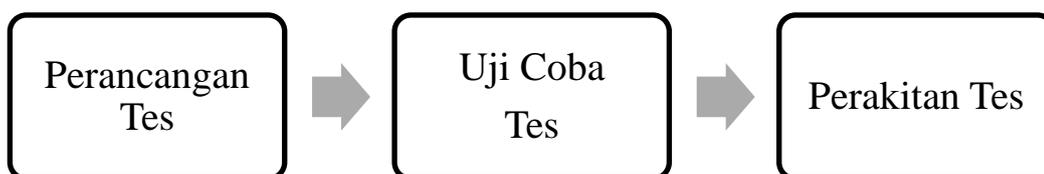


## BAB III METODE PENELITIAN

### 3.1. Desain Penelitian

Penelitian ini merupakan penelitian yang bertujuan mengembangkan suatu instrumen. Pengembangan instrumen ini menggunakan modifikasi Model Wilson, Model Oriondo dan Antonio (1998), yang dianggap paling sesuai dan efektif untuk proses pengembangan instrumen tes. Adapun tahapan yang dilalui dalam penelitian ini adalah: 1) perancangan tes, 2) uji coba tes, dan 3) perakitan tes.



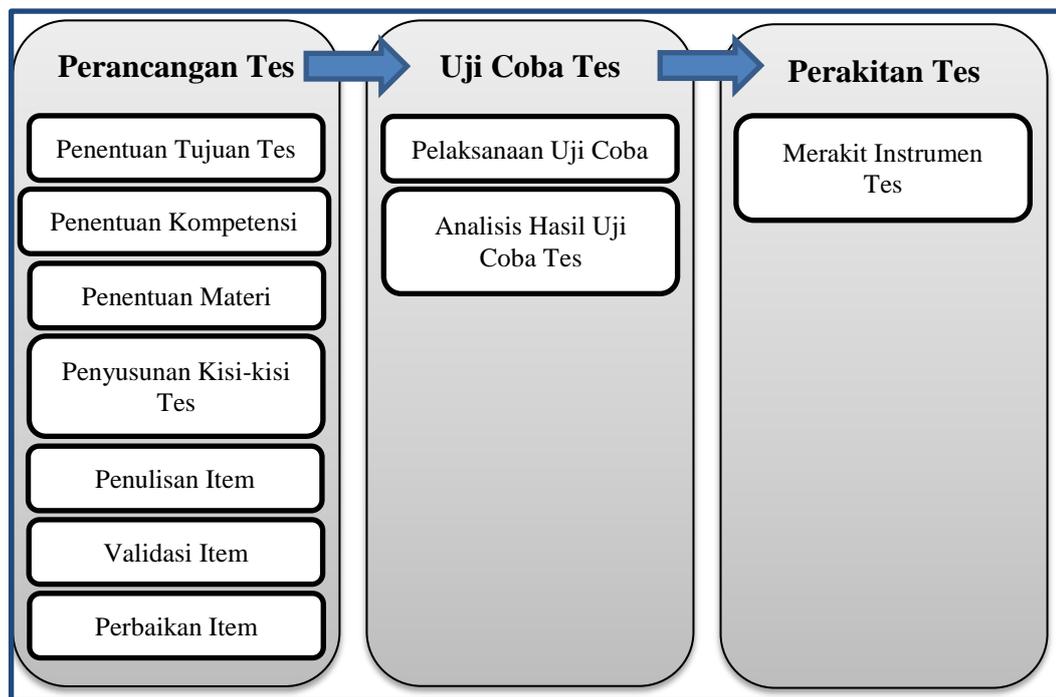
Gambar 3. 1 Tahapan Modifikasi Model Wilson, Model Oriondo & Antonio

### 3.2. Populasi dan Sampel

Populasi dalam penelitian ini yaitu peserta didik di dua SMA Negeri di Kabupaten Kuningan, Jawa Barat, dengan fokus pada peserta didik kelas XII yang telah belajar topik fluida statis. Populasi yang menjadi sasaran terdiri dari 100 peserta didik. Penelitian ini menggunakan teknik pengambilan sampel *convenience sampling*, dimana partisipan dipilih berdasarkan kemudahan akses dan kesediaan mereka untuk berpartisipasi. Teknik pengambilan sampel ini biasa digunakan saat peneliti membutuhkan sampel dengan cepat dan mudah. *Convenience sampling* menawarkan kepraktisan dalam pengumpulan data, meskipun memiliki keterbatasan dalam generalisasi hasil penelitian. Meskipun demikian, *convenience sampling* tetap menjadi teknik yang umum digunakan di berbagai bidang penelitian, seperti sosial, pendidikan, dan psikologi (Creswell & Creswell, 2018).

### 3.3. Prosedur Penelitian

Penelitian ini dilakukan berdasarkan model pengembangan modifikasi Model Wilson, Model Oriondo & Antonio, yang terdiri dari tiga tahapan, yaitu perancangan tes, uji coba tes, dan perakitan tes.



Gambar 3. 2 Tahapan Penelitian

## 1) Tahap Perancangan Tes

### a. Penentuan Tujuan Tes

Pada tahap ini, peneliti menetapkan tujuan utama dari tes diagnostik empat tingkat yang akan dikembangkan. Tujuannya adalah untuk menganalisis model mental peserta didik pada topik fluida statis. Peneliti mempertimbangkan aspek-aspek model mental yang akan digali, seperti pemahaman konseptual, kemampuan aplikasi, dan identifikasi miskonsepsi.

### b. Penentuan Kompetensi yang Akan Diujikan

Selanjutnya peneliti menganalisis kurikulum 2013, khususnya bagian Kompetensi Inti (KI) dan Kompetensi Dasar (KD) untuk tingkat SMA. Peneliti mengidentifikasi kompetensi spesifik yang berkaitan dengan pemahaman fluida statis. Kompetensi yang ditentukan mencakup berbagai aspek pemahaman fluida statis.

### c. Penentuan Materi yang Diujikan

Berdasarkan kompetensi yang telah ditetapkan, peneliti mengidentifikasi cakupan materi fluida statis yang akan diuji, yaitu tekanan hidrostatis, hukum Pascal, dan hukum Archimedes. Pemilihan topik ini didasarkan pada studi pustaka yang menunjukkan bahwa ketiga topik tersebut merupakan konsep fundamental dalam fluida statis. Selain itu, adanya miskonsepsi yang umum terjadi pada topik-topik ini, sehingga perlu dilakukan pengujian lebih lanjut.

#### **d. Penyusunan Kisi-Kisi Tes**

Kisi-kisi merupakan suatu pedoman dalam menulis atau merakit soal dengan format berbentuk matriks. Kisi-kisi ini mencakup indikator-indikator spesifik yang akan diukur untuk setiap kompetensi dan materi yang telah ditentukan. Kisi-kisi ini berfungsi sebagai panduan dalam pengembangan soal-soal tes, memastikan cakupan yang komprehensif dan proporsional.

#### **e. Penulisan Item Tes**

Setelah menyusun kisi-kisi, peneliti membuat 10 butir soal yang sesuai dengan indikator materi fluida statis. Soal-soal ini dirancang berdasarkan kisi-kisi tes yang telah ditentukan sebelumnya

#### **f. Validasi Item Tes**

Tahap berikutnya setelah penyusunan item tes adalah validasi. Tujuannya untuk memastikan item tes yang dihasilkan relevan dan dapat memenuhi kebutuhan belajar peserta didik.

Sebelum item tes diuji cobakan kepada peserta didik, dilakukan uji validitas konstruk dan isi berdasarkan pendapat para ahli (*judgement expert*) untuk mengetahui kualitas instrumen yang akan dikembangkan ini. Aspek yang akan dinilai oleh validator yaitu kesesuaian konsep dalam butir soal dengan konsep yang dikemukakan oleh ahli, relevansi butir soal dengan materi SMA, kemampuan setiap butir soal dalam menentukan konsepsi peserta didik, penggunaan bahasa yang sesuai dengan kaidah Ejaan Bahasa Indonesia, dan kesesuaian konstruksi butir soal dengan aturan penulisan pilihan ganda. Validator melakukan validasi terhadap instrumen penelitian berdasarkan aspek-aspek tersebut dengan mengisi lembar validasi yang berisi 3 kriteria validasi: Valid Tanpa Revisi (VTR), Valid Revisi (VR), dan Tidak Valid (TV). Untuk

memudahkan proses pengolahan data, setiap kriteria validasi diberi skor: Valid Tanpa Revisi (VTR=2), Valid Revisi (VR=1), dan Tidak Valid (TV=0).

#### **g. Perbaikan Item**

Berdasarkan hasil validasi, soal-soal yang memerlukan perbaikan direvisi sesuai masukan dan saran yang diberikan oleh ahli. Soal-soal yang telah direvisi kemudian dirangkai kembali menjadi instrumen tes yang layak dan siap untuk diujicobakan kepada peserta didik.

### **2) Tahap Uji Coba Tes**

#### **a. Pelaksanaan Uji Coba Tes**

Tujuan utama dari uji coba ini adalah untuk menguji kelayakan produk yang telah dikembangkan. Dari uji coba produk akan diketahui seberapa banyak kekurangan yang harus diperbaiki sehingga dapat dihasilkan sebuah produk yang mencapai sasaran dan tujuan pembelajaran, serta mendapatkan data yang dapat dianalisis karakteristik instrumen tersebut.

Uji coba dilakukan kepada 100 peserta didik kelas XI dengan memberikan soal-soal yang berkaitan dengan materi fluida statis. Data yang diperoleh dari hasil uji coba ini kemudian dianalisis untuk menilai kelayakan produk secara keseluruhan.

#### **b. Analisis Data Hasil Uji Coba**

Setelah peserta didik melaksanakan uji coba, peneliti mengolah dan menganalisis hasil jawaban peserta didik. Data hasil uji coba tersebut, kemudian diolah dan dianalisis untuk menganalisis model mental peserta didik pada materi fluida statis berdasarkan instrumen tes empat tingkat dan mengidentifikasi kualitas instrumen tes empat tingkat.

### **3) Tahap Perakitan Tes**

Setelah diperoleh hasil analisis data, dilakukan perbaikan terhadap butir-butir soal yang kurang baik. Soal-soal yang telah diperbaiki kemudian dirangkai kembali menjadi sebuah instrumen tes yang siap digunakan oleh peserta didik.

## **3.4. Instrumen Penelitian**

## 1. Lembar Soal Tes Diagnostik Empat Tingkat

Instrumen yang digunakan pada penelitian ini adalah instrumen tes empat tingkat untuk mengidentifikasi model mental peserta didik pada topik fluida statis. Instrumen tes empat tingkat ini merupakan pengembangan dari instrumen tes tiga tingkat.

Instrumen tes empat tingkat ini terdapat empat tingkatan pada setiap pertanyaan dalam instrumen ini. Pada tingkat pertama merupakan soal utama pilihan ganda dengan empat pilihan jawaban. Pada tingkat kedua merupakan soal tingkat keyakinan peserta didik dalam menjawab soal pada tingkat pertama. Pada tingkat ketiga merupakan soal yang menanyakan alasan peserta didik memilih opsi jawaban pada tingkat pertama. Pada tingkat keempat merupakan soal tingkat keyakinan alasan peserta didik dalam menjawab soal pada tingkat ketiga. Instrumen tes diagnostik empat tingkat ini bertujuan untuk mengidentifikasi model mental peserta didik pada topik fluida statis. Desain instrumen tes diagnostik empat tingkat ditunjukkan pada Tabel berikut.

Tabel 3. 1 Desain Instrumen Tes Diagnostik Empat Tingkat

Tingkat Pertama	<b>Pertanyaan:</b> <b>Mendiagnosis keadaan konsepsi peserta didik</b>
	<b>Jawaban pilihan:</b> A. Alternatif pilihan jawaban pertama B. Alternatif pilihan jawaban kedua C. Alternatif pilihan jawaban ketiga D. Alternatif pilihan jawaban keempat
Tingkat Kedua	<b>Skala penilaian keyakinan pertama</b> A. Pilihan alternatif jika “Yakin” atas jawaban yang dipilih B. Pilihan alternatif jika “Tidak Yakin” atas jawaban yang dipilih
Tingkat Ketiga	<b>Alasan:</b> A. Alternatif pilihan alasan pertama B. Alternatif pilihan alasan kedua C. Alternatif pilihan alasan ketiga D. Alternatif pilihan alasan keempat
Tingkat Keempat	<b>Skala Penilaian Keyakinan kedua:</b> A. Pilihan alternatif jika “yakin” atas alasan yang dipilih B. Pilihan alternatif jika “tidak yakin” atas alasan yang dipilih

## 2. Lembar Validasi Instrumen

Sebelum tes diagnostik empat tingkat ini digunakan untuk mengukur kemampuan peserta didik, peneliti perlu memastikan terlebih dahulu bahwa tes ini sudah dirancang dengan baik dan benar-benar mengukur apa yang akan diukur dengan melakukan tahap validasi. Dalam tahap validasi, beberapa ahli akan menilai tes berdasarkan lima indikator seperti pada tabel berikut.

Tabel 3. 2 Lembar Validasi Instrumen

No. Soal	Indikator Validasi	Validasi			Saran atau masukan
		VTR	VR	TV	
1	1. Kesesuaian konsep dalam setiap butir soal dengan konsep ilmiah yang dikemukakan para ahli				
	2. Relevansi setiap butir soal dengan materi SMA				
	3. Kemampuan setiap butir soal dalam menentukan pemahaman konsep peserta didik				
	4. Penggunaan bahasa yang sesuai dengan kaidah Ejaan Bahasa Indonesia				
	5. Kesesuaian konstruksi setiap butir soal dengan aturan penulisan soal pilihan ganda				
dst					

Keterangan:

VTR = Valid Tanpa Revisi

VR = Valid Revisi

TV = Tidak Valid

### 3.5. Teknis Analisis Data

Teknik analisis data dalam penelitian ini mencakup dua aspek utama: analisis hasil penggunaan tes dan evaluasi kualitas butir soal. Analisis hasil penggunaan tes dilakukan dengan menilai pola jawaban peserta didik melalui analisis keputusan terhadap pola jawaban peserta didik. Sementara itu, evaluasi kualitas butir soal meliputi pengujian validitas, reliabilitas, daya pembeda, serta tingkat kesukaran.

#### 1) Analisis Data Hasil Implementasi Tes Empat Tingkat

Beberapa pola jawaban peserta didik yang menandakan adanya pola model mental dapat diperoleh dari tes diagnostik empat tingkat, sebagaimana terlihat pada Tabel 3.3. Pola jawaban peserta didik ditentukan berdasarkan jawaban yang diberikan pada setiap tingkat.

Tabel 3. 3 Analisis Kategori Pola Jawaban Peserta Didik

<b>Tier-1</b>	<b>Tier-2</b>	<b>Tier-3</b>	<b>Tier-4</b>	<b>Kategori</b>	<b>Kategori</b>	<b>Skor</b>
Salah	Tidak Yakin	Salah	Yakin	<i>No Understanding (NU)</i>	<i>Initial Model (IM)</i>	0
Salah	Yakin	Salah	Tidak Yakin			
Salah	Tidak Yakin	Salah	Tidak Yakin			
Salah	Yakin	Salah	Yakin	<i>Misconception (MC)</i>	<i>Synthesis-Misconception (Sy-M)</i>	1
Benar	Yakin	Salah	Yakin	<i>Partial Understanding</i>	<i>Synthesis-Partial Understanding (Sy-P)</i>	2
Benar	Tidak Yakin	Salah	Yakin			
Benar	Yakin	Salah	Tidak Yakin			
Benar	Tidak Yakin	Salah	Tidak Yakin			
Salah	Yakin	Benar	Yakin			
Salah	Tidak Yakin	Benar	Yakin			
Salah	Yakin	Benar	Tidak Yakin			
Salah	Tidak Yakin	Benar	Tidak Yakin			
Benar	Tidak Yakin	Benar	Yakin			
Benar	Yakin	Benar	Tidak Yakin			
Benar	Tidak Yakin	Benar	Tidak Yakin			
Benar	Yakin	Benar	Yakin	<i>Sound Understanding (SU)</i>	<i>Scientific Model (SM)</i>	3

(Kaltakci-Gurel dkk., 2017)

Kriteria analisis model mental peserta didik dapat diterapkan dan dianalisis menggunakan pendekatan Rasch. Peserta didik yang tergolong dalam kategori *Initial Model* memperoleh skor '0' karena tidak mampu memberikan jawaban yang tepat pada tingkat pertama dan ketiga dari tes diagnostik. Peserta didik dalam kategori *Synthesis-Misconception* akan mendapatkan skor '1' karena tidak mampu memberikan jawaban yang tepat pada tingkat pertama dan ketiga dari tes diagnostik namun masih merasa yakin pada jawaban tingkat pertama dan ketiga. Sementara itu, peserta didik dalam kategori *Synthesis-Partial Understanding*, yang dapat memberikan jawaban yang benar pada tingkat pertama dan ketiga namun masih menunjukkan ketidakyakinan pada salah satu tingkat keyakinan akan diberikan skor '2'. Peserta didik yang termasuk dalam kategori *Scientific Model* ditandai dengan kemampuan menjawab dengan tepat dan menunjukkan keyakinan tinggi pada seluruh tingkat akan memperoleh skor tertinggi yaitu '3'.

## 2) Validitas Konstruk dan Isi

Validitas merupakan hal penting yang wajib dipertimbangkan dalam proses pemilihan atau pengembangan instrumen penelitian. Konsep validitas mengacu pada tingkat kesesuaian antara apa yang ingin diukur dengan hasil pengukuran yang diperoleh. Dengan kata lain, validitas menunjukkan sejauh mana suatu instrumen benar-benar mengukur konstruk yang hendak diukur. Proses validasi melibatkan pengumpulan dan analisis bukti untuk mendukung klaim bahwa instrumen tersebut memang valid, sebagaimana dijelaskan oleh Fraenkel dkk. (2011). Dalam penelitian ini, validitas tes empat tingkat dinilai melalui berbagai metode, baik kuantitatif maupun kualitatif. Salah satu langkah awal dalam proses validasi ini adalah melibatkan para ahli untuk memeriksa dan mengevaluasi setiap butir soal dalam tes diagnostik berdasarkan isi dan formatnya. Uji validitas konstruk dan isi dilakukan berdasarkan pendapat ahli (*judgement expert*). Hasil dari data validasi oleh ahli kemudian dianalisis menggunakan Formula Aiken. Formula Aiken digunakan untuk menentukan koefisien V dari data berskala rating secara statistik (Aiken, 1985). Formula Aiken dirumuskan sebagai berikut:

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

Dimana:

$V$  = indeks validitas item

$s = r - l_0$  = skor yang diberi penilai;  $l_0$  = skor terendah

$c$  = jumlah kategori penilaian

$n$  = jumlah validator

Tabel 3. 4 Standar Koefisien Validitas (V) Aiken

Jumlah Item (m) Atau Validator (n)	Jumlah Kategori Rating (c)					
	2		3		4	
	V	p	V	p	V	p
2						
3						
3			1,00	0,037	1,00	0,016
4					1,00	0,004
4			1,00	0,012	0,92	0,020
5			1,00	0,004	0,93	0,006
5	1,00	0,031	0,90	0,025	0,87	0,021
6			0,92	0,010	0,89	0,007
6	1,00	0,016	0,83	0,038	0,78	0,050
7			0,93	0,004	0,86	0,007
7	1,00	0,008	0,86	0,016	0,76	0,045
8	1,00	0,004	0,88	0,007	0,83	0,007
8	0,88	0,035	0,81	0,024	0,75	0,040
9	1,00	0,002	0,89	0,003	0,81	0,007
9	0,89	0,020	0,78	0,032	0,74	0,036
10	1,00	0,001	0,85	0,005	0,80	0,007
10	0,90	0,001	0,75	0,040	0,73	0,032
11	0,91	0,006	0,82	0,007	0,79	0,007
11	0,82	0,033	0,73	0,048	0,73	0,029
12	0,92	0,003	0,79	0,010	0,78	0,006

12	0,83	0,019	0,75	0,025	0,69	0,046
----	------	-------	------	-------	------	-------

(Aiken, 1985)

Ketika nilai Aiken's V hitung lebih besar dari nilai standar Aiken's V pada tabel, maka item tersebut dikatakan "valid". Berdasarkan tabel standar koefisien validitas Aiken diatas, karena instrumen yang dikembangkan divalidasi oleh 3 orang ahli dengan 3 kategori penilaian, maka diperoleh nilai Aiken's V minimal untu penelitian ini sebesar 1,00 dengan probabilitas 0,037. Sehingga pada penelitian ini soal dapat dikatakan valid ketika nilai Aiken's V = 1,00.

### 3) Validitas Empiris

Peneliti menganalisis validitas empiris menggunakan model *Rasch* dengan bantuan *software Winstep*. Skor konsepsi peserta didik dari implementasi instrumen tes dimasukkan untuk analisis ini. Peneliti mengevaluasi validitas setiap butir soal dan keseluruhan instrumen. Validitas keseluruhan dalam model Rasch ini dikenal sebagai *Unidimensionality* (Sumintono, 2016). Hasil validitas keseluruhan diperoleh dari menu output tables pada bagian item: dimensionality. Peneliti menunjukkan nilai validitas keseluruhan menggunakan *raw variance explained by measures* dan menginterpretasikan nilai tersebut berdasarkan Tabel Interpretasi Unidimensionalitas Instrumen.

Tabel 3. 5 Interpretasi Unidimensionalitas Instrumen

<i>Raw variance explained by measures</i>	<b>Interpretasi</b>
$20\% \leq \text{value} < 40\%$	Terpenuhi
$40\% \leq \text{value} < 60\%$	Sesuai
$\text{value} \geq 60\%$	Istimewa

(Sumintono, 2016)

Instrumen dianggap memiliki unidimensionalitas yang baik jika nilai *unexplained variance in 1st contrast* kurang dari 15% (Samsudin dkk., 2020). Setelah menguji validitas keseluruhan instrumen, peneliti juga melakukan uji validitas untuk setiap butir soal. Uji validitas ini diperoleh dari menu output tables pada pilihan item (*column*): *fit order*. Validitas ini menentukan apakah butir soal berfungsi dengan baik dalam pengukuran. Peneliti mengevaluasi kualitas setiap butir soal berdasarkan nilai *outfit mean square* (MNSQ), *outfit Z-standard*

(ZSTD), dan *point measure correlation* (*Pt Measure Corr*). Nilai MNSQ mengukur keacakan atau penyimpangan dalam sistem pengukuran, nilai ZSTD mengukur kemungkinan terjadinya penyimpangan, dan nilai *Pt measure corr* mengukur hubungan antara kesulitan item dengan kesulitan instrumen (Sumintono, 2016). Peneliti menggunakan Tabel berikut untuk memeriksa kualitas setiap butir soal.

Tabel 3. 6 Kriteria Nilai Outfit MNSQ, Outfit ZSTD, dan Pt Measure Corr

<i>Output Item</i>	<b>Skor</b>	<b>Keterangan</b>
<i>Outfit MNSQ</i>	$0,5 < \text{MNSQ} < 1,5$	<i>Fit</i>
<i>Outfit ZSTD</i>	$-2,0 \leq \text{ZSTD} < +20$	<i>Fit</i>
<i>Pt Measure Corr</i>	$0,4 < \text{Pt Measure Corr} < 0,85$	<i>Fit</i>

(Sumintono, 2016)

Peneliti menginterpretasikan kualitas setiap butir soal dengan menggunakan tabel interpretasi kualitas setiap butir soal.

Tabel 3. 7 Interpretasi Kualitas Setiap Butir Soal

<b>Kriteria</b>	<b>Interpretasi</b>	<b>Keterangan</b>
Tiga kriteria terpenuhi	Sangat sesuai	Valid Tanpa Revisi (VTR)
Dua kriteria terpenuhi	Sesuai	Valid Tanpa Revisi (VTR)
Satu kriteria terpenuhi	Kurang Sesuai	Valid dengan Revisi (VR)
Tiga tidak kriteria terpenuhi	Tidak Sesuai	Tidak valid (TV)

(Boone dkk., 2014; Samsudin dkk., 2020)

#### 4) Reliabilitas

Reliabilitas merujuk pada tingkat konsistensi suatu instrumen dalam mengukur sesuai dengan kriteria yang telah ditetapkan, menjawab pertanyaan apakah tes tersebut akurat dan dapat dipercaya. Uji reliabilitas butir soal dilakukan menggunakan model *Rasch*, yang mengukur reliabilitas melalui nilai *person reliability* (*p*), *item reliability* (*r*), dan *Cronbach's alpha* (KR-20). Reliabilitas yang rendah biasanya disebabkan oleh jumlah sampel yang terlalu kecil, sehingga variasi hirarki pada skala logit menjadi terbatas (Sumintono, 2016).

Tabel 3. 8 Interpretasi Person Reliability, Item Reliability, dan Cronbach Alpha

Statistik	Nilai Indeks	Interpretasi
<i>Person and Item Reliability</i>	$r \leq 0,67$	Rendah
	$0,67 < r \leq 0,80$	Cukup
	$0,80 < r \leq 0,90$	Baik
	$0,90 < r \leq 0,94$	Sangat Baik
	$r > 0,94$	Baik Sekali
<i>Cronbach Alpha (KR-20)</i>	$KR-20 \leq 0,50$	Buruk
	$0,50 \leq KR-20 < 0,60$	Jelek
	$0,60 \leq KR-20 < 0,70$	Cukup
	$0,70 \leq KR-20 < 0,80$	Bagus
	$KR-20 \geq 0,80$	Bagus Sekali

(Sumintono, 2016)

### 5) Daya Pembeda

Daya pembeda merujuk pada kemampuan suatu soal untuk membedakan antara peserta didik dengan kemampuan tinggi dan rendah. Untuk meningkatkan kualitas soal, diperlukan informasi lebih lanjut mengenai daya pembeda ini. Semakin tinggi daya pembeda suatu soal, semakin efektif soal tersebut dalam membedakan pemahaman peserta didik terhadap materi. Daya pembeda soal diukur menggunakan kriteria yang disebut *Point Measure Correlation (Pt Mean Corr)*.

Tabel 3. 9 Interpretasi Daya Pembeda

Nilai <i>Point Measure Correlation</i>	Interpretasi
<b>Pt Mean Corr &gt;0,40</b>	Sangat bagus
<b>0,30 ≤ Pt Mean Corr &lt; 0,40</b>	Bagus
<b>0,20 ≤ Pt Mean Corr &lt; 0,30</b>	Cukup
<b>0,00 ≤ Pt Mean Corr &lt; 0,20</b>	Tidak mampu mendiskriminasi
<b>Pt Mean Corr &lt; 0,00</b>	Membutuhkan pemeriksaan

(Alagumalai, S., dkk., 2005)

### 6) Tingkat Kesukaran

Tingkat kesukaran mengacu pada seberapa baik tes mengukur jumlah peserta yang menjawab dengan benar. Tes dianggap memiliki tingkat kesukaran yang tinggi jika banyak peserta didik tidak dapat menjawab dengan benar, dan sebaliknya, tingkat kesukarannya rendah jika hanya beberapa peserta didik yang tidak dapat menjawab dengan benar (Sumintono, 2016). Peneliti menyajikan interpretasi tingkat kesukaran ini pada Tabel interpretasi tingkat kesukaran, yang telah disesuaikan berdasarkan penskoran dan hasil pengolahan *item: measure*.

Tabel 3. 10 Interpretasi Tingkat Kesukaran

<b>Nilai <i>Measure</i> (M)</b>	<b>Interpretasi</b>
<b><math>M &gt; +0,31</math></b>	Sulit
<b><math>0, -31 &lt; M \leq +0,31</math></b>	Sedang
<b><math>M &lt; -0,31</math></b>	Mudah

(Sumintono, B., & Widhiarso, 2015)

## BAB IV TEMUAN

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis model mental peserta didik pada topik fluida statis menggunakan instrumen tes diagnostik empat tingkat. Melalui penelitian ini, peneliti ingin menganalisis model mental peserta didik pada topik fluida statis berdasarkan instrumen tes diagnostik empat tingkat dan menguji kualitas instrumen tes diagnostik empat tingkat untuk mengidentifikasi model mental peserta didik pada topik fluida statis. Pada penelitian ini instrumen tes diagnostik empat tingkat sebanyak 10 soal diujikan kepada peserta didik SMA kelas 12. Didapatkan responden sebanyak 100 peserta didik dari 2 sekolah berbeda.

### 4.1. Model Mental Peserta Didik pada Topik Fluida Statis

Jawaban peserta didik diolah melalui proses pengkodean (*coding*) yang dirancang untuk instrumen tes diagnostik empat tingkat. Tujuannya adalah untuk mengidentifikasi dan mengklasifikasi model mental yang dimiliki oleh peserta didik. Model mental peserta didik dikelompokkan ke dalam empat kategori, yaitu: *Scientific Model*, *Synthesisi-Partial Understanding*, *Synthesis-Misconception*, dan *Initial Model*. Teknik pengkodean yang digunakan, mengklasifikasikan model mental berdasarkan pola jawaban peserta didik. Jumlah soal terdiri dari 10 butir dengan pertimbangan untuk menjaga motivasi peserta didik dalam mengerjakan soal, sehingga uji instrumen tes dapat berjalan optimal.

Setelah proses pengkodean dan tabulasi terhadap seluruh jawaban peserta didik pada setiap pilihan ganda selesai dilakukan, diperoleh persentase yang menunjukkan proporsi peserta didik yang masuk ke dalam masing-masing kategori model mental.

Tabel 4. 1 Persentase Model Mental Peserta Didik pada Topik Fluida Statis

Sub Konsep	No Soal	Persentase Kategori Model Mental Peserta Didik (%)			
		SM	Sy-P	Sy-M	IM
Tekanan Hidrostatik	1	45	14	35	6
	2	50	16	30	4
	3	58	22	17	3

Sub Konsep	No Soal	Persentase Kategori Model Mental Peserta Didik (%)			
		SM	Sy-P	Sy-M	IM
	4	45	34	17	4
Hukum	5	42	19	29	10
Pascal	6	31	40	22	7
Hukum	7	25	20	45	20
Archimedes	8	38	32	21	9
	9	42	19	32	7
	10	46	37	14	3

Keterangan:

SM : *Scientific Model*

Sy-P : *Synthesis-Partial Understanding*

Sy-M : *Synthesis-Misconception*

IM : *Initial Model*

## 4.2. Kualitas Instrumen Tes Diagnostik Empat Tingkat

### 4.2.1. Validitas Konstruk dan Isi

Hasil validasi ahli (*judgement expert*) instrumen tes diagnostik empat tingkat pada topik fluida statis yang dilakukan kepada 1 orang dosen fisika dan 2 orang guru fisika dianalisis menggunakan Formula Aiken untuk menentukan koefisien validitas Aiken (Aiken's V). koefisien validitas Aiken yang diperoleh berdasarkan perhitungan menggunakan formula Aiken kemudian dibandingkan dengan nilai standar minimal koefisien Aiken (Aiken's V tabel). Hal tersebut dilakukan untuk menentukan validitas tiap butir soal dalam instrumen penelitian. Berdasarkan tabel standar koefisien validitas Aiken, karena instrumen yang dikembangkan divalidasi oleh 3 orang ahli dengan 3 kategori penilaian, maka diperoleh nilai Aiken's V minimal untuk penelitian ini yaitu sebesar 1,00 dengan probabilitas 0,037.

Tabel 4. 2 Hasil Uji Validitas Ahli

No. Soal	Aspek	V	Kategori	No. Soal	Aspek	V	Kategori
1	a	1	Valid	6	a	1	Valid

No. Soal	Aspek	V	Kategori	No. Soal	Aspek	V	Kategori
	b	1	Valid		b	1	Valid
	c	1	Valid		c	1	Valid
	d	1	Valid		d	1	Valid
	e	1	Valid		e	1	Valid
2	a	1	Valid	7	a	1	Valid
	b	1	Valid		b	1	Valid
	c	1	Valid		c	1	Valid
	d	1	Valid		d	1	Valid
	e	1	Valid		e	1	Valid
3	a	1	Valid	8	a	1	Valid
	b	1	Valid		b	1	Valid
	c	1	Valid		c	1	Valid
	d	1	Valid		d	1	Valid
	e	1	Valid		e	1	Valid
4	a	1	Valid	9	a	1	Valid
	b	1	Valid		b	1	Valid
	c	1	Valid		c	1	Valid
	d	1	Valid		d	1	Valid
	e	1	Valid		e	1	Valid
5	a	1	Valid	10	a	1	Valid
	b	1	Valid		b	1	Valid
	c	1	Valid		c	1	Valid
	d	1	Valid		d	1	Valid
	e	1	Valid		e	0,8	Tidak Valid

Keterangan:

V = Koefisien validitas Aiken

a = Kesesuaian konsep dalam butir soal dengan konsep yang dikemukakan oleh ahli

b = Relevansi butir soal dengan materi SMA

c = Kemampuan butir soal dalam menentukan pemahaman konsep peserta didik

d = Penggunaan bahasa yang sesuai dengan kaidah Ejaan Bahasa Indonesia

e = Kesesuaian konstruksi butir soal dengan aturan penulisan soal pilihan ganda

Dari Tabel 4.2. dapat dilihat bahwa terdapat 9 butir soal yang memiliki nilai lebih besar dari nilai standar Aiken's V tabel yang berarti "valid" dan terdapat 1 butir soal yang memiliki nilai lebih kecil dari nilai standar Aiken's V tabel yang berarti "Tidak Valid".

Butir soal nomor 10 harus diperbaiki sehingga memperoleh penilaian yang lebih baik dari penilai. Jika dalam perbaikan instrumen sudah mendapatkan nilai valid maka instrumen sudah bisa digunakan penelitian (Ihsan, 2015). Berdasarkan masukan dari validator 3 butir soal nomor 10 tetap bisa digunakan dengan catatan perlu adanya perbaikan dengan menambahkan keterangan penjelasan pada gambar yang disajikan.

#### 4.2.2. Hasil Validitas Empiris

Uji validitas empiris dilakukan dengan analisis model Rasch menggunakan software winstep. Uji validitas empiris dengan model Rasch ini dianalisis secara keseluruhan dan untuk setiap butir soal. Uji validitas secara keseluruhan yang dianalisis dengan model Rasch. Gambar berikut merupakan hasil uji validitas secara keseluruhan instrumen tes diagnostik empat tingkat pada topic Fluida Statis yang didapat dari pilihan *item:Dimensionality*.

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)				
		-- Empirical --		Modeled
Total raw variance in observations	=	21.9	100.0%	100.0%
Raw variance explained by measures	=	11.9	54.2%	53.4%
Raw variance explained by persons	=	6.2	28.3%	27.8%
Raw Variance explained by items	=	5.7	26.0%	25.6%
Raw unexplained variance (total)	=	10.0	45.8%	100.0%
Unexplned variance in 1st contrast	=	2.2	9.9%	21.6%
Unexplned variance in 2nd contrast	=	1.6	7.3%	15.9%
Unexplned variance in 3rd contrast	=	1.2	5.5%	12.0%
Unexplned variance in 4th contrast	=	1.1	4.9%	10.8%
Unexplned variance in 5th contrast	=	1.0	4.6%	10.0%

Gambar 4. 1 Hasil Uji Unidimensionalitas Instrumen

Hasil *item:dimensionality*, menunjukkan bahwa hasil keragaman (*raw variance*) data adalah 54,2% yang memenuhi syarat minimal unidimensionalitas yaitu sebesar 20%. Selain itu, varians yang tidak dijelaskan oleh kontras pertama (*unexplained variance in 1<sup>st</sup> contrast*) bernilai 9,9%, yang menyebabkan adanya

perbedaan antara nilai model dan empiris pada *Raw Variance*. Nilai ini dapat diterima karena berada di bawah nilai maksimum 15%. Dapat disimpulkan bahwa butir soal instrumen ini dapat digunakan untuk mengidentifikasi model mental peserta didik.

### 1) Uji Validitas

Salah satu aspek penting dalam proses pembuatan soal adalah validitas butir soal. Validitas butir soal dapat diuji dengan menggunakan rasch model berdasarkan kriteria *Outfit* MNSQ dan *Outfit* ZSTD yang diperoleh dari pilihan *item (column): fit order*. Kriteria ini menunjukkan apakah butir soal berperan normal dalam mengukur atau tidak.

Item STATISTICS: MEASURE ORDER															
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL			INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
7	140	100	13.89	1.51	1.49	3.1	1.30	2.0	.66	.78	51.1	51.7	S7		
5	193	100	2.02	1.51	.89	-.7	.85	-.8	.76	.72	53.2	55.2	S5		
6	195	100	1.57	1.51	1.30	1.9	1.23	1.3	.62	.72	51.1	55.2	S6		
9	196	100	1.34	1.51	1.06	.5	.99	.0	.71	.71	58.5	55.2	S9		
1	198	100	.88	1.51	.95	-.3	1.50	2.4	.68	.71	52.1	57.3	S1		
8	199	100	.65	1.51	.71	-2.1	.66	-2.0	.79	.71	64.9	57.4	S8		
2	212	100	-2.36	1.54	1.00	.1	.89	-.5	.70	.68	55.3	58.0	S2		
4	220	100	-4.26	1.55	.70	-2.2	.76	-1.1	.72	.66	64.9	59.4	S4		
10	226	100	-5.73	1.57	.78	-1.5	.87	-.5	.68	.64	64.9	60.0	S10		
3	235	100	-8.00	1.61	.89	-.7	.76	-1.0	.66	.61	61.7	60.3	S3		
MEAN			201.4	100.0	.00	1.53	.98	-.2	.98	.0	57.8	57.0			
S.D.			24.7	.0	5.68	.03	.24	1.6	.26	1.4	5.6	2.5			

Gambar 4. 2 Hasil Uji Validitas Instrumen

Hasil analisis mengenai uji validitas dapat dilihat pada Tabel 4.3.

Tabel 4. 3 Validitas Tes Diagnostik Empat Tingkat yang Dikembangkan

No. Soal	MNSQ		ZSTD		PT Measure- All Corr		Interpretasi	Keterangan
	Nilai	Ket	Nilai	Ket	Nilai	Ket		
1	1,50	<i>MisFit</i>	2,4	Fit	0,68	Fit	Sesuai	Valid
2	0,89	Fit	-0,5	Fit	0,70	Fit	Sangat sesuai	Valid
3	0,76	Fit	-1,0	Fit	0,66	Fit	Sangat sesuai	Valid
4	0,76	Fit	-1,1	Fit	0,72	Fit	Sangat sesuai	Valid
5	0,85	Fit	-0,8	Fit	0,76	Fit	Sangat sesuai	Valid
6	1,23	Fit	1,3	Fit	0,62	Fit	Sangat sesuai	Valid

7	1,30	Fit	2,0	Fit	0,66	Fit	Sangat sesuai	Valid
8	0,66	Fit	-2,0	Fit	0,79	Fit	Sangat sesuai	Valid
9	0,99	Fit	0,0	Fit	0,71	Fit	Sangat sesuai	Valid
10	0,87	Fit	-0,5	Fit	0,68	Fit	Sangat sesuai	Valid

Berdasarkan hasil uji validitas pada tabel tersebut, menunjukkan bahwa pada semua butir soal memiliki nilai MNSQ dan ZSTD yang memenuhi kriteria. Namun terdapat satu butir soal yang memiliki nilai MNSQ yang tidak memenuhi kriteria, yaitu butir soal nomor 1 yang seharusnya berada pada rentang  $0,5 < \text{MNSQ} < 1,5$ .

## 2) Uji Reliabilitas

Uji reliabilitas adalah uji untuk mengukur seberapa besar tingkat konsistensi suatu tes. Uji reliabilitas dengan menggunakan Rasch Model dianalisis dengan memperhatikan *person reliability* (p), *item reliability* (r), dan *cronbach alpha* (KR-20). Hasil uji reliabilitas instrumen ditunjukkan pada gambar berikut.

SUMMARY OF 100 MEASURED (EXTREME AND NON-EXTREME) Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	20.1	10.0	62.91	6.00				
S.D.	6.9	.0	16.48	3.56				
MAX.	30.0	10.0	98.99	18.26				
MIN.	8.0	10.0	37.95	3.99	.14	-3.0	.17	-2.8
REAL RMSE	7.25	TRUE SD	14.80	SEPARATION	2.04	Person RELIABILITY	.81	
MODEL RMSE	6.98	TRUE SD	14.93	SEPARATION	2.14	Person RELIABILITY	.82	
S.E. OF Person MEAN = 1.66								
Person RAW SCORE TO MEASURE CORRELATION = .96								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .90								
SUMMARY OF 10 MEASURED (NON-EXTREME) Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	201.4	100.0	50.00	1.53	.98	-.2	.98	.0
S.D.	24.7	.0	5.68	.03	.24	1.6	.26	1.4
MAX.	235.0	100.0	63.89	1.61	1.49	3.1	1.50	2.4
MIN.	140.0	100.0	42.00	1.51	.70	-2.2	.66	-2.0
REAL RMSE	1.60	TRUE SD	5.45	SEPARATION	3.42	Item RELIABILITY	.92	
MODEL RMSE	1.53	TRUE SD	5.47	SEPARATION	3.57	Item RELIABILITY	.93	
S.E. OF Item MEAN = 1.89								

Gambar 4. 3 Hasil Uji Reliabilitas Instrumen

Hasil analisis mengenai uji reliabilitas dapat dilihat pada Tabel 4.4.

Tabel 4. 4 Interpretasi Uji Reliabilitas Instrumen

<i>Person Reliability</i>	Interpretasi	<i>Item Reliability</i>	Interpretasi	<i>Cronbach Alpha</i>	Interpretasi
0,81	Baik	0,92	Sangat Baik	0,90	Bagus Sekali

Pada gambar ditunjukkan nilai *person reliability* sebesar 0,81 dengan interpretasi “Baik”. Hasil interpretasi baik dari *person reliability* ini menunjukkan bahwa konsistensi jawaban dari peserta didik terhadap instrumen penelitian cukup tinggi. Dengan kata lain, instrumen mampu mengukur secara konsisten pada individu yang berbeda dengan variabilitas yang rendah.

Nilai *item reliability* sebesar 0,92 dengan interpretasi “Sangat Baik”. Hasil interpretasi *item reliability* yang masuk pada kategori sangat baik ini menunjukkan bahwa setiap butir soal pada instrumen memiliki tingkat konsistensi yang sangat tinggi. Dengan reliabilitas yang tinggi, dapat disimpulkan bahwa instrumen ini layak digunakan untuk mengukur model mental peserta didik.

Kemudian ditunjukkan nilai *cronbach alpha* (KR-20) pada instrumen penelitian ini adalah 0,90, yang termasuk dalam kategori “Bagus Sekali”. Hasil ini menunjukkan bahwa instrumen memiliki tingkat kepercayaan yang sangat tinggi untuk mengukur variabel yang diukur. Nilai ini mengindikasikan bahwa setiap item dalam instrumen memiliki hubungan internal yang kuat, sehingga instrumen dapat digunakan secara konsisten.

### 3) Daya Pembeda

Daya pembeda instrumen diukur melalui nilai *PT Measure-All Corr* untuk setiap butir soal. Nilai ini mempresentasikan kemampuan setiap item dalam membedakan antara individu dengan tingkat kemampuan atau karakteristik yang berbeda.

Tabel 4. 5 Hasil Uji Daya Pembeda Tiap Butir Soal

No. Soal	PT Measure-All Corr	Interpretasi
1	0,68	Sangat bagus
2	0,70	Sangat bagus
3	0,66	Sangat bagus

No. Soal	PT Measure-All Corr	Interpretasi
4	0,72	Sangat bagus
5	0,76	Sangat bagus
6	0,62	Sangat bagus
7	0,66	Sangat bagus
8	0,79	Sangat bagus
9	0,71	Sangat bagus
10	0,68	Sangat bagus

Analisis ini menunjukkan bahwa instrumen penelitian memiliki daya pembeda yang sangat baik, sehingga dapat digunakan secara efektif untuk mendapatkan data yang akurat dan relevan.

#### 4) Tingkat Kesukaran

Tingkat kesukaran butir soal merupakan salah satu indikator penting dalam menilai kualitas perangkat tes. Menganalisis tingkat kesukaran butir soal berarti meninjau soal-soal tes berdasarkan tingkat kesukarannya, sehingga dapat dikelompokkan menjadi soal yang sangat mudah, mudah, sulit, atau sangat sulit.

Item STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFINIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item			
7	140	100	13.89	1.51	1.49	3.1	1.30	2.0	.66	.78	51.1	51.7	S7
5	193	100	2.02	1.51	.89	-.7	.85	-.8	.76	.72	53.2	55.2	S5
6	195	100	1.57	1.51	1.30	1.9	1.23	1.3	.62	.72	51.1	55.2	S6
9	196	100	1.34	1.51	1.06	.5	.99	.0	.71	.71	58.5	55.2	S9
1	198	100	.88	1.51	.95	-.3	1.50	2.4	.68	.71	52.1	57.3	S1
8	199	100	.65	1.51	.71	-2.1	.66	-2.0	.79	.71	64.9	57.4	S8
2	212	100	-2.36	1.54	1.00	.1	.89	-.5	.70	.68	55.3	58.0	S2
4	220	100	-4.26	1.55	.70	-2.2	.76	-1.1	.72	.66	64.9	59.4	S4
10	226	100	-5.73	1.57	.78	-1.5	.87	-.5	.68	.64	64.9	60.0	S10
3	235	100	-8.00	1.61	.89	-.7	.76	-1.0	.66	.61	61.7	60.3	S3
MEAN	201.4	100.0	.00	1.53	.98	-.2	.98	.0			57.8	57.0	
S.D.	24.7	.0	5.68	.03	.24	1.6	.26	1.4			5.6	2.5	

Gambar 4. 4 Hasil Uji Tingkat Kesukaran Instrumen

Pada gambar ditunjukkan data nilai *Measure* (M) dan Standar Deviasi (SD) instrumen. Nilai standar deviasi diperoleh 5,68. Interpretasi tingkat kesukaran instrumen ditunjukkan pada tabel berikut

Tabel 4. 6 Hasil Interpretasi Tingkat Kesukaran Instrumen

No. soal	Measure	Interpretasi
1	0,88	Sukar
2	-2,36	Mudah
3	-8,00	Mudah
4	-4,26	Mudah
5	2,02	Sukar
6	1,57	Sukar
7	13,89	Sukar
8	0,65	Sukar
9	1,34	Sukar
10	-5,73	Mudah

Berdasarkan tabel, tingkat kesukaran soal terbagi dalam dua kategori, yaitu Mudah dan Sukar/Sulit. Tingkat kesukaran soal dengan kriteria sukar/sulit (butir soal nomor 1, 5, 6, 7, 8, 9), dan kriteria mudah (butir soal nomor 2, 3, 4, 10).