

ANALISIS KUALITAS BUTIR SOAL TES SUMATIF MATEMATIKA BERDASARKAN TEORI TES KLASIK

Nanda Rizkiansyah¹⁾, Dian Nataria Oktaviani²⁾, Rizqi Amaliyakh Sholikhakh³⁾

¹Fakultas Keguruan dan Ilmu Pendidikan Universitas Pancasakti
email: **nandarizkiansyah11@gmail.com**

²Fakultas Keguruan dan Ilmu Pendidikan Universitas Pancasakti
email: **dian85nataria@gmail.com**

³Fakultas Keguruan dan Ilmu Pendidikan Universitas Pancasakti
email: **rizqi83as@gmail.com**

Abstrak

Penilaian hasil belajar merupakan bagian penting dalam proses pembelajaran karena digunakan untuk mengetahui tingkat pencapaian kompetensi peserta didik secara objektif. Salah satu instrumen yang umum digunakan dalam evaluasi pembelajaran adalah tes Sumatif Akhir Semester (SAS). Namun, instrumen tes yang digunakan perlu dianalisis kualitasnya agar mampu memberikan gambaran yang akurat mengenai kemampuan peserta didik. Penelitian ini bertujuan untuk mengkaji kualitas butir soal tes Sumatif Akhir Semester (SAS) Matematika kelas VIII berdasarkan Teori Tes Klasik yang meliputi reliabilitas tes, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh. Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis butir soal. Data penelitian diperoleh dari hasil pengerjaan tes oleh 96 peserta didik kelas VIII SMP Negeri 3 Talang terhadap 35 butir soal yang terdiri atas bentuk pilihan ganda, pilihan ganda kompleks, menjodohkan, isian singkat, dan uraian. Hasil penelitian menunjukkan bahwa reliabilitas tes berbeda pada setiap bentuk soal, di mana soal pilihan ganda, menjodohkan, dan uraian memiliki reliabilitas kategori tinggi, sedangkan pilihan ganda kompleks dan isian singkat memiliki reliabilitas sangat rendah. Berdasarkan tingkat kesukaran, sebagian besar butir soal berada pada kategori sedang meskipun distribusinya belum sepenuhnya seimbang. Analisis daya pembeda menunjukkan bahwa mayoritas butir soal berada pada kategori baik hingga sangat baik, meskipun masih terdapat beberapa butir dengan daya pembeda rendah. Selain itu, sebagian besar pengecoh pada soal pilihan ganda telah berfungsi dengan baik, walaupun masih ditemukan beberapa distraktor yang belum bekerja secara optimal. Hasil penelitian ini menunjukkan bahwa analisis butir soal berdasarkan Teori Tes Klasik dapat memberikan informasi penting untuk mengevaluasi dan meningkatkan kualitas instrumen penilaian sehingga mampu mengukur kemampuan peserta didik secara lebih tepat dan objektif.

Kata kunci: analisis butir soal; teori tes klasik; kualitas instrumen tes; evaluasi pembelajaran; tes sumatif matematika;

PENDAHULUAN

Evaluasi pembelajaran merupakan komponen penting dalam proses pendidikan karena berfungsi untuk mengetahui sejauh mana tujuan pembelajaran telah tercapai. Evaluasi dilakukan secara sistematis untuk memperoleh informasi yang akurat mengenai kemampuan peserta didik sehingga dapat menjadi dasar pengambilan keputusan dalam pembelajaran (Arikunto & Jabar, 2018). Dalam konteks pendidikan formal, salah satu bentuk evaluasi yang umum digunakan adalah penilaian sumatif yang dilaksanakan pada akhir periode pembelajaran. Pada jenjang sekolah menengah pertama (SMP), penilaian tersebut umumnya diwujudkan dalam bentuk Sumatif Akhir Semester (SAS) yang bertujuan untuk mengukur ketercapaian kompetensi

peserta didik selama satu semester pembelajaran.

Keakuratan informasi hasil belajar yang diperoleh dari SAS sangat dipengaruhi oleh kualitas instrumen penilaian yang digunakan. Instrumen tes yang kurang berkualitas berpotensi menghasilkan data yang bias sehingga tidak mampu menggambarkan kemampuan peserta didik secara objektif. Oleh karena itu, kualitas butir soal menjadi faktor penting dalam menentukan validitas informasi yang dihasilkan dari suatu tes (Haladyna & Rodriguez, 2016). Instrumen penilaian berkualitas seharusnya mampu menaksir ukuran kemampuan peserta didik secara tepat, konsisten, serta mampu membedakan tingkat kemampuan antar peserta didik.

Dalam praktiknya, instrumen tes yang diterapkan dalam evaluasi pembelajaran sering kali belum dianalisis secara empiris setelah digunakan. Guru cenderung memanfaatkan hasil nilai akhir tanpa menjalankan analisa lebih lanjut pada karakteristik butir soal yang menyusun instrumen tersebut. Padahal, analisa butir soal merupakan tahapan penting pada evaluasi pembelajaran untuk mengetahui kualitas setiap butir serta kualitas tes secara keseluruhan (Retnawati, 2016). Tanpa analisis tersebut, butir soal terbilang terlalu mudah atau sukar, atau tidak mampu membedakan tingkat penguasaan peserta didik berpotensi tetap digunakan dalam evaluasi berikutnya sehingga kualitas instrumen penilaian tidak mengalami perbaikan.

Analisis kualitas butir soal bisa diterapkan dengan pendekatan Teori Tes Klasik (Classical Test Theory). Teori ini menyatakan bahwa skor yang diperoleh peserta didik merupakan kombinasi antara skor sejati dan kesalahan pengukuran. Pendekatan ini banyak digunakan dalam evaluasi pendidikan karena relatif sederhana dan mudah diterapkan pada analisis hasil tes di sekolah (Retnawati, 2016). Melalui pendekatan ini, kualitas instrumen dan butir soal dapat dianalisis berdasarkan karakteristik statistik yang diperoleh dari data empirik hasil tes.

Beberapa indikator yang umum digunakan dalam analisa butir soal menurut Teori Tes Klasik meliputi reliabilitas tes, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh pada soal pilihan ganda. Reliabilitas menunjukkan konsistensi hasil pengukuran suatu instrumen, tingkat kesukaran memperlihatkan proporsi peserta didik yang mampu memberikan jawaban benar suatu butir soal, sedangkan daya pembeda memperlihatkan kapasitas suatu soal dalam mengenali perbedaan peserta didik berkemampuan tinggi maupun rendah (Arikunto, 2018). Selain itu, pada soal pilihan ganda diperlukan analisis efektivitas pengecoh guna mencari tau apakah jawaban lain selain kunci berperan secara optimal dalam menguji pemahaman peserta didik (Haladyna & Rodriguez, 2016).

Berdasarkan dokumen hasil SAS Matematika kelas VIII di SMP Negeri 3 Talang Tahun Ajaran 2025/2026, didapatkan sebaran nilai peserta didik yang cukup beragam dengan rentang nilai yang lebar antara nilai terendah

dan tertinggi. Sebagian peserta didik mendapatkan nilai di bawah Kriteria Ketercapaian Tujuan Pembelajaran (KKTP), selain itu sebagian lainnya memperoleh nilai tinggi. Kondisi ini menunjukkan adanya variasi tingkat penguasaan kompetensi peserta didik. Namun demikian, nilai akhir tersebut belum tentu sepenuhnya memperlihatkan kecakapan kognitif peserta didik dengan cara objektif apabila kualitas instrumen tes yang digunakan belum dianalisis secara empiris.

Selain itu, berlandaskan informasi guru matematika di sekolah tersebut, soal SAS yang digunakan disusun secara kolektif oleh Musyawarah Guru Mata Pelajaran (MGMP) Matematika tingkat kecamatan dan selama ini belum diterapkan analisa kualitas butir soal secara sistematis. Kondisi ini menunjukkan bahwa penggunaan instrumen penilaian belum sepenuhnya didukung oleh informasi empiris mengenai karakteristik butir soal yang digunakan dalam evaluasi pembelajaran.

Sejumlah penelitian sebelumnya menunjukkan bahwa analisis butir soal berbasis Teori Tes Klasik dapat memberikan gambaran empiris mengenai kualitas instrumen penilaian. Penelitian yang diterapkan oleh Setiawan (2020) menampilkan bahwa sebagian butir soal matematika di tingkat SMP mempunyai tingkat kesukaran yang tidak proporsional serta daya pembeda yang rendah. Penelitian lain oleh Kurniawan (2021) juga menampilkan bahwa analisa kuantitatif butir soal bisa mengidentifikasi kualitas instrumen penilaian secara lebih objektif dan memberikan dasar untuk perbaikan instrumen evaluasi pembelajaran. Namun demikian, kajian yang secara khusus menganalisis kualitas butir soal Sumatif Akhir Semester dengan berbagai bentuk soal masih belum optimal.

Berlandaskan uraian tersebut, analisa kualitas butir soal menjadi bernilai untuk digunakan guna memperoleh informasi empiris mengenai kualitas instrumen penilaian yang dipakai dalam evaluasi pembelajaran. Maka dari itu, penelitian ini mempunyai tujuan untuk menganalisis kualitas butir soal pada tes sumatif matematika berdasarkan Teori Tes Klasik yang ditinjau dari reliabilitas tes, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh pada soal pilihan ganda.

METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif dengan jenis penelitian deskriptif evaluatif. Pendekatan kuantitatif digunakan guna menganalisa kualitas instrumen dan butir soal secara statistik berlandaskan pendekatan Teori Tes Klasik (Classical Test Theory). Penelitian deskriptif evaluatif mempunyai tujuan guna menggambarkan serta mengukur kualitas suatu objek secara sistematis berdasarkan kriteria tertentu tanpa memberikan perlakuan terhadap subjek penelitian (Sugiyono, 2021). Dalam penelitian ini, objek yang dievaluasi adalah kualitas butir soal tes Sumatif Akhir Semester (SAS) mata pelajaran Matematika kelas VIII.

Penelitian dilaksanakan di SMP Negeri 3 Talang, Kabupaten Tegal, pada semester gasal Tahun Pelajaran 2025/2026 setelah pelaksanaan tes Sumatif Akhir Semester. Pemilihan lokasi penelitian didasarkan pada ketersediaan dokumen tes serta data respons peserta didik yang lengkap, sehingga memungkinkan dilakukannya analisis butir soal secara kuantitatif.

Populasi penelitian adalah seluruh peserta didik kelas VIII SMP Negeri 3 Talang Tahun Pelajaran 2025/2026 yang terdiri atas delapan kelas. Sampel penelitian ditentukan memakai teknik cluster sampling, yaitu pengambilan sampel berdasarkan kelompok yang telah terbentuk secara alami, seperti kelas (Sugiyono, 2021). Melalui teknik ini dipilih tiga kelas sebagai sampel penelitian, yaitu kelas VIII D, VIII E, dan VIII H dengan jumlah total 96 peserta didik. Jumlah sampel tersebut dinilai memadai untuk analisa butir soal berbasis Teori Tes Klasik karena analisa parameter butir membutuhkan jumlah responden yang cukup agar hasil perhitungan statistik lebih stabil (Nitko & Brookhart, 2014).

Tabel 1. berikut menunjukkan distribusi sampel penelitian.

Kelas	Jumlah Peserta Didik
VIII D	32
VIII E	32
VIII H	32
Jumlah	96

Sumber data yang diteliti terdiri dari dua jenis dokumen, yaitu naskah soal Sumatif Akhir

Semester (SAS) Matematika kelas VIII dan lembar jawaban peserta didik. Instrumen penelitian berupa dokumen soal SAS yang dianalisis kualitas butirnya berdasarkan respons peserta didik terhadap soal tersebut.

Teknik pengumpulan data yang dipakai yaitu dokumentasi, dimana teknik pengambilan data dengan menghimpun dan mempelajari dokumen yang relevan dengan kebutuhan penelitian. Dokumen yang digunakan mencakup naskah soal SAS, kunci jawaban, serta lembar jawaban peserta didik. Temuan respons peserta didik kemudian dikodekan sesuai dengan jenis soal untuk memperoleh skor setiap peserta didik dan skor pada setiap butir soal.

Prosedur penelitian dilakukan melalui beberapa tahapan, yaitu: (1) pengumpulan dokumen soal SAS dan lembar jawaban peserta didik; (2) penentuan sampel penelitian menggunakan teknik cluster sampling; (3) pengolahan data respons peserta didik dengan melakukan penskoran terhadap setiap butir soal; (4) analisa butir soal secara kuantitatif berdasarkan Teori Tes Klasik; serta (5) interpretasi hasil analisis untuk menentukan kualitas setiap butir soal dan kualitas instrumen tes secara keseluruhan.

Analisis data pada eksplorasi ilmiah ini diterapkan menggunakan pendekatan Teori Tes Klasik untuk menemukan karakteristik statistik butir soal. Analisis dilakukan terhadap berbagai bentuk soal yang ada pada instrumen SAS, yaitu pilihan ganda, pilihan ganda kompleks, menjodohkan, isian singkat, dan uraian. Parameter yang dianalisis meliputi reliabilitas tes, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh pada butir yang memiliki alternatif jawaban.

Reliabilitas tes digunakan untuk mengetahui tingkat konsistensi instrumen untuk mengukur tingkat penguasaan peserta didik. Butir tes dengan skor dikotomis, seperti pilihan ganda, pilihan ganda kompleks, menjodohkan, dan isian singkat, reliabilitas diukur dengan rumus Kuder-Richardson 20 (KR-20). Sementara itu, reliabilitas pada butir uraian dihitung menggunakan koefisien Alpha Cronbach, karena bentuk soal tersebut memungkinkan pemberian skor parsial (Arikunto, 2018).

Tingkat kesukaran butir soal mengindikasikan proporsi peserta didik yang menjawab soal dengan benar. Indeks tingkat kesukaran diukur menggunakan perbandingan

diantara jumlah peserta didik memilih jawaban benar dengan jumlah seluruh peserta tes. Untuk soal uraian, tingkat kesukaran diukur berdasarkan perbandingan antara rata-rata skor yang didapat peserta didik dengan skor tertinggi yang mungkin dicapai (Sudijono, 2015).

Daya pembeda butir soal menunjukkan tingkat penguasaan suatu soal dalam melihat perbedaan peserta didik berkemampuan tinggi dan rendah. Perhitungan daya pembeda diukur dengan melihat perbandingan proporsi jawaban benar antara kelompok atas dan kelompok bawah yang setiap kelompok diambil sebesar 27% dari jumlah peserta tes (Arikunto, 2018).

Selain itu, analisa efektivitas pengecoh diterapkan pada soal pilihan ganda dan pilihan ganda kompleks guna mengidentifikasi apakah alternatif jawaban selain kunci berfungsi dengan optimal. Pengecoh dinyatakan efektif apabila dipilih oleh minimal 5% peserta tes, sehingga mampu berperan sebagai alternatif jawaban yang logis dan tidak mudah dikenali sebagai jawaban yang salah (Haladyna & Rodriguez, 2016).

Hasil analisa butir kemudian disajikan kedalam bentuk tabel hasil analisis butir soal dan tabel rekapitulasi kategori kualitas butir soal, serta diinterpretasikan secara deskriptif untuk memberikan gambaran mengenai kualitas instrumen tes Sumatif Akhir Semester Matematika kelas VIII secara keseluruhan.

HASIL DAN PEMBAHASAN

1. Hasil Penelitian

Penelitian ini menganalisis kualitas 35 butir soal tes Sumatif Akhir Semester (SAS) Matematika kelas VIII yang terdiri atas lima bentuk soal, yaitu pilihan ganda, pilihan ganda kompleks, menjodohkan, isian singkat, dan uraian. Analisis dilakukan menerapkan pendekatan Teori Tes Klasik dengan parameter reliabilitas tes, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh. Data penelitian diperoleh melalui hasil pengerjaan tes oleh 96 peserta didik.

Reliabilitas Tes

Reliabilitas tes dianalisis guna mengukur tingkatan kekonsistenan instrumen pada kemampuan peserta didik. Hasil analisis reliabilitas berdasarkan bentuk soal disajikan pada Tabel 1 berikut.

Tabel 2. Rekapitulasi Reliabilitas Tes

Bentuk Soal	Jumlah Butir	Koefisien Reliabilitas	Kategori
Pilihan ganda	15	0,60	Tinggi
Pilihan ganda kompleks	5	0,02	Sangat rendah
Menjodohkan	5	0,64	Tinggi
Isian singkat	5	0,03	Sangat rendah
Uraian	5	0,63	Tinggi

Berdasarkan Tabel 1, tiga tipe soal berbentuk pilihan ganda, menjodohkan, dan uraian mempunyai reliabilitas tinggi sehingga dapat dikategorikan reliabel. Sebaliknya, pilihan ganda kompleks dan isian singkat memiliki reliabilitas sangat rendah sehingga belum memenuhi kriteria instrumen yang andal.

Tingkat Kesukaran Butir Soal

Analisa tingkat kesukaran dilakukan guna mencari tau sejauh mana proporsi peserta didik yang bisa menjawab suatu butir dengan benar. Secara umum, hasil analisis menunjukkan bahwa butir soal memiliki variasi tingkatan kesukaran dengan kategori yang meliputi tingkatan mudah, sedang, dan sukar.

Distribusi pada kategori tingkat kesukaran secara ringkas dapat disajikan pada Tabel 2.

Tabel 3. Distribusi Tingkat Kesukaran Butir

Bentuk Soal	Soal		
	Mudah	Sedang	Sukar
Pilihan ganda	0	9	3
Pilihan ganda kompleks	1	2	2
Menjodohkan	1	3	1
Isian singkat	1	1	3
Uraian	0	0	5

Hasil tersebut menunjukkan bahwa mayoritas soal pilihan ganda berkategori sedang, sedangkan seluruh soal uraian mempunyai kategori sukar.

Daya Pembeda Butir Soal

Daya pembeda dianalisis dengan metode kelompok atas dan kelompok bawah (27%) untuk mengetahui kapasitas tingkat butir soal dalam melihat perbedaan peserta didik dengan kemampuan tinggi dan rendah.

Distribusi kategori daya pembeda disajikan pada Tabel 3.

Tabel 4. Distribusi Daya Pembeda Butir Soal

Bentuk Soal	Sangat Baik	Baik	Cukup	Jelek
Pilihan ganda	1	8	3	2
Pilihan ganda kompleks	0	3	2	0
Menjodohkan	3	1	1	0
Isian singkat	1	1	1	2
Uraian	0	2	2	1

Berdasarkan tabel tersebut, sebagian besar butir soal mempunyai daya pembeda yang baik hingga sangat baik, meskipun masih terdapat beberapa butir dengan daya pembeda rendah.

Efektivitas Pengecoh

Analisa efektivitas pengecoh diterapkan pada soal pilihan ganda dan pilihan ganda

kompleks. Pengecoh dinyatakan efektif apabila dipilih oleh minimal 5% peserta didik (Haladyna & Rodriguez, 2016).

Hasil analisis menampilkan bahwa mayoritas pengecoh berfungsi secara efektif. Namun demikian, beberapa alternatif jawaban masih belum dipilih oleh peserta didik, dengan kata lain tidak berfungsi sebagai distraktor yang optimal.

2. Pembahasan

Reliabilitas Instrumen

Hasil analisa menampilkan bahwa reliabilitas instrumen bervariasi antar tipe soal. Soal pilihan ganda, menjodohkan, serta uraian memiliki reliabilitas tinggi, sedangkan pilihan ganda kompleks dan isian singkat memiliki reliabilitas sangat rendah. Temuan ini menunjukkan bahwa format soal dan sistem penskoran memiliki pengaruh terhadap konsistensi hasil pengukuran.

Instrumen dengan reliabilitas tinggi menunjukkan bahwa butir soal mempunyai nilai keajegan internal yang baik untuk mengukur tingkat penguasaan peserta didik (Retnawati, 2016). Sebaliknya, reliabilitas yang rendah menunjukkan bahwa butir soal belum mampu memberikan hasil pengukuran yang stabil.

Rendahnya reliabilitas pada soal pilihan ganda kompleks dan isian singkat kemungkinan dipengaruhi oleh struktur soal yang kurang jelas, variasi jawaban peserta didik, serta ketidakkonsistenan dalam sistem penskoran. Hal ini menunjukkan pentingnya penyusunan pedoman penskoran yang jelas serta revisi terhadap butir soal yang kurang konsisten.

Tingkat Kesukaran Butir Soal

Hasil analisa menampilkan bahwa sebagian besar soal pilihan ganda berkategori sedang. Secara teoretis, soal dengan tingkatan kesukaran sedang merupakan soal yang paling ideal dikarenakan mampu menyajikan informasi yang optimal mengenai tingkat penguasaan peserta didik (Retnawati, 2016).

Namun demikian, seluruh soal uraian berkategori sukar. Kondisi ini menunjukkan bahwa tingkat kompleksitas soal uraian relatif tinggi dengan demikian hanya sebagian kecil peserta didik yang bisa menjawab dengan benar. Distribusi tingkat kesukaran yang tidak seimbang dapat mengurangi efektivitas tes dalam mengukur kemampuan peserta didik secara komprehensif.

Penelitian sebelumnya juga menunjukkan bahwa variasi tingkat kesukaran menjadi salah satu indikator utama pada kualitas instrumen evaluasi pembelajaran (Wibowo & Kartowagiran, 2019). Oleh karena itu, penyusunan soal perlu mempertimbangkan proporsi tingkat kesukaran yang lebih seimbang.

Daya Pembeda Butir Soal

Hasil analisa menampilkan bahwa sebagian besar butir soal mempunyai daya pembeda pada kategori baik hingga sangat baik. Hal ini mengindikasikan bahwa instrumen SAS mampu melihat perbedaan peserta didik yang memiliki kemampuan tinggi dan rendah.

Butir soal dengan daya pembeda tinggi merupakan indikator bahwa soal tersebut efektif dalam mengidentifikasi tingkat penguasaan peserta didik (Retnawati, 2016). Sebaliknya, butir soal yang mempunyai daya pembeda rendah menunjukkan bahwa soal tersebut kurang mampu melihat perbedaan tingkat kemampuan peserta didik sehingga perlu direvisi.

Temuan ini sependapat dengan penelitian oleh Putri dan Widodo (2020) yang menampilkan bahwa butir soal dengan daya pembeda rendah umumnya disebabkan oleh ketidaksesuaian indikator soal, tingkat kesukaran yang tidak proporsional, atau redaksi soal yang kurang jelas.

Efektivitas Pengecoh

Analisis efektivitas pengecoh menunjukkan bahwa sebagian besar alternatif jawaban pada soal pilihan ganda telah berfungsi dengan baik. Pengecoh yang efektif merupakan alternatif jawaban yang mampu menarik perhatian peserta didik sehingga berfungsi sebagai distraktor yang logis (Haladyna & Rodriguez, 2016).

Namun demikian, masih ditemukan beberapa pengecoh yang tidak dipilih oleh peserta didik. Kondisi ini menunjukkan bahwa pengecoh tersebut terlalu mudah dikenali sebagai jawaban yang salah atau tidak sesuai dengan pola kesalahan yang umum dilakukan peserta didik.

Oleh karena itu, penyusunan pengecoh perlu mempertimbangkan miskonsepsi yang sering dialami peserta didik, sehingga setiap alternatif jawaban memiliki kemungkinan dipilih.

3. Implikasi Pendidikan

Hasil penelitian ini menunjukkan bahwa analisis butir soal berbasis Teori Tes Klasik

memberikan informasi penting mengenai kualitas instrumen evaluasi pembelajaran. Temuan penelitian ini dapat menjadi dasar bagi guru dalam melakukan evaluasi dan perbaikan instrumen penilaian secara sistematis.

Penerapan analisis butir soal secara berkala dapat membantu guru dalam:

1. Meningkatkan reliabilitas instrumen penilaian,
2. Menyusun distribusi tingkat kesukaran yang lebih proporsional,
3. Memperbaiki butir soal dengan daya pembeda rendah, dan
4. Merancang pengecoh yang lebih efektif.

Dengan demikian, instrumen evaluasi yang digunakan dalam pembelajaran dapat memberikan gambaran tingkat penguasaan peserta didik secara lebih tepat sasaran dan objektif.

4. Keterbatasan Penelitian

Pada penelitian ini mempunyai beberapa keterbatasan. Pertama, analisis kualitas butir soal dilakukan hanya menggunakan pendekatan Teori Tes Klasik, sehingga belum mempertimbangkan analisis yang lebih modern seperti Item Response Theory (IRT). Kedua, penelitian ini hanya dilaksanakan pada satu sekolah dengan jumlah sampel yang terbatas sehingga hasil penelitian belum bisa digeneralisasikan secara meluas.

Penelitian selanjutnya dapat mengembangkan analisis kualitas butir soal menggunakan pendekatan yang lebih komprehensif serta melibatkan sampel yang lebih besar agar memperoleh gambaran kualitas instrumen penilaian yang lebih luas.

KESIMPULAN DAN REKOMENDASI

Penelitian ini mempunyai tujuan untuk menganalisa kualitas butir soal tes Sumatif Akhir Semester (SAS) Matematika kelas VIII menggunakan pendekatan Teori Tes Klasik yang meliputi aspek reliabilitas tes, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Berlandaskan hasil analisa terhadap 35 butir soal yang dikerjakan oleh 96 peserta didik, diperoleh beberapa temuan utama terkait kualitas instrumen penilaian yang digunakan.

Secara umum, kualitas instrumen SAS menunjukkan hasil yang bervariasi pada setiap indikator analisis. Dari aspek reliabilitas, beberapa bentuk soal seperti pilihan ganda, menjodohkan, dan uraian mempunyai reliabilitas yang tergolong tinggi sehingga

dapat dikatakan cukup konsisten dalam mengukur kemampuan peserta didik. Namun demikian, tipe soal pilihan ganda kompleks serta isian singkat menunjukkan reliabilitas yang sangat rendah, yang mengindikasikan bahwa konsistensi pengukuran antarbutir belum sepenuhnya stabil. Menurut Retnawati (2016), reliabilitas merupakan indikator penting dalam menentukan konsistensi suatu instrumen evaluasi sehingga tes yang reliabel akan menunjukkan hasil pengukuran yang lebih stabil dan dapat dipercaya.

Ditinjau dari tingkat kesukaran, instrumen SAS telah memuat variasi kategori mudah, sedang, dan sukar, meskipun distribusinya belum sepenuhnya proporsional. Mayoritas soal pilihan ganda berkategori sedang, sedangkan seluruh soal uraian berada pada kategori sukar. Kondisi ini menunjukkan bahwa instrumen belum sepenuhnya mampu mengukur kemampuan peserta didik secara merata pada berbagai tingkat kemampuan. Distribusi tingkat kesukaran yang seimbang diperlukan agar instrumen mampu memberikan informasi yang optimal mengenai kemampuan peserta didik (Wibowo & Kartowagiran, 2019).

Pada aspek daya pembeda, mayoritas butir soal berada pada kategori baik hingga sangat baik sehingga mampu melihat perbedaan peserta didik yang mempunyai kemampuan tinggi dan rendah. Namun demikian, masih terdapat beberapa butir soal dengan daya pembeda rendah yang menampilkan bahwa butir tersebut belum mampu memberikan informasi secara optimal mengenai perbedaan kemampuan peserta didik. Hal ini dapat disebabkan oleh ketidaksesuaian antara indikator pembelajaran, tingkat kesukaran soal, maupun kejelasan redaksi soal.

Sementara itu, hasil analisa efektivitas pengecoh menampilkan bahwa sebagian besar alternatif jawaban pada soal tipe pilihan ganda telah berfungsi secara efektif. Meskipun demikian, masih ditemukan beberapa pengecoh yang tidak dipilih oleh peserta didik sehingga tidak berfungsi secara optimal sebagai distraktor. Hal ini menunjukkan bahwa penyusunan pengecoh belum sepenuhnya mempertimbangkan pola kesalahan atau miskonsepsi yang umum dialami peserta didik. Pengecoh yang efektif seharusnya mampu menarik perhatian peserta didik yang belum menguasai materi sehingga dapat

meningkatkan kualitas pengukuran tes (Haladyna & Rodriguez, 2016).

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa instrumen tes SAS Matematika kelas VIII yang dianalisis mempunyai kualitas yang cukup baik pada beberapa indikator, namun masih memerlukan perbaikan pada aspek reliabilitas tertentu, distribusi tingkat kesukaran, daya pembeda beberapa butir soal, serta efektivitas pengecoh. Oleh karena itu, analisis butir soal berbasis Teori Tes Klasik dapat menjadi langkah penting dalam proses evaluasi dan penyempurnaan instrumen penilaian agar lebih valid, reliabel, dan mampu mengukur tingkat penguasaan peserta didik dengan lebih akurat.

Berdasarkan hasil penelitian tersebut, beberapa rekomendasi dapat diajukan. Pertama, guru dan penyusun instrumen penilaian disarankan untuk melakukan analisis butir soal secara berkala sebelum dan sesudah tes digunakan guna memastikan kualitas instrumen yang lebih baik. Kedua, penyusunan soal perlu memperhatikan proporsi tingkat kesukaran, meningkatkan daya pembeda butir soal, serta merancang pengecoh berdasarkan kesalahan konsep yang umum dialami peserta didik. Ketiga, penelitian selanjutnya disarankan untuk mengembangkan analisis kualitas butir soal dengan cakupan sampel yang lebih luas serta mengombinasikan pendekatan Teori Tes Klasik dengan pendekatan lain seperti Teori Respons Butir (Item Response Theory) agar diperoleh gambaran kualitas instrumen penilaian yang lebih komprehensif.

Dengan demikian, hasil penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan kualitas evaluasi pembelajaran matematika di tingkat sekolah menengah pertama serta mendorong penggunaan analisis butir soal sebagai bagian dari praktik evaluasi pendidikan yang lebih sistematis dan berbasis data.

DAFTAR PUSTAKA

- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Aminatus Zuhriyah, I. (2024). *Analisis validitas dan reliabilitas butir soal sumatif akhir semester ganjil mata pelajaran Pendidikan Agama Islam*. Edukatif: Jurnal Ilmu Pendidikan.
- Arikunto, S., & Jabar, C. S. A. (2018). *Evaluasi program pendidikan: Pedoman teoretis praktis bagi mahasiswa dan praktisi pendidikan*. Jakarta: Bumi Aksara.

- Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan* (Edisi revisi). Jakarta: Bumi Aksara.
- Asri, S. A., Syahrul, & Ishak, S. (2025). Analisis kualitas butir soal ulangan akhir semester genap matematika kelas VIII SMPN Sinjai. *Pedagogy: Jurnal Pendidikan Matematika*, 10(1), 120–133.
- BSNP. (2018). Panduan penilaian oleh pendidik dan satuan pendidikan untuk SMP. Jakarta: Badan Standar Nasional Pendidikan.
- Creswell, J. W. (2018). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (5th ed.). Boston, MA: Pearson Education.
- Haladyna, T. M., & Rodriguez, M. C. (2016). *Developing and validating test items* (3rd ed.). New York: Routledge.
- Ismail, K. M., & Zubairi, A. M. (2022). *Item analysis of a reading test in Sri Lankan context using Classical Test Theory*. *International Journal of Learning, Teaching and Educational Research*.
- Kongnyuy, P., & Ndeyeh, R. (2025). An analysis of item difficulty and item discrimination of mathematics multiple choice items. *East African Journal of Education Studies*, 8(3), 505–511.
- Kurniawan, R. (2021). Analisis kualitas butir soal sumatif mata pelajaran Matematika SMP. *Jurnal Evaluasi Pendidikan*, 12(2), 145–156.
- Permendikbud Nomor 23 Tahun 2016 tentang Standar Penilaian Pendidikan. (2016). Jakarta: Kementerian Pendidikan dan Kebudayaan Republik Indonesia.
- Permendikbud Republik Indonesia Nomor 23 Tahun 2016. Standar penilaian pendidikan. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Retnawati, H. (2016). Analisis kuantitatif instrumen penelitian (Panduan peneliti, mahasiswa, dan psikometrian). Yogyakarta: Parama Publishing.
- Retnawati. (2016). *Analisis butir soal: Teori dan aplikasi* (Edisi terbaru). Yogyakarta: Parama Publishing.
- Retnawati, H., & Sunarto. (2019). Analisis kesesuaian soal uraian dan isian singkat terhadap kompetensi dasar. *Jurnal Penelitian dan Evaluasi Pendidikan*, 23(1), 85–97.
- Setiawan, D. (2020). Analisis butir soal pilihan ganda Matematika SMP berdasarkan teori tes klasik. *Jurnal Pendidikan Matematika*, 14(1), 55–66.
- Sudjana, N. (2017). *Penilaian hasil proses belajar mengajar*. Bandung: Remaja Rosdakarya.
- Sugiyono. (2022). *Metode penelitian pendidikan (Pendekatan kuantitatif, kualitatif, dan R&D)*. Bandung: Alfabeta.
- Widyaningsih, S. W. (2024). *Classical Test Theory evaluation with ITEMAN 4.3*. AL-ISHLAH: Jurnal Pendidikan.