

Analisis Kualitas Butir Soal Esai Matematika Realistik untuk Mengukur Kemampuan Pemecahan Masalah

Salsabila Az-zahra^{1*}, Kulsum Rahmayani², Sakinah³, Khafifah Nur Rangkuti⁴

^{1,2,3,4}Program Studi Tadris Matematika, Sekolah Tinggi Agama Islam Negeri Mandailing Natal, Indonesia
Email: azsalzahra@gmail.com

Abstract

This study aims to analyze the quality of realistic mathematics essay test items on the topic of Systems of Linear Equations in Two Variables (SPLDV) in terms of validity, reliability, difficulty level, and discriminating power. The test was designed to measure students' mathematical problem-solving ability. This research employed a quantitative descriptive approach involving first-semester students (academic year 2025–2026) of the Mathematics Education Program at the State Islamic College of Mandailing Natal (STAIN Mandailing Natal) as research subjects. The research instrument consisted of essay questions based on the realistic mathematics approach, which were analyzed using Microsoft Excel. The results of the analysis showed that seven out of eight items were valid with a correlation coefficient (r) ≥ 0.456 , and the reliability coefficient was 0.82, which falls into the high category. In terms of difficulty level, most of the items were classified as moderate, while the discriminating power analysis indicated that the majority of the items were in the good category. Based on these results, it can be concluded that the realistic mathematics essay test developed in this study has good quality and is feasible to be used as an assessment tool for evaluating students' mathematical problem-solving ability.

Keywords: Esai test item analysis, Realistic Mathematics Education (RME), validity, reliability, difficulty index, discrimination index

Abstrak

Penelitian ini bertujuan untuk menganalisis kualitas butir soal esai matematika realistik pada materi Sistem Persamaan Linear Dua Variabel (SPLDV) yang ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda butir soal, yang digunakan untuk mengetahui kemampuan pemecahan masalah mahasiswa. Pendekatan penelitian yang digunakan adalah deskriptif kuantitatif dengan melibatkan mahasiswa semester I (angkatan 2025-2026) program studi Tadris Matematika Sekolah Tinggi Agama Islam Negeri (STAIN) Mandailing Natal sebagai subjek penelitian. Instrumen penelitian berupa tes esai berbasis pendekatan matematika realistik yang dianalisis menggunakan perangkat lunak Microsoft Excel. Hasil analisis menunjukkan bahwa 7 dari 8 butir soal dinyatakan valid dengan nilai koefisien korelasi $r \geq 0,456$ dan reliabilitas sebesar 0,82 yang termasuk kategori tinggi. Dari segi tingkat kesukaran, mayoritas butir soal berada pada kategori sedang, sedangkan hasil analisis daya pembeda menunjukkan sebagian besar butir soal berada pada kategori baik. Berdasarkan hasil tersebut, dapat disimpulkan bahwa instrumen tes esai matematika realistik yang dikembangkan memiliki kualitas yang baik dan layak digunakan sebagai alat evaluasi kemampuan pemecahan masalah matematis mahasiswa.

Kata kunci: Analisis butir soal esai, matematika realistik, validitas, reliabilitas, tingkat kesukaran, daya pembeda

How to Cite: Az-zahra, N. H., Rahmayani, K., Sakinah, Rangkuti, K. N. (2025). Analisis Kualitas Butir Soal Esai Matematika Realistik untuk Mengukur Kemampuan Pemecahan Masalah. *Journal of Mathematics in Teaching and Learning*, 4 (1), 393-406.

PENDAHULUAN

Pembelajaran matematika memiliki peran penting dalam mengembangkan kemampuan berpikir logis, analitis, dan sistematis peserta didik. Melalui proses pembelajaran yang menekankan pada pemecahan masalah dan penalaran, peserta didik belajar menyusun langkah berpikir yang runtut dan rasional (Lewis & Manouchehri, 2020; Özerem, 2012). Namun, hasil dari survei internasional

Received November 24, 2025; Revised December 02, 2025; Accepted December 24, 2025

© 2025 The Author(s). This work is licensed under a Creative Common Attribution 4.0 International (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

menunjukkan bahwa kemampuan pemecahan masalah matematis peserta didik di Indonesia masih relatif rendah (OECD, 2023). Berdasarkan laporan *Programme for International Student Assessment (PISA)* tahun 2022, skor rata-rata matematika peserta didik Indonesia hanya mencapai 373, jauh di bawah rata-rata OECD yang sebesar 472 (OECD, 2023). Data tersebut menunjukkan kemampuan pemecahan masalah matematis perlu mendapat perhatian dalam proses pembelajaran.

Salah satu penyebab rendahnya kemampuan pemecahan masalah siswa yaitu belum optimalnya penggunaan instrumen evaluasi dalam menilai hasil belajar. Beberapa penelitian menunjukkan bahwa alat penilaian yang berfokus hanya pada hasil akhir, tanpa melihat kemampuan berpikir dan pemecahan masalah siswa (Amam, 2017; Fauji et al., 2022). Evaluasi berfungsi untuk menilai sejauh mana peserta didik memahami konsep dan mampu menggunakan kriteria validitas, reliabilitas, tingkat kesukaran, dan daya pembeda yang memadai, maka hasil penilaian tidak akan menggambarkan kemampuan peserta didik secara objektif (Himmah, 2019; Pardimin & Widodo, 2017). Dengan demikian, peningkatan kualitas instrumen penilaian menjadi hal penting agar hasil evaluasi dapat dijadikan dasar pengambilan keputusan dalam proses pembelajaran.

Berdasarkan kondisi pembelajaran yang terjadi di Indonesia, banyak guru yang masih menggunakan instrumen penilaian tanpa melalui proses uji kualitas butir soal secara empiris. Penelitian yang dilakukan oleh Alifah & Sugilar (2024) menunjukkan bahwa sekitar 40% butir soal matematika yang digunakan di tingkat SMP tergolong terlalu mudah, sedangkan 25% lainnya terlalu sulit, dan hanya 35% yang memiliki daya pembeda baik. Selanjutnya, penelitian Setiyawan & Wijayanti (2020) menunjukkan bahwa selama masa pembelajaran daring, kualitas instrumen penilaian yang digunakan guru menurun secara signifikan karena tidak melalui proses uji reliabilitas dan validitas yang memadai. Selain itu, penelitian yang dilakukan oleh Irawan & Sa'adah (2023) menemukan bahwa sebagian besar guru belum melakukan analisis validitas dan reliabilitas pada soal yang digunakan dalam tes kemampuan pemecahan masalah. Kondisi ini menyebabkan ketidaksesuaian antara tujuan pembelajaran dan hasil evaluasi. Akibatnya, kemampuan peserta didik yang sebenarnya tidak tergambarkan secara menyeluruh melalui hasil tes.

Berdasarkan ketiga penelitian tersebut, dapat disimpulkan bahwa kualitas butir soal matematika di Indonesia masih perlu mendapat perhatian serius. Sebagian besar instrumen yang digunakan belum memenuhi kriteria validitas, reliabilitas, tingkat kesukaran, dan daya pembeda yang ideal. Oleh karena itu, diperlukan penelitian yang berfokus pada analisis kualitas butir soal secara menyeluruh agar evaluasi pembelajaran matematika dapat dilakukan dengan lebih tepat dan bermakna.

Di sisi lain, salah satu pendekatan yang diyakini dapat membantu dalam pengembangan instrumen tes yang kontekstual dan bermakna yaitu *Realistic Mathematics Education* (RME), yang diperkenalkan oleh Hans Freudenthal (Gravemeijer, 1994). Pendekatan ini menekankan pentingnya konteks nyata dalam memahami konsep-konsep matematis (Prahmana, 2022). Melalui pendekatan RME, peserta didik dilatih untuk berpikir logis dengan mengaitkan konsep matematika pada situasi

yang mereka alami dalam kehidupan sehari-hari (Puspitasari & Airlanda, 2021). Pendekatan ini membantu siswa memahami konsep matematika secara bermakna karena mereka belajar dari konteks yang dekat dengan pengalaman yang dirasakan oleh peserta didik (Chavarría-Arroyo & Albanese, 2022; Muslimin et al., 2020; Zulkardi et al., 2020). Dengan demikian, RME tidak hanya berfungsi dalam pembelajaran, tetapi juga relevan dalam membangun instrumen penilaian berbasis konteks nyata.

RME menjadi pendekatan yang dikembangkan oleh Gravemeijer (1994). Menurutnya, matematika sebaiknya tidak diajarkan sebagai kumpulan rumus, melainkan sebagai aktivitas manusia yang mengaitkan matematika dengan pengalaman ataupun dunia nyata. Penelitian yang dilakukan oleh Ulya & Agustyarini (2020) menunjukkan bahwa penerapan RME dapat meningkatkan kemampuan penalaran dan pemecahan masalah matematis peserta didik sekolah dasar. Sementara Napitupulu dan Simamora (2023) melaporkan peningkatan kemampuan pemecahan masalah hingga 27% pada peserta didik SMA melalui penerapan RME. Dengan demikian, pendekatan RME menjadi pendekatan yang penting guna meningkatkan kemampuan berpikir, bernalar, dan pemecahan masalah peserta didik.

Dari sisi kognitif, kemampuan pemecahan masalah menjadi kemampuan yang perlu dimiliki siswa dalam mempelajari matematika (Midgett & Eddins, 2001). Polya (1957) menjelaskan empat tahap pemecahan masalah yaitu; memahami masalah, merencanakan strategi, melaksanakan rencana, dan memeriksa kembali hasil. Menurut Malau (2024), penerapan langkah-langkah Polya dalam penyusunan soal esai mampu menstimulasi peserta didik untuk berpikir kritis dan reflektif. Hal ini memperkuat dasar bahwa butir soal berbasis *RME* yang dikembangkan berdasarkan tahapan Polya dapat membantu peserta didik dalam mengembangkan kemampuan berpikir matematis siswa yang lebih mendalam.

Berbagai penelitian di Indonesia telah membahas analisis kualitas butir soal matematika. Son (2019) meneliti instrumen pemecahan masalah matematis dan menemukan bahwa sebagian besar butir soal telah memenuhi kriteria validitas dan reliabilitas yang baik, namun analisisnya belum mencakup konten realistik. Sedangkan, penelitian Pardimin & Widodo (2017) meninjau pentingnya daya pembeda dalam mengidentifikasi perbedaan kemampuan peserta didik, namun penelitian mereka belum mengkaji soal esai yang bersifat kontekstual. Sementara itu, penelitian yang dilakukan oleh Sriamanda et al. (2025) juga menemukan bahwa pengembangan instrumen tes berbasis kontekstual dapat menghasilkan butir soal dengan daya pembeda tinggi dan tingkat kesukaran yang proporsional. Sedangkan Penelitian internasional seperti Leung (2018) di Hong Kong menunjukkan bahwa desain tes berbasis konteks realistik dapat meningkatkan validitas hingga 30%. Adapun Wijaya, et al., (2011) menegaskan efektivitas *RME* dalam meningkatkan pemahaman konseptual peserta didik di Belanda.

Berdasarkan berbagai penelitian terdahulu, terlihat bahwa sebagian besar studi masih berfokus pada penerapan *RME* dalam proses pembelajaran, bukan pada instrumen evaluasi. Selain itu, penelitian yang secara spesifik menganalisis kualitas butir soal esai berbasis RME dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda masih terbatas. Padahal, soal esai berpotensi lebih kuat dalam mengukur proses berpikir dan kemampuan pemecahan masalah dibanding soal pilihan ganda (Toruan & Tambunan, 2024). Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut. Sehingga

penelitian ini bertujuan untuk menganalisis kualitas butir soal esai matematika realistik dalam mengukur kemampuan pemecahan masalah peserta didik yang ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda.

METODE

Penelitian ini menggunakan pendekatan deskriptif kuantitatif dengan tujuan untuk menganalisis kualitas butir soal esai matematika realistik pada materi Sistem Persamaan Linear Dua Variabel (SPLDV) yang ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Subjek penelitian terdiri atas 19 mahasiswa semester I (angkatan 2025-2026) Program Studi Tadris Matematika Sekolah Tinggi Agama Islam Negeri Mandailing Natal. Teknik pengambilan sampel dilakukan secara *purposive sampling*, dengan pertimbangan bahwa mahasiswa telah mempelajari materi SPLDV. Instrumen penelitian berupa tes esai sebanyak 8 butir soal yang berbasis pendekatan matematika realistik yang dikembangkan berdasarkan langkah pemecahan masalah Polya (1957). Setiap butir soal matematika divalidasi oleh ahli, yaitu dosen matematika, sebelum dilakukan uji coba kepada sampel.

Data hasil tes dianalisis menggunakan *Microsoft Excel*. Uji validitas butir dilakukan dengan menghitung korelasi *Product Moment Pearson*. Reliabilitas tes dihitung menggunakan koefisien *Cronbach's Alpha* dan dikategorikan berdasarkan kriteria Hinton (2004). Tingkat kesukaran butir soal dan daya pembeda dihitung menggunakan rumus yang dipaparkan Arikunto (2021). Hasil dari keempat analisis tersebut digunakan untuk menilai kualitas dan kelayakan instrumen tes esai matematika realistik sebagai alat evaluasi kemampuan pemecahan masalah matematis mahasiswa.

HASIL DAN PEMBAHASAN

Hasil Penelitian

Hasil analisis butir soal kemampuan pemecahan masalah matematika realistik materi Sistem Persamaan Linear Dua Variabel (SPLDV) berupa soal esai seperti uraian berikut ini.

Validitas Butir Soal

Uji validitas butir soal kemampuan pemecahan masalah matematika realistik ini dilakukan dengan memanfaatkan perangkat lunak *Microsoft Excel* sebagai alat bantu analisis, menggunakan rumus korelasi *product moment* sebagai dasar perhitungan untuk menilai kesesuaian antara setiap butir soal dengan skor total. Rumus korelasi *product moment*:

$$r_{xy} = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{\{n \sum X_i^2 - (\sum X_i)^2\}\{n \sum Y_i^2 - (\sum Y_i)^2\}}}$$

Keterangan:

r_{xy} = koefisien korelasi *product moment*

n = jumlah responden

X_i = skor responden pada butir soal ke-*i*

Y_i = skor total responden pada seluruh butir soal

$\sum X_i$ = jumlah keseluruhan skor butir soal dari seluruh responden

$\sum Y_i$ = jumlah keseluruhan skor total soal dari seluruh responden

$\sum X_i Y_i$ = jumlah hasil perkalian antara skor butir soal dan skor total tiap responden

Interpretasi besarnya koefisien korelasi antara skor butir soal dan skor total didasarkan pada kriteria yang dikemukakan oleh Arikunto (2021). Kriteria digunakan untuk menentukan tingkat validitas butir soal. Adapun klasifikasinya dapat dilihat pada Tabel 1.

Tabel 1. Kriteria Koefisien Korelasi *Product Moment*

Koefisien Korelasi		Interpretasi
$0,80 < r_{xy} \leq 1,00$		Sangat tinggi
$0,60 < r_{xy} \leq 0,80$		Tinggi
$0,40 < r_{xy} \leq 0,60$		Cukup
$0,20 < r_{xy} \leq 0,40$		Rendah
$r_{xy} \leq 0,20$		Sangat rendah

Valid dan tidaknya setiap butir soal ditentukan dengan membandingkan signifikansi antara r_{xy} (koefisien korelasi *product moment*) dengan r_{tabel} pada taraf signifikansi 5% dengan jumlah responden sebanyak 19 mahasiswa didik, sehingga diperoleh r_{tabel} yaitu 0,456. Hasil uji validitas yang dilakukan dengan menggunakan bantuan perangkat lunak *Microsoft Excel* seperti yang ditampilkan Gambar 1.

NO.	NAMA	SKOR MAKIMAL SOAL								SKOR	
		NOMOR SOAL									
		1	2	3	4	5	6	7	8		
PEROLEHAN NILAI											
1	Aida Fitri Azizah	3	5	5	0	0	0	0	0	13	
2	Azkiyah Salsabilah	5	5	4	3	5	1	0	0	23	
3	Dimas Ardiansyah	5	5	5	3	1	0	5	2	26	
4	Imelda Putri	3	5	5	5	5	0	5	5	33	
5	Khoirun Nisa Lubis	5	5	5	5	5	3	5	5	38	
6	Layla Fallarin	3	5	5	0	0	0	0	0	13	
7	M. Ali Amri Betubara	5	5	4	0	0	0	0	0	14	
8	Mustaqomah Brisitepu	5	5	5	5	1	0	2	2	25	
9	Nuri Wahdah Assifa	5	5	5	3	1	0	1	1	21	
10	Nurul Fatimah Pulungan	5	5	5	5	5	4	1	5	35	
11	Putri Ramadhani	3	5	5	0	0	0	0	0	13	
12	Resti Aini Hasibuan	5	5	5	2	2	0	0	0	19	
13	Reva Hariani	5	5	1	5	3	3	5	5	32	
14	Rizka Fadhilah	3	5	5	0	0	0	0	0	13	
15	Siti Alisah	5	5	1	5	3	0	5	5	29	
16	Siti Annisa Zebua	5	5	5	5	1	0	0	0	21	
17	Siti Masitoh	5	5	5	4	3	0	5	5	32	
18	Syarifah Nurul Hazanah	0	1	1	0	1	0	1	0	4	
19	Syarifah Nurul Hazani	3	4	5	1	0	0	0	0	13	
r_{xy}		0,6705	0,5057	0,0404	0,8947	0,8102	0,6100	0,7527	0,8926		
itung		3,7260	2,4171	0,1668	8,2610	5,6987	3,1742	4,7132	8,1648		
tabel (5%,19)		0,456									
Keterangan		VALID	VALID	INVALID	VALID	VALID	VALID	VALID	VALID		
Jumlah valid		7									
Jumlah tidak valid		1									

Gambar 1. Validitas Butir Soal Kemampuan Pemecahan Masalah Matematika Realistik

Berdasarkan Gambar 1, diperoleh 7 butir soal esai yang dinyatakan valid dan 1 butir soal tidak valid, yaitu soal nomor 3. Hal ini menunjukkan bahwa sebagian besar butir soal telah mampu mengukur kompetensi Sistem Persamaan Linear Dua Variabel (SPLDV) yang diharapkan. Ringkasan hasil uji validitas disajikan pada Tabel 2.

Tabel 2. Hasil Validitas Soal

No butir soal	r hitung	r tabel	Keterangan
1	0,6705	0,456	Valid
2	0,5057	0,456	Valid
3	0,0404	0,456	Tidak Valid
4	0,8947	0,456	Valid
5	0,8102	0,456	Valid
6	0,6100	0,456	Valid
7	0,7527	0,456	Valid
8	0,8926	0,456	Valid

Hasil uji validitas pada Tabel 2 menunjukkan bahwa tujuh dari delapan butir soal dinyatakan valid, sementara satu butir lainnya tidak memenuhi kriteria validitas. Soal nomor 1, 6, dan 7 memiliki kriteria validasi tinggi, soal nomor 4, 5, dan 8 memiliki kriteria sangat tinggi, soal nomor 2 memiliki kriteria cukup, dan soal nomor 3 memiliki kriteria sangat rendah. Setiap soal mempunya presentase 12,5 %. Terdapat 7 soal yang merupakan soal yang valid dengan jumlah presentase 87,5% dan ada 1 soal yang merupakan soal yang tidak valid sebesar 12,5%. Hasil tersebut menunjukan validitas yang baik.

Hasil ini selaras dengan Ramadhan et al. (2022) dan Lestari & Syafri (2021) yang mengembangkan soal esai berbasis konteks pada topik pecahan dan memperoleh 85% butir soal valid. Kedua penelitian menunjukkan bahwa konteks realistik membantu peserta didik memahami situasi soal dengan lebih baik. Tingkat validitas tinggi pada butir soal sejalan dengan Wibowo & Faizah (2021) yang mengembangkan soal kontekstual pada topik aljabar. Mereka melaporkan bahwa banyak butir soal yang valid karena disusun berdasarkan situasi sehari-hari yang akrab bagi peserta didik. Persamaan ini menunjukkan bahwa kedekatan konteks dengan pengalaman siswa atau mahasiswa menjadi faktor kunci dalam pencapaian validitas tinggi.

Secara umum, temuan penelitian ini menunjukkan bahwa sebagian besar butir soal memiliki tingkat validitas yang tinggi. Hal tersebut menandakan bahwa instrumen yang disusun sudah selaras dengan kemampuan peserta serta tujuan pembelajaran yang ingin dicapai. Adanya satu butir soal yang belum valid menunjukkan perlu adanya penyempurnaan instrumen di tahap berikutnya.

Reliabilitas Butir Soal

Reliabilitas instrumen tes esai dianalisis menggunakan koefisien *Cronbach's Alpha* yang dikemukakan oleh Lee J.Cronbach dengan rumus (Cronbach, 1951):

$$r_i = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{S^2} \right)$$

Keterangan:

r_i = koefisien reliabilitas (nilai *Cronbach's Alpha*)

k = jumlah butir soal

S_i^2 = varians dari setiap butir soal

$\sum S_i^2$ = varians total dari seluruh butir soal

Setelah diperoleh nilai koefisien *Cronbach's Alpha* dari hasil analisis, langkah selanjutnya adalah membandingkan nilai tersebut dengan kriteria interpretasi reliabilitas yang dikemukakan oleh Hinton (2004) untuk menentukan tingkat reabilitas instrumen, sebagaimana ditampilkan pada Tabel 3.

Tabel 3. Kriteria Reabilitas Butir Soal

Koefisien Reliabilitas		Interpretasi							
$0,00 \leq r_i < 0,50$		Reliabilitas rendah							
$0,50 \leq r_i < 0,70$		Reliabilitas sedang							
$0,70 \leq r_i < 0,90$		Reliabilitas tinggi							
$0,90 \leq r_i \leq 1,00$		Reliabilitas sangat tinggi							

Reabilitas butir soal yang di uji menggunakan *Microsoft Excel* dapat dilihat pada Gambar 2.

NO.	NAMA	SKOR MAKIMAL SOAL									SKOR	
		NOMOR SOAL										
		1	2	3	4	5	6	7	8			
1	Aida Fitri Azizah	3	5	5	0	0	0	0	0	13		
2	Azkiyah Salsabilah	5	5	4	3	5	1	0	0	23		
3	Dimas Ardiansyah	5	5	5	3	1	0	5	2	26		
4	Imelda Putri	3	5	5	5	5	0	5	5	33		
5	Khoirun Nisa Lubis	5	5	5	5	5	3	5	5	38		
6	Layla Fallarin	3	5	5	0	0	0	0	0	13		
7	M. Ali Amri Betubara	5	5	4	0	0	0	0	0	14		
8	Mustaqomah Brisitepu	5	5	5	5	1	0	2	2	25		
9	Nuri Wahdah Asifa	5	5	5	3	1	0	1	1	21		
10	Nurul Fatimah Pulungan	5	5	5	5	5	4	1	5	35		
11	Putri Ramadhani	3	5	5	0	0	0	0	0	13		
12	Resti Aini Hasibuan	5	5	5	2	2	0	0	0	19		
13	Reva Hariani	5	5	1	5	3	3	5	5	32		
14	Rizka Fadhilah	3	5	5	0	0	0	0	0	13		
15	Siti Alisah	5	5	1	5	3	0	5	5	29		
16	Siti Annisa Zebua	5	5	5	5	1	0	0	0	21		
17	Siti Masitoh	5	5	5	4	3	0	5	5	32		
18	Syarifah Nurul Hazanah	0	1	1	0	1	0	1	0	4		
19	Syarifah Nurul Hazani	3	4	5	1	0	0	0	0	13		

Varians item	1,88	0,87	2,20	4,78	3,77	1,59	5,14	5,25
Jumlah Var item	25,49							
Jumlah var total	90,27							
Reliabilitas	0,82							

Gambar 2. Reliabilitas Butir Soal Kemampuan Pemecahan Masalah Matematika Realistik

Hasil perhitungan menunjukkan nilai reliabilitas sebesar 0,82 sehingga dapat disimpulkan bahwa soal tes kemampuan pemecahan masalah matematis adalah reliabel, dengan kategori tinggi. Hasil uji reliabilitas ditunjukkan pada Tabel 4.

Tabel 4. Hasil Reliabilitas Butir Soal

Kemampuan Matematis	Cronbach's Alpha	Banyak Soal	Keterangan	Kriteria
TKPM	0,82	8	Reliabel	Tinggi

Pengujian reabilitas tes dilakukan dengan *Microsoft Excel* sebagai alat analisis data. Berdasarkan hasil pengujian, diperoleh koefisien reliabilitas sebesar 0,82, yang nilainya lebih besar dari kriteria minimum yaitu reliabilitas yang sebesar 0,70. Hasil tersebut menunjukkan bahwa instrumen tes memiliki tingkat konsistensi internal yang tinggi, sehingga dapat dipercaya dan layak digunakan sebagai alat ukur untuk menilai kemampuan matematika peserta didik secara konsisten.

Temuan ini sejalan dengan hasil penelitian Khasanah et al. (2023) yang mendapatkan nilai reliabilitas lebih dari 0,70 pada instrumen soal ulangan harian. Hasil ini juga selaras dengan penelitian (Hanhan et al., 2023) yang menunjukkan reliabilitas pada soal KPK dan FPB dalam bingkai soal cerita. Kedua penelitian menunjukkan bahwa soal berbasis konteks atau pemecahan masalah matematika memiliki pola jawaban yang relatif stabil di antara responden.

Dari perbandingan dengan beberapa penelitian tersebut, dapat disimpulkan bahwa reliabilitas instrumen yang dihasilkan dalam penelitian ini termasuk sangat baik. Nilai koefisien 0,82 memperlihatkan bahwa instrumen mampu memberikan hasil yang stabil dan dapat dipercaya dalam mengukur kemampuan matematika mahasiswa (Arikunto, 2021; Rahmi et al., 2024; Sudjana, 1995). Hasil ini sekaligus memperkuat temuan penelitian sebelumnya bahwa soal berbasis konteks nyata cenderung menghasilkan reliabilitas yang lebih tinggi dibandingkan soal konvensional (Ramadhan et al., 2022). Meski demikian, penelitian lanjutan disarankan untuk melibatkan jumlah responden yang lebih besar agar hasilnya semakin tepat dan dapat diandalkan.

Tingkat Kesukaran Butir Soal

Analisis tingkat kesukaran dilakukan untuk mengetahui proporsi mahasiswa yang mampu menjawab setiap butir soal esai dengan benar. Uji tingkat kesukaran butir soal kemampuan matematis ini menggunakan rumus (Johari et al., 2011).

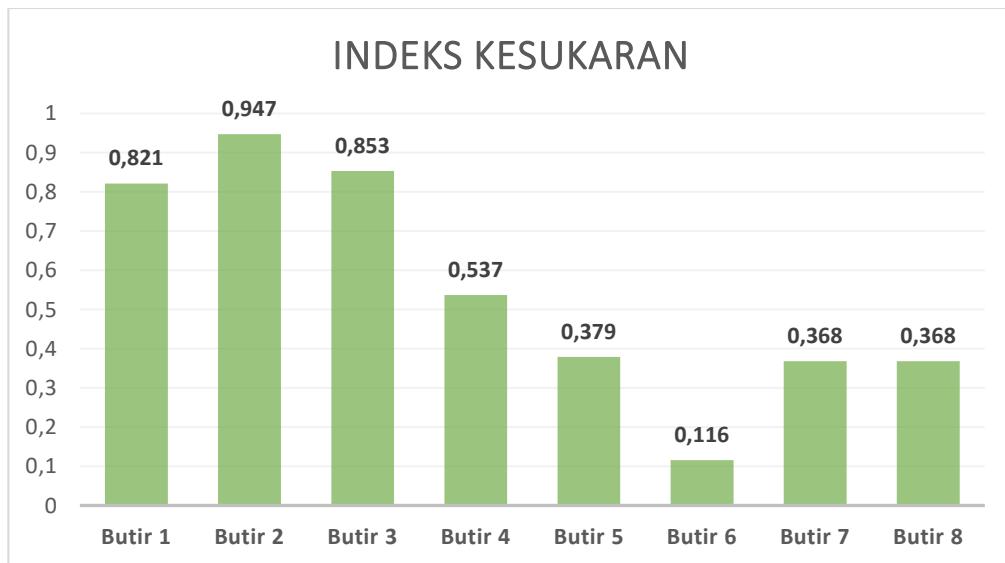
$$\text{Difficulty index} = \frac{\text{average score}}{\text{full item score}}$$

Kriteria indeks kesukaran butir soal kemampuan pemecahan masalah matematika realistik peserta didik disajikan pada Tabel 5.

Tabel 5. Kriteria Indeks kesukaran Butir Soal

No	Indeks Kesukaran	Kriteria
1	$0,00 \leq DI < 0,30$	Soal sukar
2	$0,30 \leq DI < 0,80$	Soal sedang
3	$0,80 \leq DI \leq 1,00$	Soal mudah

Hasil perhitungan indeks tingkat kesukaran butir soal kemampuan pemecahan masalah matematika realistik yang dilakukan dengan menggunakan *Microsoft Excel* seperti yang disajikan pada Gambar 3.



Gambar 3. Diagram Indeks Kesukaran Butir Soal

Gambar 3 menunjukkan bahwa soal kemampuan pemecahan masalah matematika realistik materi Sistem Persamaan Linear Dua Variabel (SPLDV) butir soal nomor 1,2,3 dalam kategori mudah, butir soal nomor 4,5,7,8 dalam kategori sedang dan nomor 6 dalam kategori sukar. Lebih rinci, soal nomor 1, 2, dan 3 memiliki kriteria mudah dengan presentase 37,5%, soal nomor 4, 5, 7, dan 8 memiliki kriteria sedang dengan presentase 50% dan soal nomor 6 memiliki kriteria sukar dengan presentase 12,5%.

Suatu butir soal dikatakan baik apabila memiliki tingkat kesukaran yang berada pada kategori sedang, yaitu tidak terlalu mudah dan tidak terlalu sulit (Sudjana, 1995). Soal yang terlalu mudah cenderung tidak mendorong peserta didik untuk berpikir dan berusaha secara optimal (Arikunto, 2010; Hayati et al., 2024). Sedangkan soal yang terlalu sulit dapat menimbulkan rasa putus asa serta menurunkan motivasi belajar karena berada di luar kemampuan peserta didik (Basri, 2020; Cholidiyah et al., 2023).

Hasil ini sejalan dengan penelitian Febriani (2016) yang menemukan bahwa proporsi soal dengan tingkat kesukaran sedang mendominasi dalam instrumen berbasis konteks kehidupan sehari-hari. Selain itu, temuan ini juga konsisten dengan hasil penelitian Khasanah et al. (2023) yang melaporkan bahwa pada instrumen soal pilihan ganda, sebagian besar butir soal berada pada kategori sedang. Berdasarkan perbandingan dengan kedua penelitian tersebut, dapat disimpulkan bahwa instrumen tes esai RME pada penelitian ini memiliki tingkat kesukaran yang seimbang dan sesuai untuk menilai kemampuan pemecahan masalah mahasiswa. Hasil ini juga memperlihatkan bahwa penggunaan konteks realistik membantu menyeimbangkan antara soal mudah, sedang, dan sukar (Arikunto, 2010; Sudjana, 1995).

Daya Pembeda Butir Soal

Analisis daya pembeda bertujuan untuk mengetahui kemampuan butir soal esai dalam membedakan mahasiswa dengan kemampuan tinggi dan rendah. Indeks daya pembeda butir soal esai dapat dihitung dengan rumus (Boopathiraj & Chellamani, 2013) yaitu:

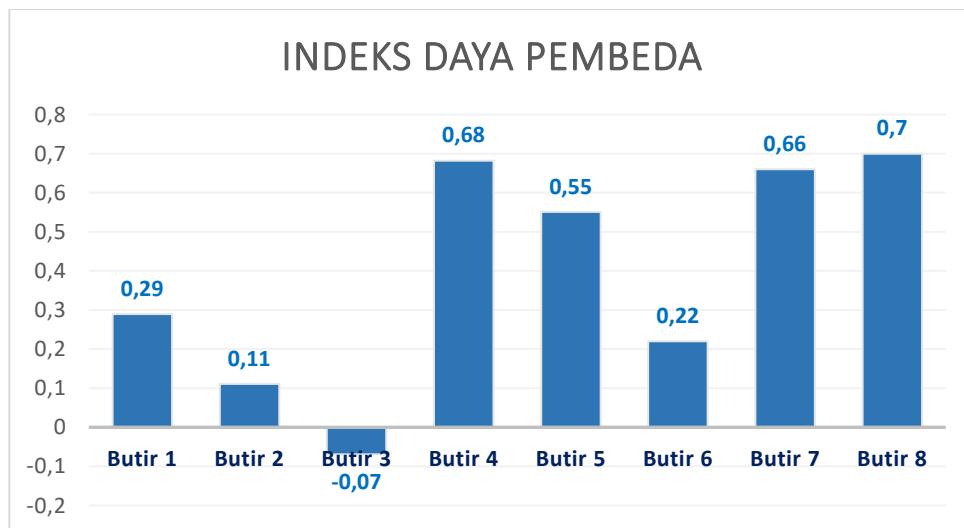
$$\text{Diskrimination Indeks} = \frac{\text{Average upper group} - \text{Average lower group}}{\text{Full item score}}$$

Kriteria indeks daya pembeda butir soal kemampuan pemecahan masalah matematis peserta didik disajikan pada Tabel 6.

Tabel 6. Indeks Daya Pembeda Butir Soal

No	Indeks Daya Pembeda	Interpretasi
1	Tanda negatif	Tidak ada daya pembeda
2	$0,00 \leq D < 0,20$	Lemah
3	$0,20 \leq D < 0,40$	Cukup
4	$0,40 \leq D < 0,70$	Baik
5	$0,70 \leq D \leq 1,00$	Baik sekali

Hasil perhitungan indeks daya pembeda butir soal kemampuan pemecahan masalah matematika realistik yang dilakukan dengan menggunakan *Microsoft Excel* seperti yang disajikan pada diagram berikut ini (lihat Gambar 4).



Gambar 4. Diagram Indeks Daya Pembeda Butir Soal

Hasil analisis menunjukkan, daya pembeda soal diproleh bahwa soal nomor 1 dan 6 memiliki kriteria lemah dengan presentase 25%, soal nomor 2 memiliki kriteria lemah dengan presentase 12,5% , soal nomor 4, 5, dan 7 memiliki kriteria baik dengan presentase 37,5% dan soal no 8 memiliki kriteria sangat baik dengan presentase nilai 12,5%. Dengan ditemukannya beberapa butir soal yang termasuk dalam kategori kurang baik dan sangat kurang baik, maka peneliti berpendapat bahwa pengajar perlu melakukan refleksi terhadap proses pembelajaran yang telah dilaksanakan.

Di sisi lain, butir soal yang memiliki nilai negatif sebaiknya tidak digunakan dan perlu dieliminasi atau direvisi kembali agar sesuai dengan kriteria soal yang baik. Hasil ini sejalan dengan penelitian Febriani (2016), yang menemukan bahwa butir soal matematika yang memiliki daya pembeda lemah sebaiknya tidak digunakan kembali. Persamaan ini terlihat juga dalam penelitian (Magdalena et al., 2021), di mana butir dengan redaksi yang ambigu atau kurang menuntut proses berpikir mendalam

memiliki daya pembeda rendah. Di sisi lain, Anas (2006) memaparkan pentingnya daya beda dalam mengukur kompetensi yang dicapai siswa agar sesuai dengan tujuan pembelajaran. Soal berdaya beda tinggi menggambarkan soal dapat memilah kemampuan siswa secara efektif (Arifin, 2017). Dengan demikian, dapat disimpulkan bahwa soal esai dalam penelitian ini secara umum memiliki daya pembeda yang baik, tetapi masih memerlukan penyempurnaan agar fungsi pembedanya lebih optimal dan hasil penilaian menjadi lebih menggambarkan kemampuan mahasiswa..

KESIMPULAN

Berdasarkan bagian hasil dan pembahasan, instrumen tes esai matematika realistik yang dikembangkan untuk mengukur kemampuan pemecahan masalah pada materi SPLDV menunjukkan kualitas yang baik secara keseluruhan. Dari delapan butir soal yang diujikan, tujuh di antaranya dinyatakan valid ($r \geq 0,456$) dan koefisien reliabilitas mencapai 0,82, yang termasuk kategori tinggi. Sebagian besar butir soal berada pada tingkat kesukaran sedang (50%), dengan daya pembeda mayoritas masuk kategori baik dan sangat baik. Temuan ini menunjukkan bahwa soal-soal berbasis pendekatan matematika realistik mampu mengukur kemampuan berpikir matematis mahasiswa. Penelitian selanjutnya disarankan melibatkan responden dalam jumlah lebih banyak dan menguji instrumen ini pada topik matematika lain agar validitas dan reliabilitasnya semakin bagus dan dapat menjadi acuan dalam asesmen soal tes matematika yang berbasis pendekatan matematika realistik.

REFERENSI

- Alifah, S. N., & Sugilar, H. (2024). Analisis tingkat kesukaran soal matematika pada kemampuan penyelesaian masalah matematis siswa kelas VII. *Gunung Djati Conference Series*, 41(2774–6585), 96–106.
- Amam, A. (2017). Penilaian kemampuan pemecahan masalah matematis siswa SMP. *Teorema: Teori Dan Riset Matematika*, 2(1), 39–46. <https://doi.org/10.25157/teorema.v2i1.765>
- Anas, S. (2006). Pengantar Statistik Pendidikan, Jakarta: PT. *Raja Grafindo Persada*.
- Arifin, Z. (2017). *Evaluasi Pembelajaran*. PT. Remaja Rosdakarya.
- Arikunto, S. (2010). *Dasar-dasar Evaluasi Pendidikan*. Bumi Aksara.
- Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan edisi 3*. Bumi aksara.
- Basri, H. (2020). *Pengembangan instrumen asesmen berpikir kritis pada materi fungsi untuk siswa SMP*. Universitas Negeri Malang.
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*, 2(2), 189–193.
- Chavarría-Arroyo, G., & Albanese, V. (2022). Contextualized Mathematical Problems: Perspective of Teachers about Problem Posing. *Education Sciences*, 13(1), 1–15. <https://doi.org/10.3390/educsci13010006>

- Cholidiyah, A. C., Sujarwo, I., & Turmudi. (2023). Analisis Kebutuhan e-LKPD Bernuansa Keislaman berbasis Problem Based Learning untuk Mendukung Penalaran Adaptif Peserta Didik. *Journal of Mathematics in Teaching and Learning*2, 2(2), 131–144.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Fauji, T., Sampoerno, P. D., & El Hakim, L. (2022). Penilaian berpikir komputasi sebagai kecakapan baru dalam literasi matematika. *Prosiding Seminar Nasional Fakultas Tarbiyah Dan Keguruan 2022*.
- Febriani, I. M. (2016). Analisis Butir Soal Ujian Akhir Semester (UAS) Bahasa Jerman Kelas X MIA 6 SMA Negeri 1 Maospati Tahun Pelajaran 2015/2016. *Laterne*, 5(2), 1–12.
- Gravemeijer, K. (1994). *Developing realistic mathematics education*. Freudenthal Institute.
- Hanan, M. P., Jannah, R. R., & Alim, J. A. (2023). Analisis Butir Soal Matematika Materi KPK Dan FPB Berbasis Cerita Di SDN 111 Pekanbaru. *Journal of Educational Learning and Innovation (ELIa)*, 3(1), 92–106. <https://doi.org/10.46229/elia.v3i1.538>
- Hayati, N., Adelina, Zahria, A., & Rangkuti, K. N. (2024). Item Analysis of Mathematics Questions on the Topic of Units of Time at 116 Percontohan State Elementary School. *Journal of Mathematics In Teaching and Learning*, 3(1), 233–242.
- Himmah, W. I. (2019). Analisis soal penilaian akhir semester mata pelajaran matematika berdasarkan level berpikir. *Journal of Medives: Journal of Mathematics Education IKIP Veteran Semarang*, 3(1), 55–63.
- Hinton, P. R. (2004). *Statistics explained: A guide for social science students* (2nd ed.). Routledge.
- Irawan, Y., & Sa'adah, R. (2023). Analisis pengembangan butir soal perbandingan bertingkat untuk meningkatkan kemampuan pemecahan masalah peserta didik. *Jurnal Indiktika*, 5(2), 222–232. <https://doi.org/10.31851/indiktika.v5i2.11592>
- Johari, J., Sahari, J., Wahab, D. A., Abdullah, S., Abdullah, S., Omar, M. Z., & Muhamad, N. (2011). Difficulty index of examinations and their relation to the achievement of programme outcomes. *Procedia-Social and Behavioral Sciences*, 18, 71–80. <https://doi.org/10.1016/j.sbspro.2011.05.011>
- Khasanah, I., Fuady, A., & Sunismi, S. (2023). Analisis Soal Ulangan Harian Matematika Bentuk Pilihan Ganda. *Mathema Journal*, 5, 110–125.
- Lestari, A., & Syafri, F. S. (2021). Pengembangan Soal Open-Ended Berbasis RME Materi Geometri (Luas Permukaan Dan Volume Balok). *Jurnal Pendidikan Matematika*, 12(1), 14–23. <https://doi.org/10.36709/jpm.v12i1.14789>
- Leung, F. K. S. (2018). Designing context-based mathematics assessments: The Hong Kong perspective. *International Journal of Science and Mathematics Education*, 16(7), 1345–1361. <https://doi.org/10.1007/s10763-017-9835-0>
- Lewis, S. T., & Manouchehri, A. (2020). Mapping Shifting Discourses in Classroom Talk Through

- Indexical Discourse Analysis: From the World of Mathematics to the World of Work. *Mathematical Modelling Education and Sense-Making*, 85–94. https://doi.org/10.1007/978-3-030-37673-4_8
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *Bintang*, 3(2), 198–214.
- Malau, A. D. M. (2024). *Efektivitas pembelajaran matematika realistik terhadap kemampuan berpikir kritis dan pemecahan masalah matematis siswa pada materi sistem persamaan*. Universitas HKBP Nommensen.
- Midgett, C. W., & Eddins, S. K. (2001). NCTM's Principles and Standards for School Mathematics: Implications for Administrators. *NASSP Bulletin*, 85(623), 43–52. <https://doi.org/10.1177/019263650108562306>
- Muslimin, Putri, R. I. I., Zulkardi, & Aisyah, N. (2020). Learning integers with realistic mathematics education approach based on Islamic values. *Journal on Mathematics Education*, 11(3), 363–384. <https://doi.org/10.22342/jme.11.3.11721.363-384>
- Napitupulu, O. T.; Simamora, R. (2023). Pengaruh pendekatan matematika realistik terhadap kemampuan pemecahan masalah dalam penyelesaian soal cerita pada kelas XI SMA Negeri 1. *Jurnal Innovative Research in Education*, 3(6), 5927–5940.
- OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning Outcomes*.
- Özerem, A. (2012). Misconceptions In Geometry And Suggested Solutions For Seventh Grade Students. *Procedia - Social and Behavioral Sciences*, 55, 720–729. <https://doi.org/10.1016/j.sbspro.2012.09.557>
- Pardimin, P., & Widodo, S. A. (2017). Analisis butir soal tes pemecahan masalah matematika. *Wacana Akademika*, 1(1), 69–76. <https://doi.org/10.30738/wa.v1i1.1084>
- Polya, G. (1957). *How to solve it: A new aspect of mathematical method* (2nd ed.). Princeton University Press. <https://doi.org/10.1515/9781400828678>
- Prahmana, R. C. I. (2022). Ethno-realistic mathematics education: The promising learning approach in the city of culture. *SN Social Sciences*, 2(257), 1–19. <https://doi.org/10.1007/s43545-022-00571-w>
- Puspitasari, R. Y., & Airlanda, G. S. (2021). Meta-Analisis Pengaruh Pendekatan Pendidikan Matematika Realistik (PMR) terhadap Hasil Belajar Siswa Sekolah Dasar. *Jurnal Basicedu*, 5(2), 1094–1103. <https://doi.org/10.31004/basicedu.v5i2.878>
- Rahmi, Y., Agustina, F., & Nasution, M. A. (2024). Analisis Butir Soal UAS Matematika di SDN 072 Sibaungbaung Kecamatan Panyabungan Utara. *Journal of Mathematics In Teaching and Learning*, 2(2), 145–155.
- Ramadhan, M. H., Zulkardi, Z., & Putri, R. I. I. (2022). Designing Learning Trajectory for Teaching Fractions Using PMRI Approach with a Chessboard Context. *SJME (Supremum Journal of*

- Mathematics Education), 6(2), 162–170.* <https://doi.org/10.35706/sjme.v6i2.5866>
- Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis kualitas instrumen untuk mengukur kemampuan pemecahan masalah siswa selama pembelajaran daring di masa pandemi. *Jurnal Ilmiah Pendidikan Matematika, 1(2), 130–139.* <https://doi.org/10.46306/lb.v1i2>
- Son, A. L. (2019). Instrumentasi Kemampuan Pemecahan Masalah Matematis: Analisis Reliabilitas, Validitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Gema Wiralodra, 10(1), 41–52.*
- Sriamanda, A., Siregar, N. A. R., & Elvi, M. (2025). Pengembangan Instrumen Tes Matematika pada Materi Statistika Berbasis Kontekstual untuk Siswa Kelas VII SMP. *Aljabar: Jurnal Ilmuan Pendidikan, Matematika Dan Kebumian, 1(3), 161–171.* <https://doi.org/https://doi.org/10.62383/aljabar.v1i3.693>
- Sudjana, N. (1995). *Penilaian hasil proses belajar mengajar.*
- Toruan, N. N. L., & Tambunan, H. (2024). Efektivitas Model Contextual Teaching and Learning terhadap Kemampuan Pemecahan Masalah dan Penalaran Matematis. *Jurnal CURERE, 8(2), 33–43.*
- Ulya, A. L., & Agustyarini, Y. (2020). Pengaruh pendekatan pendidikan matematika realistik terhadap kemampuan penalaran matematis siswa kelas V pada materi bangun ruang. *Atthiflah Journal, 7(2), 21–33.*
- Wibowo, T. E., & Faizah, S. (2021). Pengembangan Soal Tes Untuk Mengukur Kemampuan Pemecahan Masalah Siswa Pada Materi Bentuk Aljabar. *Alifmatika: Jurnal Pendidikan Dan Pembelajaran Matematika, 3(2) SE-Articles), 145–158.* <https://doi.org/10.35316/alifmatika.2021.v3i2.145-158>
- Wijaya, A., van den Heuvel-Panhuizen, M., & Doorman, M. (2011). Students' difficulties in solving context-based PISA mathematics tasks: An exploratory study in Indonesia and the Netherlands. *International Journal of Science and Mathematics Education, 9(1), 89–119.*
- Zulkardi, Z., Putri, R. I. I., & Wijaya, A. (2020). Two Decades of Realistic Mathematics Education in Indonesia BT - International Reflections on the Netherlands Didactics of Mathematics: Visions on and Experiences with Realistic Mathematics Education. In M. van den Heuvel-Panhuizen (Ed.), *International Reflections on the Netherlands Didactics of Mathematics.* (pp. 325–340). Springer International Publishing. https://doi.org/10.1007/978-3-030-20223-1_18