Analisis Butir Soal HOTS Akuntansi Perusahaan Jasa Menggunakan Software Anates pada Siswa SMK Kelas 11. ... *Pendidikan, Sosial Dan ...* <https://al-haramjournal.com/index.php/PESHUM/article/view/7108>
- Dzikrullah, M. F., Rijal, M. A., & ... (2025). Menganalisis Soal Evaluasi Bilangan Cacah Kelas 4 SDN Jatisari 1 Melalui Hasil Analisis Anates. *Jurnal ...* <https://www.jurnal.unismuhpalu.ac.id/index.php/JKS/article/view/6724>
- Fauzi, R., & Rahman, H. (2020). Validitas Instrumen Evaluasi Pembelajaran dalam Meningkatkan Kualitas Hasil Belajar Siswa. *Jurnal Evaluasi Pendidikan, 11*(1), 45–53.
- Fauziah, R., & Widodo, S. (2020). Analisis tingkat kesukaran soal sebagai indikator kualitas instrumen evaluasi pembelajaran. *Jurnal Ilmiah Pendidikan Dasar, 5*(2), 101–109.
- Fietri, W. A. (2021). Analisis Butir Soal Biologi Kelas VIII Madrasah Tsanawiyah Negeri 6 Kerinci Menggunakan Program Komputer Anates 4.0 For Windows. *Jurnal Pendidikan Biologi Undiksha*. <https://ejournal.undiksha.ac.id/index.php/JJPB/article/view/35064>
- Fiska, J. M., Hidayati, Y., Qomaria, N., & ... (2021). Analisis butir soal ulangan harian IPA menggunakan software Anates pada pendekatan teori tes klasik. *Natural Science ...* <https://journal.trunojoyo.ac.id/nser/article/view/8133>
- Ikhsaniyah, S. N., Kurnia, A. D., Zuroida, M., Pratiwi, V., & ... (2024). Analisis Butir Soal Perpajakan PPH Pasal 21 Menggunakan Software Anates pada Pendekatan Teori Tes Klasik. *PEKA*. <https://journal.uir.ac.id/index.php/Peka/article/view/19917>
- Karindi, D. R. (2024). Penggunaan ANATES 4.0. 9 dalam menguji kelayakan soal pilihan ganda sumatif matematika. *LINEAR: Journal of Mathematics Education*. <https://e-journal.metrouniv.ac.id/linear/article/view/9888>
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis butir soal ujian akhir sekolah. In *Al Asma: Journal of Islamic ...* <https://www.academia.edu/download/92430353/352718839.pdf>
- Mardiyah, H. (2024). ... APLIKASI ANATES PADA ANALISIS BUTIR SOAL PENILAIAN AKHIR JENJANG: PENGGUNAAN APLIKASI ANATES PADA ANALISIS BUTIR SOAL In *WIDYALOKA*.
- Purwati, L. M., Arianty, R., Syakilah, D. M., & ... (2021). Analisis Soal Tes Pilihan Ganda Berbasis Higher Order Thinking Skill Menggunakan Aplikasi Anates



- Windows Versi 4.0. 9 For Windows. *Jurnal Pendidikan*
<https://journal.uniga.ac.id/index.php/JP/article/view/1287>
- Rahayu, L. P., & Sukenti, D. (2024). Kualitas Soal Bahasa Indonesia Kelas XI SMAN 2 Bangko Pusako: Analisis Butir Soal. *Jurnal Onoma: Pendidikan, Bahasa, Dan* <https://e-journal.my.id/onoma/article/view/4481>
- Rani, C. A. M., Farida, A. N., Avrasta, S., & ... (2024). Analisis Butir Soal Komputer Akuntansi Kelas XI Melalui Aplikasi Anates. *Jurnal Nirta: Studi*
<https://ejournal.nlc-education.or.id/index.php/JNSI/article/view/106>
- Rufi, I. I., Akuatika, F. A., & Pratama, M. H. B. (2022). Penggunaan Aplikasi Anates pada Analisis Butir Soal Penilaian Tengah Semester Bahasa Indonesia Kelas X MIPA 8 SMA Muhammadiyah 2 Surabaya. In *Metalanguage: Jurnal Ilmu Bahasa dan Sastra*
- Sugiyono. (2022). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- Susilowati, D. A., Saputri, N., Ni'mah, K., & ... (2025). Penerapan Aplikasi Anates Untuk Analisis Soal HOTS Dalam Evaluasi Pembelajaran Spreadsheet Di SMKN 1 Kendal. *PESHUM: Jurnal* <https://al-haramjournal.com/index.php/PESHUM/article/view/7291>
- Syafiriya, D. M., Rahmawati, P. A., & ... (2024). Analisis Butir Soal HOTS untuk Menguji Pemahaman Siswa SMK Jurusan Akuntansi Menggunakan Software Anates. *Jurnal Nirta* <https://ejournal.nlc-education.or.id/index.php/JNSI/article/view/107>
- Taftazani, T. M. T., Haifa, N. H., Safira, N., & ... (2025). Analisis Kualitas Butir Soal Pemahaman Siswa Kelas 1 SD Materi Penjumlahan dan Pengurangan Menggunakan Tes Buatan Guru Berbasis Program Anates. *Jurnal*
<https://jurnal.unismuhpalu.ac.id/index.php/JKS/article/view/6666>
- YUSUF, F. W. (2024). Analisis butir soal asesmen sumatif biologi materi perubahan lingkungan dengan menggunakan anates pada kelas x sma. In *LEARNING*.
pdfs.semanticscholar.org/8d52/3bfc7b390c17aed4e7521beff10effe7e29.pdf
- Zahiroh, U. (2020). *Analisis Kualitas Butir Soal Pilihan Ganda Mata Pelajaran Kimia Pada Ujian Akhir Semester (UAS) Kelas XI MAN 2 Kepulauan Meranti*.
repository.uin-suska.ac.id. http://repository.uin-suska.ac.id/28622/2/ULFAH_ZAHIROH.pdf

