

Pengembangan dan Uji Terbatas Instrumen Diagnostik Miskonsepsi pada Materi Fluida Dinamis dengan Analisis Model Rasch

Firna Agustin, Agiva Dianda Astri, dan Tresna Galih Sukma Suryana*

Universitas Sultan Ageng Tirtayasa
Jl. Ciwaru Raya, Kota Serang, Banten 42117, Indonesia

* E-mail: tresna.galih@untirta.ac.id

Abstrak

Penelitian ini bertujuan untuk mengembangkan dan memvalidasi instrumen diagnostik berbasis *Google Form* untuk mengidentifikasi miskonsepsi siswa pada materi fluida dinamis yang meliputi konsep debit, persamaan kontinuitas, hukum Bernoulli, dan teorema Torricelli. Penelitian dilakukan dengan menggunakan metode *Research and Development* (R&D) dengan model pengembangan 4-D (*Define, Design, Develop, dan Disseminate*), namun hanya dilaksanakan hingga tahap *Develop* karena implementasi dilakukan secara terbatas. Subjek penelitian terdiri dari 35 siswa kelas XI dari salah satu Madrasah Aliyah di Kota Serang. Proses penelitian meliputi pengembangan instrumen, validasi ahli, uji coba terbatas, dan analisis data menggunakan Ministep model Rasch. Hasil penelitian menunjukkan bahwa instrumen memiliki validitas dan reliabilitas yang cukup, dengan reliabilitas instrumen sebesar 0,91 (kategori baik). Instrumen ini efektif dalam mengidentifikasi miskonsepsi siswa, terutama pada konsep hukum Bernoulli. Artikel ini dapat memberikan kontribusi dalam pengembangan evaluasi pembelajaran fisika berbasis tes diagnostik.

Kata kunci: Evaluasi, Fluida Dinamis, Miskonsepsi, Model Rasch

Abstract

This study aims to develop and validate a diagnostic instrument based on Google Forms to identify students' misconceptions on dynamic fluid material, including the concepts of discharge, continuity equations, Bernoulli's law, and Torricelli's theorem. The research was conducted using the Research and Development (R&D) method with a 4-D development model (Define, Design, Develop, and Disseminate), but was only implemented up to the Develop stage because implementation was limited. The research subjects consisted of 35 eleventh-grade students from a Madrasah Aliyah in Serang City. The research process included instrument development, expert validation, limited testing, and data analysis using the Ministep Rasch model. The results showed that the instrument had sufficient validity and reliability, with an instrument reliability of 0.91 (good category). This instrument is effective in identifying students' misconceptions, especially regarding Bernoulli's principle. This article can contribute to the development of physics learning evaluation based on diagnostic tests.

Keywords: Evaluation, Fluid Dynamics, Misconceptions, Rasch Model

PENDAHULUAN

Pembelajaran fisika bertujuan mengkaji konsep, hukum, rumus, dan fenomena yang terjadi dalam kehidupan sehari-hari melalui penerapan metode ilmiah yang didasari oleh sikap ilmiah. Untuk memahami berbagai

konsep, hukum, rumus dan peristiwa tersebut diperlukan ketelitian serta pemahaman konsep yang mendalam (Aldila et al., 2017). Untuk mendukung efektivitas proses pembelajaran, pendidik memanfaatkan beragam media pembelajaran guna memudahkan siswa dalam memahami serta mengaplikasikan konsep-konsep fisika guna memudahkan siswa secara

lebih bermakna (Ekawati et al., 2015).

Siswa masih menghadapi kesulitan dalam memahami konsep fluida dinamis, dikarenakan pembelajaran konsep fluida dinamis di sekolah cenderung masih bersifat informatif, dan menekankan pada persamaan matematis sehingga belum mampu memberikan implementasi nyata bagi siswa. Oleh karena itu, sering terjadi miskonsepsi dalam materi fluida dinamis pada siswa. Miskonsepsi merupakan pemahaman yang keliru atau tidak sejalan dengan konsep ilmiah yang diakui oleh para ahli dalam bidang tersebut (Suparno, 2013). Miskonsepsi atau kesalahan konsep merujuk pada pemahaman suatu konsep yang tidak selaras dengan penelitian ilmiah atau tidak sesuai dengan pandangan para ahli pada bidang tersebut. Miskonsepsi yang dialami siswa perlu dikenali dan dipahami oleh guru agar dapat memberikan bantuan yang tepat dalam memperbaikinya sehingga proses pembelajaran menjadi lebih efektif (Sheftyawan et al., 2018).

Penelitian terkait miskonsepsi siswa pada materi fluida dinamis telah dilakukan oleh beberapa peneliti, seperti Aprita et al. (2018), yang menyatakan bahwa sekitar 33% siswa yang mengalami kesulitan dalam memahami konsep fisika. Selain itu, penelitian oleh Ramadhani et al. (2022), menunjukkan bahwa 62,38% siswa yang masih mengalami miskonsepsi pada materi fluida dinamis dan 16,38% tidak memahami konsep tersebut dengan baik. Mengidentifikasi konsep yang dipelajari siswa menjadi hal yang penting bagi pendidik. Namun, karena bentuk miskonsepsi dapat bervariasi antara siswa serta pendekatan umum yang tidak selalu efektif, maka diperlukan suatu instrumen diagnostik yang mampu mendeteksi secara spesifik jenis miskonsepsi yang dialami oleh masing-masing siswa.

Instrumen tes seperti instrumen tes diagnostik *two-tier* menjadi solusi efektif dalam mengungkap pemahaman dan kesalahan konsep siswa, sehingga memungkinkan guru untuk merancang intervensi pembelajaran yang lebih tepat sasaran. Instrumen tes diagnostik *two-tier* adalah salah satu jenis tes diagnostik

yang memiliki soal dengan dua tingkat. Tingkat pertama berisi pertanyaan dengan lima pilihan jawaban, sementara tingkat kedua terdiri dari alasan yang merujuk pada jawaban tingkat pertama (Utami et al., 2020). Untuk mengevaluasi kualitas instrumen secara menyeluruh, analisis Rasch digunakan karena kemampuannya dalam mengevaluasi kualitas butir soal dan distribusi kemampuan siswa secara objektif. Pendekatan ini telah banyak diterapkan dalam pengembangan instrumen diagnostik di pembelajaran fisika, termasuk dalam studi literasi energi dan mental model siswa (Suryana et al., 2020, 2024).

Penelitian ini dilakukan berdasarkan urgensi untuk menyediakan instrumen evaluasi yang mampu mengidentifikasi miskonsepsi siswa dalam materi fluida dinamis. Instrumen tes diagnostik *two-tier* yang dikembangkan diimplementasikan kepada siswa di salah satu Madrasah Aliyah di Kota Serang sebagai uji coba terbatas. Fokus utama penelitian adalah mengembangkan dan memvalidasi instrumen diagnostik *two-tier* berbasis *Google Form* agar dapat mendeteksi miskonsepsi siswa dalam materi fluida dinamis. Selain itu, penelitian ini juga bertujuan untuk menguji validitas dan reliabilitas instrumen yang dikembangkan agar dapat digunakan sebagai alat diagnostik yang efektif dalam pembelajaran fisika.

METODE PENELITIAN

Penelitian ini menggunakan metode riset dan pengembangan atau *Research and Development* (R&D) dengan model pengembangan 4-D (*Define, Design, Develop, dan Disseminate*). Namun, dalam pelaksanaannya, penelitian ini hanya mencakup tiga tahap awal, yaitu *Define, Design, dan Develop*. Tahap *Disseminate* belum dilaksanakan karena penelitian masih berada pada tahap uji terbatas. Penyusunan instrumen dimulai dengan merancang soal berbentuk pilihan ganda dua tingkat berjumlah 11 soal. Proses validasi ahli dilakukan oleh ahli pada asesmen pembelajaran fisika. Hasil dari validasi ini menjadi dasar dalam perbaikan instrumen sebelum digunakan dalam uji lapangan.

Uji coba terbatas instrumen melibatkan

35 siswa kelas XI salah satu Madrasah Aliyah di Kota Serang. Jumlah tersebut sesuai untuk tahap *pilot test* dan dapat diterima untuk analisis Rasch pada fase awal pengembangan instrumen. Hal ini sesuai dengan Linacre (1994) yang menyatakan bahwa ukuran sampel berkisar 30 hingga 50 responden sudah memadai untuk menghasilkan estimasi parameter yang stabil pada tahap awal analisis Rasch.

Data yang diperoleh dari hasil uji coba terbatas dianalisis dengan menggunakan Ministep 5.9.1., sebuah perangkat lunak yang menganalisis menggunakan model Rasch. Tabel output yang digunakan dalam penelitian ini adalah *Table 1.0-Variable (Wright) maps*, *Table 13.1-Item: Measure*, *Table 3.1-Summary Statistics*, dan *Table 10-Item (Column): Fit Order*.

a. Analisis Validitas

Menurut Boone dalam Sumintono & Widhiarso (2014), kriteria yang digunakan untuk memeriksa kesesuaian butir soal yang tidak sesuai (outliers atau misfits) adalah sebagaimana Tabel 1.

Tabel 1. Kriteria Validitas

Besarnya Nilai
$0,5 < MNSQ < 1,5$
$-2,0 < ZSTD < +2,0$
$0,4 < Pt Measure Corr < 0,885$

b. Analisis Reliabilitas

Menurut Sari & Mahmudi (2024), kriteria penentuan nilai dari nilai Alpha Cronbach (mengukur reliabilitas, yaitu interaksi antara Person dan Item soal secara keseluruhan) sebagaimana Tabel 2.

Tabel 2. Kriteria nilai Alpha Cronbach

Besarnya Nilai	Kategori
$< 0,5$	Buruk
$0,5 - 0,6$	Jelek
$0,6 - 0,7$	Cukup
$0,7 - 0,8$	Bagus
$> 0,8$	Bagus Sekali

Kriteria penentuan nilai dari nilai *Person Reliability* dan *Item Reliability* (mengukur konsistensi jawaban siswa dan butir dalam pengukuran) sebagaimana Tabel 3.

Tabel 3. Kriteria nilai *Person Reliability* dan *Item Reliability*

Besarnya Nilai	Kategori
$< 0,67$	Lemah
$0,67 - 0,80$	Cukup
$0,81 - 0,90$	Bagus
$0,91 - 0,94$	Bagus Sekali
$> 0,94$	Sempurna

c. Analisis Tingkat Kesukaran

Menurut Sari & Mahmudi (2024), hasil analisis butir soal menunjukkan kriteria tingkat kesukaran terbagi dalam lima kategori, yaitu sebagaimana Tabel 4.

Tabel 4. Kriteria tingkat kesukaran

Besarnya Nilai Logit	Kategori
$< -2,00$	Sangat Mudah
$-2,00$	Mudah
$-1,00 - +1,00$	Sedang
$+2,00$	Sukar
$> +2,00$	Sangat Sukar

d. Analisis Daya Pembeda

Klasifikasi daya pembeda ditentukan berdasarkan angka indeks diskriminasi (D) butir soal. Kriteria daya pembeda soal (*discrimination*), nilai D sebagaimana Tabel 5.

Tabel 5. Kriteria daya pembeda

Besarnya Nilai D	Kategori
$< 0,2$	Jelek
$0,2 - 0,3$	Kurang Baik
$0,3 - 0,4$	Baik
$> 0,4$	Sangat Baik

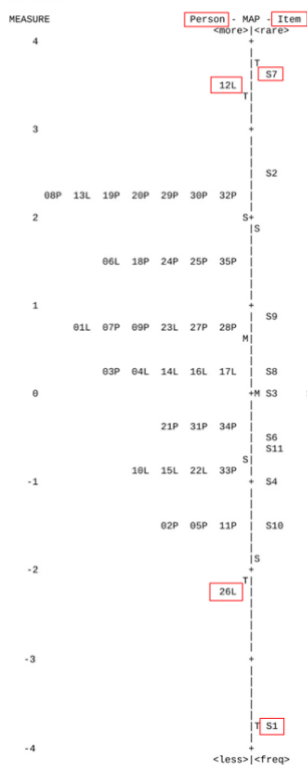
HASIL DAN PEMBAHASAN

a. Analisis Peta Wright (Person-Item Map)

Peta Wright merupakan sebuah peta yang menunjukkan penyebaran kemampuan siswa dan distribusi tingkat kesulitan soal dengan skala yang serupa. Peta Wright membantu dalam mengevaluasi sejauh mana kesesuaian antara kemampuan siswa dengan tingkat kesulitan setiap butir soal. Siswa yang berada di posisi lebih tinggi pada peta memiliki kemampuan yang lebih tinggi, sementara item yang terletak lebih tinggi menunjukkan tingkat kesulitan yang lebih tinggi.

Berdasarkan Gambar 1., siswa 12L memiliki kemampuan tertinggi, sementara 26L dengan kemampuan terendah. Butir S7 memiliki tingkat kesulitan tertinggi, yang berarti kemungkinan siswa menjawab benar sangat kecil. Sebaliknya, butir S1 merupakan soal termudah sehingga kemungkinan besar siswa menjawab benar. Hal ini menunjukkan bahwa sebagian besar siswa hanya menguasai konsep dasar debit, namun mengalami kesulitan dalam memahami hukum Bernoulli.

TABLE 1.0 C:\Users\clove\OneDrive\Documents\TUGAS\ ZOU438WS.TXTa May 07 2025
14:32:11 Ministep
INPUT: 35 Person 11 Item REPORTED: 35 Person 11 Item 2 CATS MINISTEP 5.9.1.0



Gambar 1. Peta Wright distribusi kemampuan siswa dan soal

Temuan ini sejalan dengan penelitian yang telah dilakukan oleh Jumilah et al. (2022) dan Aprita et al. (2018), yang menunjukkan bahwa miskonsepsi pada hukum Bernoulli cukup tinggi di kalangan siswa SMA. Miskonsepsi tersebut secara umum disebabkan oleh pemahaman intuitif yang keliru, seperti menganggap tekanan selalu berbanding lurus dengan kecepatan aliran fluida. Ramadhani et al. (2022) juga menemukan bahwa siswa cenderung memahami fluida dinamis secara prosedural

bukan konseptual.

b. Analisis Validitas

Uji validitas merupakan proses pengujian dengan tujuan untuk mengevaluasi apakah suatu instrumen dapat dikatakan valid atau tidak valid. Menurut Iba & Wardhana (2023), validitas mengacu pada sejauh mana instrumen tersebut mengukur variabel yang dimaksudkan dengan akurat.

ENTRY	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	INBLE	MODEL	S.E.	MNSQ	ZSTD	INFIT	OUTFIT	PTMEASUR-AL	EXACT MATCH
NUMBER											EXP.	OBSV EXP%
7	3	35	3.63			.64	1.06	.28	5.65	2.47	.03	91.4 91.4
2	7	35	2.47			.47	1.36	1.40	1.23	.55	.21	82.9 81.3
9	16	35	.86			.40	1.18	1.49	1.66	2.04	.31	52.1 68.6
8	20	35	.22			.40	1.13	.76	1.21	.77	.44	68.6 73.4
3	21	35	.06			.41	.82	-.99	-.68	-1.07	.64	74.3 74.3
5	21	35	.06			.41	.78	-1.33	-.65	-1.18	.67	80.0 74.3
6	24	35	-.45			.42	.67	-1.77	-.53	-1.34	.70	50.1 85.7
11	25	35	-.63			.43	.94	-.23	.85	-.21	.52	49.1 82.9
4	27	35	-1.03			.46	.65	-1.64	-.49	-.99	.67	46.1 88.6
10	29	35	-1.49			.50	.93	-1.15	1.09	.35	.42	41.1 85.7
1	34	35	-3.70			1.04	1.13	.43	1.46	.76	.05	197.1 97.1
MEAN	20.6	35.0	.00			.51	.98	-.16	1.41	.20		82.3 80.4
P.50	8.7	0	1.85			.18	.23	1.10	1.39	1.24		8.6 7.4

Gambar 2. Data hasil uji validitas

Berdasarkan tabel hasil analisis Uji Validitas pada Gambar 2. menggunakan Ministep Rasch, nilai MNSQ yang dihasilkan adalah 1,41 yang artinya diterima. Akan tetapi terdapat 3 butir soal yang tidak sesuai kriteria yaitu nomor 7, 9, dan 4 yang masing-masing memiliki nilai MNSQ sebesar 5,65, 1,66, dan 0,49. Begitu pun nilai ZSTD juga dapat dikatakan valid karena secara keseluruhan masih dalam rentang nilai yang diterima yakni 0,20. Namun, terdapat 2 butir soal yang tidak sesuai kriteria yaitu nomor 7 dan 9 yang masing-masing memiliki nilai ZSTD sebesar 2,47 dan 2,04. Pada Pt Measure Corr, terdapat 4 butir soal yang tidak memenuhi nilai standar yakni pada nomor 7, 2, 9, dan 1, yang masing-masing mendapat nilai 0,03, 0,21, 0,31, dan 0,05. Setelah dianalisis dari 11 butir soal, terdapat 2 butir soal yang dinyatakan tidak valid, yakni soal nomor 7 dan 9. Hal tersebut dikarenakan dari ketiga kriteria tersebut, tidak memenuhi standar yang diberikan.

Meskipun hanya soal nomor 7 dan 9 yang dinyatakan tidak valid secara keseluruhan, soal nomor 4, 2, dan 1 juga menunjukkan kelemahan pada aspek tertentu. Oleh karena itu, revisi terhadap redaksi dan distraktor pada butir-butir soal perlu dilakukan. Temuan ini mendukung pandangan Setiyawan & Wijayanti (2020), yang menyatakan bahwa validitas

instrumen harus dianalisis secara menyeluruh, tidak hanya berdasarkan satu indikator.

c. Analisis Reliabilitas

Uji reliabilitas merupakan proses yang bertujuan untuk mengevaluasi tingkat konsistensi suatu instrumen, apakah dapat dipercaya dan terus-menerus sama jika dilakukan pengukuran yang berulang. Menurut Iba & Wardhana (2023), reliabilitas mengukur sejauh mana instrumen dapat menghasilkan hasil yang serupa jika digunakan secara berulang pada subjek atau populasi yang sama.

SUMMARY OF 35 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFINIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	6.5	11.0	.61	.84	.92	-.15	1.23	.26
SEM	.4	.0	.23	.02	.09	-.20	.34	.17
P.SD	2.1	.0	1.36	.12	.50	1.14	1.98	1.00
S.SD	2.2	.0	1.38	.12	.51	1.15	2.01	1.02
MAX.	10.0	11.0	3.51	1.24	1.89	2.44	9.90	3.74
MIN.	2.0	11.0	-2.22	.73	.33	-1.37	.17	-.49

REAL RMSE .93 TRUE SD 1.00 SEPARATION 1.07 Person RELIABILITY .54
 MODEL RMSE .85 TRUE SD 1.06 SEPARATION 1.25 Person RELIABILITY .61
 S.E. OF Person MEAN = .23

Person RAW SCORE-TO-MEASURE CORRELATION = .99
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .62 SEM = 1.31
 STANDARDIZED (50 ITEM) RELIABILITY = .88

SUMMARY OF 11 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFINIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	20.6	35.0	.00	.51	.98	-.16	1.41	.20
SEM	2.8	.0	.59	.06	.07	.35	.44	.39
P.SD	8.7	.0	1.85	.18	.23	1.10	1.39	1.24
S.SD	9.2	.0	1.94	.19	.24	1.16	1.46	1.30
MAX.	34.0	35.0	3.63	1.04	1.36	1.49	5.65	2.47
MIN.	3.0	35.0	-3.70	.40	.65	-1.77	.49	-1.34

REAL RMSE .56 TRUE SD 1.76 SEPARATION 3.12 Item RELIABILITY .91
 MODEL RMSE .54 TRUE SD 1.77 SEPARATION 3.29 Item RELIABILITY .92
 S.E. OF Item MEAN = .59

Item RAW SCORE-TO-MEASURE CORRELATION = -.98
 Global statistics: please see Table 44.
 UMEAN=.0000 USCALE=1.0000

Gambar 3. Data hasil uji reliabilitas

Reliabilitas instrumen diukur melalui tiga parameter, yaitu Person Reliability, Item Reliability, dan Cronbach Alpha. Berdasarkan hasil analisis pada Gambar 3. diketahui bahwa hasil Person Reliability memiliki nilai sebesar 0,54, yang mengindikasikan konsistensi jawaban siswa masih tergolong lemah. Sebaliknya, nilai Item Item Reliability memiliki nilai sebesar 0,91 yang berarti bahwa kualitas item yang diteskan reliabilitasnya bagus sekali. Sementara itu, nilai Cronbach Alpha yang diperoleh adalah 0,62 yang berada pada kategori cukup.

Rendahnya nilai Person Reliability dapat disebabkan oleh homogenitas kemampuan siswa dan adanya miskonsepsi yang persisten. Hal ini sejalan dengan pendapat Sumintono & Widhiarso (2014), yang menyatakan bahwa reliabilitas rendah pada responden terjadi karena belum

mampu menangkap variasi pemahaman secara optimal. Sebaliknya, tingginya Item Reliability menunjukkan bahwa konstruksi soal telah mampu membedakan kemampuan siswa secara umum.

d. Analisis Tingkat Kesukaran

Analisis tingkat kesukaran butir soal merupakan proses yang bertujuan untuk mengetahui apakah sebuah soal tergolong sangat mudah, mudah, sedang, sulit, atau sangat sulit. Semakin tinggi persentase siswa yang dapat menjawab soal dengan benar, maka soal tersebut tergolong mudah. Sebaliknya, jika hanya sedikit siswa yang menjawab dengan benar, maka soal tersebut tergolong sulit (Setiyawan & Wijayanti, 2020).

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFINIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASURE-AL	EXACT MATCH	CORR. EXP.	OBSE	EXPS	Item
7	3	35	3.63	.64	1.06	.28	1.65	2.47	.03	.30	92.4	92.4	57	
2	7	35	2.47	.47	1.36	1.40	1.23	.55	.21	.42	82.9	81.3	52	
9	16	35	.86	.40	1.28	1.49	1.66	2.04	.31	.52	68.6	73.1	59	
8	20	35	.22	.40	1.13	.76	1.21	.77	.44	.52	68.6	73.4	58	
3	21	35	-.06	.41	.82	-.99	-.68	-1.07	.64	.52	74.3	74.3	53	
5	21	35	-.06	.41	.76	-1.33	.65	-1.18	.67	.52	80.0	74.3	55	
6	24	35	-.45	.42	.67	-1.77	.53	-1.34	.70	.50	85.7	76.9	56	
11	25	35	-.63	.43	.94	-.23	.85	-.21	.52	.60	82.9	77.9	51	
4	27	35	-1.03	.46	.65	-1.64	.49	-.99	.67	.46	88.6	80.4	54	
10	29	35	-1.49	.50	.93	-.15	1.09	.35	.42	.41	85.7	83.8	510	
1	34	35	-3.70	1.04	1.13	.43	1.46	.76	.05	.19	97.1	97.1	51	
MEAN	20.6	35.0	.00	.51	.98	-.16	1.41	.20			82.3	80.4		
P.SD	8.7	.0	1.85	.18	.23	1.10	1.39	1.24			8.6	7.4		

Gambar 4. Data hasil analisis tingkat kesukaran

Berdasarkan tabel hasil analisis tingkat kesukaran pada Gambar 4. menggunakan Ministep Rasch, tingkat kesukaran butir soal diurutkan dari tingkat kesukaran tinggi hingga rendah. Butir soal nomor 7 dan 2 merupakan butir soal kategori sangat sulit dengan nilai logitnya masing-masing sebesar 3,63 dan 2,47. Dari 35 siswa hanya 3 siswa yang menjawab benar pada butir soal nomor 7 dan 17 siswa yang menjawab benar pada butir soal nomor 2. Butir soal nomor 1 merupakan butir soal sangat mudah dengan nilai logitnya sebesar -3,70 dan dari 35 siswa terdapat 34 siswa yang menjawab dengan tepat. Butir soal nomor 10 dan 4 termasuk dalam kategori mudah dan butir soal nomor 11, 6, 5, 3, 8, dan 9 termasuk dalam kategori sedang.

Kesulitan pada butir soal nomor 7 dan 2 berkaitan dengan miskonsepsi pada prinsip fluida dinamis dan asas kontinuitas. Ekawati et al. (2015) menyatakan bahwa siswa cenderung memahami aliran fluida

secara intuitif tanpa memahami hubungan matematis antar variabel dalam konsep fluida dinamis. Agustini (2019) juga menemukan bahwa siswa sering gagal memahami bahwa tekanan berkurang saat kecepatan fluida meningkat.

e. Analisis Daya Pembeda

Daya pembeda soal adalah kemampuan dari sebuah butir soal untuk mengidentifikasi siswa yang telah memahami materi dengan siswa yang belum mampu memahami materi. Dengan menggunakan analisis ini, setiap butir soal dapat ditentukan apakah butir soal itu layak, perlu diperbaiki, atau sebaiknya dihapus (Arifin, 2017). Jika suatu butir soal memiliki daya pembeda yang tinggi, maka soal tersebut mampu mengidentifikasi perbedaan antara siswa yang memiliki kemampuan tinggi dan yang berkemampuan rendah.

ENTRY	TOTAL SCORE	TOTAL COUNT	MEASURE	S.E.	PTMEASUR-AL	EXACT MATCH	Item							
7	3	35	3,63	.64	1,06	.28	5,65	2,47	A	.03	.380	91,4	91,4	57
9	16	35	.86	.40	1,28	1,49	1,66	2,04	B	.31	.52	68,6	71,1	59
1	34	35	-3,70	1,04	1,13	.43	1,46	.76	C	.05	.19	97,1	97,1	51
2	7	35	2,47	.47	1,36	1,40	1,23	.55	D	.21	.42	82,9	81,3	52
8	30	35	.22	.40	1,13	.76	1,21	.77	E	.44	.52	68,6	73,4	58
10	29	35	-1,49	.50	.93	.15	1,09	.35	F	.42	.41	85,7	83,8	510
11	25	35	-.63	.43	.94	-.23	.85	-.21	G	.52	.49	82,9	77,9	511
3	23	35	.86	.41	.82	-.99	.66	-1,07	H	.64	.52	74,3	74,3	53
5	23	35	.86	.41	.76	-1,33	.65	-1,18	I	.67	.52	80,8	74,3	55
6	24	35	-.45	.42	.67	-1,77	.53	-1,34	J	.78	.58	85,7	76,9	56
4	27	35	-1,83	.46	.65	-1,64	.49	-.99	K	.67	.46	88,6	80,4	54
MEAN	20,6	35,0	.00	.51	.98	-.16	1,41	.20				82,3	80,4	
P.SD	8,7	.0	1,85	.18	.23	1,10	1,39	1,24				8,6	7,4	

Gambar 5. Data hasil uji daya pembeda

Berdasarkan tabel hasil analisis daya pembeda pada Gambar 5. menggunakan Ministep Rasch, terlihat pada nilai PTMEASUR-AL CORR. Dari hasil analisis didapatkan bahwa terdapat 2 butir soal yang berkategori jelek yaitu soal nomor 7 dan 1 dengan nilai masing-masing 0,03 dan 0,05, dan 1 butir soal berkategori kurang baik yaitu soal nomor 2 dengan nilai 0,21. Soal tersebut tidak dapat membedakan antara siswa yang tahu jawabannya dan siswa yang tidak tahu jawaban atas soal tersebut. Terdapat pula 1 soal berkategori baik yaitu soal nomor 9, dan terdapat 7 soal berkategori sangat baik yaitu soal nomor 4, 6, 5, 3, 11, 10 dan 8 dikarenakan memiliki nilai lebih dari 0.40.

Daya pembeda yang rendah menunjukkan bahwa soal tidak mampu mengidentifikasi perbedaan pemahaman siswa secara efektif. Menurut Arifin (2017)

menyatakan bahwa distraktor yang tidak dipilih oleh siswa dengan kemampuan rendah tidak berfungsi sebagai pengecoh yang efektif. Oleh karena itu, revisi terhadap opsi jawaban pada soal dengan daya pembeda jelek perlu dilakukan.

f. Analisis Distraktor

Analisis distraktor pada butir soal merupakan proses yang bertujuan untuk mengevaluasi efektivitas dan kinerja dari pilihan jawaban yang salah dalam suatu butir soal. Distraktor merupakan sebuah istilah yang digunakan untuk menyebut pilihan jawaban salah dalam instrumen soal pilihan ganda yang dirancang sebagai pengecoh. Menurut Arifin (2017), sebuah pengecoh disebut bekerja dengan baik jika siswa dari kelompok yang lebih rendah sering memilih pilihan jawaban yang mengandung pengecoh. Sebaliknya, jika siswa dari kelompok yang lebih tinggi lebih banyak yang memilih pilihan jawaban yang mengandung pengecoh, maka pengecoh atau distraktor tersebut dapat dianggap tidak berfungsi sebagaimana mestinya. Distraktor dianggap berfungsi dengan baik apabila dipilih oleh setidaknya 5% dari seluruh siswa yang mengikuti tes.

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: MEASURE ORDER

ENTRY	DATA	SCORE	DATA	ABILITY	S.E.	INFIT	OUTFIT	PTMA	Item
NUMBER	CODE	VALUE	COUNT	MEAN	P.SD	MEAN	MNSQ	CORR.	
7	E	0	9	.26		-.12	1,29	.44	.7
	D	0	3	.9		.48	.27	.19	.2
	C	0	16	.46		.91	1,22	.32	.8
	A	0	4	.11		1,13	1,30	.75	1,1
	B	1	3	.9		.75*	2,86	1,46	1,6
2	A	0	1	.3		-1,46	.00	.0	.1
	B	0	4	.11		-.93	.86	.50	.2
	D	0	6	.17		.18	1,53	.68	1,5
	C	0	17	.49		1,82	1,10	.28	1,2
	E	1	7	.20		1,18	1,01	.41	1,9
9	B	0	1	.3		-2,22	.00	.0	.1
	C	0	4	.11		-.60	.83	.48	.5
	A	0	7	.20		.63	.42	.17	1,0
	D	0	7	.20		.64	1,51	.62	1,8
	E	1	16	.46		1,07	1,31	.34	1,4
8	C	0	7	.20		-.26	.87	.35	.8
	B	0	3	.9		-.13	.68	.48	.8
	E	0	1	.3		.21	.80	.8	.7
	A	0	4	.11		.23	.67	.39	1,0
	D	1	20	.57		1,13	1,47	.34	1,5
3	B	0	4	.11		-.93	.86	.50	.5
	A	0	7	.20		-.40	1,12	.46	1,0
	C	0	2	.6		-.85	.26	.26	.7
	E	0	1	.3		.21	.80	.8	.7
	D	1	21	.60		1,33	1,08	.24	.8
5	A	0	3	.9		-1,31	.64	.46	.3
	B	0	7	.20		-.49	.65	.27	.6
	E	0	4	.11		.89	.59	.34	1,0
	C	1	21	.60		1,36	1,13	.25	.9
6	A	0	4	.11		-1,20	.59	.34	.4
	D	0	4	.11		-.74	.47	.27	.5
	C	0	3	.9		-.33	.91	.65	1,0
	B	1	24	.69		1,26	1,06	.22	.7
11	B	0	6	.17		-1,81	.92	.41	.6
	E	0	2	.6		-.62	.84	.84	.8
	C	0	2	.6		1,11	.33	.33	3,3
	A	1	25	.71		1,06*	1,16	.24	.8
4	A	0	5	.14		-1,37	.50	.25	.4
	E	0	1	.3		-.86	.80	.80	.6
	D	0	2	.6		-.34	1,12	1,12	1,2
	B	1	27	.77		1,11	1,07	.21	.6
10	A	0	3	.9		-1,71	.36	.25	.3
	D	0	2	.6		-.84	.82	.82	1,6
	B	0	1	.3		1,44	.00	.00	7,4
	C	1	29	.83		.87*	1,22	.23	.8
1	E	0	1	.3		.21	.80	.41	1,5
	B	1	24	.69		.63	1,38	.24	1,1

* Average ability does not ascend with category score

Gambar 6. Data hasil uji distraktor

Tabel 6. Tabel analisis distraktor setiap butir soal

Nomor Soal	Analisis Distraktor	Rekomendasi
1	Semua distraktor tidak berfungsi, dikarenakan 97% siswa memilih opsi B (opsi benar) dan 3% lainnya memilih opsi E	Semua distraktor tidak berfungsi, ganti semua distraktor (A, C, D, E) Opsi A diganti
2	Distraktor berfungsi dengan baik karena sebagian besar opsi telah mendominasi, namun opsi A hanya dipilih oleh 3% siswa	Opsi E diganti
3	Distraktor berfungsi dengan baik karena sebagian besar opsi telah mendominasi, namun opsi E hanya dipilih oleh 3% siswa	Opsi C dan E diganti
4	Hanya distraktor opsi A dan D yang memenuhi keberfungsian distraktor, yaitu masing-masing 14% dan 6%. Opsi C dan E berada di bawah 5%	Opsi D diganti
5	Distraktor berfungsi dengan baik karena sebagian besar opsi telah mendominasi, namun opsi D tidak dipilih oleh siswa (0%)	Opsi E diganti
6	Distraktor berfungsi dengan baik karena sebagian besar opsi telah mendominasi, namun opsi E tidak dipilih oleh siswa (0%)	-
7	Semua distraktor berfungsi, karena seluruh opsi dipilih oleh lebih dari 5% siswa	Opsi E diganti
8	Distraktor berfungsi dengan baik karena sebagian besar opsi telah mendominasi, namun opsi E hanya dipilih oleh 3% siswa	Opsi B diganti
9	Distraktor berfungsi dengan baik karena sebagian besar opsi telah mendominasi, namun opsi B hanya dipilih oleh 3% siswa	Opsi B dan E diganti
10	Hanya distraktor opsi A dan D yang memenuhi keberfungsian distraktor, yaitu masing-masing 9% dan 83%. Opsi B dan E berada di bawah 5%	Opsi D diganti
11	Distraktor berfungsi dengan baik karena sebagian besar opsi telah mendominasi, namun opsi E tidak dipilih oleh siswa (0%)	

Berdasarkan tabel hasil analisis distraktor pada Gambar 6. menggunakan Ministep Rasch. Hal ini dapat dilihat dari Persentase Count pada Tabel "Data" dari setiap butir soal untuk keberhasilan distraktor. Tabel 6 merupakan penjabaran untuk setiap butir soalnya.

Hasil penelitian menunjukkan bahwa instrumen diagnostik *two-tier* berbasis *Google Form* dapat digunakan untuk mengidentifikasi miskonsepsi siswa. Namun, penyempurnaan pada butir soal dengan validitas dan daya pembeda rendah perlu dilakukan. Guru dapat memanfaatkan simulasi interaktif seperti PhET untuk membantu siswa memahami konsep fluida yang abstrak secara visual (Aldila et al., 2017).

Temuan-temuan dalam penelitian ini

memperkuat pentingnya validasi instrumen secara menyeluruh dan penggunaan model Rasch dalam evaluasi pembelajaran berbasis diagnostik. Instrumen ini membuka peluang untuk diintegrasikan dengan strategi pembelajaran berbasis miskonsepsi dan teknologi digital. Sehingga mampu meningkatkan efektivitas pembelajaran fisika di tingkat sekolah menengah.

PENUTUP

Instrumen tes diagnostik *two-tier* berbasis *Google Form* yang dikembangkan dalam penelitian ini telah berhasil memenuhi tujuan utama, yaitu mengidentifikasi miskonsepsi siswa pada materi fluida dinamis dan menguji validitas serta reliabilitas instrumen secara digital. Hasil analisis menunjukkan bahwa instrumen memiliki

validitas yang memadai dan kualitas item yang sangat baik (Item Reliability = 0,91), meskipun konsistensi jawaban siswa masih tergolong rendah (Person Reliability = 0,54). Temuan ini menunjukkan bahwa instrumen tidak hanya layak digunakan sebagai alat evaluasi, tetapi juga sebagai sarana diagnostik yang efektif dalam mendeteksi miskonsepsi siswa khususnya pada konsep hukum Bernoulli dan teorema Torricelli. Namun demikian, keterbatasan jumlah sampel dalam uji coba terbatas menjadi salah satu kendala dalam menghasilkan generalisasi yang lebih luas. Oleh karena itu, penyempurnaan instrumen melalui revisi butir soal dan peningkatan kualitas distraktor perlu dilakukan, serta disarankan agar penelitian lanjutan melibatkan jumlah sampel yang lebih besar dan mengintegrasikan instrumen ini dengan strategi pembelajaran yang lebih inovatif agar mampu mendeteksi serta mengatasi miskonsepsi siswa secara lebih sistematis.

REFERENSI

- Agustini, A. (2019). *Identifikasi Miskonsepsi Siswa Menggunakan Tes Diagnostik Four-Tier Berbantuan Google Formulir Pada Konsep Fluida Dinamis Di SMA Negeri 1 Parung*. https://repository.uinjkt.ac.id/dspace/bitstream/123456789/49172/1/AULIA_AGUSTINI.pdf
- Aldila, W. Y., Setyarsih, W., & Kholiq, A. (2017). Penggunaan Phet Simulation Dalam Ecirr Untuk Mereduksi Miskonsepsi Siswa Pada Materi Fluida Dinamis. *Jurnal Inovasi Pendidikan Fisika*, 5(3), 161–164.
- Aprita, D. F., Supriadi, B., & Prihandono, T. (2018). Identifikasi Pemahaman Konsep Fluid Dinamis Menggunakan Four Tier Test Pada Siswa SMA. *Jurnal Pembelajaran Fisika*, 7(3), 315–321.
- Arifin, Z. (2017). Kriteria Instrumen dalam Suatu Penelitian. *Jurnal THEOREMS (The Original Research of Mathematics)*, 2(1), 28–36.
- Ekawati, Y., Haris, A., & Amin, B. D. (2015). Penerapan Media Simulasi Menggunakan PhET (Physics Education And Technology) Terhadap Hasil Belajar Fisika Peserta Didik Kelas X SMA Muhammadiyah Limbung. *Jurnal Pendidikan Fisika*, 3(1), 74–82.
- Iba, Z., & Wardhana, A. (2023). Metode Penelitian. In *Metode Penelitian Kualitatif* (Vol. 3, Issue November). Eureka Media Aksara. http://repository.unpas.ac.id/30547/5/BA_B_III.pdf
- Jumilah, Lestari, E. P., & Wasis. (2022). Introduksi Miskonsepsi dan Penyebab Miskonsepsi Peserta Didik pada Sub-materi Asas Bernoulli memakai Four-tier Diagnostic Test. *Jurnal Literasi Pendidikan Fisika*, 3(1), 20–27.
- Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7(4), 328.
- Ramadhani, N., Manullang, S. R., & Br. Simbolon, V. A. (2022). Identifikasi Kemampuan Siswa Dalam Pemecahan Masalah Miskonsepsi Pada Materi Fluida Dinamis Di Tingkat Sma. *EduFisika: Jurnal Pendidikan Fisika*, 7(2), 196–205. <https://doi.org/10.59052/edufisika.v7i2.21465>
- Sari, E. D. K., & Mahmudi, I. (2024). *Analisis Pemodelan Rasch Pada Assesment Pendidikan (Analisis dengan Menggunakan Aplikasi Winstep)* (Issue February). PT. Pena Persada Kerta Utama.
- Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis Kualitas Instrumen Untuk Mengukur Kemampuan Pemecahan Masalah Siswa Selama Pembelajaran Daring Di Masa Pandemi. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 1(2), 130–139. <https://doi.org/10.46306/lb.v1i2.26>
- Sheftyawan, W. B., Prihandono, T., & Lesmono, A. D. (2018). Identifikasi Miskonsepsi Siswa Menggunakan Four-Tier Diagnostic Test Pada Materi Optik Geometri. *Jurnal Pembelajaran Fisika*, 7(2), 147–153.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial* (Edisi Revi, Issue November). Trim Komunikata Publishing House.

- Suparno, P. (2013). *Miskonsepsi dan Perubahan Konsep dalam Pendidikan Fisika*. Gramedia Widiasarana Indonesia.
- Suryana, T. G. S., Nurdini, N., Abdurrahman, D., Oktarisa, Y., & Samsudin, A. (2024). Rasch Analysis of FOCUS for Simple Harmonic Motion: Assessing Students' Mental Model. *Wahana Pendidikan Fisika*, 9(2), 183–198. <https://doi.org/10.17509/wapfi.v9i2.76603>
- Suryana, T. G. S., Setyadin, A. H., Samsudin, A., & Kaniawati, I. (2020). Assessing Multidimensional Energy Literacy of High School Students: An Analysis of Rasch Model. *Journal of Physics: Conference Series*, 1467(1), 012034. <https://doi.org/10.1088/1742-6596/1467/1/012034>
- Utami, I., Mulyani, B., & Yamtinah, S. (2020). Identifikasi Miskonsepsi Asam-Basa dengan Two Tier Multiple Choice Dilengkapi Interview. *Jurnal Pendidikan Kimia*, 9(1), 89–97.