

Analisis Kualitas Butir Soal Matematika Realistik Pada Mahasiswa Tadris Matematika STAIN Mandailing Natal

Aisyah Husaini¹, Nanda Auliadi², Nisma Hayati^{3*}, Sofwatun Nabilah Siregar⁴

^{1,2,3,4} Program Studi Tadris Matematika, STAIN Mandailing Natal, Indonesia
Email: nismahayati14@gmail.com

Abstract

This study aims to analyze the quality of realistic mathematics test items on the topic of measures of central tendency administered to 15 third-semester students of the Tadris Mathematics Study Program at STAIN Mandailing Natal in the 2024/2025 academic year. The investigation focuses on four aspects of test quality, namely validity, reliability, item difficulty level, and discriminating power of multiple-choice questions used as an evaluation instrument. A descriptive quantitative approach is employed, with data consisting of students' test scores analyzed using product moment validity testing, reliability analysis, as well as calculations of difficulty indices and discriminating power for each item. The results indicate that only 2 out of 10 items are valid, the reliability coefficient is negative ($R_{11} = -0.653$), nine items fall into the easy category and one into the difficult category, and all items have poor discriminating power. These findings show that the realistic mathematics test instrument does not yet meet the criteria of a good assessment tool and requires comprehensive revision in terms of item construction, balance of difficulty levels, alignment of indicators with learning objectives, and improvement of discriminating power so that it can provide more accurate, reliable, and representative information on students' mastery of the material.

Keywords: realistic mathematics, item analysis, test quality

Abstrak

Penelitian ini bertujuan menganalisis kualitas butir soal matematika realistik pada materi statistika ukuran pemusatan data yang diujikan kepada 15 mahasiswa semester III Program Studi Tadris Matematika STAIN Mandailing Natal tahun akademik 2024/2025 dengan meninjau aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Pendekatan yang digunakan adalah kuantitatif deskriptif, dengan data berupa skor hasil tes yang dianalisis menggunakan uji validitas korelasi product moment, uji reliabilitas, perhitungan indeks kesukaran, serta indeks daya pembeda tiap butir soal. Hasil analisis menunjukkan bahwa dari 10 butir soal hanya 2 butir yang valid, koefisien reliabilitas bernilai negatif ($R_{11} = -0,653$), 9 butir soal berkategori mudah dan 1 sukar, serta seluruh butir memiliki daya pembeda yang jelek, sehingga instrumen belum memenuhi kriteria tes yang baik. Kondisi ini mengindikasikan perlunya revisi menyeluruh terhadap konstruksi soal, keseimbangan tingkat kesukaran, kesesuaian indikator dengan tujuan pembelajaran, dan kualitas daya pembeda agar instrumen tes matematika realistik mampu memberikan informasi yang lebih akurat, reliabel, dan representatif mengenai penguasaan materi mahasiswa.

Kata kunci: matematika realistik, analisis butir soal, kualitas tes.

How to Cite: Husaini et. al. (2026). Analisis Kualitas Butir Soal Matematika Realistik Pada Mahasiswa Tadris Matematika STAIN Mandailing Natal. *Journal of Mathematics in Teaching and Learning*, 4 (2), 440-451.

PENDAHULUAN

Pendidikan Nasional Indonesia bertujuan untuk membangun sikap, karakter, dan transformasi nilai-nilai filosofis negara untuk meningkatkan rasa nasionalisme serta mampu bersaing di kancah Internasional (Permendikbud, 2020). Tujuan Pendidikan Nasional tercantum dalam Undang-Undang Nomor 2 Tahun 2003 tentang Sistem Pendidikan Nasional pasal 3 yaitu: pendidikan Nasional berfungsi mengembangkan kemampuan dan membentuk watak dan peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa, bertujuan untuk berkembangnya potensi peserta didik agar

Received May 13, 2026; Revised May 31, 2026; Accepted June 30, 2026

menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri, serta menjadi warga negara yang demokratis dan bertanggung jawab (Sujana & Cong, 2019). Menurut Direktorat Jenderal Pendidikan Dasar dan Menengah, kualitas pendidikan yang baik dapat dilihat secara menyeluruh, tidak hanya pada hasil belajar, tetapi juga proses, lingkungan, dan dampak sosialnya (Permendikbud, 2018). Dengan kata lain, pendidikan nasional bertujuan membentuk karakter dan kemampuan peserta didik, sementara kualitas pendidikan terlihat dari proses, hasil, dan dampaknya.

Penilaian hasil belajar menilai konsep, kemampuan berpikir kritis, kreativitas, dan keterampilan siswa dalam menerapkan pengetahuan pada pendidikan (Nitko & Brookhart, 2011). Penilaian berperan merancang, mengelola, dan menginterpretasi hasil penilaian untuk mendukung perbaikan proses pembelajaran (Dreyer, 2008). Selain itu, penilaian hasil belajar juga menjadi alat evaluasi efektivitas pembelajaran, pengambilan keputusan, motivasi siswa, dan akuntabilitas pendidikan. Penilaian formatif meningkatkan pengetahuan dan mendorong tanggung-jawab siswa terhadap pembelajaran (Mihelic & Zore, 2022). Maka penilaian hasil belajar mengukur kompetensi, mendukung perbaikan pembelajaran, mendorong motivasi dan tanggung jawab siswa, serta menjadi alat evaluasi dan akuntabilitas pendidikan (Mustami, 2017; Sudjana, 1995).

Salah satu komponen utama dalam pelaksanaan asesmen formatif adalah pemberian tes yang sesuai dengan kompetensi yang telah diajarkan kepada peserta didik. Tes berfungsi untuk memperoleh informasi yang akurat mengenai tingkat pemahaman, pengetahuan, serta keterampilan siswa terhadap suatu materi pembelajaran. Menurut Saputri et al. (2023), tes yang baik harus dirancang dengan memperhatikan validitas dan reliabilitas agar hasilnya dapat dipertanggungjawabkan secara ilmiah. Tes yang dirancang secara sistematis melalui tahapan pengembangan kisi-kisi, uji coba, dan analisis butir soal akan menghasilkan instrumen yang valid serta sesuai dengan tujuan pembelajaran (Alghar et al., 2025; Musfirah et al., 2025). Oleh karena itu, penyusunan instrumen tes yang berkualitas tidak dapat dilakukan secara mendadak, melainkan harus melewati proses yang terukur dan terencana dengan baik sehingga hasil asesmen mencerminkan kemampuan siswa secara objektif.

Salah satu pendekatan penting dalam memastikan kualitas instrumen tes adalah melalui analisis butir soal. Analisis ini dilakukan untuk mengetahui sejauh mana setiap butir mampu menggambarkan kemampuan peserta didik dengan tepat dan membedakan antara siswa berkemampuan tinggi dan rendah (Asriadi & Hadi, 2021). Butir soal yang baik harus memiliki tingkat kesukaran yang proporsional, daya pembeda yang memadai, serta memenuhi kriteria validitas dan reliabilitas (Hanan et al., 2023). Hal ini sejalan dengan pandangan Akib & Ghafar (2015) bahwa uji validitas dan reliabilitas menjadi dasar utama dalam menilai keandalan sebuah instrumen evaluasi. Dengan demikian, pelaksanaan analisis butir soal menjadi langkah penting untuk memastikan bahwa setiap instrumen tes dalam asesmen formatif benar-benar berkualitas dan mampu memberikan gambaran yang akurat mengenai hasil belajar siswa (Rahmi et al., 2024).

Dapat disimpulkan penyusunan instrumen tes dalam asesmen formatif memiliki peran krusial dalam memastikan ketercapaian kompetensi belajar siswa. Instrumen yang dirancang secara sistematis dan didukung oleh proses uji validitas, reliabilitas, serta analisis butir soal akan menghasilkan alat ukur yang sah dan konsisten (Tilaar & Hasriyanti, 2019). Melalui tahapan tersebut, kualitas tes dapat dijamin karena setiap butir soal telah dianalisis tingkat kesukaran, daya pembeda, dan keakuratannya dalam menilai kemampuan peserta didik (Arikunto, 2021). Oleh sebab itu, asesmen yang berbasis pada instrumen tes berkualitas mampu memberikan gambaran autentik tentang penguasaan konsep dan keterampilan siswa secara objektif, serta menjadi dasar penting dalam peningkatan mutu pembelajaran dan pengambilan keputusan pendidikan yang lebih tepat.

Analisis butir soal menjadi langkah krusial dalam evaluasi instrumen penilaian, karena dengan analisis ini dapat dipastikan bahwa soal benar-benar mengukur kompetensi yang dimaksud secara objektif dan akurat (Arikunto, 2021). Jika analisis butir soal diabaikan, maka soal dapat mengandung kelemahan mendasar seperti cakupan materi yang tidak sesuai, distraktor yang kurang berfungsi, dan tingkat kesukaran yang tidak seimbang (Hayati et al., 2024). Kondisi ini dapat mengakibatkan hasil penilaian tidak mencerminkan kompetensi siswa secara tepat. Oleh sebab itu, melalui analisis butir soal peneliti dapat mengidentifikasi butir-butir yang tidak layak digunakan dan mengambil langkah perbaikan agar asesmen menjadi lebih adil dan akurat (Suparman et al., 2025).

Dalam ranah pendidikan matematika, analisis butir soal memiliki peranan strategis untuk meninjau sejauh mana soal mampu mengukur kemampuan berpikir matematis, penerapan konsep, dan pemecahan masalah siswa (Suparman et al., 2025). Karena pembelajaran matematika tidak hanya menuntut kemampuan menghitung, tetapi juga pemahaman konsep dan penerapan berpikir tingkat tinggi (Atikah et al., 2022), maka butir-soal matematis yang baik harus dirancang untuk mengukur berbagai aspek tersebut. Dengan demikian, analisis kualitas butir soal memungkinkan pendidik untuk melihat apakah soal sesuai dengan tujuan pembelajaran dan apakah soal dapat membedakan antar tingkat kemampuan mahasiswa.

Berdasarkan uraian tersebut, penelitian ini bermaksud untuk melakukan analisis butir soal pada instrumen tes matematika yang diimplementasikan kepada mahasiswa program studi Tadris Matematika. Instrumen tes yang digunakan dirancang dalam bentuk soal realistik yang menuntut kemampuan mahasiswa dalam memahami, menafsirkan, serta memecahkan permasalahan matematika dalam konteks kehidupan nyata. Analisis ini dilakukan untuk mengevaluasi kualitas butir soal berdasarkan kriteria validitas, reliabilitas, daya pembeda, dan tingkat kesukaran sebagaimana direkomendasikan dalam penelitian terdahulu. Dengan demikian, penelitian ini diberi judul analisis kualitas soal realistik pada mahasiswa Tadris Matematika.

METODE

Jenis penelitian ini adalah penelitian kuantitatif dengan pendekatan deskriptif. Pendekatan kuantitatif deskriptif bertujuan untuk memberikan gambaran objektif mengenai karakteristik suatu fenomena tanpa

menguji hubungan antar variabel (Creswell & Creswell, 2017). Penelitian ini difokuskan pada analisis kualitas butir soal realistik yang diberikan kepada mahasiswa. Data yang digunakan berupa hasil tes yang mencerminkan kemampuan mahasiswa dalam menyelesaikan soal-soal matematika berbasis konteks realistik. Data yang dikumpulkan meliputi lembar soal, lembar jawaban mahasiswa, serta kunci jawaban. Partisipan dalam penelitian ini yaitu 15 mahasiswa semester 3 Program Studi Tadris Matematika. Analisis data dilakukan dengan bantuan Microsoft Excel untuk mempermudah proses perhitungan statistik. Teknik analisis yang digunakan meliputi uji validitas butir soal, uji reliabilitas, analisis tingkat kesukaran, daya pembeda, sebagaimana dijelaskan oleh Arikunto (2021). Hasil analisis disajikan dalam bentuk tabel, kemudian diinterpretasikan secara deskriptif untuk menggambarkan kualitas setiap butir soal secara menyeluruh.

HASIL DAN PEMBAHASAN

Hasil Uji Validitas

Validitas butir soal merupakan tingkat ketepatan suatu item dalam mengukur kemampuan atau aspek yang memang menjadi tujuan pengukuran. Sebuah butir dikatakan valid apabila skor yang diperoleh peserta didik pada butir tersebut memiliki hubungan yang signifikan dengan skor total, sehingga menunjukkan adanya kesesuaian antara hasil pada satu item dan keseluruhan tes (Arikunto, 2021). Selain itu, validitas juga dipahami sebagai sejauh mana suatu instrumen mampu menggambarkan kemampuan yang ingin diukur secara tepat, sehingga hasil pengukuran dapat dipercaya dan digunakan dengan benar (Sugiyono, 2018). Oleh karena itu, semakin kuat hubungan antara skor butir dengan skor total, semakin tinggi pula validitas item tersebut. Rumus korelasi *product moment* digunakan untuk menghitung validitas:

$$r_{hitung} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$$

Keterangan:

r_{hitung} : Koefisien korelasi

n : Banyaknya peserta tes

X : Skor tiap butir soal

Y : Skor total setiap siswa

Untuk mengetahui valid atau tidaknya butir soal, maka harus mengetahui hasil perhitungan r_{xy} dibandingkan r_{tabel} *product moment* pada $\alpha = 0.05$

Kriteria pengujian validitasnya:

Jika $r_{hitung} \geq r_{tabel}$ maka soal tersebut valid (signifikan)

Jika $r_{hitung} < r_{tabel}$ maka soal tersebut tidak valid (tidak signifikan)

Hasil uji validitas pada instrumen test matematika materi statistika ukuran pemusatan data pada program studi tadrir matematika STAIN Mandailing Natal semester 3 (Angkatan 2024/2025), yang dianalisis secara manual dengan bantuan microsoft excel, ditampilkan pada Gambar 1.

No	Nama Mahasiswa	Butir Soal (X)										Skor Total (Y)
		1	2	3	4	5	6	7	8	9	10	
1	Adelina Amanda	1	1	1	1	1	1	1	1	0	1	9
2	Aidir Anwar	1	1	1	1	0	0	1	1	0	1	7
3	Ainun Atiqoh Pulungan	1	1	1	1	1	1	1	1	0	1	9
4	Andini Elysia Batubara	1	1	1	0	1	1	1	1	0	1	8
5	Asma Jalilah	1	1	1	1	1	1	1	1	0	0	8
6	Atikah Amelia	1	1	1	1	1	1	1	1	0	1	9
7	Fikri Anwar Nasution	1	1	1	1	1	1	1	1	0	1	9
8	Hiswani Harahap	1	1	1	1	1	1	1	1	0	1	9
9	Halimatus Sakdiah	1	1	1	1	1	1	1	1	0	1	9
10	Latifah	1	1	1	1	1	1	1	1	0	1	9
11	Mardiah Matondang	1	1	1	1	1	1	1	1	0	1	9
12	Pahrul Rozi Siregar	1	1	1	1	0	1	0	1	1	1	8
13	Selly Sofia Harahap	1	1	1	1	1	1	1	1	0	1	9
14	Sofia Sitinjak	1	1	1	1	1	1	1	1	0	1	9
15	Ummi Barokah	1	1	1	1	1	1	1	1	0	1	9
	Σx	15	15	15	14	13	14	14	15	1	14	
	Σy											130
	r hitung	0	0	0	0,299	0,767	0,747	0,299	0	0,299	0,299	
	r tabel	0,514	0,514	0,514	0,514	0,514	0,514	0,514	0,514	0,514	0,514	
		TIDAK VALID	TIDAK VALID	TIDAK VALID	TIDAK VALID	VALID	VALID	TIDAK VALID	TIDAK VALID	TIDAK VALID	TIDAK VALID	

Gambar 1. Hasil Uji Validitas terhadap instrumen tes

Gambar 1. menunjukkan hasil uji validitas terhadap instrumen tes. Dari 10 soal yang diuji, hanya dua butir soal yang memenuhi kriteria valid, yaitu butir soal nomor 5 dan nomor 6. Kedua soal tersebut dinyatakan valid karena nilai r hitung yang diperoleh lebih besar dibandingkan nilai r tabel (0.514). Sebaliknya, delapan butir soal lainnya tidak memenuhi kriteria validitas karena nilai r hitungnya lebih rendah daripada r tabel, sehingga tidak dapat dijadikan dasar pengukuran yang akurat. Dengan demikian, hanya 20% butir soal yang layak digunakan sebagai instrumen evaluasi, sedangkan sisanya perlu direvisi atau dieliminasi agar tidak memengaruhi ketepatan hasil penilaian.

Hasil tersebut menunjukkan bahwa sebagian besar butir soal belum memenuhi kualitas pengukuran yang baik. Hal ini sejalan dengan pandangan Arikunto (2021) yang menyatakan bahwa suatu butir soal dinyatakan tidak valid apabila tidak menunjukkan keterkaitan yang signifikan dengan skor total tes yang menjadi acuan utama. Hayati et al. (2024) turut menegaskan bahwa rendahnya validitas butir sering disebabkan oleh ketidaktepatan dalam perumusan indikator, ketidaksesuaian materi dengan tujuan evaluasi, serta kurang optimalnya konstruksi soal yang disusun. Selanjutnya, Mardapi (2012) menjelaskan bahwa kualitas butir sangat dipengaruhi oleh ketepatan perumusan tujuan pembelajaran, keselarasan indikator dengan kompetensi yang diukur, serta kejelasan redaksi soal agar tidak menimbulkan interpretasi ganda. Dengan demikian, hasil pada tabel tersebut menunjukkan perlunya

perbaikan yang komprehensif terhadap sebagian besar butir soal agar instrumen dapat menggambarkan kemampuan mahasiswa secara lebih akurat dan reliabel.

Hasil Uji Reliabilitas

Uji reliabilitas dilakukan untuk menilai sejauh mana suatu instrumen memiliki tingkat konsistensi yang stabil ketika digunakan dalam kondisi pengukuran yang sebanding. Azwar (2016) menyatakan bahwa reliabilitas menunjukkan kemampuan instrumen dalam menghasilkan skor yang ajeg tanpa dipengaruhi oleh kesalahan pengukuran. Sejalan dengan itu, Arikunto (2021) menegaskan bahwa instrumen yang reliabel harus memberikan hasil yang tetap ketika prosedur pengukuran diulang pada situasi yang sama. Lebih lanjut, kajian psikometrik oleh Tavakol & Dennick (2011) menjelaskan bahwa reliabilitas erat kaitannya dengan konsistensi internal antarbutir, sehingga semakin tinggi konsistensi tersebut maka semakin dapat dipercaya hasil yang diperoleh dari instrumen tersebut.

Berdasarkan pendapat tersebut dapat disimpulkan bahwa proses pengujian ini berfungsi untuk menilai kemampuan instrumen dalam menghasilkan informasi yang tetap stabil ketika digunakan pada situasi pengukuran yang serupa. Instrumen dinyatakan memiliki reliabilitas yang baik apabila nilai yang diperoleh tidak menunjukkan perubahan yang disebabkan oleh kesalahan pengukuran, melainkan benar-benar mencerminkan kondisi responden (Azwar, 2016). Keandalan semacam ini penting dipenuhi karena data yang tidak konsisten berpotensi menurunkan kualitas analisis serta melemahkan dasar penarikan kesimpulan ilmiah (Kimberlin & Winterstein, 2008). Oleh sebab itu, reliabilitas menjadi unsur kunci yang menentukan apakah sebuah instrumen layak dijadikan alat pengumpul data dalam penelitian akademik. Rumus uji reabilitas yang digunakan sebagai berikut:

$$S^2_x = \frac{n \Sigma Y^2 - (\Sigma Y)^2}{n(n-1)} \quad S^2 = \frac{n \Sigma X^2 - (\Sigma X)^2}{n(n-1)} \quad S^2_t = \Sigma S^2 \quad R_{11} = \left[\frac{m}{m-1} \right] \cdot \left[1 - \frac{S^2_x}{S^2_t} \right]$$

Keterangan:

n : Banyaknya peserta tes

m : Jumlah skor yang valid

S^2 : Varians

x : Skor tiap butir soal yang valid

t : Skor total setiap siswa

Kriteria pengujian validitasnya:

Jika $r_{11} \geq 0,6$ maka soal tersebut reabel

Jika $r_{11} < 0,6$ maka soal tersebut tidak reabel

Hasil uji reabilitas pada instrumen test matematika materi statistika ukuran pemusatan data pada program studi tadrir matematika STAIN Mandailing Natal semester 3 (Angkatan 2024/2025), yang dianalisis secara manual dengan bantuan microsoft excel, ditampilkan pada Gambar 2.

No	Nama Mahasiswa	Total Skor Valid (Y)	Butir Soal Valid (X)		Skor Total (Y^2)
			X5	X6	
1	Adelina Amanda	2	1	1	4
2	Aidir Anwar	0	0	0	0
3	Ainun Atiqoh Pulungan	2	1	1	4
4	Andini Elysia Batubara	2	1	1	4
5	Asma Jalilah	2	1	1	4
6	Atikah Amelia	2	1	1	4
7	Fikri Anwar Nasution	2	1	1	4
8	Hiswani Harahap	2	1	1	4
9	Halimatus Sakdiah	2	1	1	4
10	Latifah	2	1	1	4
11	Mardiah Matondang	2	1	1	4
12	Pahrul Rozi Siregar	1	0	1	1
13	Selly Sofia Harahap	2	1	1	4
14	Sofia Sitinjak	2	1	1	4
15	Ummi Barokah	2	1	1	4
Σ		27	13	14	53
S^2		0,314	0,123	0,067	
S^2_t					0,19
R_{11}					-0,653

Gambar 2. Hasil Uji Reabilitas terhadap instrument tes

Tabel uji reliabilitas tersebut memuat data skor dari dua butir instrumen, yaitu X5 dan X6, yang diberikan kepada 15 responden. Skor dari masing-masing butir kemudian dijumlahkan sehingga menghasilkan total skor valid (Y) dan skor total keseluruhan (Y^2). Sebagian besar responden menunjukkan pola jawaban yang relatif serupa, ditandai oleh banyaknya skor 1 pada kedua butir dan skor total yang umumnya bernilai 4. Akumulasi skor pada bagian akhir tabel memperlihatkan bahwa X5 memiliki jumlah skor 13, X6 berjumlah 14, dan total seluruh skor responden mencapai 53. Nilai-nilai ini kemudian digunakan dalam perhitungan varians butir dan varians total sebagai dasar penilaian reliabilitas. Namun, hasil kalkulasi menunjukkan bahwa koefisien reliabilitas (R_{11}) bernilai $-0,653$. Nilai negatif ini menandakan bahwa kedua butir instrumen tidak bergerak searah dengan skor total, sehingga konsistensi internalnya tidak terpenuhi. Dengan demikian, berdasarkan data pada tabel, instrumen belum dapat dinyatakan reliabel dan memerlukan perbaikan atau penyusunan ulang sebelum digunakan dalam pengukuran berikutnya.

Hasil Analisis Tahap Kesukaran Soal

Analisis taraf kesukaran bertujuan untuk mengetahui sejauh mana suatu butir soal mudah atau sulit dijawab oleh mahasiswa. Arikunto (2021) menegaskan bahwa pengelompokan butir soal ke dalam kategori mudah, sedang, dan sulit diperlukan untuk memastikan keseimbangan kualitas instrumen evaluasi. Sebuah butir yang baik pada umumnya berada pada tingkat kesukaran sedang, karena butir yang terlalu mudah tidak mampu membedakan kemampuan mahasiswa, sementara butir yang terlalu sulit cenderung tidak memberikan informasi yang memadai mengenai penguasaan materi (Sudjana, 1990). Selain itu, Sudijono & Anas (2013) menjelaskan bahwa taraf kesukaran yang proporsional akan menghasilkan instrumen yang lebih representatif terhadap variasi kemampuan mahasiswa. Dengan demikian, analisis taraf kesukaran menjadi langkah penting dalam memastikan bahwa setiap butir soal dapat berfungsi optimal dalam mengukur kemampuan secara objektif dan akurat. Rumus untuk mengukur taraf kesukaran soal yang digunakan yaitu:

$$P = \frac{Np}{N}$$

Keterangan:

P : Tingkat kesukaran

Np : Jumlah siswa yang menjawab benar

N : Jumlah siswa yang mengikuti tes

Hasil analisis taraf kesukaran pada instrumen test matematika materi statistika ukuran pemusatan data pada program studi tadaris matematika STAIN Mandailing Natal semester 3 (Angkatan 2024/2025), yang dianalisis secara manual dengan bantuan microsoft excel, ditampilkan pada Gambar 3.

No Soal	Banyak Siswa yang Menjawab Benar (NP)	Jumlah Siswa yang Mengikuti Tes (N)	Indeks Kesukaran (P) = $\frac{NP}{N}$	Keterangan
1	15	15	1	Mudah
2	15	15	1	Mudah
3	15	15	1	Mudah
4	14	15	0,93	Mudah
5	13	15	0,86	Mudah
6	14	15	0,93	Mudah
7	14	15	0,93	Mudah
8	15	15	1	Mudah
9	1	15	0,06	Sukar
10	14	15	0,93	Mudah

Gambar 3. Hasil Analisis Daya Pembeda terhadap instrument tes

Berdasarkan hasil analisis taraf kesukaran pada Gambar 3, diperoleh bahwa sebagian besar butir soal termasuk dalam kategori mudah. Terdapat empat butir soal yang memiliki indeks kesukaran 1,00, yaitu butir nomor 1, 2, 3, dan 8. Selanjutnya, terdapat empat butir soal dengan indeks kesukaran 0,93, yaitu butir nomor 4, 6, 7, dan 10. Kemudian, satu butir soal memiliki indeks kesukaran 0,86, yaitu butir nomor 5. Sementara itu, hanya satu butir soal yang berada pada kategori sukar, yaitu butir nomor 9 dengan indeks kesukaran 0,66. Dengan demikian, dapat disimpulkan bahwa dari 10 butir soal pilihan ganda, 9 soal termasuk kategori mudah dan 1 soal termasuk kategori sukar. Secara proporsional, komposisi tersebut menunjukkan bahwa 90% butir soal berada pada tingkat kesukaran mudah.

Temuan ini menunjukkan bahwa instrumen yang digunakan belum memiliki variasi kesukaran yang proporsional. Arikunto (2021) menyatakan bahwa komposisi soal yang baik hendaknya memuat butir dengan tingkat kesukaran mudah, sedang, dan sukar secara seimbang agar mampu menggambarkan kemampuan peserta tes secara lebih akurat. Sementara itu, Alghar et al. (2025) menekankan bahwa soal mudah memang diperlukan dalam jumlah tertentu untuk memberikan motivasi kepada peserta tes, namun dominasi soal mudah dapat mengurangi daya beda instrumen. Sejalan dengan itu, penelitian Febriani (2016) juga menemukan bahwa instrumen dengan dominasi soal mudah cenderung tidak layak karena tidak mencerminkan sebaran kesukaran yang ideal.

Hasil Analisis Daya Pembeda Soal

Daya pembeda adalah kemampuan suatu butir soal untuk membedakan antara peserta didik yang memiliki kemampuan tinggi dan peserta didik yang memiliki kemampuan rendah dalam menguasai materi (Rahmi et al., 2024). Pada daya pembeda semakin tinggi nilai indeksnya, maka akan semakin baik pula kemampuan soal untuk membedakan. Untuk menganalisis daya pembeda yaitu menggunakan rumus berikut:

$$D = \frac{BA}{JA} - \frac{BB}{JB}$$

Keterangan:

- D : Daya Pembeda
- BA : Jumlah kelas atas yang benar
- BB : Jumlah kelas atas bawah yang benar
- JA : Jumlah responden kelas atas
- JB : Jumlah responden kelas bawah

Hasil analisis daya pembeda pada instrumen test matematika materi statistika ukuran pemusatan data pada program studi tadris matematika STAIN Mandailing Natal semester 3 (Angkatan 2024/2025), yang dianalisis secara manual dengan bantuan microsoft excel, ditampilkan pada Gambar 4.

Nomor Soal	BA	BB	JA	JB	D	Keterangan
1	8	7	8	7	0	Jelek
2	8	7	8	7	0	Jelek
3	8	7	8	7	0	Jelek
4	7	7	8	7	- 0,125	Jelek
5	7	6	8	7	0,018	Jelek
6	7	7	8	7	-0,125	Jelek
7	8	6	8	7	0.142	Jelek
8	8	7	8	7	0	Jelek
9	0	1	8	7	-0,142	Jelek
10	7	7	8	7	-0,125	Jelek

Gambar 4. Hasil Analisis Daya Pembeda terhadap instrument tes

Hasil analisis daya pembeda pada Gambar 4 menunjukkan bahwa 0% soal tergolong baik, 0% soal tergolong cukup, 100% soal tergolong dalam kategori jelek, dan 0% soal tergolong sangat baik. Karena banyaknya soal yang berkategori jelek menunjukkan butir soal yang digunakan kurang efektif dalam mengukur perbedaan kemampuan siswa. Hal ini mungkin disebabkan karena soal yang terlalu mudah untuk kalangan mahasiswa, sehingga seluruh mahasiswa baik yang berkemampuan tinggi maupun yang berkemampuan rendah dapat menyelesaikan soal tersebut dengan baik.

Dalam analisis butir soal, daya pembeda digunakan untuk mengukur sejauh mana sebuah soal dapat membedakan antara siswa berkemampuan tinggi dan rendah (Solichin, 2017). Namun, ketika tingkat kesukaran terlalu rendah (soal sangat mudah), nilai daya pembeda bisa melemah karena siswa dari semua kelompok hampir selalu menjawab benar (Alghar et al., 2025). Arikunto (2021) dalam beberapa kajian evaluasi pendidikan menyatakan bahwa soal ideal bukanlah yang sangat mudah maupun sangat sukar karena soal yang terlalu mudah tidak merangsang peserta didik untuk mempertinggi usaha memecahkannya. Selain itu, penelitian Pengembangan Instrumen Kognitif menunjukkan bahwa soal sangat mudah dapat menghasilkan korelasi butir skor total yang rendah atau bahkan negatif, karena butir tersebut gagal mengukur kemampuan sejati siswa (Imamuddin et al., 2022; Mustami, 2017).

SIMPULAN

Berdasarkan hasil analisis validitas, reliabilitas, taraf kesukaran, dan daya pembeda, dapat disimpulkan bahwa instrumen tes matematika materi statistika ukuran pemusatan data pada Program Studi Tadris Matematika STAIN Mandailing Natal Semester 3 Tahun Akademik 2024/2025 belum memenuhi kriteria instrumen evaluasi yang baik. Dari sepuluh butir soal yang diuji, hanya dua butir yang dinyatakan valid, sementara sisanya tidak memenuhi batas minimal korelasi yang dipersyaratkan. Hasil uji reliabilitas juga menunjukkan bahwa koefisien reliabilitas bernilai negatif ($R_{11} = -0,653$), menandakan bahwa konsistensi internal antarbutir sangat rendah sehingga instrumen belum dapat

dikatakan reliabel. Selain itu, mayoritas butir soal tergolong mudah sehingga tidak mampu menggambarkan variasi kemampuan mahasiswa secara optimal. Analisis taraf kesukaran menunjukkan bahwa 90% butir soal berada pada kategori mudah dan hanya satu soal berada pada kategori sukar. Pada daya pembeda soal, seluruh butir soal berada pada kategori jelek dan tidak mampu membedakan kemampuan mahasiswa berkemampuan tinggi dan rendah.

REFERENSI

- Alghar, M. Z., & Jamaluddin. (2024). Ethnomodelling: Fractal Geometry On The Door Ornament Of The Sumenep Palace Using The Lindenmayer System. *Euclid*, 11(1), 1–16. <https://doi.org/10.33603/x5gk7n46>
- Azzahra, S., Mislaini, & Alghar, M. Z. (2023). Eksplorasi Deret Kuadrat dan Barisan Aritmatika pada Batik Madura Motif Bha'jit. *Journal of Matematics In Teaching and Learning*, 2(1), 86–96.
- Boden, M. A., & Edmonds, E. A. (2009). What is generative art? *Digital Creativity*, 20(1–2), 21–46. <https://doi.org/10.1080/14626260902867915>
- Boutayeb, A., Twizell, E. H., Achouayb, K., & Chetouani, A. (2004). A mathematical model for the burden of diabetes and its complications. *BioMedical Engineering OnLine*, 3(1), 20. <https://doi.org/10.1186/1475-925X-3-20>
- Boyko, E. J., Karuranga, S., Magliano, D. J., Saedi, P., & Sun, H. (2021). International diabetes federation (IDF) diabetes atlas 10th edition [internet]. In *International Diabetes Federation*. www.diabetesatlas.org.
- Christakis, N. A., & Fowler, J. H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine*, 357(4), 370–379. <https://doi.org/10.1056/NEJMs066082>
- Duun-Henriksen, Anne Katrine, Schmidt, Signe, Røge, Rikke Meldgaard, Møller, Jonas Bech, Nørgaard, Kirsten, Jørgensen, John Bagterp, & Madsen, Henrik. (2013). Model Identification Using Stochastic Differential Equation Grey-Box Models in Diabetes. *Journal of Diabetes Science and Technology*, 7(2), 431–440. <https://doi.org/10.1177/193229681300700220>
- Greenfield, G. R. (2006). Art by computer program == programmer creativity. *Digital Creativity*, 17(1), 25–35. <https://doi.org/10.1080/14626260600665694>
- Hariadi, Y., Lukman, M., & Destiarmand, A. H. (2013). Batik fractal: Marriage of art and science. *ITB J. Vis. Art & Des*, 4(1), 84–93.
- Hill, A. L., Rand, D. G., Nowak, M. A., & Christakis, N. A. (2010). Infectious Disease Modeling of Social Contagion in Networks. *PLOS Computational Biology*, 6(11), e1000968. <https://doi.org/10.1371/journal.pcbi.1000968>
- Hruschka, D. J., Brewis, A. A., Wutich, A., & Morin, B. (2011). Shared norms and their explanation for the social clustering of obesity. *American Journal of Public Health*, 101(S1), S295–S300.
- Iliinsky, N., & Steele, J. (2011). *Designing data visualizations: Representing informational*

Relationships. “O’Reilly Media, Inc.”

- Kaya’, K., Darmawati, & Ekawati, D. (2021). Model Matematika pada Penyakit Diabetes Melitus dengan Faktor Genetik dan Faktor Sosial. *Journal of Mathematics: Theory and Applications*, 3(1), 23–30.
- McCormack, J., Bown, O., Dorin, A., McCabe, J., Monro, G., & Whitelaw, M. (2014). Ten Questions Concerning Generative Computer Art. *Leonardo*, 47(2), 135–141. https://doi.org/10.1162/LEON_a_00533
- Romadiastri, Y. (2017). Batik Fraktal: Perkembangan Aplikasi Geometri Fraktal. *Delta: Jurnal Ilmiah Pendidikan Matematika*, 1(2), 158–164. <https://doi.org/10.31941/delta.v1i2.484>
- Rukmawati, M., Gita, R. S. D., Emyus, A. Z., & Agustin, I. H. (2025). Incorporating the fractal geometry on batik patterns based on natural leaf prints under RBL-STEM learning activities to enhance students’ skills. In I. H. Agustin, A. I. Kristiana, R. Nisviasari, E. Y. Kurniawati, & Dafik (Eds.), *AIP Conference Proceedings* (Vol. 3372, Number 1). American Institute of Physics. <https://doi.org/10.1063/5.0299317>
- Soewondo, P., Ferrario, A., & Tahapary, D. L. (2013). Challenges in diabetes management in Indonesia: a literature review. *Globalization and Health*, 9(1), 63. <https://doi.org/10.1186/1744-8603-9-63>
- Sudrajat, S., Warniasih, K., & Purwaningsih, D. (2026). Eksplorasi Etnomatematika pada Motif Batik Ganasan Khas Subang Sebagai Sumber Pembelajaran Matematika. *Jurnal Pendidikan Matematika*, 13(1), 35–50. <https://doi.org/10.18592/jpm.vi.20558>
- Ulfah, J. (2014). Model Matematika Untuk Penyakit Diabetesmellitus Tanpa Faktor Genetik Dengan Perawatan. *Unnes Journal of Mathematics*, 3(1), 25–32.
- Zheng, Y., Ley, S. H., & Hu, F. B. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nature Reviews Endocrinology*, 14(2), 88–98. <https://doi.org/10.1038/nrendo.2017.151>