

Analisis Butir Soal pada Olimpiade Matematika Tingkat Mahasiswa di STAIN Mandailing Natal

Muhammad Zia Alghar¹, Rina Sari Lubis², Aslamiah³, Nisa Cahya Pertiwi Lubis⁴

^{1,2,3,4}*Sekolah Tinggi Agama Islam Negeri Mandailing Natal, Jl. Prof. Andi Hakim Nasution, Mandailing Natal, Sumatera Utara*

e-mail: muhammadzia1904@gmail.com¹, rinalubis44@gmail.com²,
aslamiahpulungan@gmail.com³, nisacahya22@gmail.com⁴

ABSTRAK

Penelitian ini bertujuan mengetahui kualitas butir soal pada Olimpiade Matematika dalam kegiatan *Geometric Competition 2025* di STAIN Mandailing Natal. Fokus penelitian dilakukan pada sepuluh soal pilihan ganda yang digunakan dalam kompetisi tersebut. Penelitian ini menggunakan pendekatan kuantitatif dengan jenis analisis deskriptif. Data penelitian berasal dari lembar jawaban 64 peserta yang merupakan mahasiswa yang mengikuti kompetisi. Data dianalisis menggunakan statistik klasik yang meliputi uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Hasil analisis menunjukkan bahwa 60% soal berada dalam kategori valid. Hasil uji reliabilitas menunjukkan butir soal belum mencapai batas yang ditetapkan untuk kategori reliabel. Hasil tingkat kesukaran menunjukkan 100% soal berada pada kategori sukar yang menuntut kemampuan tinggi peserta. Hasil daya pembeda menunjukkan banyak soal berada pada kategori lemah dan sedang. Sedangkan hasil analisis distraktor menunjukkan bahwa 87.5% distraktor berfungsi dengan baik. Penelitian ini menunjukkan bahwa kualitas butir soal dalam *Geometric Competition 2025* masih perlu diperbaiki. Butir soal perlu peninjauan ulang agar dapat menilai kemampuan matematis peserta secara lebih tepat.

Kata Kunci: analisis butir soal, olimpiade, matematika, mahasiswa.

ABSTRACT

This study aims to determine the quality of questions in the Mathematics Olympiad in the 2025 Geometric Competition at STAIN Mandailing Natal. The focus of the study was on ten multiple-choice questions used in the competition. This study used a quantitative approach with descriptive analysis. The research data came from the answer sheets of 64 participants who were students participating in the competition. The data were analyzed using classical statistics, including validity, reliability, difficulty level, discriminating power, and distractor effectiveness tests. The results of the analysis showed that 60% of the questions were in the valid category. The reliability test results showed that the questions had not reached the threshold set for the reliable category. The difficulty level results showed that 100% of the questions were in the difficult category, requiring high ability from participants. The discriminating power results showed that many questions were in the weak and moderate categories. Meanwhile, the distractor analysis results showed that 87.5% of the distractors functioned well. This study shows that the quality of the questions in the 2025 Geometric Competition still needs improvement. The questions need to be reviewed in order to assess the mathematical abilities of the participants more accurately.

Keywords: item analysis, olympiad, mathematics, students.

PENDAHULUAN

Olimpiade matematika merupakan ajang kompetisi yang dirancang untuk mengembangkan kemampuan berpikir tingkat tinggi dan menumbuhkan budaya akademik yang kompetitif (Nieto-Said & Sánchez-Lamoneda, 2022). Kompetisi ini tidak hanya mengukur kemampuan kognitif peserta didik, tetapi juga mendorong kepercayaan diri dan motivasi dalam belajar matematika (Amado & Carreira, 2018; Keldibekova, 2021). Lebih lanjut, banyak institusi pendidikan memanfaatkan olimpiade untuk menemukan bakat-bakat unggul yang berpotensi mengembangkan keilmuan matematika lebih lanjut (Kenderov, 2022). Beberapa penelitian menegaskan bahwa kompetisi matematika mampu meningkatkan kemampuan pemecahan masalah dan kreativitas matematis di tingkat mahasiswa (de Losada & Taylor, 2022; Grugnetti & Jaquet, 2005; Marushina, 2021). Oleh karena itu, olimpiade matematika memegang peran penting dalam pengembangan akademik di tingkat sekolah menengah hingga ranah perguruan tinggi.

Soal-soal dalam olimpiade matematika biasanya memiliki karakteristik yang berbeda dari soal pada evaluasi kelas reguler. Penyusunan soal olimpiade umumnya mengacu pada kemampuan berpikir tingkat tinggi atau *Higher Order Thinking Skills* (HOTS) yang menuntut peserta untuk menganalisis, mengevaluasi, dan mencipta solusi (Saul & Vaderlind, 2022; Tohir, 2019). Tingkat kesulitan soal olimpiade dirancang untuk menguji ketelitian, ketekunan, dan kemampuan berpikir siswa dengan bentuk soal non-rutin (Tohir, 2019). Tujuan penyusunan soal olimpiade yaitu untuk menyeleksi peserta dengan kemampuan terbaik, bukan sekadar menilai penguasaan materi (Grugnetti & Jaquet, 2005; Jacinto & Carreira, 2012; Toh, 2015). Oleh sebab itu, pola penyelesaian soal olimpiade sangat berbeda jauh dari ujian biasa.

Karakteristik soal olimpiade berbeda dari soal ujian tengah semester, ujian akhir semester, maupun kuis harian. Soal UTS atau UAS biasanya disusun berdasarkan capaian pembelajaran yang terstandar, sedangkan soal olimpiade disusun untuk memetakan kemampuan matematika peserta (Saat, 2020). Asesmen terhadap soal olimpiade menuntut analisis yang lebih mendalam, karena tingkat variasi kesulitan yang lebih ekstrem (Nieto-Said & Sánchez-Lamoneda, 2022). Meski demikian, teknik analisis dengan analisis butir soal seperti indeks kesukaran, daya pembeda, dan efektivitas distraktor tetap dapat diterapkan untuk mengukur kualitas soal olimpiade (Saat, 2020; Tilaar & Hasriyanti, 2019).

Analisis butir soal merupakan prosedur sistematis untuk mengevaluasi kualitas setiap pertanyaan dalam suatu tes (Arikunto, 2021). Analisis ini bertujuan mengetahui sejauh mana soal dapat berfungsi sesuai tujuan pengukuran yang telah dirancang. Dalam praktiknya, analisis butir soal mencakup penghitungan indeks kesukaran, daya pembeda, efektivitas distraktor, reliabilitas, dan validitas (Hamimi et al., 2020; Patandean, 2020). Para ahli

menyatakan bahwa analisis butir soal penting dilakukan untuk menjamin instrumen dapat memberikan hasil yang valid (Aiken, 1985; Arikunto, 2021; Gokbulut & Kus, 2019).

Distraktor atau pengecoh merupakan pilihan jawaban yang sengaja disusun untuk mengidentifikasi pola kesalahan peserta (Haladyna & Rodriguez, 2021). Distraktor yang baik harus mampu menarik perhatian peserta yang belum menguasai materi dan tetap dapat dihindari oleh peserta yang lebih kompeten. Pada soal olimpiade, fungsi distraktor sering kali lebih kompleks. Hal ini karena peserta menghadapi alternatif jawaban yang mirip dan membutuhkan penalaran mendalam (Asamoah & Ocansey, 2019; Toh, 2015). Beberapa penelitian menyatakan bahwa distraktor pada soal HOTS cenderung tidak berfungsi apabila peserta memiliki kemampuan yang terlalu rendah atau terlalu tinggi (Maryani et al., 2021). Oleh karena itu, efektivitas distraktor pada soal olimpiade perlu dianalisis secara khusus.

Adapun indeks kesukaran mendeskripsikan proporsi peserta yang mampu menjawab suatu soal dengan benar (Arikunto, 2021). Pada tes ujian sumatif, soal yang baik biasanya memiliki tingkat kesukaran moderat, karena dapat membedakan peserta dengan kemampuan yang beragam. Namun, pada olimpiade matematika, indeks kesukaran sering kali lebih tinggi. Hal ini bertujuan untuk menyeleksi peserta terbaik (Grugnetti & Jaquet, 2005). Beberapa penelitian menunjukkan bahwa tes seleksi pada kompetisi matematika biasanya memiliki tingkat kesukaran yang ekstrem (Keldibekova, 2021; Nieto-Said & Sánchez-Lamonedá, 2022). Sehingga, indeks kesukaran soal olimpiade perlu menyesuaikan dengan tujuan olimpiade.

Daya beda merupakan ukuran kemampuan sebuah soal dalam membedakan peserta berkemampuan tinggi dari peserta berkemampuan rendah (Haladyna & Rodriguez, 2021). Pada tes reguler, soal dengan daya pembeda tinggi sangat diperlukan untuk memperoleh hasil evaluasi yang adil dan representatif. Namun pada olimpiade, hasil daya beda bisa berbeda. Hal ini karena sebagian besar peserta memiliki motivasi yang tinggi, tetapi kemampuannya beragam (Amado & Carreira, 2018). Beberapa penelitian menunjukkan bahwa soal berdaya beda 'luar biasa' lebih umum dijumpai pada tes kompetitif seperti olimpiade dan seleksi akademik (Mahendra et al., 2020; Maryani et al., 2021). Oleh karena itu, analisis daya beda dalam konteks olimpiade perlu mendapatkan perhatian khusus.

Selain itu, uji reliabilitas dan validitas menjadi aspek penting dalam analisis butir soal. Uji ini berperan dalam memastikan hasil yang valid dan mengukur kemampuan yang diinginkan (Arikunto, 2021). Pada tes sumatif, reliabilitas yang tinggi menjadi indikator bahwa soal bisa digunakan untuk menilai capaian pembelajaran dengan konsisten (Khoshaim & Rashid, 2016). Dalam konteks olimpiade, reliabilitas tidak selalu menjadi tuntutan utama. Hal ini karena tes bersifat selektif dan hanya digunakan satu kali (Haladyna & Rodriguez, 2021). Meskipun demikian, validitas isi tetap penting karena tes harus mencerminkan kemampuan

matematis yang menjadi tujuan olimpiade (Siregar & Panjaitan, 2022; Toh, 2015). Oleh sebab itu, analisis reliabilitas dan validitas relevan untuk menilai kualitas soal olimpiade.

Di sisi lain, beberapa penelitian telah membahas analisis butir soal dalam konteks pendidikan matematika. Penelitian oleh Susetyawati & Wicaksono (2023) mengevaluasi kualitas soal matematika sekolah menengah dan menemukan bahwa banyak soal yang belum memenuhi standar daya pembeda. Penelitian oleh Amiruddin et al. (2020) menganalisis Ujian Akhir Semester pada mata pelajaran matematika dan menemukan bahwa distraktor banyak yang tidak berfungsi. Sedangkan studi Asamoah & Ocansey (2019) mengkaji analisis butir soal pada kompetisi matematika di tingkat sekolah dan memaparkan tingkat kesukaran yang sangat tajam. Sementara itu, penelitian Kurniawan (2019) membahas analisis reliabilitas soal matematika dan menegaskan pentingnya validitas isi pada butir soal. Beberapa penelitian tersebut menunjukkan analisis butir soal menjadi penting saat penyusunan tes matematika.

Meskipun banyak penelitian yang membahas analisis butir soal matematika, sebagian besar fokusnya masih berada pada instrumen evaluasi kelas seperti UTS, UAS, atau ujian harian. Masih sedikit penelitian yang secara khusus menganalisis kualitas soal kompetisi matematika di tingkat perguruan tinggi. Lebih spesifik, penelitian yang mengkaji efektivitas distraktor, daya beda, dan indeks kesukaran pada soal olimpiade mahasiswa masih jarang ditemukan. Oleh karena itu, penelitian ini mengisi celah tersebut dengan menelaah butir soal olimpiade matematika yang digunakan pada kompetisi matematika di tingkat mahasiswa.

Penelitian ini menganalisis butir-butir soal yang digunakan pada Olimpiade Matematika "Geometric Competition 2025", sebuah kegiatan yang diselenggarakan oleh Program Studi Tadris Matematika STAIN Mandailing Natal. Kompetisi ini diikuti oleh mahasiswa dari berbagai program studi, sehingga menggambarkan kemampuan matematis pada populasi yang heterogen. Selain itu, kehadiran peserta dari luar prodi matematika menghasilkan variasi kemampuan yang menarik untuk dievaluasi melalui analisis butir soal. Dengan demikian, hasil penelitian ini tidak hanya memberikan informasi mengenai kualitas soal olimpiade, tetapi juga memberikan gambaran mengenai profil kemampuan matematis mahasiswa di lingkungan STAIN Mandailing Natal.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian deskriptif. Jenis penelitian deskriptif bertujuan untuk menggambarkan karakteristik butir soal tanpa menguji korelasi antar variabel. Data penelitian ini berupa hasil pengerjaan peserta terhadap 10 soal pilihan ganda yang diujikan dalam Olimpiade Matematika Geometric Competition II Tahun 2025 di Sekolah Tinggi Agama Islam Negeri Mandailing Natal.

Sumber data berasal dari lembar jawaban seluruh peserta yang mengikuti olimpiade, yaitu 64 mahasiswa dari berbagai program studi di STAIN Mandailing Natal. Pengumpulan data dilakukan melalui dokumentasi, yaitu dengan mengumpulkan lembar jawaban asli dan merekapnya ke dalam file excel untuk memudahkan proses analisis. Setiap pilihan jawaban dikonversi menjadi data numerik sehingga dapat dianalisis secara statistik. Data ini menjadi dasar untuk menganalisis indeks kesukaran, daya beda, reliabilitas, validitas, dan distraktor.

Analisis data dilakukan dengan teknik analisis butir soal yang meliputi indeks kesukaran, daya pembeda, efektivitas distraktor, uji validitas, dan uji reliabilitas pada seluruh item tes. Indeks kesukaran dan daya pembeda dihitung dengan rumus standar analisis tes yang umum digunakan dalam evaluasi pendidikan. Uji efektivitas distraktor dilakukan dengan melihat distribusi pemilihan opsi oleh peserta. Adapun uji reliabilitas dianalisis menggunakan koefisien *Cronbach Alpha*. Prosedur penelitian dilakukan melalui beberapa tahapan, (1) identifikasi data tes, (2) pengkodean jawaban ke dalam format numerik, (3) perhitungan statistik tiap butir, (4) interpretasi hasil analisis, dan (5) penyusunan kesimpulan terkait kualitas soal. Prosedur penelitian dilakukan secara sistematis agar hasil penelitian dapat menggambarkan kualitas butir soal.

HASIL DAN PEMBAHASAN

Hasil Analisis Distraktor (Pengecoh)

Analisis distraktor dilakukan untuk menilai sejauh mana opsi jawaban selain kunci dapat menjalankan fungsinya sebagai pengecoh yang efektif. Pengecoh dianggap efektif apabila lebih banyak dipilih oleh peserta tes dari kelompok bawah daripada oleh peserta tes dari kelompok atas. Sebaliknya, jika pengecoh lebih sering dipilih oleh peserta tes kelompok atas, maka pengecoh tersebut tidak berfungsi dengan baik.

Dalam praktik analisis butir soal, pengecoh juga disebut berkualitas apabila sedikitnya dipilih oleh 5% peserta tes. Efektivitas pengecoh dianalisis menggunakan patokan perbandingan antara jumlah peserta kelompok atas (NA) dan kelompok bawah (NB), yaitu: pengecoh berfungsi apabila $NA < NB$, tidak berfungsi apabila $NA = NB$, dan menyesatkan apabila $NA > NB$. Lebih lanjut, Tabel 1 menyajikan hasil analisis distraktor untuk soal olimpiade matematika pada kegiatan Geometric Competition tahun 2025 di STAIN Mandailing Natal, yang dianalisis menggunakan Microsoft Excel.

Tabel 1. Hasil analisis distractor (pengecoh) pada butir soal

Kategori	Banyaknya	Persentase opsi jawaban	Persentase pengecoh
Kunci Jawaban	10	20%	0%
Pengecoh Berfungsi	35	70%	87,50%
Pengecoh tidak berfungsi	5	10%	12,50%
Pengecoh Menyesatkan	0	0%	0%

Berdasarkan Tabel 1, sebagian besar pengecoh (70%) berfungsi efektif. Hal ini menunjukkan opsi jawaban selain kunci jawaban dapat mengalihkan perhatian peserta olimpiade dengan kemampuan rendah, sesuai dengan tujuan utama dari pengecoh. Kemudian sebanyak 5 pengecoh (10%) tidak berfungsi efektif. Temuan ini menunjukkan bahwa sebagian besar pengecoh mampu mengalihkan perhatian peserta berkemampuan rendah sebagaimana tujuan soal olimpiade. Sementara itu, sebagian kecil pengecoh tidak berfungsi karena lebih sering dipilih peserta berkemampuan tinggi, yang mengindikasikan bahwa pengecoh tersebut tidak cukup kuat membedakan kemampuan peserta.

Hasil analisis ini selaras dengan penelitian Febriani (2016) bahwa pengecoh yang berfungsi dengan baik menandakan siswa yang belum cermat dalam menjawab soal. Beberapa hal yang menyebabkan pengecoh berfungsi dengan baik yaitu ketidaksiapan peserta untuk mengikuti olimpiade dengan maksimal, keterbatasan waktu pengerjaan yang membuat peserta tidak teliti dalam memahami soal, pilihan jawaban yang mirip sehingga membingungkan peserta, dan tingkat soal olimpiade yang menyulitkan peserta (Anas, 2006; Arikunto, 2021; Khasanah et al., 2023; Suzana, 2018).

Hasil Analisis Tingkat Kesukaran

Analisis tingkat kesukaran dilakukan untuk mengetahui sejauh mana sebuah butir soal tergolong mudah, sedang, atau sulit (Arikunto, 2021). Tingkat kesukaran yang ideal adalah tingkat kesukaran sedang, namun untuk konteks kompetisi matematika, soal yang tergolong sulit tetap dinilai wajar karena olimpiade menekankan aspek seleksi kemampuan tinggi (Saul & Vaderlind, 2022). Rumus yang digunakan untuk mengukur tingkat kesukaran yaitu

$$P = \frac{\bar{S}}{S_{maks}}$$

keterangan:

P : indeks kesukaran

S : skor rata-rata pada butir soal

S_{maks} : skor maksimum untuk butir soal tersebut

Tabel 2 menunjukkan hasil analisis tingkat kesukaran soal olimpiade matematika pada kegiatan Geometric Competition tahun 2025 di STAIN Mandailing Natal, yang telah dianalisis menggunakan aplikasi Microsoft Excel.

Tabel 2. Hasil analisis tingkat kesukaran butir soal

Tingkat Kesukaran	Nomor Soal	Banyaknya	Persentase
Sulit	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10	100%
Sedang	0	0	0%
Mudah	0	0	0%

Berdasarkan Tabel 2, 10 soal (100%) atau seluruh soal berada dalam kategori sulit, 0 soal (0%) termasuk dalam kategori sedang, dan 0 soal (0%) termasuk dalam kategori mudah. Dengan kata lain, tidak ada soal pada tingkat sedang dan mudah. Temuan ini menandakan bahwa seluruh soal memiliki tingkat tantangan yang tinggi, sehingga layak digunakan dalam kompetisi matematika yang menuntut kemampuan analitis dan penalaran tingkat lanjut. Temuan ini selaras dengan penelitian Febriani (2016) yang menunjukkan 47,5% soal berada pada tingkat sulit. Sedangkan Mania et al. (2020) yang menunjukkan 20%, 28%, dan 22% berada pada tingkat sangat sukar, sukar, dan mudah sehingga kualitas soal yang disajikan kurang baik.

Soal dalam kategori sulit berjumlah 10 butir (100%) dan mencakup nomor 1, 2, 3, 4, 5, 6, 7, 8, 9, dan 10. Soal dengan tingkat kesukaran sulit sebaiknya dikurangi kuantitasnya, diberikan beberapa soal tingkat sedang, namun tetap mendominasi untuk tujuan kompetisi matematika (Keldibekova, 2021; Nieto-Said & Sánchez-Lamonedas, 2022). Adapun hasil ini tidak selaras dengan persentase tingkat kesukaran yang memiliki proporsi proporsi 3-4-3 atau 30% sulit, 40% sedang, dan 30% mudah (Arikunto, 2021). Dengan demikian, meskipun hasil analisis menunjukkan soal-soal Geometric Competition memenuhi karakteristik soal olimpiade, variasi tingkat kesukaran dapat dipertimbangkan pada penyusunan soal tahun berikutnya.

Hasil Analisis Daya Beda

Daya beda digunakan untuk mengetahui kemampuan suatu soal dalam membedakan peserta berkemampuan tinggi dan rendah (Anas, 2006). Indeks daya beda yang tinggi menunjukkan soal yang efektif dalam mendeteksi perbedaan kemampuan peserta. Sebaliknya, soal dengan daya beda rendah tidak mampu menjalankan fungsi selektifnya (Arikunto, 2021; Suzana, 2018). Daya beda dihitung berdasarkan perbandingan jumlah peserta kelompok atas dan bawah yang menjawab benar. Untuk menganalisis daya beda, digunakan rumus berikut

$$DB = \frac{BaB}{Ba} - \frac{BbS}{Bb}$$

Keterangan:

DB sebagai indeks daya beda

BaB sebagai banyak peserta yang menjawab benar di kelompok atas

Ba sebagai banyak peserta di kelompok atas

BbS sebagai banyak peserta yang menjawab benar di kelompok bawah

Bb sebagai banyak peserta di kelompok bawah.

Tabel 3 menunjukkan hasil analisis daya beda kesukaran soal olimpiade matematika pada kegiatan Geometric Competition tahun 2025 di STAIN Mandailing Natal, yang telah dianalisis menggunakan aplikasi Microsoft Excel.

Tabel 3. Hasil analisis daya beda pada butir soal

Tingkat Daya Beda	Kategori	Nomor Soal	Banyaknya	Persentase
0,71-100	Sangat Kuat	4	1	10%
0,41-0,71	Kuat	6	1	10%
0,21-0,40	Sedang	7, 8, 9, 10	4	40%
0-0,20	Lemah	1, 2, 3, 5	4	40%
< 0	Sangat Lemah	-	0	0%

Tabel 3 menunjukkan bahwa 10% soal memiliki daya beda sangat kuat, 10% tergolong kuat, 40% berada pada kategori sedang, dan 40% sisanya tergolong lemah. Tidak terdapat soal dengan daya beda sangat lemah. Adanya butir soal berkategori sangat kuat dan kuat menunjukkan adanya beberapa soal yang mampu membedakan kemampuan peserta secara efektif. Namun, dominasi kategori sedang dan lemah mengindikasikan bahwa sebagian besar soal belum optimal dalam mengidentifikasi perbedaan kemampuan peserta.

Rendahnya daya beda kemungkinan dipengaruhi oleh dua faktor. Pertama, soal yang terlalu sulit mengakibatkan baik peserta berkemampuan tinggi maupun rendah sama-sama mengalami kesulitan. Hal ini mengakibatkan perbedaan performa tidak terlihat secara signifikan. Kedua, profil peserta olimpiade yang didominasi oleh siswa dengan kemampuan dasar yang belum merata dapat menyebabkan sebagian besar peserta tidak mampu mencapai jawaban yang benar.

Hasil analisis daya beda ini diperkuat oleh Magdalena et al. (2021) yang memaparkan soal-soal dengan daya beda lemah sering kali menunjukkan ketidaksesuaian antara kompetensi siswa dengan tingkat kesulitan soal. Febriani (2016) menyatakan bahwa soal dengan daya beda terlalu lemah tidak layak digunakan dalam evaluasi pembelajaran. Menurut Anas (2006), daya beda sangat penting untuk memastikan bahwa soal dapat mengukur kompetensi yang sesuai dengan tujuan pembelajaran. Soal dengan daya beda tinggi menunjukkan bahwa soal tersebut mampu membedakan kemampuan siswa secara efektif dan relevan (Anas, 2006; Arifin, 2017).

Hasil Uji Validitas

Validitas butir soal diartikan sebagai ukuran sejauh mana suatu soal dapat mengukur apa yang seharusnya diukur. Kevalidan butir soal terpenuhi apabila memiliki korelasi signifikan dengan skor total. Skor pada butir soal mendukung peningkatan atau penurunan skor total. Dengan kata lain suatu butir soal memperoleh nilai validitas yang baik jika skor butir

soal memiliki kesejajaran dengan skor total. Rumus korelasi *product moment* digunakan untuk menghitung validitas (Sudjiono, 2005).

$$r_{hitung} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y - (\Sigma Y)^2\}}}$$

Keterangan

r_{hitung} : koefisien korelasi

N : banyak peserta

X : skor pada butir soal

Y : total skor

Hasil uji validitas butir soal olimpiade matematika pada kegiatan Geometric Competition tahun 2025 di STAIN Mandailing Natal, yang dianalisis dengan aplikasi Ms. Excel, ditampilkan Tabel 4.

Tabel 4. Hasil uji validitas butir soal

No	Validitas	Nomor Butir Soal	Banyaknya	Persentase
1.	$\geq 0,3$ (Valid)	2, 6, 7, 8, 9, 10	6	60%
2.	$< 0,3$ (Tidak Valid)	1, 3, 4, 5	4	40%

Tabel 4 menunjukkan bahwa dari 10 butir soal, hanya 6 soal (60%) yang memenuhi kriteria validitas ($r_{hitung} \geq 0,3$), sementara 4 soal (40%) dinyatakan tidak valid ($r_{hitung} < 0,3$) (Azwar, 2016). Soal yang valid dapat digunakan sebagai bahan evaluasi karena telah memenuhi kriteria dan tidak menyimpang dari tujuan pengukuran. Sebaliknya, soal yang tidak valid perlu direvisi atau bahkan dihindari penggunaannya di kompetisi matematika mendatang karena tidak memenuhi kriteria pengukuran yang sesuai.

Temuan ini selaras dengan Tilaar et al. (2020) yang menunjukkan 40% butir soal pada ujian akhir semester berada dalam kategori tidak valid. Penelitian Khasanah et al. (2023) menunjukkan 47% butir soal ujian harian berada dalam kelompok tidak valid. Di sisi lain, Arifin (2017) memaparkan tiga aspek yang mempengaruhi validitas butir soal, diantaranya aspek instrumen tes, aspek penskoran, dan aspek jawaban siswa. Selain itu, validitas soal dapat ditingkatkan dengan memastikan siswa membaca soal secara teliti, menjawab dengan cermat, dan memeriksa kembali jawaban mereka sebelum menyerahkan lembar jawaban (Gronlund, 1998).

KESIMPULAN DAN SARAN

Berdasarkan temuan penelitian, peneliti menyimpulkan bahwa kualitas soal olimpiade matematika pada kegiatan Geometric Competition tahun 2025 di STAIN Mandailing Natal yang dikaji menunjukkan kualitas yang belum sepenuhnya baik dan perlu diperbaiki. Hasil validitas menunjukkan 60% soal berada dalam kategori valid. Reabilitas butir soal menunjukkan bahwa

butir soal belum memenuhi kriteria reliabel. Tingkat kesukaran butir soal memiliki kualitas yang baik untuk soal olimpiade dengan 100% butir soal tergolong sukar. Daya beda yang digunakan pada soal perlu diperbaiki, karena persentasenya didominasi pada kategori lemah (40%) dan sedang (40%). Sementara itu, analisis distraktor menunjukkan hasil yang cukup baik, di mana sebanyak 87,50% pengecoh berfungsi efektif dalam mengalihkan perhatian peserta berkemampuan rendah. Secara keseluruhan, butir soal yang digunakan dalam kegiatan Geometric Competition tahun 2025 di STAIN Mandailing Natal masih perlu direvisi dan dikaji ulang agar lebih efektif dalam mengevaluasi kemampuan matematis peserta.

DAFTAR RUJUKAN

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Amado, N., & Carreira, S. (2018). Emotions experienced by participants in a mathematical problem solving competition. *Educatio Siglo XXI*, 36(3), 73–100.
- Amiruddin, K., Mania, S., Ichiana, N. N., & Majid, A. F. (2020). Analisis Butir Soal Ujian Akhir Sekolah (UAS) Mata Pelajaran Matematika. *Alauddin Journal of Mathematics Education*, 2(2), 207–217. <https://doi.org/10.24252/ajme.v2i2.17364>
- Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan edisi 3*. Bumi aksara.
- Asamoah, D. B., & Ocansey, M. K. K. (2019). Item Discrimination and Distractor Analysis: A Technical Report on Thirty Multiple Choice Core Mathematics Achievement Test Items. *International Journal of Research and Scientific Innovation (IJRSI)*, 6(9), 24–33.
- de Losada, M. F., & Taylor, P. J. (2022). Perspectives on mathematics competitions and their relationship with mathematics education. *ZDM - Mathematics Education*, 54(5), 941–959. <https://doi.org/10.1007/s11858-022-01404-z>
- Gokbulut, Y., & Kus, S. (2019). Cartoon to solve teaching problem on mathematics. *International Journal of Evaluation and Research in Education*, 8(1), 145–150. <https://doi.org/10.11591/IJERE.V8I1.17609>
- Grugnetti, L., & Jaquet, F. (2005). A mathematical competition as a problem solving and a mathematical education experience. *Journal of Mathematical Behavior*, 24(3–4), 373–384. <https://doi.org/10.1016/j.jmathb.2005.09.012>
- Haladyna, T. M., & Rodriguez, M. C. (2021). Using Full-information Item Analysis to Improve Item Quality. *Educational Assessment*, 26(3), 198–211. <https://doi.org/10.1080/10627197.2021.1946390>
- Hamimi, L., Zamharirah, R., & Rusydy, R. (2020). Analisis Butir Soal Ujian Matematika Kelas VII Semester Ganjil Tahun Pelajaran 2017/2018. *MATHEMA: Jurnal Pendidikan Matematika*,

- 2(1), 57–66. <https://doi.org/10.33365/jm.v2i1.459>
- Jacinto, H., & Carreira, S. (2012). Problem solving in and beyond the classroom: perspectives and products from participants in a web-based mathematical competition 1. *12th International Congress on Mathematical Education*, 8–15.
- Keldibekova, A. O. (2021). The mathematical competence of Olympiad participants as an indicator of the quality of mathematical training level. In *Perspectives of Science and Education* (Vol. 51, Issue 3, pp. 169–187). <https://doi.org/10.32744/pse.2021.3.12>
- Kenderov, P. S. (2022). Mathematics competitions: an integral part of the educational process. *ZDM - Mathematics Education*, 54(5), 983–996. <https://doi.org/10.1007/s11858-022-01348-4>
- Khoshaim, H. B., & Rashid, S. (2016). Assessment of the assessment tool: Analysis of items in a Non-MCQ mathematics exam. *International Journal of Instruction*, 9(1), 119–132. <https://doi.org/10.12973/iji.2016.9110a>
- Kurniawan, D. D. (2019). Analisis Butir Soal Ujian Akhir Semester Matematika dengan Teori Respon Butir. *Briliant: Jurnal Riset Dan Konseptual*, 4(2), 215–224. <https://doi.org/10.28926/briliant.v4i2.316>
- Mahendra, I. W. E., Parmithi, N. N., Hermawan, E., Juwana, D. P., & Gunartha, I. W. (2020). Teachers' formative assessment: Accessing students' high order thinking skills (HOTS). *International Journal of Innovation, Creativity and Change*, 12(12), 180–202.
- Marushina, A. (2021). Mathematics competitions: what has changed in recent decades. *ZDM - Mathematics Education*, 53(7), 1591–1603. <https://doi.org/10.1007/s11858-021-01275-w>
- Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitriawanati, M. (2021). HOTS Multiple Choice and Essay Questions: A Validated Instrument to Measure Higher-Order Thinking Skills of Prospective Teachers. *Journal of Turkish Science Education*, 18(4), 674–690.
- Nieto-Said, J. H., & Sánchez-Lamonedá, R. (2022). A curriculum for mathematical competitions. *ZDM - Mathematics Education*, 54(5), 1043–1057. <https://doi.org/10.1007/s11858-022-01389-9>
- Patandean, A. H. (2020). Analisis Butir Soal Matematika Berdasarkan Aspek Kognitif TIMSS. *Jurnal Keguruan Dan Ilmu Pendidikan*, 7(2), 137–146.
- Saat, N. A. (2020). Summative Test Items Analysis Using Classical Test Theory (CTT)/ Analisis Item Kertas Peperiksaan Sumatif Menggunakan Teori Ujian Klasik (TUK). *Sains Humanika*, 12(2–2), 75–83. <https://doi.org/10.11113/sh.v12n2-2.1788>
- Saul, M., & Vaderlind, P. (2022). Outreach by the International Mathematical Olympiad to the mathematics education community. In *ZDM - Mathematics Education* (Vol. 54, Issue 5, pp. 997–1007). <https://doi.org/10.1007/s11858-022-01381-3>

- Siregar, H. N. I., & Panjaitan, A. (2022). A Systematic Literature Review : the Role of Classical Test Theory and Item Response Theory in Item Analysis to Determine the Quality of Mathematics Tests. *Jurnal Fibonacci: Jurnal Pendidikan Matematika*, 2(2), 29–48. <https://doi.org/10.24114/jfi.v2i2.31056>
- Susetyawati, M. E., & Wicaksono, B. (2023). Analisis Butir Soal Tes Matematika Materi Trigonometri. *Jurnal Derivat: Jurnal Matematika Dan Pendidikan Matematika*, 10(3), 129–138. <https://doi.org/10.31316/jderivat.v10i3.5218>
- Tilaar, A. L. F. F., & Hasriyanti. (2019). Analisis Butir Soal Semester Ganjil Mata Pelajaran Matematika pada Sekolah Menengah Pertama. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 8(1), 57–68. <https://doi.org/10.15408/jp3i.v8i1.13068>
- Toh, T. L. (2015). Mathematics Competition Questions: Their Pedagogical Values and an Alternative Approach of Classification. In *Selected Regular Lectures from the 12th International Congress on Mathematical Education* (pp. 807–819). https://doi.org/10.1007/978-3-319-17187-6_45
- Tohir, M. (2019). Keterampilan Berpikir Kreatif Siswa dalam Menyelesaikan Soal Olimpiade Matematika berdasarkan Level Metakognisi. *Alifmatika: Jurnal Pendidikan Dan Pembelajaran Matematika*, 1(1), 1–14. <https://doi.org/10.35316/alifmatika.2019.v1i1.1-14>