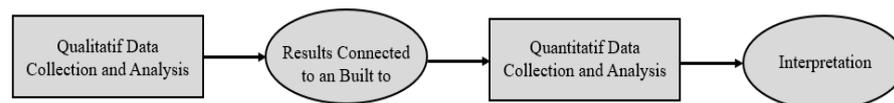


BAB III METODE PENELITIAN

3. 1. Metode dan Desain Penelitian

Metode penelitian yang akan digunakan dalam penelitian ini yaitu metode campuran (*mixed methods*) dengan *exploratory sequential design*. Desain ini terdiri dari dua tahap yaitu tahap kualitatif dan tahap kuantitatif. Adapun tahapan *exploratory sequential design* dapat dilihat pada gambar 3.1.



Gambar 3.1 *Exploratory Sequential Design*

Exploratory sequential design digunakan ketika pengukuran atau instrumen tidak tersedia, variabelnya tidak diketahui, atau tidak ada kerangka kerja atau teori pemandu (Green, et al., 2007). Dalam desain penelitian ini data kualitatif didapatkan dan dianalisis, kemudian data tersebut digunakan untuk membuat instrumen kuantitatif yang akan digunakan untuk mengeksplorasi masalah penelitian (Creswell, et al., 2011; Onwuegbuzie, Bustamante, and Nelson 2010).

Kemudian desain ini dikombinasikan dengan tahapan pengembangan instrumen yang dilakukan oleh Caleon & Subramaniam (2009) untuk instrumen *four-tier wave diagnostic instrument (4WADI)*. Tahapan inti dalam pengembangan instrumen ini terdiri pendefinisian dan batasan konten penelitian; eksplorasi dan identifikasi konten; validasi konten dan uji coba instrumen; serta pengembangan instrumen final dan uji validitas.

Pada tahap pendefinisian dan batasan konten penelitian dilakukan analisis kurikulum, standar kompetensi lulusan dan buku teks fisika kelas sebelas yang selanjutnya diturunkan dalam bentuk indikator pencapaian kompetensi. Kemudian pada tahapan eksplorasi dilakukan tinjauan literatur, mengamati proses pembelajaran, meninjau laporan hasil ujian tahun-tahun sebelumnya, mempelajari kesalahan dalam menjawab lembar kerja siswa/tes kelas dan wawancara individu. Setelah tahap eksplorasi, berikutnya informasi yang didapatkan dibuat menjadi “*pilot project*” instrumen dengan pertanyaan terbuka pada tingkat ke tiga dan lima.

Setelah uji coba instrumen selesai, tahapan terakhir yaitu membuat instrumen final *five-tier word association diagnostic instrument* kemudian dilakukan uji lapangan untuk mengetahui validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan profil konsepsi peserta didik pada materi fluida statis.

3. 2. Partisipan Penelitian

Partisipan dalam penelitian ini yaitu siswa kelas XI dengan kriteria telah mempelajari materi fluida statis. Adapun partisipan yang terlibat dalam penelitian ini terdiri dari siswa kelas XI yang berada di tiga sekolah di tiga wilayah yang berbeda yaitu SMA Negeri 3 Cimahi, SMA Negeri 1 Cisolok dan SMAN 1 Pusakanagara. Adapun jumlah partisipannya yaitu sebanyak 170 peserta didik kelas sebelas dengan sebaran seperti terlihat dalam tabel 3.1 berikut ini.

Tabel 3.1 Partisipan Penelitian

Sekolah	Jumlah Sampel	
		Laki-laki
SMAN 3 Cimahi	Perempuan	32
SMAN 1 Cisolok	Laki-laki	9
	Perempuan	24
SMAN 1 Pusakanagara	Laki-laki	15
	Perempuan	64

3. 3. Instrumen Penelitian

Instrumen yang digunakan dalam penelitian ini terdiri dari lembar observasi, instrumen *four-tier*, lembar wawancara, lembar uji validitas, instrumen *five-tier word association diagnostic instrument (FTWADI)*. Untuk penjelasan keempat instrumen tersebut yaitu sebagai berikut:

3.3.1. Lembar Observasi

Observasi kualitatif merupakan kegiatan peneliti turun langsung kelapangan untuk mengamati perilaku dan aktivitas individu-individu di lokasi penelitian (Creswell, 2019). Dalam penelitian ini observasi dilakukan untuk mengamati proses pembelajaran, perilaku peserta didik, proses penilaian dan hasil tes peserta didik. Tujuan dari lembar observasi ini untuk mengetahui karakteristik pemahaman

konsep peserta didik yang meliputi proses pembelajaran di kelas dan proses penilaian.

Tabel 3.2 Lembar Observasi

Aspek	Ya	Tidak	Keterangan
Pembelajaran menggunakan kurikulum 2013.			
Menggunakan buku sumber sebagai bahan rujukan.			
Menggunakan sumber pustaka lain selain buku sumber.			
Menggunakan media pembelajaran yang beragam.			
Menggunakan metode pembelajaran yang menuntun siswa untuk aktif dalam pembelajaran.			
Menggunakan LKPD dalam pembelajaran di kelas.			
Menggunakan tes pilihan ganda dalam evaluasi pembelajaran.			
Menggunakan tes essay dalam evaluasi pembelajaran.			
Guru melakukan tes diagnostik kepada peserta didik.			
Rata-rata hasil belajar siswa diatas KKM.			
Guru pernah mengukur struktur kognitif peserta didik.			
Guru mengetahui struktur kognitif peserta didik.			

3.3.2. Instrumen *four-tier open ended question* (4TOEQ)

Instrumen *four-tier open ended question* (4TOEQ) merupakan tes diagnostik yang terdiri dari empat tingkat pertanyaan untuk mengidentifikasi konsepsi peserta

Upar Suparta, 2023

PENGEMBANGAN FIVE-TIER WORD ASSOCIATION DIAGNOSTIC INSTRUMENT (FTWADI) UNTUK MENGIDENTIFIKASI KONSEPSI SISWA PADA MATERI FLUIDA STATIS

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

didik. Pada tingkat pertama berisi pertanyaan terbuka mengenai konsep fluida statis. Tingkat kedua merupakan tingkat keyakinan peserta didik dalam menjawab tingkat pertama. Tingkat ketiga berisi alasan peserta didik dalam menjawab tingkat pertama dengan sifat jawaban terbuka. Tingkat keempat merupakan tingkat keyakinan peserta didik dalam menjawab pertanyaan pada tingkat ketiga. Instrumen ini selanjutnya akan dikembangkan menjadi instrumen *five-tier word association diagnostic instrument* (FTWADI).

3.3.3. Lembar pedoman wawancara

Wawancara merupakan cara yang dipakai untuk memperoleh informasi melalui kegiatan interaksi sosial antara peneliti dengan yang diteliti (Slamet, dalam Edi, F 2016). Dalam penelitian ini wawancara digunakan untuk mengumpulkan informasi atau data kualitatif baik dari guru maupun peserta didik. Kemudian informasi yang diperoleh akan dijadikan acuan dalam pembuatan instrumen *five-tier test*. Adapun pedoman wawancara yang akan dilakukan adalah sebagai berikut.

Tabel 3.3 Kisi-Kisi Lembar Pedoman Wawancara Untuk Guru

No	Indikator
1	Kurikulum yang digunakan dalam pembelajaran di sekolah
2	Proses pembelajaran yang dilakukan di kelas
3	Jenis instrumen yang digunakan untuk mengukur ketercapaian pembelajaran
4	Konsepsi peserta didik
5	Jenis tes yang digunakan dalam mengukur konsepsi peserta didik
6	Bentuk tes yang digunakan dalam evaluasi pembelajaran

3.3.4. Instrumen Uji Validitas

Uji validitas digunakan untuk mengetahui valid tidaknya instrumen yang sudah dibuat. Suatu instrumen bisa digunakan dalam penelitian ketika validitas dari instrumen tersebut sudah dinyatakan valid. Valid berarti instrumen bisa digunakan

untuk mengukur sesuai dengan apa yang seharusnya diukur. Uji validitas ahli dalam penelitian ini meliputi validasi konten, validitas konstruk, dan kesesuaian bahasa yang digunakan. Adapun kisi-kisi uji validasi ahli dapat dilihat pada tabel berikut.

Tabel 3.4 Kisi-Kisi Uji Validitas Ahli

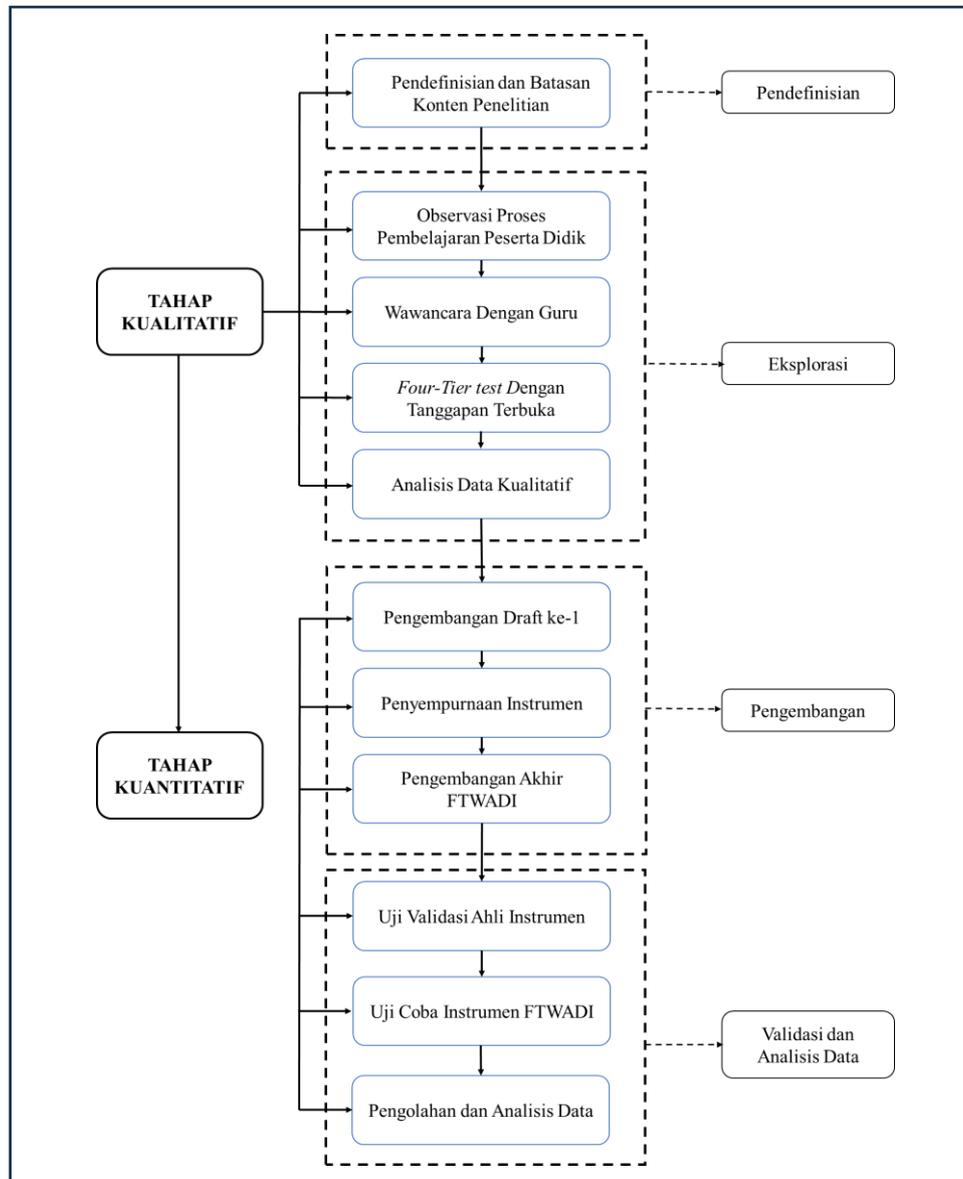
Aspek	Indikator
Konten	Kesesuaian isi materi dengan indikator pencapaian kompetensi.
	Kebenaran konsep pada tiap pertanyaan.
	Kesesuaian pertanyaan dengan indikator pencapaian kompetensi.
	Kesesuaian materi dengan jenjang pendidikan.
	Mendukung pemahaman konsep.
	Ketepatan cakupan soal.
Konstruksi	Rumusan butir soal tidak menimbulkan penafsiran ganda.
	Rumusan soal memiliki panjang yang relatif sama.
	Gambar maupun tabel terlihat jelas dan sesuai dengan soal.
	Rumusan soal tidak mengarah ke jawaban yang benar.
Bahasa	Kalimat bersifat komunikatif dan efektif.
	Ketepatan penggunaan bahasa sesuai dengan EYD.
	Kalimat mudah dipahami dan tidak memiliki penafsiran ganda.

3.3.5. Instrumen *five-tier word association diagnostic Instrument (FTWADI)*

Instrumen *five-tier test* merupakan tes diagnostik yang terdiri dari lima tingkat pertanyaan untuk mengidentifikasi konsepsi peserta didik pada materi fluida statis. Pada tingkat pertama berisi pertanyaan pilihan ganda yang berfungsi untuk mengukur konsep peserta didik terhadap materi fluida statis. Tingkat kedua merupakan tingkat keyakinan peserta didik dalam menjawab tingkat pertama. Tingkat ketiga berisi alasan peserta didik dalam menjawab tingkat pertama. Tingkat keempat merupakan tingkat keyakinan peserta didik dalam menjawab pertanyaan pada tingkat ketiga. Tingkat kelima merupakan tes asosiasi kata yang bertujuan untuk mengetahui struktur kognitif peserta didik.

3.4. Prosedur Penelitian

Prosedur dalam penelitian ini didasarkan pada desain *The Exploratory Sequential* yang dipetakan secara rinci pada Gambar 3.2.



Gambar 3.2 Prosedur Penelitian

3.4.1. Tahap Kualitatif

1. Pendefinisian

Pendefinisian dan batasan konten penelitian dilakukan dengan tujuan untuk menentukan ruang lingkup penelitian. Pendefinisian konten penelitian meliputi pemilihan topik penelitian, merumuskan masalah penelitian, menentukan tujuan penelitian, membuat pertanyaan penelitian, menentukan populasi dan sampel

Upur Suparta, 2023

PENGEMBANGAN FIVE-TIER WORD ASSOCIATION DIAGNOSTIC INSTRUMENT (FTWADI) UNTUK MENGIDENTIFIKASI KONSEPSI SISWA PADA MATERI FLUIDA STATIS

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

penelitian. Data yang diperoleh dari tahapan ini akan mendukung atau membangun tahapan selanjutnya yakni tahap kuantitatif.

2. Eksplorasi

Pada tahapan eksplorasi kegiatan yang dilakukan yaitu observasi proses pembelajaran; melakukan tes pilihan ganda bertingkat dengan tanggapan terbuka; wawancara guru & peserta didik; dan analisis data yang ditemukan.

3.4.2. Tahap Kuantitatif

1. Pengembangan

Pada tahap ini, pengembangan instrumen *five-tier* mulai dikembangkan berdasarkan hasil analisis data kualitatif yang telah diperoleh. Pengembangan dilakukan dengan menambahkan tes asosiasi kata pada tingkat kelima dan mengubah pertanyaan instrumen 4TOEQ menjadi pertanyaan tertutup.

2. Validasi dan Analisis Data

Pada tahap ini, instrumen yang telah dikembangkan dilakukan uji validitas ahli untuk menilai konten, konstruk dan bahasa. Setelah instrumen dinyatakan layak, maka langkah berikutnya yaitu dilakukan pengambilan dan analisis data untuk mengetahui validitas dan reliabilitas instrumen serta konsepsi dan struktur kognitif peserta didik.

3. 5. Analisis Data

3.5.1. Analisis Data Hasil Observasi

Pengolahan data hasil observasi dilakukan dengan cara mengkategorikan data yang diperoleh kedalam beberapa kategori. Kategori tersebut akan memiliki pola jawaban “Ya” dan “Tidak” serta keterangan yang menyertainya. Setelah terlihat kecenderungan data yang diperoleh selanjutnya dilakukan analisis data dan ditarik kesimpulan.

3.5.2. Analisis Data 4TOEQ

Data yang diperoleh dari tahap ini akan diolah secara kualitatif untuk mengetahui konsepsi peserta didik pada materi fluida statis. Sebelum diinterpretasikan, terlebih dahulu dilakukan pengkodean data yang diadopsi dari penelitian Aminudin et al, (2019) untuk instrumen *four-tier*. Pengkodean dilakukan dalam enam kategori yaitu *Sound Understanding (SU)*, *Partial Positif (PP)*, *Partial*

Negatif (PN), *Misconception (MC)*, *Not Understanding (NU)*, dan *No Coded (NC)*. Selain pengkategorian konsepsi peserta didik, data dalam tahap ini juga akan digunakan untuk membuat instrumen FTWADI. Pengolahan data dilakukan dengan cara menyajikan setiap level konsepsi untuk setiap butir soalnya dihitung dengan menggunakan persamaan berikut.

$$\text{Level konsepsi (\%)} = \frac{\sum n}{N} \times 100\%$$

Dengan n adalah jumlah peserta didik pada level tertentu, dan N adalah jumlah peserta didik yang melaksanakan tes.

Setelah persentase level konsepsi untuk setiap item didapatkan, berikutnya adalah menyeleksi butir soal yang memiliki persentase miskonsepsi paling besar untuk selanjutnya dikembangkan menjadi instrumen FTWADI. Selain itu alasan-alasan peserta didik dalam menjawab pertanyaan pada tingkat ketiga dijadikan opsi pilihan ganda dalam pertanyaan tingkat ketiga pada instrumen FTWADI. Terakhir, asosiasi kata yang diperoleh dalam instrumen ini dikelompokkan menjadi tujuh kategori yang selanjutnya akan dijadikan acuan dalam membuat pedoman penskoran tes asosiasi kata.

3.5.3. Analisis Data Wawancara

Data hasil wawancara yang diperoleh dalam penelitian ini akan diolah dengan teknik pengolahan data kualitatif yang terdiri dari tiga langkah yaitu reduksi data, penyajian data, dan penarikan kesimpulan. Kemudian data yang diperoleh akan digunakan untuk pertimbangan dalam pengembangan instrumen 4TOEQ menjadi FTWADI.

3.5.4. Analisis Data Uji validitas Ahli

Pengolahan data untuk instrumen validitas ahli bertujuan untuk menentukan apakah instrumen tersebut sudah layak untuk digunakan atau tidak. Adapun kisi-kisi lembar validasi yang digunakan adalah sebagai berikut.

Butir soal	Indikator	Skor					Keterangan
		1	2	3	4	5	
Butir 1	Butir soal yang dibuat sesuai dengan konsep yang akan diukur.						
	Butir soal dibuat untuk mengetahui pemahaman konsep peserta didik						
	Menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia.						
	Bahasa yang digunakan mudah dimengerti oleh peserta didik.						
	Pilihan jawaban dituliskan secara jelas, homogen, masuk akal dan panjang rumusan relatif sama.						
	Hanya ada satu kunci jawaban yang paling benar						

Gambar 3.3 Kisi-Kisi Lembar Validasi Instrumen 4TOEQ

Setelah dilakukan validasi oleh ahli/pakar, data yang diperoleh kemudian dilakukan pengolahan data dengan menggunakan persamaan validitas Aiken. Adapun persamaan untuk validitas Aiken adalah sebagai berikut.

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

Dengan V adalah indeks Aiken; S = r - I_o, dimana r merupakan skor yang diberikan expert, I_o merupakan skala penilaian terendah; n adalah banyaknya judgment; dan c adalah skala penilaian tertinggi. Setelah nilai dari indeks Aiken diperoleh selanjutnya dilakukan pengklasifikasian validitasnya. Adapun kategori validitas Aiken ditunjukkan dalam tabel di bawah ini.

Tabel 3.5 Kategori Validitas Aiken

Indeks Aiken (V)	Kategori Validitas
$0 \leq V < 0,4$	Kurang Valid (rendah)
$0,4 \leq V < 0,8$	Cukup Valid (sedang)
$0,8 \leq V \leq 1$	Sangat Valid (tinggi)

(Retnawati, 2016)

3.5.5. Analisis Instrumen Fier-tier Test

Setelah dilakukan uji instrumen pada sampel penelitian yang sudah ditentukan berikutnya adalah mengolah dan menganalisis data hasil penelitian. Namun sebelum pengolahan data dilakukan terlebih dahulu dilakukan pengkodean untuk mengetahui level konsepsi peserta didik dan memudahkan dalam pengolahan data. Pengkodean ini diadopsi dari penelitian Aminudin et al, (2019) untuk instrumen *four-tier* yang dikembangkan menjadi pengkodean untuk instrumen *five-*

Upur Suparta, 2023

PENGEMBANGAN FIVE-TIER WORD ASSOCIATION DIAGNOSTIC INSTRUMENT (FTWADI) UNTUK MENGIDENTIFIKASI KONSEPSI SISWA PADA MATERI FLUIDA STATIS

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

tier seperti yang disajikan dalam tabel 3.4. Pengkodean dilakukan dalam enam kategori yaitu *Sound Understanding (SU)*, *Partial Positif (PP)*, *Partial Negatif (PN)*, *Misconception (MC)*, *Not Understanding (NU)*, dan *No Coded (NC)*.

Lima dari enam kategori tersebut masing-masing dibedakan lagi menjadi tiga level kognitif yaitu *Low* (rendah), *Middle* (sedang), dan *High* (Tinggi). Level kognitif tersebut didasarkan pada proses kognitif yaitu generalisasi, sintesis dan abstraksi. Dalam tes asosiasi kata level kognitif dinilai dari struktur kognitif peserta didik dengan kuat/ lemahnya asosiasi kata yang mereka miliki (Bahar, Johnstone & Sutcliffe, 1999; Derman & Eilks, 2016). Pengkategorian level kognitif diadopsi dari penelitian Turkkkan (2017) dengan membandingkan frekuensi asosiasi kata yang dimiliki peserta didik dengan jumlah frekuensi asosiasi kata untuk stimulus yang diberikan.

Tabel 3.6 Kategori Konsepsi Peserta Didik

No	Tier					Kategori
	1	2	3	4	5	
1	B	Y	B	Y	High	SU-Hi
					Middle	SU-Mi
					Low	SU-Lo
					Mc	SU-Mc
2	B	TY	B	Y	High	PP-Hi
3	B	Y	B	TY	Middle	PP-Mi
4	B	TY	B	TY	Low	PP-Lo
					Mc	PP-Mc

No	Tier					Kategori
	1	2	3	4	5	
5	B	Y	S	Y		
6	B	TY	S	Y		
7	B	Y	S	TY	High	PN-Hi
8	B	TY	S	TY	Middle	PN-Mi
9	S	Y	B	Y	Low	PN-Lo
10	S	TY	B	Y	Mc	PN-Mc
11	S	Y	B	TY		
12	S	TY	B	TY		
13	S	Y	S	Y	High	MC-Hi
					Middle	MC-Mi
					Low	MC-Lo
					Mc	Mc-Mc
14	S	Y	S	TY	High	NU-Hi
					Middle	NU-Mi
					Low	NU-Lo
					Mc	NU-Mc
15	S	TY	S	Y		
16	S	TY	S	TY		
17	-					NC

Keterangan:

B : Benar

S : Salah

Y : Yakin

TY : Tidak yakin

NU : *Not Understanding*

SU : *Sound understanding*

PP : *Partial positif*

PN : *Partial negatif*

MC : *Misconception*

NC : *No Coded*

Upar Suparta, 2023

**PENGEMBANGAN FIVE-TIER WORD ASSOCIATION DIAGNOSTIC INSTRUMENT (FTWADI) UNTUK
MENGIDENTIFIKASI KONSEPSI SISWA PADA MATERI FLUIDA STATIS**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Kemudian untuk memudahkan pengolahan data setiap level konsepsi dan asosiasi kata diberikan skor yang diadopsi dari penelitian yang dilakukan oleh Kaltakci-Gurel et al. (2017). Penskoran dibuat dalam dua kategori yaitu tingkat satu sampai empat dan tingkat lima. Untuk tingkat satu sampai empat, SU diberikan skor '1' karena siswa dapat menjawab benar pada tingkat pertama, tingkat ketiga, dan mereka menjawab yakin pada tingkat keyakinan jawaban. PP diberikan skor '2' karena mereka menjawab benar pada tingkat pertama dan ketiga tetapi menjawab tidak yakin pada salah satu tingkat keyakinan jawaban. PN diberikan skor '3' untuk mereka yang menjawab benar pada tingkat pertama atau tingkat ketiga. NU diberikan skor '4' untuk mereka yang menjawab salah pada tier pertama dan ketiga. MC diberikan skor '5' karena mereka menjawab salah pada tingkat pertama dan ketiga tetapi menjawab yakin pada tingkat keyakinan jawaban. NC tidak diberikan skor karena mereka tidak menjawab satu atau lebih tingkat jawaban.

Kemudian untuk tingkat lima, Hi diberikan skor '1' karena siswa memiliki asosiasi kata yang tinggi, Mi diberikan skor '2' karena siswa memiliki asosiasi kata yang sedang, Lo diberikan skor '1' yakni untuk siswa yang memiliki asosiasi kata pada level rendah dan Mc diberikan skor '4' karena siswa mengalami miskonsepsi pada asosiasi kata mereka. Adapun pemberian skornya dapat dilihat dari tabel 3.5.

Tabel 3.7 Skor Untuk Instrumen FTWADI

Kategori konsepsi		Asosiasi kata	
SU	1	Hi	1
PP	2	Mi	2
PN	3	Lo	3
NU	4	Mc	4
MC	5	Nc	-
NC	-		

Setelah dilakukan pengkodean dan penskoran, selanjutnya dilakukan pengolahan data menggunakan *Many Facet Rasch Measurement* (MFRM) untuk mengetahui validitas, reliabilitas, tingkat kesukaran dan daya pembeda. *The Many-Facet Rasch Model* dikembangkan oleh Mike Linacre (1989/1992) dalam penelitian

disertainya bersama Ben Wright di Chicago. Sejak saat itu, telah banyak digunakan di banyak konteks pengukuran (Wind & Hua, 2021). MFRM adalah suatu metode pengujian yang menggunakan skala peringkat yang melibatkan beberapa validator untuk menghasilkan penilaian yang lebih mendalam (Sumintono & Widhiarso, 2014; Zahir & Sumintono, 2017). MFRM memiliki keunggulan karena setiap hasil validasi dapat dipaparkan berdasarkan metode penilaian menggunakan skala peringkat, sehingga para ahli dapat menjelaskan secara tegas tentang skala peringkat yang mereka gunakan (Ariantara, 2022).

Dalam Many Facet Rasch Measurement (MFRM), tabel hasil data menyajikan beberapa nilai yang memiliki makna penting. Konsep "Item fit" dalam tabel ini mengindikasikan apakah butir soal berfungsi secara normal dalam melakukan pengukuran atau tidak (Sumintono & Widhiarso 2015). Nilai "infit" lebih menitikberatkan pada respons dari peserta yang memiliki kemampuan yang mendekati tingkat kesulitan suatu item. Sementara itu, "outfit" tidak memihak pada siapa pun dan lebih peka terhadap pengaruh jawaban yang tidak wajar menurut Bond dan Fox (2015). Nilai "MNSQ" yang merupakan ukuran keacakan dalam sistem pengukuran. Nilai ini mencerminkan jumlah penyimpangan yang ada (Sumintono & Widhiarso, 2015). Nilai "ZSTD" yang menggambarkan seberapa besar kemungkinan terjadinya penyimpangan menurut (Bond & Fox 2015). Kemudian data MFRM, "Pt Measure Corr" menunjukkan hubungan antara tingkat kesulitan masing-masing item dengan tingkat kesulitan instrumen secara keseluruhan (Smiley, 2015).

Untuk melakukan pengolahan data menggunakan *Many Facet Rasch Measurement* (MFRM) ini dibantu dengan program Minifac yang dikembangkan oleh Winstep. Prinsip pada pengolahan ini adalah data mentah dari *Microsoft Excel* diubah menjadi sebuah program pengkodean khusus untuk dianalisis dengan *inter-rater* (Zahir & Sumintono, 2017).

Uji unidimensionalitas (*unidimensionality*) instrumen adalah ukuran penting untuk mengevaluasi apakah instrumen yang dikembangkan mampu mengukur apa yang seharusnya diukur berdasarkan tujuan yang diharapkan (Sumintono & Widhiarso, 2015). Untuk nilai validitas pada tingkat instrumen

ditentukan dengan melihat unidimensional untuk nilai *variance explained by Rasch measures*. Kriteria *unidimensionality* pada *Rasch model* dijelaskan pada Tabel 3.8.

Tabel 3.8 Kriteria *Unidimensionality*

<i>Variance explained by Rasch measures</i>	Interpretasi
$20\% \leq X \leq 40\%$	Cukup
$40\% < X \leq 60\%$	Baik
$X > 60\%$	Sangat Baik

(Sumintono & Widhiarso, 2015)

Uji validitas bertujuan untuk mengukur ketepatan suatu instrumen yang digunakan dalam penelitian agar penelitian dapat dilaksanakan sesuai tujuan yang telah ditetapkan. Item instrumen dinyatakan valid berdasarkan *Rasch Model* menurut Sumintono & Widhiarso (2015) apabila memenuhi minimal dua dari tiga kriteria berikut.

- 1) Nilai *Outfit Mean Square* (MNSQ), untuk menguji konsistensi jawaban dengan tingkat kesulitan butir pernyataan;
- 2) Nilai *Outfit Z-Standard* (ZSTD), untuk mendeskripsikan *how much* hasil measure merupakan butir outfit, tidak mengukur atau terlalu mudah, atau terlalu sulit;
- 3) Nilai *Point Measure Correlation* (Pt Measure Corr), untuk mendeskripsikan *how good* butir pernyataan dipahami, direspon beda, atau membingungkan dengan item lainnya.

Untuk butir dinyatakan *fit* berdasarkan kriteria penerimaan *Outfit MnSq*, *Outfit ZStd* dan *PT Measure Correlation* yang dapat dilihat pada Tabel 3.9.

Tabel 3.9 Kriteria Penerimaan *Outfit Mnsq*, *Outfit Zstd* dan *PT Measure Correlation*

Nilai	Kriteria
<i>Outfit MnSq</i>	$0,5 < MnSq < 1,5$
<i>Outfit ZStd</i>	$-2,0 < ZStd < +2,0$
<i>PT Measure Correlation</i>	$0,4 < PT Measure Correlation < 0,85$

Menurut Sumintono & Widhiarso (2015), uji reliabilitas item dengan menggunakan *Rasch Model* menunjukkan keajegan item instrumen dalam

melakukan pengukuran. Suatu instrumen asesmen dikatakan mempunyai reliabilitas atau taraf kepercayaan yang tinggi jika instrumen asesmen tersebut memberikan hasil yang tetap (Arikunto, 2016). Hasil pengukuran itu harus tetap sama jika pengukurannya diberikan pada subjek yang sama meskipun oleh orang yang berbeda, waktu yang berbeda dan tempat yang berbeda pula. Hasil asesmen tidak terpengaruh oleh pelaku, situasi dan kondisi. Nilai *item reliability* dalam *Rasch model* memiliki kriteria yang terdapat pada Tabel 3.10.

Tabel 3.10 Kriteria Item Reliability

<i>Item Reliability</i>	Interpretasi
$0,94 \leq \text{nilai}$	Istimewa
$0,91 \leq \text{nilai} < 0,94$	Sangat Baik
$0,81 \leq \text{nilai} < 0,90$	Baik
$0,67 \leq \text{nilai} < 0,80$	Cukup
$\text{nilai} < 0,67$	Lemah

(Sumintono & Widhiarso, 2015)

Tingkat kesukaran menyatakan seberapa sukar atau mudahnya suatu soal. Secara kuantitatif, tingkat kesulitan dinyatakan dengan indeks kesukaran (Arikunto, 2016). Tingkat kesulitan butir pada *Rasch model* pada dasarnya sama dengan taraf kesukaran teori klasik, yaitu perbandingan antara jumlah jawaban benar dengan jumlah soal yang diujikan (*odd-ratio*). Hanya saja yang membedakan adalah nilai peluang itu kemudian diskalakan dengan memasukkan fungsi logaritma. Hasil estimasi logit dari *odd-ratio* inilah yang disebut logit atau *item measure*. Logits merupakan satuan pengukuran tingkat interval yang sesuai dengan skor total yang telah mengalami transformasi eksponensial (Ludlow dalam Rasch, 1960). Logit didefinisikan secara matematis sebagai log alami dari rasio peluang (*odds ratio*) (Ludlow dalam Cox, 1970). Rasio peluang menunjukkan frekuensi relatif dari suatu peristiwa yang terjadi versus frekuensi relatif dari peristiwa yang tidak terjadi (Ludlow & Haley, 1995). Misalnya, peluang suatu item dijawab benar adalah $\frac{2}{3}$ berarti bahwa dari 3 responden, item dijawab benar oleh 2 responden. Log natural dari $\frac{2}{3}$ adalah $-0,40$ (logit negatif), nilai tersebut menunjukkan bahwa item tersebut mudah. Sebaliknya, nilai logit yang tinggi menunjukkan item tersebut sukar

(Sumintono & Widhiarso, 2015). Kriteria tingkat kesukaran dalam *Rasch model* terdapat pada Tabel 3.11.

Tabel 3.11 Kriteria Tingkat Kesukaran

Tingkat Kesukaran (TK)	Interpretasi
$X > 0,50$	Sangat Sukar
$0,00 < X < 0,50$	Sukar
$-0,50 < X < 0,00$	Mudah
$X < -0,50$	Sangat Mudah

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara peserta didik yang berkemampuan tinggi (pandai) dengan peserta didik yang berkemampuan rendah (tidak pandai) (Arikunto, 2016). Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminasi. Daya diskriminasi Rasch atau nilai korelasi skor butir dan skor Rasch (*Point Measure Correlation*) pada prinsipnya sama dengan daya diskriminasi item yang diukur dengan pendekatan teori klasik. Hanya saja jika pada teori klasik komputasinya menggunakan skor mentah, untuk *Rasch model* yang digunakan adalah skor *Point Measure Correlation*.

Nilai *Point Measure Correlation* (*Pt Measure Corr.*) yang tinggi mengindikasikan bahwa peserta didik dengan abilitas rendah menjawab butir dengan salah dan peserta didik dengan abilitas tinggi menjawab butir dengan benar. Sementara nilai *Point Measure Correlation* negatif mengindikasikan butir soal yang menyesatkan karena peserta didik dengan kemampuan rendah mampu menjawab butir dengan benar dan peserta didik dengan kemampuan tinggi justru menjawab salah. Soal-soal dengan nilai korelasi negatif harus diperiksa untuk melihat apakah kunci jawaban salah, perlu direvisi, atau dihapus (Smiley, 2015).

Seperti pada teori klasik, nilai korelasi skor butir dan skor Rasch yang ideal adalah positif serta tidak mendekati nol. Alagumalai, Curtis, & Hungi (2005) mengklasifikasikan daya pembeda berdasarkan nilai *Point Measure Correlation* dalam *Rasch model* seperti yang ditunjukkan pada Tabel 3.12.

Tabel 3.12 Kriteria Daya Pembeda

Daya Pembeda	Interpretasi
>0,40	Sangat Baik
0,30–0,39	Baik
0,20-0,29	Cukup
0,00-0,19	Kurang
<0,00	Mebutuhkan pemeriksaan butir

(Alagumalai, Curtis, & Hungi, 2005)