

**PENGEMBANGAN INSTRUMEN PENILAIAN BERBASIS *ADAPTIVE*
ASSESSMENT UNTUK MENGUKUR KETERAMPILAN BERPIKIR
KRITIS PADA MATERI FLUIDA STATIS**

(Tesis)

Oleh

**ERMA NOVELA
NPM 2423022004**



**FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026**

ABSTRAK

PENGEMBANGAN INSTRUMEN PENILAIAN BERBASIS *ADAPTIVE ASSESSMENT* UNTUK MENGUKUR KETERAMPILAN BERPIKIR KRITIS PADA MATERI FLUIDA STATIS

Oleh

ERMA NOVELA

Penelitian ini bertujuan untuk mengembangkan dan mendeskripsikan validitas, reliabilitas empiris, serta kepraktisan instrumen penilaian berbasis *Adaptive Assessment* untuk mengukur keterampilan berpikir kritis siswa pada materi fluida statis. Penelitian menggunakan pendekatan *Research and Development* (R&D) dengan model 4D. Instrumen yang dikembangkan berupa soal pilihan ganda beralasan berdasarkan indikator berpikir kritis Ennis dan dianalisis menggunakan *Partial Credit Model* (PCM) dan model Rasch. Uji coba terbatas melibatkan 159 responden, sedangkan uji coba luas melibatkan 62 responden. Hasil penelitian menunjukkan bahwa media penilaian memiliki validitas sebesar 91,3% dengan kategori sangat valid dan validitas isi berada pada nilai V Aiken > 0.92 . Seluruh butir soal memenuhi kriteria *item fit* serta memiliki reliabilitas tinggi (KR-20 = 0,90; reliabilitas person = 0,89; reliabilitas item = 0,98). Tingkat kesukaran butir berada pada rentang $-1,04$ hingga $0,95$ logit dan 93% butir soal bebas dari bias. Kepraktisan instrumen memperoleh skor rata-rata 91,43% dari siswa dan 80,66% dari guru. Dengan demikian, instrumen penilaian adaptif yang dikembangkan dinyatakan valid, reliabel, dan praktis untuk mengukur keterampilan berpikir kritis siswa dalam pembelajaran fisika, khususnya pada materi fluida statis.

Kata kunci : *Adaptive Assessment*, Berpikir Kritis, Fluida Statis, Instrumen Penilaian, Item Respon Theory

ABSTRACT

DEVELOPMENT OF *ADAPTIVE ASSESSMENT*-BASED ASSESSMENT INSTRUMENTS TO MEASURE CRITICAL THINKING SKILLS IN STATIC FLUID MATERIALS

By

ERMA NOVELA

This study aims to develop and describe the validity, empirical reliability, and practicality of *Adaptive Assessment*-based assessment instruments to measure students' critical thinking skills on static fluid materials. The research used a *Research and Development* (R&D) approach with a 4D model. The instrument developed in the form of a reasoned multiple-choice question based on Ennis' critical thinking indicator and analyzed using *the Partial Credit Model* (PCM) and the Rasch model. The limited trial involved 159 respondents, while the broad trial involved 62 respondents. The results showed that the assessment media had a validity of 91.3% with a very valid category and the validity of the content was in >0.92 . All question items met the criteria of *item fit* and had high reliability ($KR-20 = 0.90$; person reliability = 0.89; item reliability = 0.98). The difficulty level of the items was in the range of -1.04 to 0.95 logits and 93% of the questions were free of bias. The practicality of the instrument obtained an average score of 91.43% from students and 80.66% from teachers. Thus, the *Adaptive Assessment* instrument developed is declared valid, reliable, and practical to measure students' critical thinking skills in physics learning, especially in static fluid materials.

Keywords: Adaptive Assessment, Critical Thinking, Static Fluid, Assessment Instruments, Response Theory Items

**PENGEMBANGAN INSTRUMEN PENILAIAN BERBASIS *ADAPTIVE*
ASSESSMENT UNTUK MENGUKUR KETERAMPILAN BERPIKIR
KRITIS PADA MATERI FLUIDA STATIS**

Oleh

ERMA NOVELA

Tesis

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
MAGISTER PENDIDIKAN FISIKA**

Pada

**Jurusan Pendidikan Matematika dan Ilmu Pengetahuan Alam
Fakultas Keguruan dan Ilmu Pendidikan**



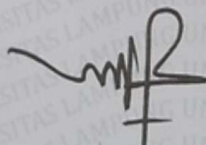
**FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026**

Judul Tesis : **PENGEMBANGAN INSTRUMEN
PENILAIAN BERBASIS *ADAPTIVE*
ASSESSMENT UNTUK MENGUKUR
KETERAMPILAN BERPIKIR KRITIS
PADA MATERI FLUIDA STATIS**

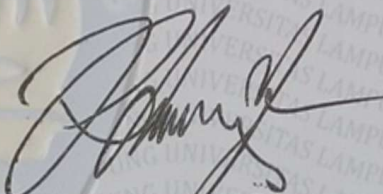
Nama Mahasiswa : **Erma Novela**
NPM : **2423022004**
Program Studi : **Magister Pendidikan Fisika**
Jurusan : **Pendidikan MIPA**
Fakultas : **Keguruan dan Ilmu Pendidikan**

MENYETUJUI

1. Komisi Pembimbing



Dr. Viyanti, M.Pd.
NIP 19800330 200501 2 001



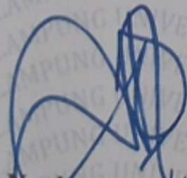
Dr. Fatkhur Rohman, M.Pd.
NIP 19910716 20242 1 001

2. Mengetahui Ketua Jurusan Pendidikan MIPA

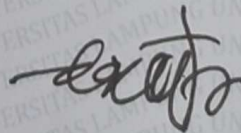
Ketua Jurusan Pendidikan MIPA

Ketua Program Studi

Magister Pendidikan Fisika



Dr. Nurchanurawati, M.Pd.
NIP 19670808 199103 2 001



Prof. Dr. I Wayan Distrik, M.Si
NIP 19631215 199102 1 001

MENGESAHKAN

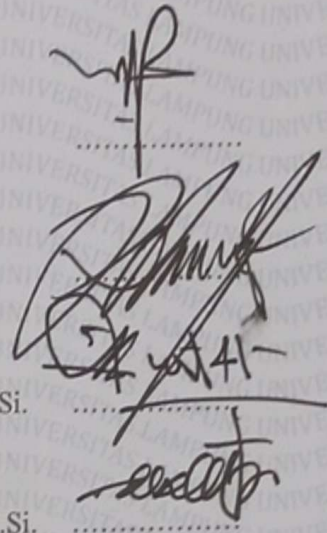
1. Tim Penguji

Ketua : Dr. Viyanti, M.Pd.

Sekretaris : Dr. Fatkhur Rohman, M.Pd.

Penguji Anggota : Prof. Dr. Kartini Herlina, M.Si.

Prof. Dr. I Wayan Distrik, M.Si.



2. Dekan Fakultas Keguruan dan Ilmu Pendidikan



Dr. Albet Maydiantoro, M.Pd.

NIP 19870504 201404 1 001

3. Direktur Program Pascasarjana



Prof. Dr. Ir. Murhadi, M.Si.

NIP 19640326 198902 1 001

Tanggal Lulus Ujian Tesis : 13 Februari 2026

SURAT PERNYATAAN

Yang bertanda tangan dibawah ini adalah :

Nama : Erma Novela
NPM : 2423022004
Fakultas/Jurusan : KIP/Pendidikan MIPA
Program Studi : Magister Pendidikan Fisika
Alamat : PMK. II Desa Bakhu, Kecamatan Batu Ketulis,
Kabupaten Lampung Barat

Dengan ini menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar magister disuatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebut dalam daftar pustaka.

Bandar Lampung, 23 Februari 2026



Erma Novela
2423022004

RIWAYAT HIDUP

Nama lengkap peneliti adalah Erma Novela, dilahirkan di Bakhu pada tanggal 22 Februari 2002. Peneliti adalah putri dari pasangan Bapak Pauzan dan Ibu Erlina.

Peneliti mengawali pendidikan formal pada tahun 2007 sebagai siswi di TK Darma Wanita dan lulus pada tahun 2008. Peneliti melanjutkan pendidikan di SD Negeri 1 Bakhu dan lulus pada tahun 2014. Kemudian peneliti melanjutkan pendidikan di SMP Negeri 1 Belalau dan lulus pada tahun 2017. Selanjutnya, pada tahun 2017 peneliti melanjutkan pendidikan di SMA Negeri 1 Belalau dan lulus pada tahun 2020.

Pada tahun 2020 peneliti diterima sebagai mahasiswa Pendidikan Fisika FKIP Universitas Bengkulu. Penulis lulus sarjana pada tahun 2023, kemudian melanjutkan studi di Magister Pendidikan Fisika Universitas Lampung pada tahun 2024.

MOTTO

*“Maka apabila engkau telah selesai (dari suatu urusan), tetaplah bekerja keras
(untuk urusan yang lain)”*
(Q.S. Al-Insyirah (94): 7)

*“Allah menghendaki kemudahan bagimu, dan tidak menghendaki kesukaran
bagimu.”*
(Q.S. Al Baqarah (2) : 185)

“Berjuang, bertahan, lalu tumbuh.”
- Erma Novela

PERSEMBAHAN

Puji syukur kehadiran Allah SWT yang selalu melimpahkan rahmat dan hidayah-Nya. sehingga penulis mampu menyelesaikan tesis ini tepat pada waktunya. Dengan tulus dan penuh rasa syukur, penulis mempersembahkan karya tulis ini kepada :

1. Kedua orang tua, Ayah Pauzan dan Ibu Erlina sebagai wujud terimakasih dan baktiku atas doa, pengorbanan, kasih sayang, dan cinta yang diberikan.
2. Kakak tersayang, Cerli Anjarsari
3. Sahabat-sahabat tercinta
4. Keluarga besar Magister Pendidikan Fisika
5. Almamater tercinta Universitas Lampung

SANWACANA

Puji syukur kehadiran Allah SWT yang selalu melimpahkan rahmat dan hidayah-Nya, sehingga penulis mampu menyelesaikan tesis ini sebagai salah satu syarat memperoleh gelar Sarjana Pendidikan di FKIP Universitas Lampung. Pada kesempatan ini penulis mengucapkan terimakasih kepada :

1. Prof. Dr. Ir. Lusmeilia Afriani, D.E.A.IPM., selaku Rektor Universitas Lampung;
2. Prof. Dr. Ir. Murhadi, M.Si., selaku Direktur Pascasarjana Universitas Lampung;
3. Dr. Albet Maydiantoro, M.Pd., selaku Dekan FKIP Universitas Lampung;
4. Dr. Nurhanurawati, M.Pd., selaku Ketua Jurusan Pendidikan MIPA;
5. Prof. Dr. I Wayan Distrik, M.Si., selaku Ketua Program Studi Magister Pendidikan Fisika Universitas Lampung sekaligus Dosen Pembahas 2 dan Validator;
6. Dr. Viyanti, M.Pd., selaku Dosen Pembimbing Akademik dan Dosen Pembimbing I, atas kesediaan, kesabaran, dan keikhlasannya memberikan arahan, bimbingan, serta motivasi selama menyelesaikan tesis;
7. Dr. Fatkhur Rahman, M.Pd., selaku Dosen Pembimbing II, yang selalu memberikan arahan, bimbingan, dan motivasi selama menyelesaikan tesis;
8. Prof. Dr. Kartini Herlina, M.Si., selaku Dosen Pembahas I sekaligus Validator, atas kesabaran beliau dalam memberikan bimbingan, arahan serta motivasi kepada peneliti selama menyelesaikan tesis;
9. Bapak dan Ibu dosen Magister Pendidikan Fisika Universitas Lampung yang telah memberikan ilmu dan membimbing penulis selama melaksanakan pendidikan di Universitas Lampung;

10. Dr. Rangga Firdaus, M.Kom., dan Deni Anggraini, M.Pd. selaku validator produk. Terimakasih atas kritik dan saran yang terhadap produk yang dikembangkan.
11. Dewan guru serta siswa-siswi SMAN 14 Bandar Lampung, SMAN 1 Belalau, dan SMA Global Madani atas bantuan dan kerjasamanya
12. Lutfi Nur Rohman yang telah membantu dalam proses pengembangan media. Terima kasih atas kontribusi, dukungan, dan kerja sama yang diberikan sehingga media ini dapat terselesaikan dengan baik.
13. Sahabat-sahabat tercinta, Aisyah, Anggen, Arizka, Meliya, Puji, Sela, dan Sonia yang dengan ketulusan telah memberikan kebersamaan, dukungan, doa, dan semangat tanpa henti, menemani penulis dalam berbagai proses hingga tesis ini akhirnya dapat diselesaikan;
14. Rekan-rekan Magister Pendidikan Fisika angkatan 2024 atas kebersamaan dan dukungan selama perkuliahan dan proses penyelesaian tesis ini;
15. Kepada semua pihak yang terlibat dalam membantu penyelesaian penyusunan tesis ini.

Penulis berharap semoga tesis ini dapat bermanfaat dan dapat menjadi tambahan referensi untuk penelitian selanjutnya.

Bandar Lampung, 27 januari 2026

Erma Novela

DAFTAR ISI

Halaman

DAFTAR ISI.....	xiii
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xv
I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian	7
1.4 Manfaat Penelitian	7
1.5 Ruang Lingkup Penelitian.....	8
II. TINJAUAN PUSTAKA	9
2.1 Deskripsi Teori.....	9
2.1.1 Instrumen Penilaian	9
2.1.2 Keterampilan Berpikir Kritis	11
2.1.3 <i>Multimedia</i>	15
2.1.4 <i>Adaptive Assessment</i>	16
2.1.5 Teori Respons Butir (<i>Item Response Theory</i>)	18
2.1.6 Fluida Statis	20
2.2 Penelitian yang Relevan	26
III. METODE PENELITIAN	30
3.1 Desain Penelitian.....	30
3.2 Subjek Penelitian.....	31
3.3 Prosedur Pengembangan	31
3.4 Instrumen Penelitian.....	39
3.5 Teknik Analisis data.....	43
IV. HASIL DAN PEMBAHASAN	51
4.1 Hasil Penelitian	51
4.2 Pembahasan.....	106
V. SIMPULAN DAN SARAN	113
5.1 Kesimpulan	113
5.2 Saran.....	114
DAFTAR PUSTAKA	115

DAFTAR TABEL

Tabel	Halaman
1. Indikator keterampilan berpikir kritis menurut Ennis.....	12
2. Indikator berpikir kritis menurut Facione	13
3. Indikator Berpikir Kritis menurut Angelo	14
4. Penelitian yang relevan	26
5. Kisi-kisi instrumen validasi media <i>Adaptive Assessment</i>	40
6. <i>Skala likert</i> Angket Validitas Media.....	41
7. Kisi-kisi validasi instrumen soal	41
8. <i>Skala likert</i> Angket Validitas Instrumen Penilaian	42
9. Kisi-kisi Angket Respon	42
10. <i>Skala likert</i> Uji Kepraktisan.....	43
11. Interpretasi Uji Validitas	45
12. Kriteria <i>Alpha Cronbach</i>	47
13. Kriteria <i>Person Reliability</i> dan <i>Item Reliability</i>	48
14. Interval Keterampilan Berpikir Kritis Siswa	49
15. Interpretasi Uji Kepraktisan.....	50
16. Sebaran Soal yang dikembangkan	71
17. Hasil Uji Validitas Media	72
18. Perbaikan Berdasarkan Saran Validator	73
19. Jumlah Responden Uji Coba.....	75
20. Interval Kemampuan Keterampilan Berpikir Kritis.....	87
21. Hasil Uji Kepraktisan siswa.....	103
22. Hasil Uji Kepraktisan Guru.....	105

DAFTAR GAMBAR

Gambar	Halaman
1. Model Pengembangan 4D	30
2. Prosedur Pengembangan	32
3. Diagram proses pelaksanaan Adaptive Assessment	36
4. Hasil Analisis Kebutuhan Guru	52
5. Soal yang biasa guru gunakan dalam pembelajaran	54
6. Hasil Analisis Kebutuhan Siswa	56
7. Rancangan Media <i>Adaptive Assessment</i>	59
8. Memasukkan halaman website pada browser	60
9. Tampilan Beranda Utama	61
10. Menu Daftar	61
11. Menu Login.....	62
12. Tampilan Dashboard Guru	62
13. Menu Tambah Soal	63
14. Buat Tes Baru.....	64
15. Rekap Tes.....	65
16. Menu Daftar Siswa.....	66
17. Menu Login Siswa	67
18. Tampilan Halaman Siswa	67
19. Tampilan Kode Akses	68
20. Tampilan Soal pada Media	69
21. Hasil Validitas Empiris menggunakan Winstep	76
22. Hasil <i>Cronbach's Alpha</i> (KR-20) dan <i>Person Reliability</i>	78
23. Hasil <i>Item Reliability</i>	79
24. Hasil Analisis Tingkat Kesulitan Butir	81
25. Analisis peta <i>Wright (Person-Item Map)</i>	83
26. Hasil <i>Test Information Function</i>	84
27. <i>Differential Item Functioning (DIF)</i>	86
28. Pengkategorian Keterampilan Berpikir Kritis.....	89
29. Soal Merumuskan Masalah.....	91
30. Soal Indikator Memberikan Argumen	93
31. Soal Indikator Melakukan Deduktif.....	96
32. Soal indikator melakukan induksi.....	98

33. Soal indikator Melakukan Evaluasi	100
34. Soal Indikator Memutuskan dan Melaksanakan	102
35. Hasil Pengerjaan Siswa 1	109
36. Hasil Pengerjaan Siswa 2	110

I. PENDAHULUAN

1.1 Latar Belakang

Pendidikan di abad ke-21 telah memasuki era baru yang ditandai dengan perubahan signifikan dalam cara berpikir, belajar, dan mengajar (Permana, 2024). Era ini menuntut siswa untuk memiliki keterampilan abad ke-21 yang tidak hanya berfokus pada penguasaan materi, tetapi juga mencakup keterampilan *Critical thinking, Creativity, Collaboration, Communication, Culture/Citizenship*, dan *Character education/connectivity* (Shabrina & Astuti, 2020). Keterampilan abad ke-21 merupakan kompetensi penting yang harus dimiliki setiap individu agar mampu menghadapi tantangan kehidupan, pekerjaan, dan perkembangan teknologi (Redhana, 2020).

Keterampilan berpikir kritis telah diakui sebagai elemen esensial yang berperan penting dalam mendukung keberhasilan individu dalam belajar, bekerja, dan menjalani kehidupan pada abad ke-21 (Roudlo, 2020). *World Economic Forum* tahun 2020 menyoroti bahwa kemampuan berpikir kritis dan keterampilan pemecahan masalah termasuk di antara sepuluh kompetensi utama yang harus dimiliki individu di masa depan (Saputra, 2024). Pentingnya keterampilan berpikir kritis juga tercermin dari hasil evaluasi pendidikan di Indonesia, yaitu Indonesia mengalami penurunan kualitas pendidikan berdasarkan hasil survei *PISA (Programme for International Student Assessment)* tahun 2018 (Pamungkas *et al.*, 2018). Salah satu komponen PISA adalah keterampilan berpikir kritis (Vincent-Lancrin *et al.*, 2019). Nilai rata-rata sains siswa Indonesia masih di bawah rata-rata negara-negara *OECD (Organisation for Economic Co-operation and Development)*, yang mengindikasikan bahwa keterampilan berpikir kritis siswa Indonesia masih rendah (Kemendikbud, 2019). Peringkat Indonesia dalam PISA 2022 mengalami kenaikan, namun skor rata-rata siswa justru menurun pada

semua subjek yang dinilai, seperti membaca, matematika, dan sains sehingga peningkatan peringkat tersebut tidak sepenuhnya mencerminkan perbaikan kualitas pembelajaran (OECD, 2023). Oleh karena itu, perlu dilakukan upaya peningkatan mutu pendidikan dan pengembangan keterampilan berpikir kritis (Yanto Zebua *et al.*, 2024).

Keterampilan berpikir kritis merupakan proses berpikir reflektif yang logis dan berorientasi pada pengambilan keputusan yang harus dipercayai dan dilakukan (Astutik *et al.*, 2020). Salah satu kemampuan utama dari pemikir kritis adalah keterampilan menilai kredibilitas sumber informasi, yang di antaranya ditunjukkan melalui kemampuan memberikan alasan yang kuat (Ennis, 2011). Kemampuan ini berkaitan erat dengan dimensi “*assessment*” dalam keterampilan berpikir kritis sebagaimana dijelaskan oleh Perkins & Murphy (2006). Indikator keterampilan berpikir kritis yang dikemukakan oleh Ennis (2009), yakni merumuskan masalah, memberikan argumen, melakukan deduksi, melakukan induksi, melakukan evaluasi, memutuskan dan melaksanakan. Keterampilan berpikir kritis telah menjadi salah satu fokus utama dalam pembelajaran fisika (Ertikanto *et al.*, 2023). Salah satu kompetensi dalam pembelajaran fisika sebagaimana tercantum dalam Permendikbud Nomor 64 Tahun 2013 adalah proses pembelajaran fisika maupun asesmen perlu diarahkan untuk menumbuhkan keterampilan berpikir kritis (Asniar *et al.*, 2022).

Materi dalam pembelajaran fisika SMA yang memiliki potensi tinggi untuk menumbuhkan keterampilan berpikir kritis adalah fluida statis (Puspita & Jatmiko, 2013). Materi ini merupakan bagian dari capaian pembelajaran Fisika Fase F dan mencakup konsep-konsep seperti tekanan, hukum Pascal, hukum Archimedes, dan kapilaritas, yang erat kaitannya dengan fenomena dalam kehidupan sehari-hari (Ringo *et al.*, 2023). Konsep fluida statis tidak mudah dipelajari oleh siswa, kesulitan yang dialami siswa dalam mempelajari konsep fluida statis dapat ditemukan pada beberapa subtopik fluida statis (Rosdiana & Kusairi, 2019). Beberapa contoh kesulitan-kesulitan tersebut diantaranya adalah terdapat beberapa siswa yang menganggap bahwa tekanan hidrostatik dipengaruhi oleh volume zat cair atau bentuk wadah, bukan dipengaruhi oleh kedalaman atau

massa jenis zat cair (Berek *et al.*, 2016). Penelitian lain menunjukkan bahwa ketika siswa menyelesaikan masalah tentang aplikasi hukum Archimedes, pada pokok bahasan terapung, melayang dan tenggelam, siswa lebih fokus pada konsep berat sebagai faktor yang memengaruhi keadaan benda ketika berada di dalam zat cair dan mengabaikan faktor massa jenis yang dimiliki oleh zat cair (Petrosino & Mann, 2017). Konsep fluida statis tidak mudah dipahami oleh siswa karena menuntut kemampuan analisis yang tinggi dan keterkaitan logis antarkonsep (Harizah *et al.*, 2019). Rendahnya kemampuan berpikir kritis siswa menjadi salah satu faktor yang menyebabkan kesulitan dalam memahami materi ini (Rahmatin *et al.*, 2024). Pengembangan keterampilan berpikir kritis secara optimal penting bagi siswa agar dapat memahami berbagai konsep pembelajaran dengan lebih baik (Rosdiana & Kusairi, 2019).

Keterampilan berpikir kritis sering disebut sebagai elemen terpenting dalam pembelajaran fisika, tetapi pada kenyataannya penilaian terhadap keterampilan berpikir kritis sering kali diabaikan (Mundilarto & Ismoyo, 2017). Alasan utama mengapa penilaian keterampilan berpikir kritis dalam pembelajaran fisika sering terabaikan adalah karena kurangnya ketersediaan tes yang secara khusus mengukur keterampilan berpikir kritis dalam konteks fisika (Abidin *et al.*, 2019). Permasalahan kedua berkaitan dengan ketidaktepatan format tes dalam menilai keterampilan berpikir kritis dan mencakup keseluruhan materi yang diajarkan (Shaaban & Ghaith, 2014). Esay dianggap sebagai bentuk tes yang paling komprehensif, format ini memiliki sejumlah kelemahan, seperti keterbatasan dalam mengukur kemampuan berpikir tingkat tinggi, boros waktu dan biaya, tingkat subjektivitas yang tinggi, serta sulit untuk ditentukan validitas dan reliabilitasnya (Khan, 2015). Tes pilihan ganda memiliki sejumlah keunggulan, seperti mudah diterapkan di kelas dengan jumlah siswa besar, bersifat objektif, mampu mencakup materi yang luas, dan mudah dalam proses penilaiannya (Istiyono *et al.*, 2014). Pilihan ganda dinilai kurang mendalam dan tidak memungkinkan untuk mengamati proses berpikir siswa secara jelas (Winarti *et al.*, 2017). Salah satu alternatif modifikasi dari tes pilihan ganda adalah pilihan ganda beralasan (*reasoning multiple-choice*) (Istiyono *et al.*, 2014). Tes ini terdiri

dari pilihan jawaban dan alasan yang mendukungnya, yang menuntut siswa untuk menentukan alasan yang mendukung pilihan jawabannya (Fajrianthi *et al.*, 2016).

Permasalahan selanjutnya, penggunaan teori tes klasik masih cukup banyak dijumpai di Indonesia dalam analisis tes dan estimasi kemampuan peserta (Fajrianthi *et al.*, 2016). Teori ini memiliki sejumlah kelemahan, salah satunya adalah karakteristik butir soal yang sangat bergantung pada peserta tes (Mukti & Istiyono, 2018). Suatu soal, misalnya, akan dianggap mudah jika dikerjakan oleh peserta yang pandai, tetapi akan terasa sulit bagi peserta dengan kemampuan rendah (Bichi & Talib, 2018). Skor yang mencerminkan kemampuan peserta pun menjadi bergantung pada tes yang digunakan, sehingga tingkat kemampuan yang sebenarnya tidak dapat diketahui secara akurat (Awopeju & Afolabi, 2016).

Penerapan teori respons butir (*Item Response Theory/IRT*), khususnya model *Rasch (IPL)*, menjadi alternatif yang tepat karena memungkinkan pengukuran kemampuan peserta secara lebih adil serta estimasi kesukaran butir yang objektif dan independen dari karakteristik peserta tes (Hambleton *et al.*, 2012). Penelitian oleh Lestari *et al.*, (2023) menunjukkan bahwa instrumen *two-tier test* berbasis berpikir kritis yang dikembangkan valid dan reliabel, serta sesuai dengan model Rasch dan IRT, sehingga efektif digunakan untuk mengukur keterampilan berpikir kritis siswa SMA secara lebih akurat.

Permasalahan keempat adalah penggunaan tes yang tidak memperhatikan perbedaan kemampuan siswa karena menganggap setiap siswa memiliki standar yang sama (Abidin *et al.*, 2019). Pendekatan ini menjadi kurang efektif saat digunakan untuk mengukur kemampuan peserta dengan tingkat sangat rendah maupun sangat tinggi (Ling *et al.*, 2017). Penyebabnya adalah banyaknya butir soal yang tidak mampu membedakan variasi kemampuan peserta tes (Aybek & Demirtasli, 2017). Salah satu solusi yang tepat adalah dengan menerapkan *Adaptive Assessment* (Abidin *et al.*, 2019). *Adaptive Assessment* dirancang agar lebih efisien dalam mengukur kemampuan siswa (Khotimah & Mindyarto, 2021).

Adaptive Assessment mampu menyesuaikan tingkat kesulitan soal secara dinamis sesuai dengan kompetensi siswa, tetapi juga memberikan gambaran yang lebih

akurat mengenai kemampuan dan proses belajar mereka secara individual (Davey, 2011). *Adaptive Assessment* menyesuaikan tingkat kesulitan soal dengan kemampuan siswa, jika siswa menyelesaikan soal dengan baik, tingkat kesulitan soal berikutnya meningkat, sedangkan jika mengalami kesulitan, sistem menyesuaikannya ke tingkat yang lebih sesuai (Ross *et al.*, 2018). *Adaptive Assessment* berbasis komputer dikenal dengan istilah *Computerized Adaptive Testing* (CAT), yang menawarkan efisiensi lebih baik karena dapat menghindari soal yang terlalu mudah atau terlalu sulit, sehingga durasi tes dapat dipersingkat tanpa mengorbankan akurasi pengukuran (Rezaie & Golshan, 2015) .

Analisis kebutuhan dilakukan dengan menyebar instrumen analisis kebutuhan dengan menggunakan *Google Form*. Berdasarkan hasil analisis kebutuhan terhadap 24 guru fisika dan 37 siswa di Provinsi Lampung diperoleh hasil bahwa jenis soal yang digunakan guru adalah berbentuk soal esay dan pilihan ganda, Hasil data menunjukkan bahwa lebih dari 50% guru belum melaksanakan penilaian keterampilan berpikir kritis siswa berbasis indikator, sehingga pelaksanaan asesmen belum dilakukan secara sistematis dalam mengukur capaian setiap indikator berpikir kritis. Meskipun guru menyatakan mengetahui konsep berpikir kritis, dalam praktiknya mereka masih cenderung menggunakan taksonomi Bloom pada level rendah (C1–C3), yaitu mengingat, memahami, dan menerapkan, sehingga indikator berpikir kritis tingkat tinggi (C4–C6) jarang muncul dalam soal yang diberikan. Sebanyak 66,7% guru menyatakan bahwa perlu untuk melakukan analisis terhadap kualitas soal, namun guru belum melakukan analisis terhadap kualitas soal dan 62,5% guru masih menggunakan soal yang sama untuk semua siswa. Selain itu, guru belum pernah menggunakan asesmen adaptif, namun sebanyak 66,7% sudah mengetahui tentang asesmen adaptif, dan 66,7% guru sudah cukup familiar dengan platform digital untuk asesmen dan sudah pernah menggunakannya. Hal ini menunjukkan potensi guru untuk mengintegrasikan teknologi dalam pembelajaran, sejalan dengan tuntutan abad ke-21, di mana guru dituntut tidak hanya menguasai konten, tetapi juga mampu memanfaatkan teknologi untuk mendukung pembelajaran dan memonitor kemajuan siswa (Varas *et al.*, 2023; Zou *et al.*, 2025)

Sementara itu, sebagian besar siswa menyatakan belum pernah mengikuti asesmen adaptif, namun siswa menunjukkan ketertarikan yang tinggi yaitu 86.5% siswa percaya bahwa asesmen yang menyesuaikan soal dengan kemampuan mereka akan sangat membantu proses belajar. Temuan ini menunjukkan bahwa baik guru maupun siswa membutuhkan asesmen yang tidak hanya relevan dengan kemampuan berpikir kritis, tetapi juga responsif terhadap kebutuhan belajar individu. Penelitian sebelumnya telah menunjukkan potensi *Adaptive Assessment* dalam mengukur kemampuan berpikir kritis secara efisien dan akurat yaitu penelitian Abidin *et al.* (2019) dan Istiyono *et al.* (2019) mengembangkan *Computerized Adaptive Test* (CAT) berbasis *Item Response Theory* (IRT) untuk keterampilan berpikir kritis fisika. Tes tersebut terbukti valid, handal, serta efektif untuk digunakan.

Uraian diatas menunjukkan perlunya instrumen penilaian yang valid, reliabel, praktis, dan efektif untuk mengukur keterampilan berpikir kritis siswa secara lebih akurat, terutama pada materi fluida statis. Instrumen ini dirancang dengan pendekatan *Adaptive Assessment* agar mampu menyesuaikan tingkat kesulitan soal secara dinamis sesuai kemampuan siswa yang menekankan pemahaman konseptual dan pengalaman belajar reflektif. Penelitian ini mengangkat judul "Pengembangan Instrumen Penilaian Berbasis *Adaptive Assessment* untuk Mengukur Keterampilan Berpikir Kritis pada Materi Fluida Statis" sebagai respon terhadap kebutuhan tersebut.

1.2 Rumusan Masalah

Berdasarkan uraian diatas, rumusan masalah dalam penelitian ini adalah

1. Bagaimana instrumen penilaian berbasis *Adaptive Assessment* yang valid untuk mengukur keterampilan berpikir kritis pada materi fluida statis ?
2. Bagaimana instrumen penilaian berbasis *Adaptive Assessment* yang valid dan reliabel secara empiris untuk mengukur keterampilan berpikir kritis pada materi fluida statis ?

3. Bagaimana instrumen penilaian berbasis *Adaptive Assessment* yang praktis mengukur keterampilan berpikir kritis pada materi fluida statis ?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dalam penelitian ini adalah

1. Mendeskripsikan instrumen penilaian berbasis *Adaptive Assessment* yang valid untuk mengukur keterampilan berpikir kritis pada materi fluida statis.
2. Mendeskripsikan instrumen penilaian berbasis *Adaptive Assessment* yang valid dan reliabel secara empiris untuk mengukur keterampilan berpikir kritis pada materi fluida statis
3. Mendeskripsikan instrumen penilaian berbasis *Adaptive Assessment* yang praktis untuk mengukur keterampilan berpikir kritis pada materi fluida statis.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberi manfaat bagi peningkatan kualitas dan pengembangan pembelajaran fisika materi fluida baik bagi guru/pendidik, siswa, sekolah, maupun peneliti, antara lain yaitu:

1. Bagi Peneliti
Mendorong pengembangan instrumen penilaian berbasis *Adaptive Assessment* untuk mengukur keterampilan berpikir kritis pada materi fluida statis
2. Bagi siswa
Membantu siswa untuk mengukur keterampilan berpikir kritis pada pembelajaran fisika khususnya materi fluida statis.
3. Bagi guru
Membantu guru dalam mengukur keterampilan berpikir kritis pada pembelajaran fisika khususnya pada materi fluida statis.
4. Bagi dunia Pendidikan

Memberikan sumbangan pemikiran dalam upaya mengukur keterampilan berpikir kritis siswa pada pembelajaran fisika.

1.5 Ruang Lingkup Penelitian

Ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini adalah pengembangan instrumen penilaian untuk mengukur keterampilan berpikir kritis siswa. Instrumen soal yang dikembangkan adalah soal pilihan ganda beralasan.
2. *Adaptive Assessment* dalam penelitian ini dibatasi pada penerapan penilaian adaptif berbasis *Item Response Theory (IRT)* model 1PL (*Rasch*), di mana pemilihan dan penyajian butir soal disesuaikan dengan estimasi kemampuan siswa berdasarkan respon terhadap butir sebelumnya.
3. Keterampilan yang diukur adalah keterampilan berpikir kritis berdasarkan indikator Ennis (2009), meliputi kemampuan merumuskan masalah, memberikan argumen, melakukan deduksi, melakukan induksi, melakukan evaluasi, serta memutuskan dan melaksanakan.
4. Validitas instrumen soal mengacu pada sejauh mana butir soal mampu mengukur keterampilan berpikir kritis sesuai konstruk yang ditetapkan, yang ditinjau melalui validitas isi, konstruk, dan bahasa berdasarkan penilaian ahli, sedangkan kevalidan media *Adaptive Assessment* mengacu pada kesesuaian media dengan tujuan penilaian yang ditinjau dari aspek isi, tampilan, kesesuaian dengan konsep penilaian adaptif, serta keterpakaian media dalam pembelajaran.
5. Validitas empiris mengacu pada tingkat kesesuaian butir soal dengan model pengukuran, sedangkan reliabilitas mengacu pada konsistensi hasil pengukuran instrumen ketika digunakan pada subjek yang memiliki karakteristik serupa.
6. Kepraktisan mengacu pada sejauh mana instrumen dan media *Adaptive Assessment* mudah digunakan, dipahami, dan dilaksanakan dalam proses pembelajaran..

II. TINJAUAN PUSTAKA

2.1 Deskripsi Teori

2.1.1 Instrumen Penilaian

Penilaian merupakan proses sistematis untuk mengumpulkan, menganalisis, dan menginterpretasikan informasi mengenai capaian belajar siswa dengan tujuan untuk mendukung pengambilan keputusan dalam pembelajaran (Rosidin, 2017). Penilaian sebagai proses pengumpulan dan pengolahan informasi untuk mengetahui kebutuhan belajar, capaian perkembangan, dan hasil belajar siswa (Kemdikbudristek, 2022). Penilaian didasarkan pada hasil pengukuran yang dilakukan menggunakan instrumen atau alat ukur tertentu (Inanna *et al.*, 2021). Instrumen penilaian adalah alat yang dirancang untuk memperoleh data terkait kemampuan siswa, baik dalam ranah kognitif, afektif, maupun psikomotorik (Syamsinar *et al.*, 2022). Instrumen harus dianalisis dan dikembangkan dengan memperhatikan validitas, reliabilitas, dan kepraktisan (Badriyah *et al.*, 2018). Menurut Arifin (2017) instrumen penilaian yang baik harus memenuhi beberapa prinsip sebagai berikut:

- a. Spesifik, maknanya instrumen yang dikembangkan didesain secara khusus untuk objek yang dievaluasi.
- b. Relevan, maknanya instrumen harus dikembangkan sesuai kompetensi dasar dan indikator yang telah dirumuskan.
- c. Representatif, maknanya setiap materi dalam instrumen harus mewakili semua materi dalam indikator.
- d. Valid, maknanya instrumen tersebut benar-benar dapat mengukur yang seharusnya diukur.
- e. Reliabel, maknanya instrumen tersebut harus memiliki hasil yang konsisten.

- f. Diskriminatif, maknanya instrumen harus dapat menunjukkan perbedaan antara kelas atas dan kelas bawah.
- g. Proporsional, maknanya tingkat kesulitan instrumen harus bervariasi antara mudah, sedang, dan sulit.
- h. Praktis, maknanya instrumen yang dikembangkan memudahkan penggunaan baik secara teknis maupun administratif.

Instrumen terdiri atas instrumen tes dan instrumen non-tes (Fauziyah *et al.*, 2023). Instrumen tes terdiri dari tes subjektif dan tes objektif, sedangkan instrumen non tes terdiri dari penilaian unjuk kerja, penilaian sikap, penilaian proyek, penilaian produk dan penilaian diri (Mahendra *et al.*, 2024). Tes subjektif atau tes esai adalah jenis tes yang mengharuskan siswa merangkai kata berdasarkan pengetahuan yang dimiliki untuk menjawab soal dalam bentuk uraian, sehingga memungkinkan siswa menjelaskan secara bebas sesuai pemikirannya dan memberikan informasi tentang sejauh mana pengetahuan yang mereka kuasai, karena setiap jawaban dapat berbeda tergantung pada pemahaman masing-masing (Magdalena *et al.*, 2023). Bentuk tes esai memiliki kelemahan dalam proses penilaian, yaitu adanya kemungkinan munculnya subjektivitas saat memeriksa hasil jawaban siswa (Mukti & Istiyono, 2018).

Tes objektif adalah salah satu instrumen tes yang menuntut siswa memilih jawaban dari opsi yang tersedia atau memberikan jawaban singkat, dengan pemeriksaan yang dilakukan secara seragam untuk semua siswa, serta mencakup bentuk soal seperti pilihan ganda, benar-salah, menjodohkan, dan isian singkat (Khaerudin, 2016). Tes objektif bentuk *multiple choice item* sering dikenal dengan istilah tes objektif bentuk pilihan ganda, yaitu salah satu bentuk tes obyektif yang terdiri atas pertanyaan atau pernyataan yang sifatnya belum selesai, dan untuk menyelesaikannya harus dipilih salah satu atau lebih dari beberapa kemungkinan jawaban yang telah disediakan pada tiap-tiap butir soal yang bersangkutan (Sudijono, 2009). Pilihan ganda merupakan bentuk soal yang terdiri atas pernyataan atau informasi yang belum lengkap dan harus dilengkapi dengan memilih satu jawaban paling tepat dari beberapa pilihan (*options*) yang

disediakan, di mana hanya satu merupakan kunci jawaban yang benar, sementara sisanya berfungsi sebagai pengecoh (*distractor*) (Mulyadi, 2010).

Tes pilihan ganda memiliki keunggulan dibandingkan jenis tes lainnya karena efisien dan mudah dilaksanakan, mampu mengukur berbagai aspek kemampuan kognitif dalam waktu yang relatif singkat, serta memungkinkan proses koreksi yang lebih sederhana dan objektif (Septy *et al.*, 2024). Kekurangan dari tes pilihan ganda yaitu superfisialitas yaitu tidak selalu mengukur pemahaman mendalam atau kemampuan berpikir kritis (Mukti & Istiyono, 2018). Pilihan ganda yang dapat digunakan untuk mengukur keterampilan berpikir kritis siswa adalah dengan memberikan pilihan ganda beralasan (Furqaan *et al.*, 2023). Pilihan ganda beralasan dikembangkan untuk menganalisis pemahaman dan kesulitan belajar siswa, karena bentuk ini disusun guna mendeteksi kemungkinan penyebab kesulitan tersebut, dengan menambahkan alasan sebagai alternatif jawaban guna mengatasi kelemahan pada pilihan ganda biasa, khususnya kecenderungan siswa menebak jawaban (Wardani *et al.*, 2013). Bentuk soal ini menuntut para peserta didik menentukan jawaban pada tes serta menambahkan argumentasi terkait pemilihan jawabannya (Samaduri, 2022).

Berdasarkan uraian tersebut, maka peneliti akan mengembangkan instrumen penilaian berupa tes pilihan ganda beralasan yang ditujukan untuk mengukur keterampilan berpikir kritis siswa. Instrumen ini disusun dengan memperhatikan prinsip validitas, reliabilitas, dan kepraktisan. Bentuk pilihan ganda beralasan dipilih karena dapat mengungkap alasan di balik pilihan siswa, sehingga memberikan informasi lebih mendalam tentang proses berpikir mereka serta mengurangi kecenderungan menebak. Dengan demikian, instrumen ini diharapkan mampu memberikan gambaran yang lebih akurat dan komprehensif terhadap kemampuan berpikir kritis siswa.

2.1.2 Keterampilan Berpikir Kritis

Berpikir merupakan suatu proses aktif dalam mengelola dan mengubah informasi yang tersimpan dalam memori untuk membentuk konsep, melakukan penalaran, serta menyelesaikan masalah (Nadlifah *et al.*, 2022). Berpikir kritis adalah satu

keterampilan berpikir yang penting untuk dikembangkan dalam pembelajaran sains (Supena *et al.*, 2021). Keterampilan ini termasuk dalam kategori keterampilan abad ke-21 yang sangat dibutuhkan oleh siswa guna menghadapi dan menyelesaikan persoalan dalam kehidupan sehari-hari (Jannah & Pahlevi, 2020). Menurut Arini & Juliadi, (2018) berpikir kritis adalah kemampuan seseorang dalam menarik kesimpulan secara tepat berdasarkan penalaran yang sistematis dan logis, sehingga mampu menghasilkan berbagai alternatif solusi dari pemikirannya. Sementara itu, menurut Ruggiero (2012) berpikir kritis melibatkan aktivitas menilai yang didefinisikan sebagai proses mengevaluasi klaim atau argumen untuk menilai mana yang memiliki nilai atau kebenaran dan mana yang tidak.

Siswa yang menguasai keterampilan berpikir kritis memiliki kemampuan untuk mencari, memahami, dan menilai informasi atau pernyataan yang relevan dengan cara yang logis dan rasional ketika menghadapi situasi pemecahan masalah atau saat membuat keputusan. Selain itu, mereka juga menunjukkan kecenderungan berpikir secara eksploratif dan reflektif selama proses pembelajaran (Shaw *et al.*, 2020).

Terdapat beberapa indikator berpikir kritis menurut para ahli, salah satunya adalah indikator berpikir kritis dari Ennis (2009), sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Indikator keterampilan berpikir kritis menurut Ennis

Keterampilan Berpikir Kritis	Indikator
Merumuskan Masalah (<i>Formulating Problems</i>)	Merumuskan pertanyaan yang memberikan arah untuk memperoleh jawaban
Memberikan Argumen (<i>Giving Arguments</i>)	Menyampaikan argumen dengan alasan yang tepat Menunjukkan persamaan dan perbedaan Menyusun argumen secara utuh
Melakukan Deduksi (<i>Deducting</i>)	Melakukan deduksi secara logis Menggunakan kondisi logis Melakukan interpretasi terhadap pertanyaan
Melakukan Induksi (<i>Inducting</i>)	Melakukan investigasi atau pengumpulan data

Keterampilan Berpikir Kritis	Indikator
Melakukan Evaluasi (Evaluating)	Membuat generalisasi dari data, menyusun tabel, dan grafik
	Menarik kesimpulan yang berkaitan dengan hipotesis
	Memberikan penilaian berdasarkan fakta
	Berdasarkan prinsip atau aturan
Memutuskan dan Melaksanakan (Deciding and Implementing)	Memberikan alternatif
	Memilih kemungkinan solusi yang tersedia
	Menentukan kemungkinan solusi yang akan dilaksanakan

Tabel 1 menyajikan berbagai indikator berpikir kritis beserta sub-indikator yang mendukung pengembangan keterampilan berpikir kritis siswa. Setiap indikator memiliki sub-indikator yang membantu mengarahkan proses berpikir kritis secara lebih terstruktur dan terperinci.

Indikator keterampilan berpikir kritis menurut Facione (1990) terdiri dari enam indikator, yaitu seperti ditunjukkan pada Tabel 2.

Tabel 2. Indikator berpikir kritis menurut Facione

No.	Indikator	Sub-Indikator
1.	<i>Interpretation</i>	Mengelompokkan atau mengkategorikan Mendefinisikan Menyatakan makna secara jelas
2.	<i>Analysis</i>	Menguji ide Mengenali argumen Menganalisis argument
3.	<i>Evaluation</i>	Menilai kredibilitas pertanyaan Menilai kredibilitas argument atau opini berdasarkan pertimbangan induktif dan deduktif Menentukan hasil dari suatu pertimbangan
4.	<i>Inference</i>	Menilai kualitas pertanyaan Menyimpulkan alternative
5.	<i>Explanation</i>	Menyatakan hasil Menjelaskan prosedur Menyampaikan argument secara utuh
6.	<i>Self Regulation</i>	Memantau diri sendiri Melakukan perbaikan

Enam keterampilan berpikir kritis pada Tabel 2 dapat diukur dengan menggunakan instrumen yang dikembangkan melalui sub-indikator dan pertanyaan yang mengacu pada keterampilan berpikir kritis. Berpikir kritis merupakan kemampuan yang mencakup beberapa aspek penting, seperti menganalisis, mensintesis, mengenali dan memecahkan masalah, menyimpulkan, serta mengevaluasi. Indikator berpikir kritis juga ada yang merujuk pada pendapat Angelo (1995) yang terdiri atas lima indikator tersebut disajikan pada Tabel 3.

Tabel 3. Indikator Berpikir Kritis menurut Angelo

Indikator Berpikir Kritis	Penjelasan
<i>Analyzing</i> (Menganalisis)	Kemampuan untuk menguraikan struktur menjadi komponen-komponen agar dapat memahami organisasi atau keterkaitannya. Bertujuan memahami konsep dan mengklasifikasikan ke dalam bagian yang lebih kecil dan rinci.
<i>Synthesizing</i> (Mensintesis)	Kemampuan yang berlawanan dengan menganalisis; yaitu menggabungkan bagian-bagian terpisah menjadi suatu struktur atau kesatuan baru.
<i>Recognizing and Solving Problem</i> (Mengenali dan Memecahkan Masalah)	Kemampuan aplikatif terhadap berbagai definisi; menuntut pemahaman terhadap ide pokok dari suatu teks bacaan hingga mampu merancang suatu konsep.
<i>Concluding</i> (Menyimpulkan)	Aktivitas penalaran berdasarkan definisi atau pengetahuan yang telah diperoleh untuk membentuk definisi baru atau menyimpulkan suatu pemikiran.
<i>Evaluating or Judging</i> (Mengevaluasi/Menilai)	Kemampuan untuk berpikir secara kritis dalam menentukan nilai dari sesuatu berdasarkan berbagai kriteria yang ada.

Indikator ketrampilan berpikir kritis yang digunakan dalam penelitian ini adalah indikator ketrampilan berpikir kritis menurut Ennis (2009). Indikator yang dikembangkan oleh Ennis mencakup dimensi keterampilan berpikir kritis secara komprehensif, yang sangat sesuai untuk mengidentifikasi proses berpikir siswa. Selain itu, indikator Ennis cukup operasional dan fleksibel untuk diintegrasikan ke

dalam model *Adaptive Assessment*, karena setiap indikator dapat dikembangkan menjadi butir soal yang bertingkat kesulitannya. Dengan demikian, pemilihan indikator ini mendukung pengembangan instrumen penilaian yang tidak hanya adaptif, tetapi juga valid dalam mengukur keterampilan berpikir kritis secara menyeluruh.

2.1.3 Multimedia

Multimedia interaktif merupakan media pembelajaran yang memadukan berbagai unsur seperti teks, gambar, audio, animasi, dan video dalam satu sistem digital untuk mendukung proses belajar yang lebih efektif (Mayer, 2024). Penggunaan multimedia interaktif dalam pembelajaran terbukti mampu meningkatkan aktivitas belajar dan keterlibatan siswa dibandingkan metode konvensional (Clark & Mayer, 2016). Multimedia tidak hanya berfungsi sebagai penyaji informasi visual, tetapi juga sebagai sarana interaksi yang memungkinkan siswa berpartisipasi secara aktif melalui navigasi dan umpan balik sistem (Mayer, 2009).

Penelitian menunjukkan bahwa penggunaan multimedia interaktif memberikan pengaruh signifikan terhadap hasil belajar kognitif siswa karena penyajian materi menjadi lebih menarik dan mudah dipahami (Rahmawati & Dewi, 2019). Integrasi multimedia dalam pembelajaran sains juga dapat membantu meningkatkan literasi sains serta kemampuan memahami konsep yang bersifat abstrak, karena kombinasi visual dan interaktivitas mempermudah proses konstruksi pengetahuan siswa (Islamyati & Manuaba, 2021). Selain itu, media berbasis multimedia interaktif mampu meningkatkan motivasi belajar dan respon positif siswa terhadap kegiatan pembelajaran karena memberikan pengalaman belajar yang lebih variatif (Rofiqoh & Khairani, 2024).

Dengan demikian, multimedia interaktif tidak hanya berfungsi sebagai alat bantu visual, tetapi juga sebagai lingkungan belajar digital yang mendukung aktivitas kognitif, interaksi pengguna, serta pengalaman belajar yang lebih adaptif dan personal. Hal ini menjadi relevan dalam pengembangan instrumen penilaian berbasis *Adaptive Assessment* yang memerlukan tampilan digital interaktif untuk memberikan respons dan umpan balik secara dinamis kepada siswa.

2.1.4 *Adaptive Assessment*

Adaptive Assessment menjadi semakin relevan dalam dunia pendidikan karena kemampuannya menyesuaikan tingkat kesulitan soal secara dinamis sesuai dengan kemampuan masing-masing siswa (Ozyurt *et al.*, 2012). Pada pelaksanaannya, soal berikutnya akan diberikan dengan tingkat kesulitan yang lebih tinggi apabila siswa mampu menjawab soal sebelumnya dengan benar. Sebaliknya, apabila

siswa memberikan jawaban yang salah, sistem akan menyesuaikan dengan menyajikan soal yang lebih mudah sehingga tetap sesuai dengan kemampuan siswa (Khotimah & Mindyarto, 2021). Mekanisme tersebut memungkinkan pemilihan butir soal berlangsung secara adaptif selama pelaksanaan tes, sehingga setiap individu menerima butir soal yang sesuai dengan tingkat kemampuannya masing-masing (Bimpeh & West, 2022).

Konsep tes adaptif pertama kali dikembangkan pada awal abad ke-20 oleh Alfred Binet dalam upaya mengukur kecerdasan anak (Burr *et al.*, 2023). Strategi yang diterapkan Binet memiliki beberapa karakteristik utama, yaitu penggunaan satu butir awal sebagai titik masuk tes, pemberian skor secara langsung setelah setiap butir diselesaikan untuk menentukan soal selanjutnya, serta adanya kriteria penghentian tes yang memungkinkan setiap individu menyelesaikan tes dengan jumlah soal yang berbeda (Weiss & Bock, 1983). Karakteristik tes adaptif tersebut kemudian berkembang seiring dengan pemanfaatan teknologi komputer, sehingga pelaksanaannya menjadi lebih sistematis dan efisien.

Adaptive Assessment memungkinkan estimasi kemampuan siswa dilakukan secara berkelanjutan setelah setiap butir soal dijawab. Tingkat kesulitan soal akan berubah mengikuti pola jawaban siswa, sehingga tes tidak bersifat statis dan sama bagi semua individu (Finkelman *et al.*, 2014). Tes yang digunakan termasuk dalam kategori tes kognitif yang mampu memberikan informasi valid mengenai apa yang dapat dilakukan oleh siswa. Kondisi tersebut berpotensi meningkatkan motivasi siswa untuk menunjukkan kemampuan terbaiknya. Hasil tes tidak akan optimal apabila siswa tidak termotivasi dalam mengerjakannya. Oleh karena itu,

Adaptive Assessment diasumsikan mampu mendorong siswa untuk lebih terlibat dan berupaya maksimal selama pelaksanaan tes (Wise, 2014).

Pelaksanaan *Adaptive Assessment* memerlukan tahapan yang terstruktur, mulai dari penyusunan butir soal, penentuan karakteristik butir, hingga penetapan kriteria penghentian tes. Setiap butir soal tidak dapat langsung digunakan, melainkan harus melalui proses kalibrasi terlebih dahulu untuk memperoleh informasi tingkat kesukaran dan karakteristik butir (Zheng, 2016). Proses ini bertujuan agar soal yang disajikan benar-benar sesuai dengan kemampuan siswa dan tidak menimbulkan bias dalam pengukuran (Huang, 2018). Oleh karena itu, pengembangan dan penyelenggaraan *Adaptive Assessment* membutuhkan perhatian khusus serta perencanaan yang matang.

Adaptive Assessment memiliki beberapa keunggulan, antara lain mampu mengukur kemampuan siswa secara lebih efisien dan akurat, menyajikan soal yang sesuai dengan kemampuan individu, serta memberikan hasil penilaian secara otomatis dan objektif. Selain itu, tes adaptif memungkinkan integrasi dengan berbagai bentuk asesmen lain sehingga lebih fleksibel untuk dikembangkan. Keunggulan tersebut menjadikan *Adaptive Assessment* layak digunakan sebagai alternatif penilaian modern dalam pendidikan (Bimpeh & West, 2022).

Meskipun demikian, *Adaptive Assessment* juga memiliki beberapa keterbatasan. Penggunaan komputer dapat menimbulkan kecemasan bagi siswa yang kurang terbiasa dengan teknologi. Selain itu, siswa tidak dapat mengulang atau melewati soal yang telah disajikan, sehingga setiap respons memiliki konsekuensi terhadap jalannya tes. Penyusunan bank soal yang besar dan berkualitas juga membutuhkan waktu serta sumber daya yang tidak sedikit (Ling *et al.*, 2017b). Kendati demikian, keterbatasan tersebut tidak menghilangkan keunggulan *Adaptive Assessment* sebagai sistem penilaian yang inovatif dan relevan dengan kebutuhan pendidikan masa kini.

Penerapan *Adaptive Assessment* terbukti memberikan dampak positif terhadap keterlibatan dan partisipasi siswa dalam pembelajaran (Adeoye *et al.*, 2025). Asesmen yang disesuaikan dengan kebutuhan dan kemampuan individu mampu

meningkatkan keterlibatan aktif siswa serta menjaga motivasi belajar tetap tinggi (El-Sabagh, 2021). Penyajian tugas yang seimbang, tidak terlalu mudah maupun terlalu sulit, membentuk rasa percaya diri dan kepemilikan terhadap proses belajar. Umpan balik yang diberikan secara langsung juga membantu siswa memahami capaian belajarnya dan menyesuaikan strategi belajar secara lebih efektif (Halkiopoulous & Gkintoni, 2024; Zhang & Huang, 2024).

Berdasarkan kajian tersebut, pemilihan *Adaptive Assessment* dalam penelitian ini didasarkan pada kemampuannya untuk mengakomodasi perbedaan kemampuan siswa, serta menghasilkan pengukuran kemampuan yang lebih efisien dan akurat. Pendekatan ini dinilai sesuai untuk digunakan sebagai dasar pengembangan sistem asesmen keterampilan berpikir kritis siswa, khususnya pada materi fisika, sehingga diharapkan mampu memberikan gambaran kemampuan siswa secara lebih objektif dan komprehensif.

2.1.5 Teori Respons Butir (*Item Response Theory*)

Teori Respons Butir (*Item Response Theory* atau IRT) merupakan pendekatan modern dalam pengukuran pendidikan dan psikometri yang digunakan untuk menganalisis hubungan antara kemampuan laten individu dan respons yang diberikan terhadap butir tes. IRT juga dikenal sebagai Teori Ciri Laten (*Latent Trait Theory*) dan dikembangkan sebagai upaya untuk mengatasi keterbatasan teori tes klasik. Pada teori tes klasik, parameter butir seperti tingkat kesulitan, daya pembeda, dan reliabilitas tes bersifat bergantung pada kelompok responden dan karakteristik tes yang digunakan, sehingga hasil pengukuran cenderung kurang stabil ketika diterapkan pada populasi yang berbeda (Naga, 2012).

Kemampuan yang diukur dipandang sebagai variabel laten yang tidak dapat diamati secara langsung, tetapi dapat diestimasi melalui pola respons peserta tes terhadap butir soal dalam kerangka IRT. Variabel laten tersebut disebut sebagai *theta* (θ), yang merepresentasikan tingkat kemampuan individu. Semakin tinggi nilai *theta* yang dimiliki seseorang, maka semakin besar probabilitas individu tersebut untuk memberikan respons yang benar terhadap suatu butir tes. Dengan demikian, respons peserta tes merupakan data yang dapat diamati (*observable*),

sedangkan kemampuan yang mendasarinya bersifat tidak tampak (*unobservable*) (Baker & Kim, 2017).

IRT menggunakan model probabilistik untuk menjelaskan hubungan antara kemampuan laten dan respons terhadap butir tes. Salah satu model IRT yang paling dikenal adalah model Rasch. Model ini mengasumsikan bahwa probabilitas seorang peserta tes menjawab benar suatu butir ditentukan oleh selisih antara kemampuan individu dan tingkat kesulitan butir tersebut. Model Rasch awalnya digunakan untuk data dikotomi, yaitu respons benar dan salah, dan dikenal memiliki keunggulan dalam menghasilkan pengukuran yang bersifat objektif serta invarian terhadap kelompok responden tertentu (Embretson & Reise, 2013).

Seiring dengan perkembangan kebutuhan pengukuran yang lebih kompleks, IRT juga dikembangkan untuk menganalisis data politomus, yaitu data dengan lebih dari dua kategori skor. Model politomus memungkinkan pemberian skor parsial terhadap respons peserta tes, sehingga pengukuran kemampuan menjadi lebih rinci dan akurat. Beberapa model politomus yang banyak digunakan antara lain *Partial Credit Model* (PCM), *Graded Response Model* (GRM), *Generalized Partial Credit Model* (GPCM), dan *Rating Scale Model* (RSM). Model-model tersebut banyak digunakan untuk menganalisis item dengan *skala likert* maupun soal uraian, karena seluruhnya berlandaskan pada asumsi bahwa respons peserta tes dipengaruhi oleh tingkat kemampuan laten yang (Embretson & Reise, 2013).

Penerapan analisis IRT mensyaratkan terpenuhinya beberapa asumsi dasar agar hasil pengukuran valid dan dapat diinterpretasikan secara akurat. Asumsi pertama adalah unidimensi, yaitu tes mengukur satu kemampuan atau atribut dominan. Meskipun dalam praktiknya sulit memperoleh konstruk yang sepenuhnya unidimensi, asumsi ini dianggap terpenuhi apabila terdapat satu dimensi utama yang lebih dominan dibandingkan dimensi lainnya. Pengujian unidimensi umumnya dilakukan melalui analisis faktor dengan melihat nilai eigen pada matriks korelasi antarbutir (Koğar & Koğar, 2018). Asumsi kedua adalah independensi lokal, yang menyatakan bahwa respons peserta terhadap suatu butir tidak dipengaruhi oleh respons terhadap butir lain setelah kemampuan laten

dikendalikan (Naga, 2012). Asumsi ketiga adalah invariansi parameter, yaitu karakteristik butir dan estimasi kemampuan peserta tes relatif stabil ketika diterapkan pada kelompok yang berbeda dengan tingkat kemampuan yang setara (Sudaryono, 2011).

Selain pemenuhan asumsi dasar, analisis IRT juga memerlukan uji kecocokan model (*model fit*) sebelum dilakukan estimasi parameter kemampuan. Uji kecocokan model bertujuan untuk memastikan bahwa model IRT yang digunakan sesuai dengan karakteristik data. Model dikatakan sesuai apabila nilai statistik yang dihasilkan memenuhi kriteria kecocokan yang ditetapkan (Alagoz, 2000; Baker & Kim, 2017). Penggunaan lebih dari satu model IRT dan membandingkan hasil kecocokan model juga disarankan untuk memperoleh model yang paling tepat (Hambleton *et al.*, 1991).

Berdasarkan uraian tersebut, IRT dipilih dalam penelitian ini karena mampu memberikan pengukuran kemampuan yang lebih akurat, objektif, dan stabil dibandingkan teori tes klasik. IRT memungkinkan analisis karakteristik butir dan kemampuan peserta tes secara terpisah, serta menghasilkan estimasi parameter yang relatif bebas dari pengaruh kelompok responden. Selain itu, penggunaan model politomus dalam IRT memungkinkan pengukuran kemampuan berpikir kritis secara lebih mendalam melalui pemberian skor parsial pada respons siswa, sehingga sesuai dengan karakteristik instrumen dan tujuan penelitian yang berfokus pada peningkatan kualitas pengukuran kemampuan berpikir kritis siswa.

2.1.6 Fluida Statis

Materi fluida statis yang dipelajari untuk siswa SMA kelas XI pada fase F membahas mengenai materi tekanan hidrostatik, prinsip Archimedes, hukum Pascal, tegangan permukaan, dan Kapilaritas. Fluida merupakan zat yang dapat mengalir dan mengalami perubahan bentuk secara terus-menerus ketika diberikan gaya geser sekecil apa pun. Berbeda dengan benda padat yang mampu mempertahankan bentuknya, fluida tidak memiliki bentuk tetap dan akan menyesuaikan diri dengan bentuk wadahnya. Berdasarkan karakteristik geraknya, fluida dibedakan menjadi dua jenis, yaitu fluida statis dan fluida dinamis. Fluida

statis adalah fluida yang berada dalam keadaan diam atau tidak mengalami perpindahan makroskopik, sedangkan fluida dinamis adalah fluida yang bergerak atau mengalir dengan kecepatan tertentu.

Pembahasan fluida statis menjadi penting dalam fisika karena banyak fenomena alam dan aplikasi teknologi yang bergantung pada sifat-sifat fluida dalam keadaan diam, seperti tekanan air di bendungan, sistem hidrolik, kapal laut, dan prinsip kerja kapal selam. Oleh karena itu, pemahaman mengenai tekanan, tekanan hidrostatis, hukum hidrostatis, hukum Pascal, serta hukum Archimedes menjadi dasar penting dalam mempelajari fluida statis.

1) Tekanan

Tekanan didefinisikan sebagai gaya yang bekerja tegak lurus pada suatu permukaan per satuan luas permukaan tersebut. Secara matematis, tekanan dirumuskan sebagai

$$P = \frac{F}{A}$$

Keterangan :

P = Tekanan (Pa)
F = Gaya (N)
A = luas permukaan (m²)

Dari persamaan ini dapat diketahui bahwa semakin besar gaya yang diberikan, maka semakin besar tekanan yang dihasilkan, sedangkan semakin besar luas permukaan maka tekanan akan semakin kecil. Tekanan memiliki sifat bekerja ke segala arah dan selalu tegak lurus terhadap bidang sentuh. Sifat ini menyebabkan tekanan fluida dapat diteruskan ke seluruh bagian fluida secara merata.

2) Tekanan Hidrostatis

Tekanan hidrostatis adalah tekanan yang disebabkan oleh berat fluida akibat pengaruh gravitasi bumi. Semakin dalam suatu titik dari permukaan fluida,

semakin besar tekanan yang dialami titik tersebut karena semakin besar berat fluida di atasnya.

Tekanan hidrostatik dirumuskan sebagai

$$P = \rho g h$$

Keterangan :

P = Tekanan hidrostatik (Pa)

ρ = massa jenis fluida (kg/m^3)

g = percepatan gravitasi (m/s^2)

h = kedalaman (m).

Tekanan hidrostatik tidak bergantung pada bentuk wadah, tetapi hanya bergantung pada kedalaman dan massa jenis fluida. Fenomena ini dapat diamati pada dinding bendungan yang semakin tebal di bagian bawah, serta pada penyelam yang merasakan tekanan semakin besar saat menyelam lebih dalam.

Hukum pokok hidrostatika menyatakan bahwa semua titik yang terletak pada kedalaman yang sama di dalam fluida sejenis memiliki tekanan hidrostatik yang sama. Hal ini berarti bahwa tekanan fluida hanya bergantung pada kedalaman, bukan pada posisi horizontal atau bentuk wadah. Hukum ini menjadi dasar dalam perancangan bejana berhubungan, tangki air, dan sistem distribusi fluida.

3) Hukum Pascal

Hukum Pascal menyatakan bahwa tekanan yang diberikan pada fluida dalam ruang tertutup akan diteruskan sama besar ke segala arah. Secara matematis dinyatakan dengan

$$P_1 = P_2$$

atau

$$\frac{F_1}{A_1} = \frac{F_2}{A_2}$$

Prinsip ini dimanfaatkan dalam sistem hidrolik, seperti dongkrak hidrolik, rem hidrolik, mesin pengepres, dan alat pengangkat kendaraan. Dengan memperbesar luas penampang piston keluaran, gaya kecil yang diberikan pada piston masukan dapat menghasilkan gaya besar.

4) Hukum Archimedes

Hukum Archimedes menyatakan bahwa benda yang dicelupkan sebagian atau seluruhnya ke dalam fluida akan mengalami gaya ke atas yang besarnya sama dengan berat fluida yang dipindahkan oleh benda tersebut.

Gaya apung dirumuskan sebagai

$$F_a = \rho V g$$

Keterangan :

F_a = Gaya apung (N)

ρ = massa jenis fluida

V = volume benda yang tercelup

g = percepatan gravitasi.

Benda dikatakan mengapung jika gaya apung lebih besar daripada berat benda ($F_a > W$), sehingga sebagian benda berada di atas permukaan fluida. Pada kondisi ini massa jenis benda lebih kecil daripada massa jenis fluida. Benda dikatakan melayang jika gaya apung sama dengan berat benda ($F_a = W$), sehingga benda berada di dalam fluida tanpa naik atau turun. Benda dikatakan tenggelam jika gaya apung lebih kecil daripada berat benda ($F_a < W$), sehingga benda bergerak ke dasar fluida.

5) Tegangan permukaan

Tegangan permukaan merupakan salah satu sifat khas zat cair yang muncul akibat adanya gaya tarik-menarik antarpartikel pada permukaan zat cair.

Tegangan permukaan dapat didefinisikan sebagai kecenderungan permukaan zat cair untuk menegang sehingga bertindak seolah-olah ditutupi oleh suatu lapisan elastis tipis. Fenomena ini menyebabkan zat cair berusaha memperkecil luas

permukaannya semaksimal mungkin. Oleh karena itu, tetesan air cenderung berbentuk mendekati bola karena bentuk bola memiliki luas permukaan minimum untuk volume tertentu.

Secara mikroskopis, tegangan permukaan dapat dijelaskan melalui gaya kohesi antarpartikel zat cair. Partikel yang berada di dalam zat cair mengalami gaya tarik-menarik yang sama besar ke segala arah sehingga resultan gaya yang bekerja padanya bernilai nol. Sebaliknya, partikel yang berada di dekat atau tepat di permukaan zat cair tidak memperoleh gaya tarik dari atas, sehingga mengalami resultan gaya yang arahnya menuju ke dalam fluida. Kondisi ini menyebabkan permukaan zat cair seolah-olah tertarik ke arah dalam dan membentuk suatu selaput elastis yang tegang.

Akibat adanya tegangan permukaan, sejumlah tertentu zat cair akan cenderung membentuk permukaan sekecil mungkin, mempertahankan kestabilan tetesan, serta mampu menopang benda ringan seperti jarum atau serangga air di permukaan air. Tegangan permukaan secara kuantitatif dapat dinyatakan sebagai perbandingan antara gaya yang bekerja pada permukaan zat cair dengan panjang garis tempat gaya tersebut bekerja. Tegangan permukaan dirumuskan sebagai

$$\gamma = \frac{F}{d} = \frac{F}{2L}$$

Keterangan =

γ = Tegangan permukaan (N/m)

F = Gaya tegangan permukaan (N)

d = Panjang total permukaan yang mengalami gaya (m)

L = Panjang kawat atau garis kontak (m).

Dalam percobaan klasik, sebuah kawat lurus yang dicelupkan secara horizontal ke dalam larutan sabun akan mengalami gaya tarik ke atas akibat tegangan permukaan. Karena larutan sabun memiliki dua permukaan (permukaan atas dan bawah), maka gaya tegangan permukaan bekerja sepanjang dua kali panjang kawat. Fenomena ini membuktikan bahwa permukaan zat cair mampu menghasilkan gaya yang nyata dan terukur.

6) Kapilaritas

Kapilaritas merupakan peristiwa naik atau turunnya permukaan zat cair di dalam suatu pipa sempit atau celah sempit, seperti pada pipa kapiler, pori-pori kain, atau celah dinding. Fenomena ini terjadi akibat interaksi antara gaya kohesi antartpartikel zat cair dan gaya adhesi antara partikel zat cair dengan dinding pembatas.

Pada zat cair seperti air, gaya adhesi antara molekul air dan molekul dinding kaca lebih besar daripada gaya kohesi antar molekul air itu sendiri, sehingga air akan membasahi dinding dan naik di dalam pipa kapiler. Sebaliknya, pada zat cair seperti raksa, gaya kohesi lebih besar daripada gaya adhesi, sehingga raksa tidak membasahi dinding dan justru turun membentuk permukaan cekung ke bawah.

Fenomena kapilaritas memiliki peran penting dalam kehidupan sehari-hari, seperti proses naiknya air dari akar menuju daun pada tumbuhan, peresapan air pada tanah, serta kerja sumbu pada lampu minyak.

Sudut kontak merupakan sudut yang terbentuk antara garis singgung permukaan zat cair dengan dinding wadah pada titik pertemuan zat cair dan dinding. Sudut ini mencerminkan perbandingan antara gaya adhesi dan gaya kohesi. Jika sudut kontak kurang dari 90° , zat cair membasahi dinding (seperti air pada kaca). Jika sudut kontak lebih dari 90° , zat cair tidak membasahi dinding (seperti raksa pada kaca).

Besarnya kenaikan atau penurunan zat cair dalam pipa kapiler dapat dirumuskan sebagai

$$h = \frac{2\gamma \cos \theta}{\rho g r}$$

Keterangan :

h = tinggi kenaikan atau penurunan zat cair (m)

γ = tegangan permukaan (N/m)

θ = sudut kontak (derajat)

ρ = massa jenis zat cair (kg/m^3)
 g = percepatan gravitasi (m/s^2)
 r = jari-jari pipa kapiler (m).

Persamaan ini menunjukkan bahwa semakin kecil jari-jari pipa kapiler dan semakin besar tegangan permukaan, maka semakin besar pula kenaikan zat cair di dalam pipa tersebut.

Konsep tegangan permukaan dan kapilaritas merupakan bagian penting dari materi fluida statis yang sering menimbulkan miskonsepsi pada siswa, terutama dalam membedakan peran gaya kohesi, adhesi, dan gaya apung. Banyak siswa menganggap bahwa kapilaritas disebabkan oleh tekanan udara atau gaya apung semata, padahal fenomena tersebut terutama dipengaruhi oleh interaksi molekuler.

2.2 Penelitian yang Relevan

Tabel 4. Penelitian yang relevan

No	Nama Penulis/Jurnal	Nama Jurnal	Judul Artikel	Hasil
1.	(Abidin et al., 2019c)	<i>International Journal of Evaluation and Research in Education (IJERE)</i>	<i>A computerized adaptive test for measuring the physics critical thinking skills</i>	Penelitian ini mengembangkan tes adaptif berbasis komputer (CAT-PhysCriTS) untuk mengukur kemampuan berpikir kritis fisika siswa SMA kelas 11. Tes ini menggunakan format <i>two-tier multiple-choice</i> dan dinyatakan sangat layak serta valid oleh para ahli. Hasil tes sejalan dengan prestasi akademik siswa, sehingga CAT-PhysCriTS dinilai efektif, efisien, dan cocok untuk penilaian berskala besar.

2.	<i>(Istiyono et al., 2019)</i>	<i>International Journal of Instruction</i>	<i>Developing IRT-Based Physics Critical Thinking Skill Test: A CAT to Answer 21st Century Challenge</i>	Hasil penelitian menunjukkan bahwa pengembangan tes kemampuan berpikir kritis fisika berbasis CAT dan IRT tersebut valid dan terpercaya. Instrumen yang terdiri dari 52 item ini mampu mengukur kemampuan berpikir kritis siswa dari tingkat rendah hingga tinggi dengan reliabilitas mencapai 0,75. Selain itu, tes ini efektif dalam mengurangi kecurangan karena setiap siswa mendapatkan soal yang berbeda sesuai kemampuan mereka, dan dapat menampilkan hasil secara cepat dan akurat
3.	<i>(Istiyono et al., 2020)</i>	<i>European Journal of Educational Research</i>	<i>Developing of computerized adaptive testing to measure physics higher order thinking skills of senior high school students and its feasibility of use</i>	Penelitian ini berhasil mengembangkan sistem <i>Computerized Adaptive Testing</i> (PhysTHOTS-CAT) untuk mengukur keterampilan HOTS fisika siswa kelas X. Instrumen menggunakan model 4D dan dinyatakan valid serta layak berdasarkan penilaian ahli, guru, dan siswa. CAT mampu menampilkan soal adaptif sesuai kemampuan siswa dan memberikan hasil pengukuran yang akurat. Secara keseluruhan, PhysTHOTS-CAT dinilai feasible (82,28%) dan efektif digunakan untuk

				mengukur HOTS fisika siswa SMA.
4.	(Istiyono et al., 2018)	<i>Jurnal Cakrawala Pendidikan</i>	<i>The Developing of Creative Thinking Skills Test Based on Modern Test Theory in Physics of Senior High Schools</i>	Instrumen PhysCreTHOTS berhasil dikembangkan menggunakan teori tes modern (PCM) untuk mengukur Creative Thinking Skills Fisika siswa SMA. Semua butir soal dinyatakan valid, fit dengan model, dan memiliki tingkat kesukaran yang baik. Reliabilitas instrumen sebesar 0,71 sehingga cukup tinggi dan layak digunakan. Instrumen efektif untuk mengukur kemampuan berpikir kreatif siswa pada rentang kemampuan sedang hingga tinggi.

Berdasarkan tabel kajian terdahulu, terlihat bahwa berbagai instrumen *Adaptive Assessment* dan tes berbasis teori pengukuran modern telah berhasil dikembangkan untuk mengukur kemampuan berpikir kritis, *higher order thinking skills* (HOTS), maupun berpikir kreatif dalam pembelajaran fisika. Instrumen tersebut antara lain CAT-PhysCriTS yang dikembangkan oleh Abidin et al. (2019), instrumen *Computerized Adaptive Test* (CAT) berbasis *Item Response Theory* (IRT) oleh Istiyono et al. (2019), PhysTHOTS-CAT oleh Istiyono et al. (2020), serta PhysCreTHOTS (2018) yang dikembangkan menggunakan *Partial Credit Model* (PCM) untuk mengukur keterampilan berpikir kreatif. Secara umum, penelitian-penelitian tersebut memberikan kontribusi signifikan dalam penguatan pengukuran kemampuan berpikir tingkat tinggi melalui penerapan pendekatan modern seperti IRT, PCM, dan sistem asesmen adaptif. Namun

demikian, pengembangan instrumen tersebut masih berada pada cakupan materi fisika yang bersifat umum dan belum difokuskan secara spesifik pada satu topik tertentu, sehingga pengembangan instrumen adaptif yang lebih kontekstual dan mendalam masih memiliki peluang yang besar.

Oleh karena itu, kebaruan penelitian ini terletak pada pengembangan instrumen penilaian keterampilan berpikir kritis berbasis *Adaptive Assessment* yang secara khusus diarahkan pada materi fluida statis. Fokus ini memberikan nilai tambah karena konsep fluida statis bersifat kontekstual dan sangat dekat dengan fenomena kehidupan sehari-hari, sehingga relevan untuk mengukur kemampuan berpikir kritis siswa secara autentik. Selain itu, penelitian ini menggunakan indikator berpikir kritis yang lebih komprehensif dan sesuai dengan karakteristik analisis pada fluida statis, meliputi kemampuan merumuskan masalah, memberikan argumen, melakukan deduksi, melakukan induksi, melakukan evaluasi, serta memutuskan dan melaksanakan solusi.

Kebaruan lain yang membedakan penelitian ini dari penelitian sebelumnya terletak pada aspek implementasi teknologi. Jika penelitian-penelitian terdahulu umumnya diuji coba menggunakan perangkat komputer, instrumen dalam penelitian ini dirancang dan diuji coba menggunakan perangkat *handphone*, sehingga lebih sesuai dengan karakteristik penggunaan teknologi oleh siswa saat ini dan mendukung fleksibilitas akses lintas perangkat. Selain itu, sistem *Adaptive Assessment* yang dikembangkan dilengkapi dengan mekanisme pengendalian kecurangan, yaitu dengan menghentikan tes secara otomatis ketika siswa keluar dari laman website asesmen. Fitur ini berfungsi untuk menjaga integritas proses penilaian dan meningkatkan keakuratan hasil pengukuran kemampuan berpikir kritis siswa.

Dengan demikian, penelitian ini menawarkan kontribusi baru berupa integrasi antara *Adaptive Assessment*, indikator berpikir kritis yang komprehensif, konteks materi fluida statis yang spesifik, serta implementasi teknologi berbasis *mobile* yang dilengkapi dengan sistem pencegahan kecurangan, yang belum ditemukan secara eksplisit pada penelitian-penelitian sebelumnya.

III. METODE PENELITIAN

3.1 Desain Penelitian

Jenis penelitian ini adalah penelitian pengembangan R&D (*Research and Development*). Penelitian ini dilakukan untuk mengukur keterampilan berpikir kritis siswa pada materi fluida statis. Adapun produk yang dikembangkan dalam penelitian ini adalah instrumen penilaian berbasis *Adaptive Assessment*. Model Pengembangan yang digunakan adalah model pengembangan 4D, model ini dipilih karena memiliki langkah-langkah sistematis yang dapat digunakan secara prosedural dalam mengembangkan instrumen penilaian (Saputra *et al.*, 2022). Model Pengembangan menurut (Sivasailan, 1974) adalah sebagai berikut :



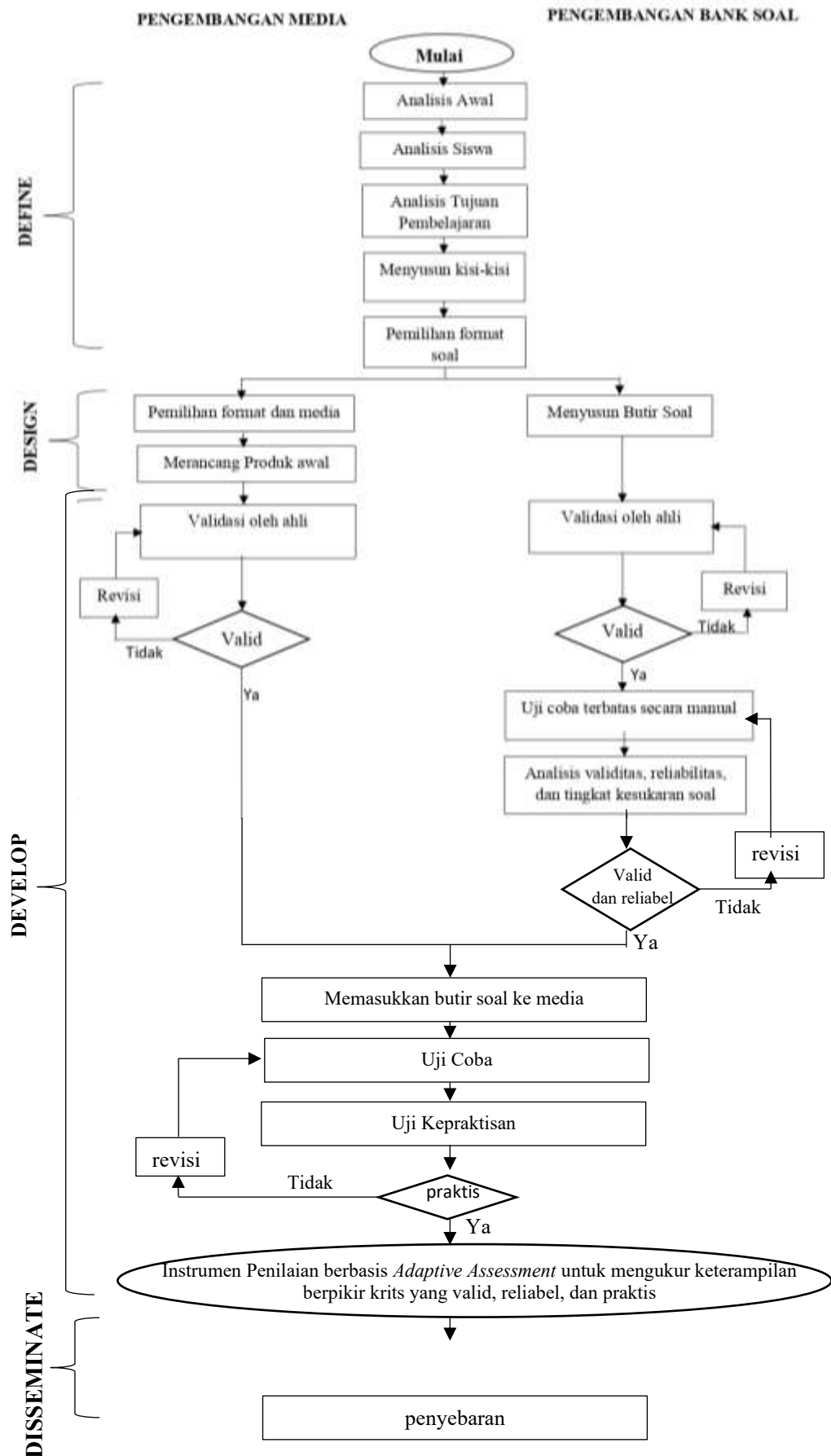
Gambar 1. Model Pengembangan 4D

3.2 Subjek Penelitian

Penelitian ini melibatkan dua subjek, yaitu subjek penelitian dan subjek uji coba. Teknik pengambilan sampel pada subjek uji coba menggunakan metode *purposive sampling* dengan pemilihan sekolah berdasarkan kriteria tertentu. Subjek dalam penelitian pengembangan ini adalah instrumen penilaian berbasis *Adaptive Assessment* yang bertujuan untuk mengukur keterampilan berpikir kritis pada materi fluida statis. Adapun subjek uji coba dibagi menjadi empat kelompok, yaitu kelompok pertama terdiri dari 24 guru dan 37 siswa yang berperan dalam tahap analisis kebutuhan. Kelompok kedua mencakup 3 dosen ahli dan 1 guru yang berfungsi sebagai validator dalam menilai tingkat validitas instrumen dan 2 dosen ahli dan 1 guru media *Adaptive Assessment* yang dikembangkan. Kelompok ketiga mencakup 159 siswa untuk uji validitas dan reliabilitas empiris yang berasal dari 3 Sekolah yaitu SMAN 14 Bandar Lampung, SMA Global Madani Bandar Lampung, dan SMAN 1 Belalau. Sementara itu, kelompok ke empat yaitu 62 siswa dari 3 Sekolah yaitu SMAN 14 Bandar Lampung, SMA Global Madani Bandar Lampung, dan SMAN 1 Belalau yang dilibatkan dalam uji coba untuk uji kepraktisan instrumen tersebut.

3.3 Prosedur Pengembangan

Berdasarkan model pengembangan 4-D yang digunakan. Berikut ini merupakan prosedur yang akan dilakukan dalam pengembangan produk menggunakan langkah model 4-D, seperti pada gambar 2 berikut



Gambar 2. Prosedur Pengembangan

3.3.1 *Define*

Tahap pendefinisian dilakukan untuk melihat gambaran kondisi di lapangan pada proses pembelajaran dengan tujuan menetapkan dan mendefinisikan syarat-syarat pembelajaran yang diawali dengan analisis permasalahan hingga analisis tujuan dari batasan materi yang akan dikembangkan. Adapun Langkah-langkah pada tahap ini yaitu :

1. Analisis Awal

Analisis awal dilakukan untuk mengidentifikasi masalah utama yang dihadapi dalam pembelajaran sehingga dapat meningkatkan mutu praktik pembelajaran. Analisis ini akan didapatkan gambaran alternatif penyelesaian masalah utama yang memudahkan dalam spesifikasi produk yang dikembangkan. Analisis yang dilakukan mencakup keterampilan berpikir kritis siswa, instrumen penilaian yang digunakan dalam pembelajaran, serta terkait *Adaptive Assessment*. Pada diagram pengembangan, langkah ini merupakan tahap awal baik dalam pengembangan media maupun pengembangan bank soal, yang menjadi dasar seluruh proses desain dan pengembangan selanjutnya. Analisis dilakukan dengan menggunakan instrumen analisis kebutuhan melalui Google Form.

2. Analisis Siswa

Analisis siswa bertujuan mengidentifikasi karakteristik siswa yang menjadi target pengembangan instrumen penilaian, untuk memperoleh gambaran masalah yang dihadapi dalam proses pembelajaran. Analisis kebutuhan dilakukan melalui penyebaran angket secara daring menggunakan Google Form kepada siswa di Provinsi Lampung.

3. Analisis tujuan pembelajaran

Tahap ini bertujuan merumuskan tujuan pembelajaran yang harus dicapai dengan mengacu pada Capaian Pembelajaran dan Alur Tujuan Pembelajaran. Hasil dari analisis ini menjadi acuan dalam menentukan tujuan pengembangan serta untuk menyusun kisi-kisi soal yang akan dikembangkan.

3. Menyusun kisi-kisi

Tahap penyusunan kisi-kisi merupakan langkah awal dalam pengembangan bank soal keterampilan berpikir kritis fisika. Penyusunan kisi-kisi didasarkan pada hasil analisis tujuan pembelajaran, analisis karakteristik siswa, serta capaian pembelajaran pada materi fluida statis. Penyusunan ini bertujuan untuk memastikan keterwakilan setiap indikator keterampilan berpikir kritis yang akan diukur, kesesuaian materi dengan kurikulum, serta keterpaduan antara indikator, bentuk soal, dan tingkat kognitif yang diharapkan. Setiap butir soal dipetakan berdasarkan indikator berpikir kritis, materi pokok, konteks soal, dan tingkat kesukaran yang direncanakan sehingga menjadi acuan sistematis dalam penyusunan butir soal. Keberadaan kisi-kisi yang jelas dan terstruktur diharapkan mampu menghasilkan butir soal yang memiliki validitas isi yang baik serta dapat mengukur keterampilan berpikir kritis siswa secara komprehensif.

4. Pemilihan Format Soal

Format instrumen yang digunakan adalah soal pilihan ganda beralasan berbasis skenario kontekstual, yang disusun secara bertingkat sesuai indikator keterampilan berpikir kritis. Format ini dipilih karena mendukung prinsip asesmen adaptif, mudah digunakan melalui perangkat digital, serta memungkinkan penilaian otomatis sekaligus tetap menggali alasan berpikir siswa secara mendalam.

3.3.2 *Design*

Setelah tahap pendefinisian yang dilakukan selanjutnya adalah tahap perancangan (*design*). Tahap perancangan ini bertujuan untuk merancang instrumen penilaian berpikir kritis pada materi fluida statis. Tahap ini meliputi :

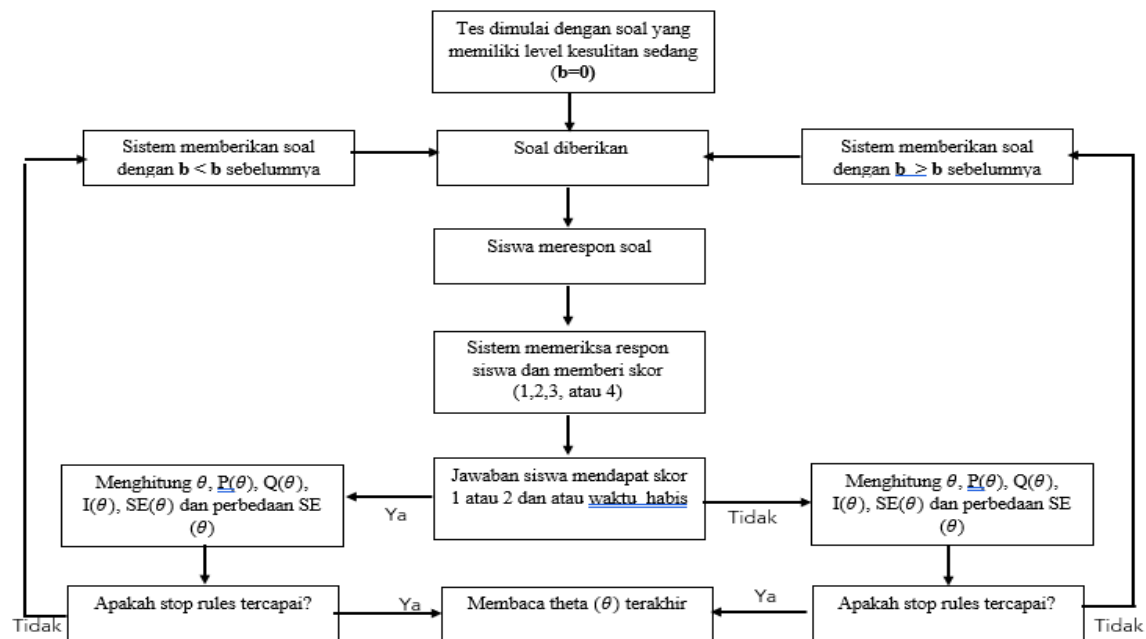
1. Pemilihan Format dan Media

Media yang digunakan untuk menyajikan instrumen adalah media digital yang dibuat menggunakan media digital website karena mampu menyajikan soal secara adaptif sesuai respons siswa, menyajikan soal secara interaktif, memberikan

umpan balik otomatis, serta merekam dan menyimpan hasil jawaban untuk dianalisis lebih lanjut oleh guru. Kriteria utama dalam pemilihan media ini mencakup beberapa aspek penting, yaitu: (1) kemampuan adaptif untuk menyesuaikan alur dan tingkat kesulitan soal dengan kemampuan individu siswa; (2) kemudahan penggunaan (*user-friendly*) baik bagi guru sebagai pengelola maupun siswa sebagai pengguna; (3) kemampuan media untuk memberikan umpan balik secara real-time sebagai bagian dari proses pembelajaran yang reflektif; dan (4) kemampuan menyimpan dan mengelola data hasil asesmen secara otomatis dan terstruktur. Berdasarkan kriteria tersebut, dipilihlah aplikasi penilaian berbasis web yang memungkinkan sistem secara otomatis menyesuaikan jalur soal berdasarkan indikator keterampilan berpikir kritis

2. Merancang Produk Awal

Rancangan awal instrumen penilaian disusun berdasarkan hasil analisis kebutuhan dan tujuan pengembangan. Instrumen yang dikembangkan berupa soal pilihan ganda beralasan, dengan struktur jalur adaptif yang mengacu pada indikator berpikir kritis. Pada tahap ini, disusun terlebih dahulu blueprint instrumen. Setiap item dirancang memiliki dua komponen utama yaitu pilihan jawaban dan alasan yang mendasari pilihan tersebut. Jawaban siswa terhadap keduanya menjadi dasar untuk menentukan arah soal selanjutnya dalam sistem. Selain itu, dirancang pula alur adaptif soal, yang memetakan pergerakan siswa dari satu tingkat soal ke tingkat berikutnya berdasarkan akurasi dan kualitas alasan yang diberikan. Prototipe awal disusun menggunakan platform digital dengan tampilan antarmuka sederhana, agar mudah diakses oleh siswa dan mendukung navigasi yang jelas selama proses asesmen. Berikut adalah diagram proses pelaksanaan media *Adaptive Assessment* yang dikembangkan pada Gambar 3.



Gambar 3. Diagram proses pelaksanaan *Adaptive Assessment*

Gambar 3 menggambarkan alur pelaksanaan *Adaptive Assessment* yang dimulai dengan pemberian soal pada tingkat kesulitan sedang ($b = 0$) untuk memperoleh estimasi awal kemampuan siswa secara seimbang. Setelah soal diberikan, siswa merespons pertanyaan dan sistem secara otomatis memeriksa jawaban tersebut lalu memberikan skor 1, 2, 3, atau 4 sesuai dengan kualitas respons berdasarkan rubrik keterampilan berpikir kritis. Skor yang diperoleh kemudian digunakan sebagai dasar penyesuaian tingkat kesulitan soal berikutnya. Apabila siswa memperoleh skor rendah, seperti 1 atau 2, maka sistem akan memilih soal dengan tingkat kesulitan lebih rendah dari sebelumnya ($b < b$ sebelumnya). Sebaliknya, jika performa siswa lebih baik, sistem akan memberikan soal dengan tingkat kesulitan yang lebih tinggi ($b > b$ sebelumnya), sehingga proses penilaian berlangsung secara adaptif mengikuti kemampuan siswa.

Setiap kali siswa memberikan respons, sistem menghitung parameter estimasi kemampuan menggunakan pendekatan teori respons butir, meliputi nilai kemampuan sementara (θ), peluang menjawab benar $P(\theta)$, peluang salah $Q(\theta)$, fungsi informasi $I(\theta)$, serta *standard error* pengukuran $SE(\theta)$. Nilai *standard error* digunakan untuk melihat tingkat ketelitian estimasi kemampuan siswa. Setelah perhitungan dilakukan, sistem memeriksa aturan berhenti (*stop rules*).

Pada diagram ini, tes akan dihentikan apabila nilai *standard error* sudah kurang dari atau sama dengan 0,01 yang menunjukkan estimasi kemampuan sudah sangat stabil, atau ketika waktu pengerjaan telah habis (Zakwandi *et al.*, 2024). Jika salah satu kondisi tersebut terpenuhi, sistem membaca nilai θ terbesar sebagai hasil akhir kemampuan siswa. Namun, jika aturan berhenti belum tercapai, sistem kembali memberikan soal baru yang telah disesuaikan dengan estimasi kemampuan terbaru, sehingga proses adaptif berlangsung secara berulang hingga kriteria penghentian terpenuhi.

3. Menyusun Butir Soal

Pada tahap ini dilakukan penyusunan butir soal yang didasarkan atas hasil perumusan pada tahap *define* yang mengacu pada indikator berpikir kritis yaitu merumuskan masalah, memberikan argumen, melakukan deduksi, melakukan induksi, melakukan evaluasi, dan memutuskan dan melaksanakan.

3.3.3 Develop

Tahap pengembangan adalah tahap untuk menghasilkan produk pengembangan instrumen penilaian berpikir kritis yang sudah direvisi berdasarkan masukan para ahli. Tahap ini terdiri dari dua Langkah yaitu *expert appraisal* (penilaian ahli) yang disertai revisi dan *developmental testing* (uji coba pengembangan).

1. Validasi Oleh Ahli

Validasi yang dilakukan dalam penelitian ini terdiri dari validasi instrumen penilaian dan validasi media *Adaptive Assessment*. Penilaian dilakukan oleh beberapa ahli yang terdiri dari masing-masing 2 dosen dan 1 guru fisika.

a. Validasi Media

Validasi media dilakukan untuk menilai kelayakan isi, tampilan, kesesuaian dengan konsep adaptif dan keterpakaian media dalam proses pembelajaran. Validasi bertujuan untuk memastikan bahwa media yang dikembangkan telah sesuai dengan tujuan pembelajaran, akurat secara konsep, serta menarik dan mudah digunakan oleh siswa.

Revisi dilakukan terhadap butir soal dan aspek media yang memperoleh skor rendah (1 atau 2), serta yang disertai komentar meskipun mendapat skor tinggi. Butir yang telah memenuhi kriteria kelayakan (skor 3 atau 4 dan tanpa saran perbaikan) tidak direvisi.

b. Validasi Instrumen

Instrumen divalidasi menggunakan lembar penilaian yang mencakup aspek substansi isi, konstruksi soal, dan kebahasaan, dengan *skala likert* empat poin. Selain itu, lembar penilaian dilengkapi dengan kolom komentar untuk masukan kualitatif. Hasil penilaian dianalisis secara kuantitatif untuk menentukan tingkat validitas isi instrumen dan secara kualitatif untuk menyempurnakan redaksi dan konteks soal.

c. *Developmental Testing* (Uji Coba Pengembangan)

Langkah selanjutnya adalah *developmental testing*. Subjek uji coba adalah siswa yang dipilih secara purposif. Data hasil uji coba tes yang dialkukan secara manual dianalisis menggunakan *Rasch Model* untuk menilai validitas, reliabilitas, serta distribusi tingkat kesukaran soal. Hasil uji coba ini menjadi dasar untuk revisi instrumen sebelum dilanjutkan ke uji coba kelompok besar. Setelah soal dinyatakan valid, reliabel, dan uji tingkat kesukaran, bank soal dimasukkan kedalam media *Adaptive Assessment* yang telah dikembangkan.

Selanjutnya, dilakukan uji kelompok besar. Uji coba kelompok besar bertujuan untuk menilai kepraktisan instrumen dalam mengukur keterampilan berpikir kritis secara adaptif. Subjek pada tahap ini adalah siswa yang telah mengikuti pembelajaran materi fluida statis. Instrumen disajikan dalam sistem digital yang secara otomatis menyesuaikan tingkat kesulitan soal berdasarkan respons peserta. Uji kepraktisan dilakukan dengan menyebar angket respon kepada siswa yang telah menggunakan instrumen penilaian berbasis *Adaptive Assessment*. Hasil uji coba besar menjadi dasar untuk penyempurnaan akhir instrumen sebelum digunakan dalam implementasi yang lebih luas.

3.3.4 Disseminate

Setelah dilakukan revisi produk pada tahap pengembangan, instrumen penilaian berbasis *Adaptive Assessment* untuk mengukur keterampilan berpikir kritis siswa pada materi fluida statis dinyatakan valid, reliabel, dan praktis. oleh karena itu selanjutnya dapat dilakukan penyebaran instrumen dilakukan melalui publikasi hasil penelitian dalam jurnal ilmiah terakreditasi.

3.4 Instrumen Penelitian

Instrumen penelitian yang akan digunakan yaitu instrumen analisis kebutuhan, angket validitas, angket respon, dan instrumen penilaian keterampilan berpikir kritis. Kemudian skala dalam penelitian ini terdiri dari skala validasi, skala uji kepraktisan dengan penskoran menggunakan *skala likert*.

3.4.1 Instrumen Analisis Kebutuhan

Instrumen analisis kebutuhan yang digunakan dalam penelitian ini dilakukan penyebaran dengan menggunakan google form yang terdiri atas 18 pertanyaan untuk guru dan 15 pertanyaan untuk siswa, yang terdiri atas 4 indikator yaitu jenis dan strategi asesmen yang digunakan, penilaian terhadap indikator berpikir kritis, penggunaan dan persepsi terhadap asesmen digital, kebutuhan terhadap *Adaptive Assessment*.

3.4.2 Angket validitas

1. Angket Validitas Media

Media *Adaptive Assessment* yang telah kembangkan selanjutnya divalidasi oleh ahli media. Tujuan dari validasi untuk melihat kelayakan media sebelum dilakukan uji coba. Angket validitas ini menggunakan skala angka yang akan diubah menjadi persentase kelayakan berdasarkan isi, tampilan, kesesuaian dengan konsep adaptif dan keterpakaian media dalam proses pembelajaran dengan kisi-kisi pada Tabel 5.

Tabel 5. Kisi-kisi instrumen validasi media *Adaptive Assessment*

Aspek Penilaian	Indikator Penilaian	Nomor Butir
Desain Tampilan	Desain antarmuka menarik dan sesuai usia siswa SMA	1
	Tata letak halaman terstruktur rapi dan mudah dinavigasi	2
	Konsistensi warna, font, ikon, dan grafik	3
	Elemen visual mendukung fokus dan tidak berlebihan	4
Kemudahan Penggunaan	Navigasi dan tombol kontrol mudah digunakan	5
	Petunjuk penggunaan jelas dan mudah dipahami	6
	Media dapat digunakan tanpa pelatihan khusus	7
Fungsionalitas Adaptif	Media dapat diakses pada berbagai perangkat	8
	Media menyesuaikan soal berdasarkan jawaban siswa	9
	Mekanisme adaptif berjalan responsif	10
	Jalur soal sesuai kemampuan siswa	11
	Media mencatat respons siswa secara akurat	12
Kinerja Teknis	Proses loading cepat dan tidak mengganggu	13
	Tidak terdapat bug atau error	14
	Data hasil tes tersimpan otomatis dan aman	15
	Media berjalan stabil tanpa gangguan	16
Pelaporan Hasil	Laporan hasil mencakup skor akhir siswa	17
	Tampilan hasil mudah diakses siswa dan guru	18
	Media menyajikan umpan balik setelah tes	19
	Media dapat digunakan daring, luring, dan hybrid	20
Kelayakan Implementasi	Media mendukung asesmen formatif dan sumatif	21
	Media fleksibel untuk berbagai topik pelajaran	22
	Media mudah dipelihara dan diperbarui	23

Data yang diperoleh dari angket validitas ini berupa data kuantitatif dengan menggunakan *skala likert* yang diadaptasi dari Ratumanan & Laurent (2011) dengan 4 tingkatan, yaitu 1, 2, 3, dan 4 seperti pada Tabel 6.

Tabel 6. *Skala likert* Angket Validitas Media

Pilihan jawaban	Angka
Sangat Setuju	4
Setuju	3
Kurang Setuju	2
Tidak Setuju	1

2. Angket Validitas Instrumen Penilaian

Instrumen penilaian berpikir kritis yang telah disusun oleh peneliti selanjutnya divalidasi oleh dosen ahli. Tujuan dari penilaian validasi produk untuk melihat kelayakan soal-soal tersebut sebelum dilakukan uji coba. Angket validitas ini menggunakan skala angka yang akan diubah menjadi persentase kelayakan berdasarkan isi, konstruk, dan bahasa dengan kisi-kisi pada Tabel 7.

Tabel 7. Kisi-kisi validasi instrumen soal

Komponen	Indikator / Jenis Persyaratan
Isi	Butir soal sesuai dengan indikator
	Hanya ada satu kunci jawaban benar
	Isi materi sesuai dengan tujuan pengukuran
	Isi materi sesuai dengan tingkat kelas/jenjang pendidikan
	Butir pengecoh berfungsi dengan baik
	Pokok soal dirumuskan dengan jelas
	Pilihan jawaban dirumuskan dengan jelas
	Pokok soal tidak mengarah ke jawaban benar
	Tidak ada bentuk negatif ganda
	Pilihan jawaban homogen
Konstruk	Panjang pilihan jawaban relatif sama
	Antar-butir soal tidak bergantung satu sama lain
	Pilihan dalam bentuk angka/waktu diurutkan
	Rumusan bahasa komunikatif
Bahasa	Menggunakan kalimat gramatikal (baku)
	Menggunakan kalimat tidak bermakna ganda
	Kalimat tidak mengandung unsur SARA
	Menggunakan kosakata baku/umum

Data yang diperoleh dari angket validitas ini berupa data kuantitatif dengan menggunakan *skala likert* yang diadaptasi dari Ratumanan & Laurent (2011) dengan 4 tingkatan, yaitu 1, 2, 3, dan 4 seperti pada Tabel 8.

Tabel 8. *Skala likert* Angket Validitas Instrumen Penilaian

Pilihan jawaban	Angka
Sangat Setuju	4
Setuju	3
Kurang Setuju	2
Tidak Setuju	1

3.4.3 Angket Respon

Angket Respon bertujuan untuk mengetahui penilaian dari praktisi terhadap instrumen penilaian berpikir kritis pada materi fluida statis yang telah dikembangkan. Instrumen penilaian yang telah di evaluasi oleh guru fisika di sekolah akan dilanjutkan dengan uji coba berupa kelompok kecil terhadap siswa untuk mengetahui respon siswa yang juga sebagai pengguna terhadap instrumen penilaian yang dikembangkan dengan kisi-kisi pada Tabel 9.

Tabel 9. Kisi-kisi Angket Respon

Aspek yang Dinilai	Indikator
Kelayakan butir soal	Butir soal mudah dipahami siswa
	Materi tes sesuai dengan pembelajaran
Kelayakan produk <i>Media Adaptive Assessment</i>	Tes menampilkan butir-butir soal yang mengukur kemampuan berpikir kreatif
	Butir-butir tes jarang ditemui dalam tes atau pembelajaran sebelumnya
Keefektifan pemberian tes dengan menggunakan <i>Media Adaptive Assessment</i>	Produk <i>Media Adaptive Assessment</i> sangat menarik
	Penggunaan <i>Media Adaptive Assessment</i> dalam pemberian tes merupakan hal baru dan layak digunakan
	Penggunaan <i>Media Adaptive Assessment</i> mempermudah proses tes
	Pemberian tes menggunakan <i>Media Adaptive Assessment</i> efektif
	Daya tarik dan keberlanjutan penggunaan <i>Media Adaptive Assessment</i>
	Kualitas umpan balik dan kejujuran tes

Data yang diperoleh digunakan untuk menilai kelayakan instrumen penilaian berpikir kritis dengan model 4D pada materi fluida statis yang dikembangkan. Data tersebut diperoleh dari sistem penskoran menggunakan *Skala likert* yang

diadaptasi dari Ratumanan & Laurent (2011) dengan 4 tingkatan yaitu 1, 2, 3, dan 4 seperti pada Tabel 10.

Tabel 10. *Skala likert Uji Kepraktisan*

Pilihan jawaban	Angka
Sangat Setuju	4
Setuju	3
Kurang Setuju	2
Tidak Setuju	1

3.4.4 Instrumen Penilaian Berpikir Kritis

Instrumen penilaian berpikir kritis yang dikembangkan akan terdiri atas soal pilihan ganda beralasan yang memuat indikator berpikir kritis menurut Ennis (2009), yang terdiri atas : merumuskan masalah, memberikan argumen, melakukan deduksi, melakukan induksi, melakukan evaluasi, memutuskan dan melaksanakan. Karakteristik butir soal ditinjau berdasarkan validitas empiris, reliabilitas tes, dan indeks kesulitan butir. Validitas empiris dari butir soal ditentukan berdasarkan kesesuaian antara butir soal dengan model penskoran. Kesesuaian butir (item fit) diperoleh dari hasil uji coba terbatas. Jawaban siswa dianalisis dengan scoring menurut (Abidin *et al.*, 2019) sebagai berikut:

- 1) skor 1 jika jawaban salah dan alasan salah
- 2) skor 2 jika jawaban benar tetapi alasan salah
- 3) skor 3 jika jawaban salah tetapi alasan benar
- 4) skor 4 jika jawaban dan alasan keduanya benar

3.5 Teknik Analisis data

3.5.1 *Partial Credit Model*

Analisis kemampuan berpikir kritis peserta tes dilakukan menggunakan pendekatan *Item Response Theory* (IRT) dengan model *Partial Credit Model* (PCM). Model ini sesuai untuk instrumen dengan kategori skor politomus karena setiap butir soal memberikan beberapa tingkat skor berdasarkan kualitas jawaban peserta. Estimasi kemampuan peserta (θ) diperoleh dengan mempertimbangkan indeks kesukaran tiap butir dan kategori jawaban yang dipilih, menghasilkan

pengukuran yang lebih objektif dan stabil. Penggunaan PCM juga memberikan keunggulan berupa hasil pengukuran yang bebas dari pengaruh bentuk tes serta lebih akurat dalam menggambarkan kemampuan berpikir kritis peserta. Rumus PCM menurut Istiyono (2019) untuk menghitung θ disajikan sebagai berikut.

$$(\theta) = bi + \frac{1}{D_{ai}} \ln (0,5(1 + \sqrt{(1 + 8c_i)})$$

Keterangan :

θ = Tingkat kemampuan peserta tes

bi = indeks kesukaran butir ke- i

D = Faktor penskalaan (bernilai 1.7)

ai = Indeks daya butir ke- i (1 PL model Rasch digunakan nilai 1)

ci = Indeks guessing ke- i (1 PL model Rasch digunakan nilai 0)

Probabilitas benar dapat dihitung menggunakan rumus berikut:

Probabilitas pada kategori 0 (skor 1):

$$P_{i0}(\theta) = \frac{1}{1 + \exp(\theta_n - b_{i1}) + \exp[(\theta_n - b_{i1}) + (\theta_n - b_{i2})] + \exp[(\theta_n - b_{i1}) + (\theta_n - b_{i2}) + (\theta_n - b_{i3})]}$$

Probabilitas pada kategori 1 (skor 2) :

$$P_{i1}(\theta) = \frac{\exp(\theta_n - b_{i1})}{1 + \exp(\theta_n - b_{i1}) + \exp[(\theta_n - b_{i1}) + (\theta_n - b_{i2})] + \exp[(\theta_n - b_{i1}) + (\theta_n - b_{i2}) + (\theta_n - b_{i3})]}$$

Probabilitas pada kategori ke 2 (skor 3):

$$P_{i2}(\theta) = \frac{\exp(\theta_n - b_{i1}) + \exp(\theta_n - b_{i2})}{1 + \exp(\theta_n - b_{i1}) + \exp[(\theta_n - b_{i1}) + (\theta_n - b_{i2})] + \exp[(\theta_n - b_{i1}) + (\theta_n - b_{i2}) + (\theta_n - b_{i3})]}$$

Probabilitas pada kategori ke 3 (skor 4):

$$P_{i3}(\theta) = \frac{\exp(\theta_n - b_{i1}) + \exp(\theta_n - b_{i2}) + \exp(\theta_n - b_{i3})}{1 + \exp(\theta_n - b_{i1}) + \exp[(\theta_n - b_{i1}) + (\theta_n - b_{i2})] + \exp[(\theta_n - b_{i1}) + (\theta_n - b_{i2}) + (\theta_n - b_{i3})]}$$

Menghitung Probabilitas salah :

$$Q = 1 - P$$

Menghitung Informasi butir :

$$I = P \times Q$$

Menghitung *Standar error* :

$$SE = \frac{1}{\sum_{i=1}^N Q \cdot P}$$

3.5.2 Uji Validitas

1. Uji Validitas Media

Data yang diperoleh melalui uji validitas media menggunakan angket validitas media ini berupa data kuantitatif dengan menggunakan *skala likert*. Hasil dari skor tersebut kemudian dianalisis dengan menggunakan perhitungan menurut Sudjana, (2005) dimana :

$$P = \frac{\text{Jumlah skor yang diperoleh}}{\text{Jumlah skor maksimum}} \times 100$$

Keterangan :

P = Persentase Kelayakan

Perolehan persentase kelayakan media selanjutnya dikategorikan sesuai dengan kriteria uji validitas yang ada di Tabel 11.

Tabel 11. Interpretasi Uji Validitas

Persentase	Kriteria
81,26% - 100%	Sangat Valid
62,51% - 81,25%	Valid
43,76% - 62,50%	Kurang valid
25%-43,75%	Tidak valid

(Diani & Hartati, 2018).

Berdasarkan kriteria pada Tabel 11, media dinyatakan valid apabila persentase kelayakan yang diperoleh berada pada rentang 62,51%–81,25%. Kriteria ini

digunakan sebagai acuan dalam menentukan tingkat validitas media yang dikembangkan.

2. Uji Validitas Instrumen Penilaian

Data yang diperoleh setelah dilakukan uji validitas menggunakan Angket validitas instrumen penilaian ini berupa data kuantitatif dengan menggunakan *skala likert*. Hasil dari skor tersebut kemudian dianalisis dengan menggunakan perhitungan menurut Aiken, (1985) dimana :

$$V = \frac{\sum s}{[n(c - 1)]}$$

Keterangan :

V = Indeks Aiken

$s = r - I_0$

I_0 = skala terkecil

Berdasarkan tabel nilai kritis Aiken dengan empat validator dan empat kategori skala, nilai minimum koefisien Aiken yang dapat diterima adalah sebesar 0,92. Oleh karena itu, suatu butir soal dinyatakan valid apabila memiliki nilai $V \geq 0,92$. Butir soal dengan nilai $V < 0,92$ direvisi berdasarkan masukan validator sebelum digunakan pada tahap uji coba.

3. Uji Validitas dan Reliabilitas Instrumen Tes

Data validitas dan realibitas instrumen tes dilakukan setelah uji kelompok kecil, hasil dianalisis menggunakan Rasch Model. Uji validitas empirik dalam penelitian ini menggunakan model *Rasch* dengan *software Winstep 4.7.0*. Adapun parameter yang digunakan untuk mengetahui ketepatan atau kesesuaian responden dan butir pertanyaan menurut Boone *et al.*, (2014) antara lain:

- a) Nilai *outfit mean square* (MNSQ) yang diterima: $0,5 < \text{MNSQ} < 1,5$
- b) Nilai *outfit Z-standars* (ZSTD) yang diterima: $-2,0 < \text{MNSQ} < +2,0$
- c) Nilai *outfit Point Measure Correlation* (Pt Mean Corr) yang diterima: $0,4 < \text{Pt Measure Corr} < 0,85$

Nilai *outfit means-square*, *outfit z-standard*, dan *point measure correlation* adalah kriteria yang digunakan untuk melihat tingkat kesesuaian butir pertanyaan (*item*

fit). Dari ketiga parameter menurut Boone *et al.*, (2014) tersebut, maka dapat ditarik kesimpulan bahwa:

- a) Apabila nilai dari MNSQ, ZSTD, Pt-Mean Corr sesuai kriteria, maka butir pertanyaan dinyatakan valid.
- b) Apabila nilai dari ZSTD dan Pt. Mean Corr memenuhi kriteria tetapi nilai MNSQ tidak memenuhi kriteria maka butir soal masih dinyatakan valid.
- c) Apabila nilai dari MNSQ dan Pt. Mean Corr memenuhi kriteria tetapi nilai ZSTD tidak memenuhi kriteria maka butir soal masih dinyatakan valid.
- d) Apabila nilai dari MNSQ dan Pt. Mean Corr tidak memenuhi kriteria tetapi nilai ZSTD memenuhi kriteria maka butir soal masih dinyatakan valid.
- e) Apabila nilai dari MNSQ dan ZSTD tidak memenuhi kriteria tetapi nilai Pt Mean Corr memenuhi kriteria maka butir soal cenderung tidak valid dan harus diperbaiki.
- f) Apabila nilai dari MNSQ, ZSTD, dan Pt-Mean Corr tidak memenuhi kriteria maka butir pertanyaan tidak valid sehingga perlu diperbaiki dan harus diganti ataupun harus dibuang.

Selanjutnya yaitu uji reliabilitas yang bertujuan untuk mengetahui kekonsistenan instrumen yang dikembangkan. Untuk mencapai hal tersebut, maka dilakukan uji reliabilitas menggunakan *model Rasch* dengan berbantuan *software Winstep 4.7.0* dengan menggunakan formula *alpha Cronbach*. Pada penelitian ini menggunakan dua analisis reliabilitas yaitu *item reliability* dan *person reliability*. Kriteria *Alpha Cronbach* terdapat pada Tabel 12.

Tabel 12. Kriteria *Alpha Cronbach*

Nilai	Kriteria
>0,80	Bagus Sekali
0,71 – 0,80	Bagus
0,61 – 0,70	Cukup
0,51-0,60	Jelek
<0,50	Buruk

(Sumintono & Widhiarso, 2015).

Berdasarkan Tabel 12 , instrumen yang dikembangkan terkategori reliabel jika memperoleh minimal skor *Alpha Cronbach* sebesar 0,61 dengan kriteria cukup.

Tabel 13. Kriteria *Person Reliability* dan *Item Reliability*

Nilai	Kriteria
>0,94	Istimewa
0,91 – 0,94	Bagus sekali
0,81 – 0,90	Bagus
0,67-0,80	Cukup
<0,67	Lemah

(Sumintono & Widhiarso, 2015).

Berdasarkan Tabel 13, butir soal yang dikembangkan terkategori dapat berfungsi dengan normal untuk melakukan pengukuran jika *item reliability* dan *person reliability* memperoleh minimal skor 0,67 dengan kriteria cukup.

Setelah dilakukan analisis validitas dan reliabilitas dengan menggunakan Model Rasch melalui perangkat lunak Winstep versi 4.7.0, langkah selanjutnya adalah menganalisis tingkat kesukaran tiap butir soal. Tingkat kesukaran ditunjukkan oleh nilai logit (*Measure*) yang dihasilkan dari output. Nilai ini mencerminkan posisi item dalam skala kemampuan responden. Soal dalam kategori baik jika masih dalam kategori $-2 \leq b \leq 2$ (Hambleton & Swaminathan 1985). Jika mendekati -2 soal dikatakan mudah, sementara jika mendekati +2 soal dikatakan sulit.

Selanjutnya dilakukan analisis *Test Information Function* (TIF) untuk mengetahui tingkat ketepatan instrumen dalam mengukur kemampuan pada berbagai level kemampuan laten siswa. Analisis TIF dilakukan menggunakan Model Rasch melalui perangkat lunak Winsteps versi 4.7.0 dengan melihat kurva informasi tes yang dihasilkan. Kurva TIF menunjukkan besarnya informasi yang diberikan instrumen pada setiap rentang kemampuan dalam skala logit. Semakin tinggi nilai informasi pada suatu rentang kemampuan, maka semakin baik ketepatan pengukuran instrumen pada rentang tersebut (Sumintono & Widhiarso, 2015).

Analisis *Differential Item Functioning* (DIF) dilakukan untuk mengidentifikasi apakah terdapat butir yang berpotensi bias terhadap kelompok tertentu. Analisis DIF bertujuan untuk memastikan bahwa setiap butir memiliki peluang yang sama untuk dijawab benar oleh siswa yang memiliki kemampuan setara, tanpa

dipengaruhi oleh faktor kelompok seperti jenis kelamin atau kategori lainnya. Butir dinyatakan tidak terindikasi bias apabila nilai probabilitas (Sig.) lebih besar dari 0,05 (Sumintono & Widhiarso, 2015).

3.5.3 Keterampilan Berpikir Kritis

Hasil penelitian untuk setiap peserta tes dikategorikan menjadi beberapa tingkat kemampuan (Azwar, 2010), terdapat pada Tabel 14:

Tabel 14. Interval Keterampilan Berpikir Kritis Siswa

Interval	Kategori
$M_i + 1,5 \text{ sbi} < \theta$	Sangat Tinggi
$M_i + 0,5 \text{ sbi} < \theta \leq M_i + 1,5 \text{ sbi}$	Tinggi
$M_i - 0,5 \text{ sbi} < \theta \leq M_i + 0,5 \text{ sbi}$	Sedang
$M_i - 1,5 \text{ sbi} < \theta \leq M_i - 0,5 \text{ sbi}$	Rendah
$0 < \theta \leq M_i - 1,5 \text{ sbi}$	Sangat Rendah

Keterangan:

$M_i = 1/2 (\theta \text{ maksimum} + \theta \text{ minimum})$

$\text{sbi} = 1/6 (\theta \text{ maksimum} - \theta \text{ minimum})$

θ = kemampuan peserta tes

3.5.4 Uji Kepraktisan

Uji kepraktisan ini dilakukan dengan memberikan angket kepada siswa.

Pemberian angket kepraktisan ini bertujuan untuk mengetahui tanggapan siswa yang dapat dijadikan tolak ukur kualitas instrumen yang telah dikembangkan.

Perolehan data angket respon selanjutnya dianalisis dengan menghitung nilai rata-rata dari setiap jawaban untuk mengetahui kepraktisan instrumen penilaian menggunakan rumus Sudjana, (2005) berikut.

$$P = \frac{\sum n}{\sum n_{maks}} \times 100$$

Keterangan :

P = Persentase kriteria suatu produk (%)

$\sum n$ = Jumlah skor jawaban dari tiap aspek

$\sum n_{maks}$ = Jumlah skor maksimal dari tiap aspek

Analisis kriteria kepraktisan dilakukan dengan meninjau antara kriteria skala kepraktisan produk dengan persentase nilai rata-rata suatu produk. Interval kriteria kepraktisan ditinjau dari angket yang dijelaskan pada Tabel 15.

Tabel 15. Interpretasi Uji Kepraktisan

Persentase	Tingkat Kepraktisan
75,01% - 100%	Sangat Praktis
50,01% - 75,00%	Praktis
25,01% - 50,00%	Kurang Praktis
0,00 % - 25,00%	Sangat Kurang Praktis

(Riduwan, 2004).

Berdasarkan Tabel 15, produk dinyatakan layak digunakan apabila memperoleh persentase kepraktisan minimal berada pada kategori praktis yaitu 50,01%–75,00%. Apabila persentase yang diperoleh berada di atas rentang tersebut, maka tingkat kepraktisan produk termasuk kategori sangat praktis, yang menunjukkan kualitas kepraktisan yang lebih baik.

V. SIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan pada tahap pengembangan instrumen penilaian berpikir kritis fisika berbasis *Adaptive Assessment* pada materi fluida statis, dapat disimpulkan beberapa hal sebagai berikut.

1. Penelitian pengembangan ini menghasilkan instrumen penilaian berbasis *Adaptive Assessment* yang dirancang untuk mengukur keterampilan berpikir kritis siswa. Instrumen ini disusun dengan sistem adaptif, sehingga tingkat kesulitan soal menyesuaikan dengan respons siswa pada butir sebelumnya. Keterampilan berpikir kritis yang diukur meliputi kemampuan menganalisis permasalahan, mengevaluasi informasi, menarik kesimpulan logis, serta memberikan alasan yang tepat berdasarkan konsep yang dipelajari. Indikator keterampilan berpikir kritis dioperasionalkan ke dalam butir soal yang disusun secara bertingkat, sehingga siswa dituntut menunjukkan proses berpikir yang mendalam dan sistematis. Proses penilaian menggunakan pedoman penskoran yang objektif dan terstruktur. Berdasarkan hasil validasi para ahli, instrumen dinyatakan sangat valid ditinjau dari aspek konstruk, materi, dan bahasa.
2. Instrumen penilaian berbasis *Adaptive Assessment* tersebut dinyatakan valid secara empiris dan reliabel berdasarkan hasil uji coba. Analisis data menunjukkan bahwa butir-butir soal memenuhi kriteria validitas empiris serta reliabilitas yang baik. Dengan demikian, instrumen ini layak digunakan untuk mengukur keterampilan berpikir kritis siswa dalam pembelajaran.
3. Dari segi kepraktisan, instrumen berbasis *Adaptive Assessment* dinilai sangat praktis untuk diterapkan dalam proses pembelajaran. Hal ini dibuktikan melalui hasil penilaian kepraktisan yang berada pada kategori sangat tinggi. Tingginya

tingkat kepraktisan menunjukkan bahwa sistem adaptif mudah digunakan oleh guru dan siswa, efisien dalam pelaksanaan, serta mampu memberikan informasi yang akurat mengenai tingkat keterampilan berpikir kritis masing-masing siswa.

5.2 Saran

Berdasarkan simpulan hasil penelitian yang telah diperoleh, peneliti menyampaikan beberapa saran sebagai berikut.

1. Guru fisika disarankan untuk memanfaatkan instrumen penilaian berbasis *Adaptive Assessment* sebagai alternatif alat evaluasi dalam mengukur keterampilan berpikir kritis siswa. Penggunaan instrumen ini dapat membantu guru memperoleh gambaran kemampuan siswa secara lebih akurat karena tingkat kesulitan soal menyesuaikan dengan respons siswa, sehingga mendukung pembelajaran yang berorientasi pada pengembangan kemampuan berpikir tingkat tinggi.
2. Peneliti selanjutnya disarankan untuk mengembangkan instrumen penilaian berbasis *Adaptive Assessment* pada materi fisika lainnya atau mengintegrasikannya dengan pendekatan dan teknologi yang berbeda. Selain itu, pengembangan juga dapat diarahkan untuk mengukur keterampilan abad ke-21 lainnya, seperti kemampuan pemecahan masalah, kreativitas, maupun literasi sains.
3. Guru diharapkan dapat memanfaatkan hasil pengukuran keterampilan berpikir kritis siswa sebagai dasar dalam merancang tindak lanjut pembelajaran

DAFTAR PUSTAKA

- Abidin, A. Z., Istiyono, E., Fadilah, N., & Dwandaru, W. S. B. (2019c). A computerized adaptive test for measuring the physics critical thinking skills. *International Journal of Evaluation and Research in Education*, 8(3), 376–383. <https://doi.org/10.11591/ijere.v8i3.19642>
- Adeoye, M. A., Prastikawati, E. F., & Riwayatinationsih, R. (2025). Beyond Traditional Testing: Exploring the Efficacy of *Adaptive Assessments* in Shaping Future Learners. *Journal of Nonformal Education*, 11(1), 153–169. <https://doi.org/10.15294/jone.v11i1.19934>
- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Ali, M., Ikbali, M. S., & Jusman. (2023). Determining the Surface Tension of a Liquid And the Drop Comparison Method. *JPF (Jurnal Pendidikan Fisika) Universitas Islam Negeri Alauddin Makassar*, 11(1), 143–150. <https://doi.org/10.24252/jpf.v11i1.34113>
- Angelo, T. A. (1995). Classroom assessment for critical thinking. *Teaching of Psychology*, 22, 6-7.
- Arifin, Z. 2016. *Evaluasi Pembelajaran (Prinsip, Teknik, dan. Prosedur)*, Cetakan Kedelapan, Jakarta: Rosda Karya
- Arini, W., & Juliadi, F. (2018). Analisis Kemampuan Berpikir Kritis Pada Mata Pelajaran Fisika Untuk Pokok Bahasan Vektor Siswa Kelas X Sma Negeri 4 Lubuklinggau, Sumatera Selatan. *Berkala Fisika Indonesia*, 10.
- Asniar, N., & Khaerudin. (2022). Analisis Keterampilan Berpikir Kritis Dalam Pembelajaran Fisika Siswa Di Sman 11 Makassar. *Jurnal Sains Dan Pendidikan Fisika (JSPF)*.
- Astutik, S., Mahardika, I. K., Supeno, Indrawati, Sudarti, & Zakaria, A. F. (2020). Development of meaningful investigation laboratory (mil) learning model to improve critical thinking skills in physics learning. *IOP Conference Series: Earth and Environmental Science*, 485(1). <https://doi.org/10.1088/1755-1315/485/1/012112>

- Awopeju, O. A. , & Afolabi, E. R. (2016). Comparative Analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of Senior School Certificate Mathematics Examination. *European Scientific Journal, ESJ*, 12(28), 263. <https://doi.org/10.19044/esj.2016.v12n28p263>
- Aybek, E. C., & Demirtasli, R. N. (2017). Computerized adaptive test (Cat) applications and item response theory models for polytomous items. *International Journal of Research in Education and Science*, 3(2), 475–487. <https://doi.org/10.21890/ijres.327907>
- Azwar, S. 2010. Metode Penelitian. Yogyakarta: Pustaka Pelajar.
- Badriyah, L. N., Thamrin, A. G., & Nurhidayati, A. (2018). `Analisis Instrumen Penilaian Hasil Belajar. *IJCEE*.
- Baker, F. B., & Kim, S.-H. (2004). *Item Response Theory* (F. B. Baker & S.-H. Kim, Eds.). CRC Press. <https://doi.org/10.1201/9781482276725>
- Baker, F. B., & Kim, S.-H. (2017). *The Basics of Item Response Theory Using R*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-54205-8>
- Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, 62(8), 750–762. <https://doi.org/10.1119/1.17449>
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Berek, F. X., Sutopo, S., & Munzil, M. (2016). Concept enhancement of junior high school students in hydrostatic pressure and archimedes law by predict-observe-explain strategy. *Jurnal Pendidikan IPA Indonesia*, 5(2), 230–238. <https://doi.org/10.15294/jpii.v5i2.6038>
- Bessas, N., Tzanaki, E., Vavougiou, D., & Plagianakos, V. P. (2024). Diagnosing students' misconception in Hydrostatic Pressure through a 4-tier test. *Heliyon*, 10(23), e40425. <https://doi.org/10.1016/j.heliyon.2024.e40425>
- Bichi, A. A., & Talib, R. (2018). Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development. *International Journal of Evaluation and Research in Education (IJERE)*, 7(2), 142–151. <https://doi.org/10.11591/ijere.v7.i2.pp142-151>
- Bimpeh, Y., & West, D. (2022). *Adaptive testing: an introduction Assessment Research and Innovation*.

- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Netherlands. <https://doi.org/10.1007/978-94-007-6857-4>
- Burr, S. A., Gale, T., Kisielewska, J., Millin, P., Pêgo, J. M., Pinter, G., Robinson, I. M., & Zahra, D. (2023). A narrative review of adaptive testing and its application to medical education. *MedEdPublish*, 13, 221. <https://doi.org/10.12688/mep.19844.1>
- Clark, R. C., Mayer, R. E., & Thalheimer, W. (2003). E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning. *Performance Improvement*, 42(5), 41–43. <https://doi.org/10.1002/pfi.4930420510>
- Davey.(2011). *A Guide to Computer Adaptive Testing Systems*. the Council of Chief State School Officers, Washington, DC. www.ccsso.org
- Diani, R., & Hartati, N. S. (2018). Flipbook berbasis literasi Islam: Pengembangan media pembelajaran fisika dengan 3D pageflip professional. *Jurnal Inovasi Pendidikan IPA*, 4(2), 234–244. <https://doi.org/10.21831/jipi.v4i2.20819>
- El-Sabagh, H. A. (2021). Adaptive e-learning environment based on learning styles and its impact on development students' engagement. *International Journal of Educational Technology in Higher Education*, 18(1), 53. <https://doi.org/10.1186/s41239-021-00289-4>
- Embretson, S. E., & Reise, S. P. (2013). *Item Response Theory*. Psychology Press. <https://doi.org/10.4324/9781410605269>
- Ennis, R. H. (2009). Critical thinking assessment. *Theory Into Practice*, 32(3), 179–186. <https://doi.org/10.1080/00405849309543594>
- Ertikanto, C., Viyanti, & Hardini, S. N. (2023). Effectiveness of Guided Inquiry Learning Model on Thermodynamics Content in LMS-Based Blended Learning with Regard to Students' Higher Order Thinking Skills. *Jurnal Penelitian Pendidikan IPA*, 9(12), 11280–11288. <https://doi.org/10.29303/jppipa.v9i12.4026>
- Facione, P.A. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction, Research Findings and Recommendations*. California: California State University, Fullerton.
- Facione, P. A. (2020). Critical Thinking: What It Is and Why It Counts. *Proposal for a Cluster of Excellence: Vol. XXVIII (Issue 1)*.

- Fajrianthi, F., Hendriani, W., & Septarini, B. G. (2016). Pengembangan Tes Berpikir Kritis Dengan Pendekatan Item Response Theory. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(1), 45–55. <https://doi.org/10.21831/pep.v20i1.6304>
- Fauziyah, A., As Sakinah, Z., Mariyanto, & Juansah, dase E. (2023). Instrumen Tes dan Non Tes pada Penelitian. *Pendas : Jurnal Ilmiah Pendidikan Dasar*.
- Febriani Yudhittara, R., & Hindarto, N.. (2017). Identifikasi Miskonsepsi Menggunakan Cri Dan Penyebabnya Pada Materi Mekanika Fluida Kelas Xi Sma. *Unnes Physics Education Journal* , 6(2). <http://journal.unnes.ac.id/sju/index.php/upej>
- Finkelman, M., Kim, W., Weissman, A., & Cook, R. (2014). Cognitive Diagnostic Models and Computerized Adaptive Testing: Two New Item-Selection Methods That Incorporate Response Times. *Journal of Computerized Adaptive Testing*, 59–76. <https://doi.org/10.7333/1412-0204059>
- Firdausi, Firdausi, *et al.* "Students' Conceptual Errors on Hydrostatics Pressure and Bouyancy Theory." *Jurnal Pendidikan Sains Universitas Negeri Malang*, vol. 7, no. 1, 23 Mar. 2019, doi:10.17977/jps.v7i1.12501.
- Fitria, A., Anggreini, A., Rosidin, U., Viyanti, V., & Suana, W. (2025). Effectiveness Problem Based Learning On Students Critical Thinking Skills To Support Esd. *EduFisika: Jurnal Pendidikan Fisika*, 10(2), 249–258. <https://doi.org/10.59052/edufisika.v10i2.44181>
- Furqaan, F. M., Rasyid, A., & Gaffar, A. A. (2023). Pengembangan Soal-Soal Pilihan Ganda Beralasan Menggunakan Google Form. *PedagogiBiologi*, 15–23. <https://ejournal.unma.ac.id/index.php/bp/>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Halkiopoulous, C., & Gkintoni, E. (2024). Leveraging AI in E-Learning: Personalized Learning and *Adaptive Assessment* through Cognitive Neuropsychology—A Systematic Analysis. *Electronics*, 13(18), 3762. <https://doi.org/10.3390/electronics13183762>
- Halpern, D. F. (2014). *Thought and knowledge: An introduction to critical thinking* (5th ed.). Psychology Press.
- Hambleton, R. K. ., Swaminathan, Hariharan., & Rogers, H. Jane. (2012). *Fundamentals of item response theory*. SAGE Publications, Inc.

- Hambleton, Ronald. K., & Swaminathan, H. (1985). *Items response theory: Principles and application*. Kluwer-Nijhoff Publish.
- Harizah, Z., Kusairi, S., & Latifah, E. (2019). Penguasaan Konsep Fluida Statis Siswa SMA. *Jurnal Pendidikan : Teori, Penelitian, Dan Pengembangan*. <http://journal.um.ac.id/index.php/jptpp/>
- Huang, H.-Y. (2018). Effects of Item Calibration Errors on Computerized Adaptive Testing under Cognitive Diagnosis Models. *Journal of Classification*, 35(3), 437–465. <https://doi.org/10.1007/s00357-018-9265-y>
- Inanna, Rahmatullah, & Hasan, M. (2021). *Evaluasi Pembelajaran: Teori dan Praktek*. Tahta Media Grup.
- Irawati, R. K., Sofianto, E. W. N., Assidiqi, H., & Rahmawati, I. (2022). Pascal or Archimedes: Which Misconception is Higher? *Journal of Physics: Conference Series*, 2392(1), 012030. <https://doi.org/10.1088/1742-6596/2392/1/012030>
- Islamyati, M. P., & Manuaba, I. B. S. (2021). Multimedia Interactive Learning in Science Subjects for Grade Fourth Elementary School Students. *Indonesian Journal Of Educational Research and Review*, 4(2), 201. <https://doi.org/10.23887/ijerr.v4i2.39664>
- Istiyono, E. (2020). Computer Adaptive Test as The Appropriate Model to Assess Physics Achievement in 21 st Century. *Advances in Social Science, Education and Humanities Research*.
- Istiyono, E., Dwandaru, W. B., & Rahayu, F. (2018). The Developing of Creative Thinking Skills Test Based on Modern Test Theory in Physics of Senior High Schools. *Jurnal Cakrawala Pendidikan*. <https://doi.org/10.21831/cp.v37i2.19233>
- Istiyono, E., Dwandaru, W. S. B., Lede, Y. A., Rahayu, F., & Nadapdap, A. (2019). Developing IRT-based physics critical thinking skill test: A CAT to answer 21st century challenge. *International Journal of Instruction*, 12(4), 267–280. <https://doi.org/10.29333/iji.2019.12417a>
- Istiyono, E., Dwandaru, W. S. B., Setiawan, R., & Megawati, I. (2020). Developing of computerized adaptive testing to measure physics higher order thinking skills of senior high school students and its feasibility of use. *European Journal of Educational Research*, 9(1), 91–101. <https://doi.org/10.12973/eu-jer.9.1.91>

- Istiyono, E., Mardapi, D., & Suparno, S. (2014). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (Pysthots) Siswa Sma. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 18(1), 1–12. <https://doi.org/10.21831/pep.v18i1.2120>
- Jannah, K., & Pahlevi, T. (2020). Pengembangan Instrumen Penilaian Berbasis Higher Order Thinking Skills Berbantuan Aplikasi “Kahoot!” Pada Kompetensi Dasar Menerapkan Penanganan Surat Masuk dan Surat Keluar Jurusan OTKP di SMK Negeri 2 Buduran. *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*. <https://journal.unesa.ac.id/index.php/jpap>
- Kamilah, D. S., Muki, B. G., Aviyanti, L., & Suhandi, A. (2025). Review of misconceptions in physics among Indonesian high school students: Diagnosis, causes, and remediation. *Momentum: Physics Education Journal*, 9(2), 251–263. <https://doi.org/10.21067/mpej.v9i2.11187>
- Kemendikbud. (2019). *Pendidikan Di Indonesia Belajar dari PISA 2018*. Kemendikbud.
- Khaerudin. (2016). Teknik Penskoran Tes Obyektif Model Pilihan Ganda. *Jurnal Madaniyah*.
- Khan, G. (2015). Using Item analysis on essay types questions given in summative examination of medical college students: Facility value, Discrimination index. *International Journal of Research in Medical Sciences*, 1. <https://doi.org/10.5455/2320-6012.ijrms20150131>
- Khotimah, H., & Mindyarto, B. N. (2021). Pengembangan Computerized Adaptive Test (CAT) Pada Materi Getaran, Gelombang dan Bunyi. *Unnes Physics Education Journal*.
- Khusaini, K., Azizah, N., Suwasono, P., Yogihati, C. I., & Andriani, A. D. (2025). Rasch Model Application: Identification of Students’ Conceptual Understanding of Static Fluid. *Jurnal Pendidikan Fisika Indonesia*, 21(1), 54–68. <https://doi.org/10.15294/jpfi.v21i1.15023>
- Koğar, E. Y., & Koğar, H. (2018). Examination Of Dimensionality And Latent Trait Scores On Mixed-Format Tests. *PUPIL: International Journal of Teaching, Education and Learning*, 2(1), 29–49. <https://doi.org/10.20319/pijtel.2018.21.2949>
- Lawson, A. E. (2004). The Nature and Development of Scientific Reasoning: A Synthetic View. *International Journal of Science and Mathematics Education*, 2(3), 307–338. <https://doi.org/10.1007/s10763-004-3224-2>

- Lestari, N., Nisa, C., Mutazam, M., & Soeharto, S. (2023). Developing Two Tier Test of Critical Thinking Ability using Rasch Model. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, 8(2), 259. <https://doi.org/10.26737/jipf.v8i2.3737>
- Ling, G., Attali, Y., Finn, B., & Stone, E. A. (2017a). Is a Computerized Adaptive Test More Motivating Than a Fixed-Item Test? *Applied Psychological Measurement*, 41(7), 495–511. <https://doi.org/10.1177/0146621617707556>
- Magdalena, I., Aqmarani, A., Nurhalisa, N., & Syahra, N. P. (2023). Perbandingan Penggunaan Tes Objektif dan Tes Subjektif terhadap Hasil Belajar. *YASIN*, 3(4), 710–720. <https://doi.org/10.58578/yasin.v3i4.1316>
- Mardapi, D. (2017). *Pengukuran Penilaian dan Evaluasi Pendidikan Edisi 2*. Yogyakarta: Nuha Medika
- Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, 36(1), 8. <https://doi.org/10.1007/s10648-023-09842-1>
- McNeill, K. L., & Krajcik, J. (2009). Synergy Between Teacher Practices and Curricular Scaffolds to Support Students in Using Domain-Specific and Domain-General Knowledge in Writing Arguments to Explain Phenomena. *Journal of the Learning Sciences*, 18(3), 416–460. <https://doi.org/10.1080/10508400903013488>
- Mukti, S. T., & Istiyono, E. (2018). Instrument for Assessing the Critical Thinking Ability of X Grade High School Students on Biology Learning. *BIOEDUKASI: Jurnal Pendidikan Biologi*, 11, 107. <https://doi.org/10.20961/bioedukasi-uns.v11i2.21624>
- Mulyadi. (2010). *Evaluasi Pendidikan: Pengembangan Model Evaluasi Pendidikan Agama di Sekolah*. UIN Maliki Press.
- Mundilarto, & Ismoyo, H. (2017). Effect of problem-based learning on improvement physics achievement and critical thinking of senior high school student. *Journal of Baltic Science Education*, 16(5), 761–779. <https://doi.org/10.33225/jbse/17.16.761>
- Nadlifah, Zahriani, N., & Latif, M. A. (2022). *Perkembangan Kognitif ANAK USIA DINI Teori dan Aplikasinya*.
- Naga, D. S. (2012). *Teori skor pada pengukuran mental*. PT Nagarani Citrayasa.

- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020. <https://doi.org/10.1002/tea.20035>
- OECD. (2023). *PISA 2022 Results (Volume I)*. OECD. <https://doi.org/10.1787/53f23881-en>
- Ozyurt, H., Ozyurt, O., Baki, A., & Guven, B. (2012). An Application of Individualized Assessment in Educational Hypermedia: Design of Computerized Adaptive Testing System and its Integration Into UZWEBMAT. *Procedia - Social and Behavioral Sciences*, 46, 3191–3196. <https://doi.org/10.1016/j.sbspro.2012.06.035>
- Pamungkas, Z. S., Aminah, N. S., & Nurosyid, F. (2018). Students Critical Thinking Skill in Solving Scientific Literacy using a Metacognitive Test Based on Scientific Literacy. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 7(2), 161–169. <https://doi.org/10.24042/jipfalbiruni.v7i2.2909>
- Perkins, C., & Murphy, E. (2006). Identifying and Measuring Individual Engagement in Critical Thinking in online Discussion. *Educational Technology & Society*, 298–307.
- Permana, C. (2024). *Belajar dan Pembelajaran di Abad 21*. Yayasan Kita Menulis.
- Petrosino, A. J., & Mann, M. J. (2017). The Challenges of Understanding Fluid in Fluid Density. *Journal of Continuing Education and Professional Development*. <https://doi.org/10.7726/jcepd.2017.1003>
- Puspita, A. T., & Jatmiko, B. (2013). Implementasi Model Pembelajaran Inkuiri Terbimbing (Guided Inquiry) Terhadap Keterampilan Berpikir Kritis Siswa Pada Pembelajaran Fisika Materi Fluida Statis Kelas Xi Di Sma Negeri 2 Sidoarjo. *Jurnal Inovasi Pendidikan Fisika*, 02(03), 121–125.
- Putri, R. E., & Subekti, H. (2021). Analisis Miskonsepsi Menggunakan Metode Four- Tier Certainty Of Response Index: Studi Eksplorasi Di Smp Negeri 60 Surabaya. *Pensa E-Jurnal: Pendidikan Sains*, 9(2), 220–226.
- Rahmatin, J. A., Respasari, B. N., & Syukur, A. (2024). Efektivitas Model Pembelajaran Cinqase Untuk Meningkatkan Kemampuan Berpikir Kritis Siswa Pada Materi Fluida Statis. *Journal of Classroom Action Research*, 6(2). <https://doi.org/10.29303/jppipa.v6i2.8222>
- Rahmawati, A. S., & Dewi, R. P. (2019). Penggunaan Multimedia Interaktif (MMI) Sebagai Media Pembelajaran Dalam Meningkatkan Prestasi Belajar Fisika.

- Jurnal Pendidikan Fisika Dan Teknologi*, 5(1), 50–58.
<https://doi.org/10.29303/jpft.v5i1.958>
- Ratumanan, T. G., & Laurent, T. (2011). *Penilaian Hasil Belajar pada Tingkat Satuan Pendidikan (Edisi 2)*. Unesa University Press.
- Redhana, W. (2020). Mengembangkan Keterampilan Abad ke-21 dalam Pembelajaran Kimia. *Jurnal Inovasi Pendidikan Kimia*.
- Rezaie, M., & Golshan, M. (2015). Computer Adaptive Test (CAT): Advantages and Limitations. *International Journal of Educational Investigations Available Online @ Wwww.Ijeionline.Com*, 2(5), 128–137. www.ijeionline.com
- Riduwan. (2004). *Pengantar Statsitika Sosial*. Alfabeta.
- Ringo, E. S., Kusairi, S., & Latifah, E. (2023). Profil Kemampuan Pemecahan Masalah Siswa SMA pada Materi Fluida Statis. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 178–187.
<http://journal.um.ac.id/index.php/jptpp/>
- Rofiqoh, A., & Khairani, I. (2024). Peran Media Interaktif Dalam Meningkatkan Motivasi Belajar Siswa pada Mata Pelajaran SKI di Madrasah Ibtidaiyah. *Jurnal AL-AZHAR INDONESIA SERI HUMANIORA*, 9(1).
<https://doi.org/10.36722/sh.v9i1.2828>
- Rosdiana, S. R., & Kusairi, S. (2019). Kemampuan Berpikir Kritis Siswa SMA pada Materi Fluida Statis. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 4, 731–737. <http://journal.um.ac.id/index.php/jptpp/>
- Rosidin, U. (2017). *Asesmen Pembelajaran Sains Berbasis Pendekatan Keterampilan Proses Sains dan Berpikir Tingkat Tinggi*. Aura Publisher.
- Ross, B., Chase, A. M., Robbie, D., Oates, G., & Absalom, Y. (2018). Adaptive quizzes to increase motivation, engagement and learning outcomes in a first year accounting unit. *International Journal of Educational Technology in Higher Education*, 15(1). <https://doi.org/10.1186/s41239-018-0113-2>
- Roudlo, M. (2020). Kemampuan Berpikir Kritis dan Kemandirian Belajar Melalui Model Pembelajaran Flipped Classroom dengan Pendekatan STEM. *SEMINAR NASIONAL PASCASARJANA 2*.
- Ruggiero. (2012). *Philosophy and Religion Beyond Feelings: A Guide to Critical Thinking 9th Edition Philosophy and Religion*. The McGraw-Hili Companies.
<http://www.primisonline.com>

- Samadun, S., & Dwikoranto, D. (2022). Improvement of Student's Critical Thinking Ability sin Physics Materials Through The Application of Problem-Based Learning. *IJORER : International Journal of Recent Educational Research*, 3(5), 534–545. <https://doi.org/10.46245/ijorer.v3i5.247>
- Samaduri, A. (2022). Analisis Pemahaman Konsep Siswa Yang Diukur Menggunakan Tes Pilihan Ganda Beralasan Pada Mata Pelajaran Biologi. *JURNAL PENDIDIKAN GLASSER*, 6(1), 109. <https://doi.org/10.32529/glasser.v6i1.1466>
- Saputra, H. (2024). BERSATU: Jurnal Pendidikan Bhinneka Tunggal Ika Penguatan Kemampuan Siswa Dalam Menghadapi Era Society 5.0 Melalui Pembelajaran Matematika. *BERSATU: Jurnal Pendidikan Bhinneka Tunggal Ika*, 387–302. <https://doi.org/10.51903/bersatu.v2i2.640>
- Saputra, I. G. E., Jampel, I. N., & Parwata, I. G. L. A. (2022). Pengembangan Instrumen Penilaian Kemampuan Berpikir Kritis Siswa Kelas VIII SMP pada Materi Getaran dan Gelombang. *JURNAL PENDIDIKAN DAN PEMBELAJARAN SAINS INDONESIA (JPPSI)*.
- Saputra, O., Setiawan, A., & Rusdiana, D. (2019). Identification of student misconception about static fluid. *Journal of Physics: Conference Series*, 1157, 032069. <https://doi.org/10.1088/1742-6596/1157/3/032069>
- Septy, D., I*, I., Agustin, E. W., Hakim, L., & Pratiwi4, V. (2024). Edukatif: Jurnal Ilmu Pendidikan Analisis Butir Soal Pilihan Ganda pada Siswa SMK Menggunakan Anates. *Jurnal Ilmu Pendidikan*, 6. <https://doi.org/10.31004/edukatif.v6i6.7780>
- Shaaban, K. A., & Ghaith, G. (2014). The Theoretical Relevance and Efficacy of Using Cooperative Learning in the ESL/EFL Classroom Assessment of critical thinking skills through reading comprehension. *International Journal of Language Studies*, 8(2), 117–140. <https://www.researchgate.net/publication/277801105>
- Shabrina, A., & Astuti, U. P. (2022). The Integration of 6Cs of the 21 st Century Education into English Skills: Teachers' Challenges and Solutions. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 7, 28–37. <http://journal.um.ac.id/index.php/jptpp/>
- Shaw, A., Liu, O. L., Gu, L., Kardonova, E., Chirikov, I., Li, G., Hu, S., Yu, N., Ma, L., Guo, F., Su, Q., Shi, J., Shi, H., & Loyalka, P. (2020). Thinking critically about critical thinking: validating the Russian HEIghten® critical thinking

- assessment. *Studies in Higher Education*, 45(9), 1933–1948.
<https://doi.org/10.1080/03075079.2019.1672640>
- Sudaryono. (2011). Implementasi Teori Responsi Butir (Item Response Theory) Pada Penilaian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan Dan Kebudayaan*, 17(6), 719–732. <https://doi.org/10.24832/jpnk.v17i6.62>
- Sudijono, A. (2009). *Pengantar Evaluasi Pendidikan*. PT. Raja Grafindo Persada.
- Sudjana. (2005). *Metode Statistika*. Tarsito.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*.
- Supena, I., Darmuki, A., & Hariyadi, A. (2021). The influence of 4C (constructive, critical, creativity, collaborative) learning model on students' learning outcomes. *International Journal of Instruction*, 14(3), 873–892.
<https://doi.org/10.29333/iji.2021.14351a>
- Syamsinar, S., Sukmawati, & Rosleny B. (2022). Pengembangan Instrumen Penilaian Berbasis Kinerja Dalam Pembelajaran Matematika Siswa. *Edum Journal*, 5(1).
- Tiagarajan, S. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Indiana University.
- Tiruneh, D. T., De Cock, M., Weldelessie, A. G., Elen, J., & Janssen, R. (2017). Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism. *International Journal of Science and Mathematics Education*, 15(4), 663–682.
<https://doi.org/10.1007/s10763-016-9723-0>
- Toulmin, S. E. (2003). *The Uses of Argument*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511840005>
- Varas, D., Santana, M., Nussbaum, M., Claro, S., & Imbarack, P. (2023). Teachers' strategies and challenges in teaching 21st century skills: Little common understanding. *Thinking Skills and Creativity*, 48, 101289.
<https://doi.org/10.1016/j.tsc.2023.101289>
- Vincent-Lancrin, S., González-Sancho, C., Bouckaert, M., de Luca, F., Fernández-Barrerra, M., Jacotin, G., Urgel, J., & Vidal, Q. (2019). *Fostering Students' Creativity and Critical Thinking*. OECD. <https://doi.org/10.1787/62212c37-en>

- Viyanti, V., Cari, C., Sunarno, W., & Prasetyo, Z. K. (2020). Reconstructing of Higher Order Thinking Skill through Enriching Student's Argumentation Skills. *Jurnal Pendidikan Progresif*, 10(2), 327–335. <https://doi.org/10.23960/jpp.v10.i2.202016>
- Viyanti, V., Cari, C., Sunarno, W., & Prasetyo, Z. K. (2017). Level of Skill Argued Students on Physics Material. *Journal of Physics: Conference Series*, 895, 012043. <https://doi.org/10.1088/1742-6596/895/1/012043>
- Wardani, Y., Rosidin, U., & Viyanti. (2013). Perbandingan Hasil Belajar Menggunakan Bentuk Tes Ditinjau Dari Kemampuan Berpikir Kritis. *Jurnal Pembelajaran Fisika Universitas Lampung*.
- Weiss, D. J. (1982). Improving Measurement Quality and Efficiency with Adaptive Testing. *Applied Psychological Measurement*, 6(4), 473–492. <https://doi.org/10.1177/014662168200600408>
- Weiss, D. J. ., & Bock, R. Darrell. (1983). *New horizons in testing : latent trait test theory and computerized adaptive testing*. Academic Press.
- Wenno, I. H., Limba, A., & Silahoy, Y. G. M. (2022). The development of physics learning tools to improve critical thinking skills. *International Journal of Evaluation and Research in Education*, 11(2), 863–869. <https://doi.org/10.11591/ijere.v11i2.21621>
- Winarno, W. (2013). Pengembangan Computerized Adaptive Testing (Cat) Menggunakan Metode Pohon Segitiga Keputusan. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 16(2), 574–592. <https://doi.org/10.21831/pep.v16i2.1132>
- Winarti, Cari, Suparmi, Sunarno, W., & Istiyono, E. (2017). Development of two tier test to assess conceptual understanding in heat and temperature. *Journal of Physics: Conference Series*, 795, 012052. <https://doi.org/10.1088/1742-6596/795/1/012052>
- Wise, S. L. (2014). The Utility of Adaptive Testing in Addressing the Problem of Unmotivated Examinees. *Journal of Computerized Adaptive Testing*.
- Zakwandi, R., Istiyono, E., & Dwandaru, W. S. B. (2024). A two-tier computerized adaptive test to measure student computational thinking skills. *Education and Information Technologies*, 29(7), 8579–8608. <https://doi.org/10.1007/s10639-023-12093-w>

- Zebua, J. Y., Zega, Y., Niat Telaumbanua, Y., Studi, P., Matematika, P., & Nias, U. (2024). Analisis Kemampuan Berpikir Kritis Siswa dalam Menyelesaikan Soal Matematika. *Didaktika: Jurnal Kependidikan*, 13. <https://jurnaldidaktika.org>
- Zhang, Z., & Huang, X. (2024). Exploring the impact of the adaptive gamified assessment on learners in blended learning. *Education and Information Technologies*, 29(16), 21869–21889. <https://doi.org/10.1007/s10639-024-12708-w>
- Zheng, Y. (2016). Online Calibration of Polytomous Items Under the Generalized Partial Credit Model. *Applied Psychological Measurement*, 40(6), 434–450. <https://doi.org/10.1177/0146621616650406>
- Ziva Mahendra, K., Rohmah Diah, N., Laura Utami, V., & Marhadi, H. (2024). Analisis Penilaian Teknik Tes Dan Non Tes Pada Kelas Iv Di Sd Negeri 122 Pekanbaru. In *Jurnal Ilmiah Kajian Multidisipliner* (Vol. 8, Issue 5).
- Zou, Y., Kuek, F., Feng, W., & Cheng, X. (2025). Digital learning in the 21st century: trends, challenges, and innovations in technology integration. *Frontiers in Education*, 10. <https://doi.org/10.3389/educ.2025.1562391>
- Zukhruf, K. D., Khaldun, I., & Ilyas, S. (2016). Remediasi Miskonsepsi dengan Menggunakan Media Pembelajaran Interaktif Pada Materi Fluida Statis. *Jurnal Pendidikan Sains Indonesia*, 4, 56–68.
- Zumbo, Bruno D., (1999). *A Handbook on the Theory and Methods of Differential Item Functioning*, Canada: Directorate of Human Resources.