

BAB III

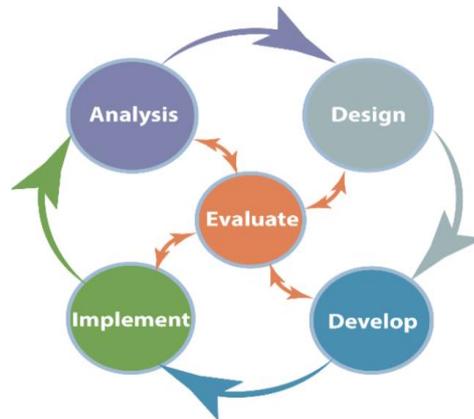
METODE PENELITIAN

Bab ini menyajikan metode penelitian yang digunakan dalam penelitian ini. Adapun paparan pada bagian ini meliputi desain dan rancangan penelitian, populasi dan sampel, instrumen penelitian, prosedur penelitian, teknik analisis data dari karakteristik PhyLeM-ReT, kuantitas dan kualitas perubahan konsepsi, dan profil keterampilan argumentasi peserta didik

3.1 Desain dan Rancangan Penelitian

Metode penelitian yang digunakan adalah pendekatan Penelitian Desain dan Pengembangan (*Design and Development Research*) (Richey & Klein, 2007). DDR adalah studi sistematis yang mengkaji desain, pengembangan, dan proses evaluasi dengan tujuan untuk membangun landasan empiris dalam menciptakan pembelajaran, produk, alat non-instruksional, serta model baru atau yang disempurnakan untuk mengatur perkembangan mereka (Richey & Klein, 2007). Richey & Nelson (2004) mengidentifikasi dua tipe DDR: (1) Tipe pertama berfokus pada perancangan dan evaluasi suatu produk atau program tertentu, dengan tujuan memperoleh pemahaman tentang proses pengembangan dan mempelajari kondisi yang mendukung implementasi program tersebut; (2) Tipe kedua berfokus pada pengkajian program pengembangan model yang telah dilaksanakan, dengan tujuan memperoleh gambaran tentang prosedur perancangan dan evaluasi yang efektif.

Dalam penelitian ini, peneliti menggunakan tipe 1, yaitu *Product and Tool Research* dalam konteks DDR, yang terdiri dari lima tahapan dalam proses pengembangannya, yakni analisis, desain, pengembangan, penerapan, serta evaluasi seperti ditunjukkan pada Gambar 3.1 (Richey, Rita C. & Klein, 2007). Produk yang dimaksud dalam penelitian ini adalah Media Pembelajaran Fisika berbasis *Rebuttal Texts* (PhyLeM-ReT) melalui model *Predict Observe Explain Apply Write* (POEAW), yang ditujukan untuk mengubah konsepsi dan melatih keterampilan argumentasi peserta didik pada materi usaha dan energi.



Gambar 3.1. Skema desain penelitian ADDIE

3.2 Populasi dan Sampel

Berdasarkan materi yang dipilih oleh peneliti, yaitu usaha dan energi, populasi dalam penelitian ini adalah seluruh peserta didik kelas X di salah satu SMA swasta di Kota Bandung yang menerapkan kurikulum merdeka. Sampel penelitian terdiri dari peserta didik kelas X. Teknik pengambilan sampel yang dilakukan adalah sampling jenuh, dimana seluruh anggota populasi kelas X digunakan sebagai sampel. Hal ini terjadi dikarenakan jumlah peserta didik dalam satu angkatan yang relatif sedikit yaitu berjumlah 30 orang.

3.3 Instrumen Penelitian

Instrumen penelitian yang digunakan terdiri dari instrumen tes dan non-tes. Instrumen tes meliputi tes diagnostik *four-tier open ended* pada topik usaha dan energi, yang bertujuan untuk mengetahui profil konsepsi peserta didik secara langsung di lapangan. Pada *tier-1*, terdapat soal pilihan ganda mengenai usaha dan energi, *tier-2* mencakup tingkat keyakinan terhadap jawaban di *tier-1*, sedangkan *tier-3* terdiri dari pertanyaan terbuka yang meminta alasan di balik pilihan jawaban pada *tier-1*, dan *tier-4* menilai tingkat keyakinan terhadap *tier-3*. Instrumen ini kemudian dikembangkan menjadi *Work and Energy in Four-tier Test (WonE-ForT)*, yang digunakan untuk mengidentifikasi konsepsi peserta didik. WonE-ForT terdiri dari empat *tier*: *tier-1* menyajikan soal pilihan ganda tentang konsep usaha dan energi, *tier-2* mencakup tingkat keyakinan terhadap jawaban di *tier-1*, *tier*

ketiga memberikan alasan terhadap *tier-1* dalam bentuk pilihan ganda yang diambil dari jawaban peserta didik dalam studi pendahuluan, dan *tier-4* menilai tingkat keyakinan terhadap *tier-3*. WonE-ForT digunakan pada *pre-test* dan *post-test* untuk mengukur tingkat pemahaman konsepsi peserta didik. Contoh WonE-ForT dapat dilihat pada Gambar 3.2 di bawah ini.

12.1 Indra melakukan percobaan secara simulasi melalui aplikasi PHET pada pemain *skate board* yang sedang meluncur. Diperoleh data ketinggian, energi potensial dan energi kinetik seperti ditunjukkan pada Tabel 1.

Tabel 1. Data energi potensial dan energi kinetik

Ketinggian (m)	Ep (J)	Ek (J)
5,83	3426,9	43
4,26	2506,6	963,3
2,77	1628,9	1841
0,04	25,5	3444,4

Grafik energi mekanik terhadap ketinggian yang tepat berikut ini adalah ...

Tier-1

A. B. C. D. E.

12.2 Tingkat keyakinan Anda terkait pilihan jawaban pertanyaan 12.1 : **Tier-2**
 A. Yakin B. Tidak yakin

12.3 Alasan Anda terkait pilihan jawaban pertanyaan 12.1 : **Tier-3**
 A. Semakin tinggi kedudukan suatu benda maka energi mekaniknya semakin besar.
 B. Semakin rendah kedudukan suatu benda maka energi mekaniknya semakin besar.
 C. Semakin besar energi kinetik maka energi mekaniknya semakin besar pula.
 D. Semakin besar energi potensial maka energi mekaniknya semakin besar pula.
 E. Energi mekanik merupakan penjumlahan energi kinetik dan energi potensial serta nilainya tetap.

12.4 Tingkat keyakinan Anda terkait pilihan jawaban alasan 12.3 : **Tier-4**
 A. Yakin B. Tidak yakin

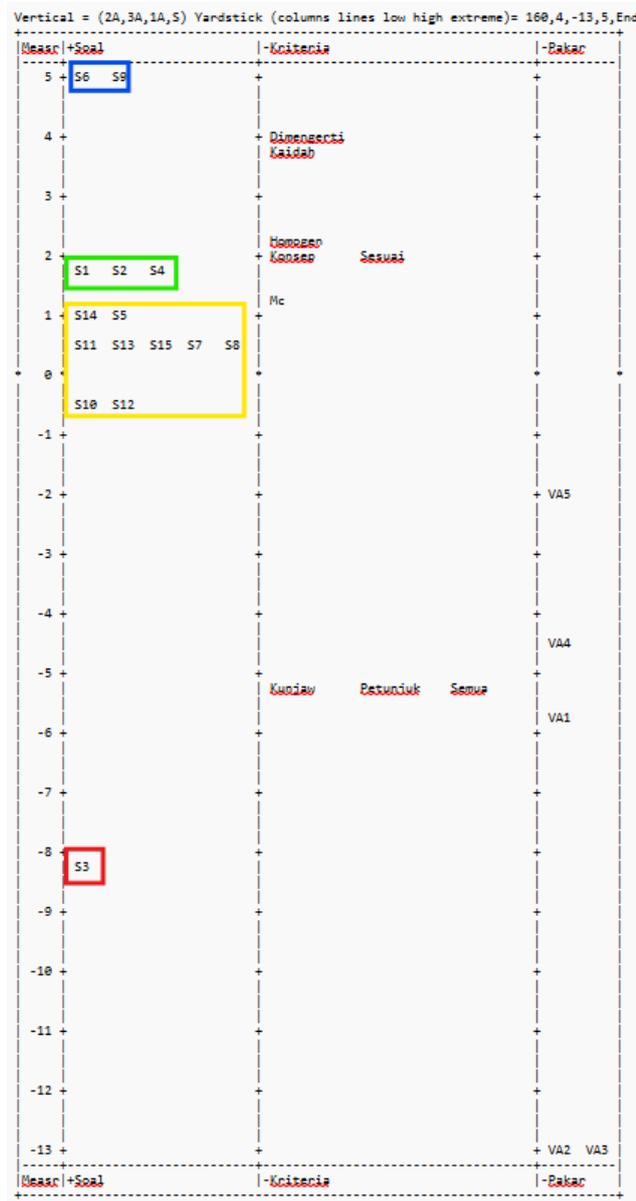
Gambar 3.2. Contoh WonE-ForT

Untuk analisis instrumen akan dijabarkan sebagai berikut:

3.3.1 Uji Validitas Isi (*Content Validity*)

Uji validitas isi dilakukan oleh lima orang ahli terhadap instrumen tes yang telah dikembangkan (judgment ahli). Validator untuk soal *four-tier* tentang usaha dan energi terdiri dari dua dosen ahli dan tiga guru fisika. Beberapa aspek yang dinilai dalam proses validasi isi dirinci menjadi sembilan pernyataan (Lampiran B1). Setiap ahli memberikan *checklist* terhadap masing-masing pernyataan dengan kategori “valid tanpa revisi” yang diberi skor 3, “valid dengan revisi” yang diberi

skor 2, atau “tidak valid” dengan skor 1. Hasil validasi dari para ahli kemudian dianalisis menggunakan uji rater dengan bantuan aplikasi Minifac. Hasil uji rater untuk validitas isi ditunjukkan dalam Gambar 3.3



Gambar 3.3. Hasil validitas isi soal *four-tier* usaha dan energi

Berdasarkan Gambar 3.3, bagian kiri menunjukkan soal (S1-S15), bagian tengah menunjukkan indikator validasi, dan bagian paling kanan adalah validator ahli (VA1-VA5). Soal yang dinilai terdiri dari 15 butir soal *four-tier* dengan sebaran multirepresentasi, termasuk pernyataan, gambar, tabel, dan simbol matematis. Indikator validasi ditulis dalam bentuk kode sebagai berikut: Kode “Mc”

menunjukkan “butir soal yang dibuat sesuai dengan pernyataan miskonsepsi”. Kode “Sesuai” mengacu pada “kesesuaian konsep dalam butir soal dengan konsep yang dikemukakan oleh para ahli”. Kode “Konsep” berarti “butir soal dibuat untuk mengetahui konsep peserta didik”. Kode “Kaidah” menunjukkan “penggunaan bahasa yang sesuai dengan kaidah bahasa Indonesia”. Kode “Dimengerti” berarti “bahasa yang digunakan mudah dipahami oleh peserta didik”. Kode “Homogen” menunjukkan “pilihan jawaban dan alasan yang homogen serta logis dari segi materi”. Kode “Kunjaw” berarti “hanya terdapat satu kunci jawaban”. Kode “Petunjuk” menunjukkan bahwa “soal tidak memberikan petunjuk jawaban menuju jawaban yang benar”. Kode “Semua” berarti “pilihan jawaban tidak menggunakan pernyataan ‘semua jawaban benar’ atau ‘semua jawaban salah’”.

Sesuai dengan data hasil uji rater yang ditunjukkan dalam Gambar 3.3, terdapat satu soal dengan kode S3 yang dinyatakan tidak valid. Validator ahli menilai bahwa soal yang terdapat pada kotak biru memenuhi sembilan indikator. Sementara itu, soal yang berada di kotak hijau dan kuning dianggap valid, tetapi memerlukan beberapa perbaikan berdasarkan saran yang diberikan oleh para ahli.

Setelah divalidasi oleh dosen ahli, soal-soal tersebut akan diuji coba kepada peserta didik untuk melakukan analisis terhadap butir soal, yang meliputi uji validitas konstruk, uji reliabilitas, daya beda, dan tingkat kesukaran. Berikut adalah cara analisis butir soal:

3.3.2 Uji Validitas Konstruk (*Construct Validity*)

Validitas merupakan ukuran yang menunjukkan sejauh mana suatu instrumen tes dapat dikatakan valid. Untuk memastikan data yang valid, instrumen yang digunakan untuk evaluasi harus memenuhi kriteria validitas. Dalam analisis Rasch, uji validitas konstruk disebut sebagai unidimensionalitas instrumen (*item unidimensionality*) (Sumintono & Widhiarso, 2014). Unidimensionalitas instrumen adalah hal yang penting untuk mengevaluasi apakah instrumen tes yang dikembangkan mampu mengukur apa yang seharusnya diukur, sehingga instrumen tersebut dapat dianggap valid.

Persyaratan agar suatu instrumen dapat dianggap valid dalam analisis Rasch melibatkan penggunaan analisis komponen utama (*principal component analysis*)

dari *standardized residual variance* (dalam satuan Eigenvalue) dan *raw variance explained by measures* (Sumintono & Widhiarso, 2014). Interpretasi validitas instrumen ini didasarkan pada *variance explained by measures* yang ditunjukkan pada Tabel 3.1 berikut:

Tabel 3.1. Interpretasi *item unidimensionality*

Nilai Raw Variance Explained By Measures	Interpretasi
Nilai < 20%	Perlu Perbaikan
20% ≤ Nilai < 40%	Terpenuhi
40% ≤ Nilai < 60%	Sesuai
60% ≤ Nilai	Istimewa

(Sumintono & Widhiarso, 2014)

Nilai *eigen* dan *observed* dalam *unexplained variance 1 st contrast* digunakan sebagai alat untuk menilai validitas instrumen. Sebuah instrumen dianggap valid jika *Eigenvalue* kurang dari 3 dan nilai *observed* kurang dari 15%. *Eigenvalue* dapat dihitung berdasarkan karakteristik yang dijelaskan oleh persamaan 3.1 berikut:

$$\det (A - \lambda I) = 0 \quad (3.1)$$

Dalam persamaan tersebut, A adalah matriks korelasi, λ adalah nilai eigen, dan I adalah matriks identitas. Jika kriteria tersebut tidak dipenuhi, maka dapat disimpulkan bahwa item dalam instrumen tes memiliki masalah dan tidak sesuai, sehingga tidak dapat secara akurat mengukur apa yang seharusnya diukur. Hasil dari pengolahan data mengenai uji validitas konstruk disajikan pada Tabel 3.2.

Tabel 3.2. Hasil Analisis Uji Validitas Konstruk

Sub-materi	Raw Variance Explained By Measures	Interp.	<i>unexplained variance 1 st contrast</i>		Interp.	Kesimpulan
			<i>Eigenvalue</i>	<i>Observed</i>		

Konsep usaha	51,1 %	Sesuai	2,9780	9,7%	Terpenuhi	Valid
Energi (EK, EP, EM)	50,7 %	Sesuai	4,4669*	9,2%	Tidak Terpenuhi	Valid
Konservatif & non-konservatif	51,4 %	Sesuai	2,6710	8,1%	Terpenuhi	Valid

Berdasarkan nilai *observed* pada *unexplained variance contrast*, tidak ditemukan adanya kecenderungan ketidaksesuaian pada butir soal, yang menunjukkan bahwa instrumen tersebut dapat digunakan dengan baik. Namun, nilai *eigenvalue* yang melebihi 3 mengindikasikan adanya butir soal yang bermasalah, sehingga perlu dilakukan analisis lebih mendalam dengan menggunakan *item fit order* untuk menilai apakah butir soal tersebut sebaiknya dipertahankan atau perlu diganti. Meskipun demikian, secara keseluruhan, *raw variance explained by measures* menunjukkan bahwa semua soal dalam instrumen tersebut dapat dianggap valid.

3.3.3 Analisis Tingkat Kesesuaian (*Fit-Statistic*)

Selain melakukan uji validitas konstruk, proses validitas juga dapat dilakukan menggunakan *fit-statistic* yang berasal dari analisis Rasch. Kualitas setiap item soal dapat dianalisis melalui urutan kecocokan item (*item fit order*) dengan cara meninjau nilai yang didapat dari *outfit mean square* (MNSQ), *Z-Standard* (ZSTD), dan *Point Measure Correlation* (*PT Measure Corr.*). kriteria untuk masing-masing parameter *outfit* tersebut dipaparkan pada Tabel 3.3.

Tabel 3.3. Kriteria *Outfit* MNSQ, ZSTD, dan *PT Measure Corr*

Kriteria	Nilai
<i>Outfit Mean Square</i> (MNSQ)	$0,5 < \text{MNSQ} < 1,5$
<i>Z-Standard</i> (ZSTD)	$-2,0 < \text{ZSTD} < +2,0$
<i>Point Measure Correlation</i> (<i>PT Measure Corr.</i>)	$0,4 < \text{PTMeasure Corr} < 0,85$

Kemudian hasil nilai *outfit* dan *PT Measure Corr.* tersebut diinterpretasikan berdasarkan kriteria pada Tabel 3.4 sebagai berikut:

Tabel 3.4. Interpretasi Kualitas Item Soal

Kriteria Nilai <i>Fit-Statistic</i>	Interpretasi
Ketiga kriteria nilai terpenuhi	Sangat Sesuai
Dua dari tiga kriteria nilai terpenuhi	Sesuai
Satu dari tiga kriteria nilai terpenuhi	Kurang Sesuai
Semua kriteria nilai tidak terpenuhi	Tidak Sesuai

Setelah dilakukan interpretasi terhadap masing-masing item, item soal yang akan digunakan dalam penelitian ini dapat mencerminkan kualitasnya sebagai alat ukur. Adapun kualitas item soal dari hasil analisis *fit-statistic* dijelaskan pada Tabel 3.5 berikut.

Tabel 3.5. Hasil Analisis Data Fit-Statistic dilihat dari Misfit-Order

Sub-materi	Nomor dan Konstruksi Soal	Nilai <i>Outfit</i>		PT	Interp.
		MNSQ	ZSTD	<i>Measure Corr.</i>	
Konsep usaha	1 (P, Ga)	0,72	-0,81	0,56	Sangat Sesuai
	2 (Ga)	1,43	1,27	0,51	Sangat Sesuai
	5 (P, Ga)	0,46*	-1,80	0,48	Sesuai
	7 (P, Ga, SM)	1,43	1,47	0,06*	Sesuai
Energi (EK, EP, EM)	3 (P, Ga, SM)	0,58	-1,33	0,60	Sangat Sesuai
	4 (Ga)	1,21	0,72	0,10*	Sesuai
	9 (P, Ga)	1,08	0,34	0,56	Sangat Sesuai
	11 (T, P, Gr)	1,64*	1,55	0,12*	Kurang Sesuai
	12 (P, Ga, Gr)	1,04	0,23	0,43	Sangat Sesuai
	14 (P, Ga)	1,23	0,76	0,23*	Sesuai
Konservatif & non- konservatif	6 (P, Ga)	0,57	-1,30	0,61	Sangat Sesuai
	8 (P, Ga)	1,14	0,54	0,48	Sangat Sesuai
	10 (P, Ga, SM)	1,41	1,16	0,40*	Sesuai

13 (P)	1,13	0,54	0,50	Sangat Sesuai
--------	------	------	------	---------------

Keterangan:

P = Pernyataan

G = Gambar (Ga) dan Grafik (Gr)

T = Tabel

SM = Simbol Matematik (Simbol, Angka, dan Persamaan matematis)

Interpretasi kualitas soal berdasarkan analisis *fit-statistic* (item fit order) pada sub-materi konsep usaha, energi, gaya konservatif & non-konservatif dikategorikan menjadi tiga tingkatan, yaitu “sangat sesuai”, “sesuai”, dan “kurang sesuai”. Berdasarkan interpretasi tersebut, semua soal bisa digunakan untuk penelitian. Namun, item soal yang “kurang sesuai” masih dapat digunakan tanpa diganti, meskipun diperlukan revisi item pada deskriptor dan opsi yang ada.

3.3.4 Tingkat Kesukaran Butir Soal (*Item Measure*)

Uji item soal lainnya bertujuan untuk menentukan tingkat kesukaran dari setiap butir soal. Evaluasi terhadap tingkat kesukaran ini digunakan untuk menilai sejauh mana item soal tersebut bisa dianggap mudah, sedang, atau sukar. Untuk menentukan tingkat kesukaran, analisis Rasch dilakukan dengan menggunakan aplikasi Ministeps.

Tingkat kesukaran setiap butir soal (*item measure*) dapat dianalisis dengan lebih mendalam melalui *output* yang tersedia di bagian item, *measure* (ME), serta standar deviasi (SD). Analisis dilakukan dengan membandingkan nilai ME untuk masing-masing item dengan nilai SD, dan berdasarkan perbandingan ini, kriteria nilai ME terhadap SD dapat ditetapkan. Dengan demikian, kriteria untuk setiap item dapat diperoleh, memudahkan penilaian. Interpretasi mengenai tingkat kesukaran setiap butir soal kemudian dapat diacu dari Tabel 3.6.

Tabel 3.6. Interpretasi Tingkat Kesukaran Item Soal

Kriteria	Interpretasi
$ME < -1SD$	Mudah
$-1SD \leq ME \leq +1SD$	Sedang
$ME > +1SD$	Sukar

Item measure masing-masing soal dipaparkan pada Tabel.3.7 berikut.

Tabel 3.7. Hasil Analisis Tingkat Kesukaran Soal

Sub-materi	Nomor dan Konstruksi Soal	<i>Measure</i> (ME)	<i>Standart</i>	Interp.
			<i>Deviasi</i> (SD)	
Konsep usaha	1 (P, Ga)	0,12	0,26	Sedang
	2 (Ga)	-0,01	0,26	Sedang
	5 (P, Ga)	0,30	0,26	Sukar
	7 (P, Ga, SM)	-0,41	0,26	Mudah
Energi (EK, EP, EM)	3 (P, Ga, SM)	0,13	0,37	Sedang
	4 (Ga)	-0,12	0,37	Sedang
	9 (P, Ga)	-0,32	0,37	Sedang
	11 (T, P, Gr)	0,73	0,37	Sukar
	12 (P, Ga, Gr)	-0,41	0,37	Mudah
	14 (P, Ga)	-0,02	0,37	Sedang
Konservatif & non- konservatif	6 (P, Ga)	0,45	0,40	Sukar
	8 (P, Ga)	-0,14	0,40	Sedang
	10 (P, Ga, SM)	0,27	0,40	Sedang
	13 (P)	-0,59	0,40	Mudah

Keterangan:

P = Pernyataan

G = Gambar (Ga) dan Grafik (Gr)

T = Tabel

SM = Simbol Matematik (Simbol, Angka, dan Persamaan matematis)

Berdasarkan hasil analisis tingkat kesukaran soal ditinjau dari nilai logit *measure* (ME) untuk masing-masing item yang dibandingkan dengan nilai SD, soal-soal tersebut dapat dikelompokkan ke dalam kategori mudah, sedang, dan sukar pada setiap sub-materi.

3.3.5 Uji Reliabilitas

Reliabilitas merupakan ukuran yang digunakan untuk mengevaluasi sejauh mana suatu instrumen pengumpul data dapat dianggap efektif. Sebuah instrumen

dinyatakan baik jika mampu secara konsisten memberikan data yang akurat dan sesuai dengan realitas. Konsistensi di sini berarti bahwa ketika sebuah tes dilakukan berulang kali pada waktu yang berbeda, hasil skor yang diperoleh sebaiknya relatif sama atau tidak mengalami perubahan yang signifikan.

Dalam penelitian ini, uji reliabilitas dilakukan melalui analisis Rasch dengan menggunakan aplikasi Ministeps. Untuk menentukan reliabilitas, analisis dilakukan dengan merangkum statistik yang mencakup nilai *Cronbach Alpha*, *item reliability*, dan *person reliability*. *Person reliability* menggambarkan tingkat kemampuan peserta didik, sementara *item reliability* merefleksikan kualitas dari pertanyaan yang ada dalam tes. Nilai *Cronbach Alpha* sendiri menunjukkan hubungan antara *reliability item* dan *reliability individu*, dan berfungsi untuk menilai konsistensi dari setiap butir soal yang diberikan. Setelah nilai *Cronbach Alpha* (α) diperoleh, hasil tersebut akan diinterpretasikan berdasarkan acuan yang terdapat dalam Tabel 3.8 berikut:

Tabel 3.8. Interpretasi Uji Reliabilitas berdasarkan Nilai *Cronbach Alpha* (α)

Nilai <i>Cronbach Alpha</i>	Interpretasi
$\alpha < 0,5$	Jelek sekali
$0,5 \leq \alpha < 0,6$	Jelek
$0,6 \leq \alpha < 0,7$	Cukup
$0,7 \leq \alpha < 0,8$	Bagus
$0,8 \leq \alpha$	Bagus sekali

(Sumintono & Widhiarso, 2014)

Selain itu, uji reliabilitas juga ditentukan berdasarkan nilai *item reliability* dan *person reliability*. Hasil dari masing-masing nilai tersebut diinterpretasikan sesuai dengan Tabel 3.9.

Tabel 3.9. Interpretasi Nilai *Item Reliability* dan *Item Person*

Nilai <i>Item Reliability</i> dan <i>Person Reliability</i>	Interpretasi
Nilai $< 0,67$	Lemah

$0,67 \leq \text{Nilai} < 0,8$	Cukup
$0,8 \leq \text{Nilai} < 0,9$	Bagus
$0,9 \leq \text{Nilai} < 0,94$	Bagus Sekali
$0,94 \leq \text{Nilai}$	Istimewa

(Sumintono & Widhiarso, 2014)

Berdasarkan dari interpretasi nilai *Cronbach Alpha*, *item reliability*, dan *person reliability* dapat diketahui reliabel atau tidak item yang akan digunakan untuk penelitian. Apabila semua indikator masuk pada kategori “Bagus”, maka instrument sangat reliabel. Tabel 3.10 menunjukkan hasil interpretasi nilai *Cronbach Alpha*, *item reliability*, dan *person reliability* pada setiap sub-materi.

Tabel 3.10. Hasil Analisis Uji Reliabilitas Soal

Seb Materi	<i>item reliability</i>	Interp.	<i>person reliability</i>	Interp.	<i>Cronbach Alpha</i>	Interp.	Kesimpulan
Konsep usaha	0,91	Bagus Sekali	0,68	Cukup	0,69	Cukup	Reliabel
Energi (EK, EP, EM)	0,91	Bagus Sekali	0,76	Cukup	0,79	Bagus	Reliabel
Konservatif & non-konservatif	0,90	Bagus Sekali	0,78	Cukup	0,79	Bagus	Reliabel

Soal yang telah diuji coba menunjukkan bahwa mereka dapat dianggap reliabel atau konsisten. Hal ini dapat dilihat dari nilai *Cronbach Alpha* (α) untuk setiap sub-materi, yang tergolong dalam kategori "bagus" dan "cukup." Nilai *Cronbach Alpha* (α) ini merefleksikan hubungan antara reliabilitas item dan reliabilitas individu. Reliabilitas item diinterpretasikan sebagai "bagus sekali," sementara reliabilitas individu dinilai sebagai "cukup." Sebuah nilai *person reliability* yang dikategorikan "cukup" menggambarkan bahwa konsistensi jawaban peserta didik cukup konsisten, yang mungkin disebabkan adanya variasi dalam kemampuan di antara peserta didik. Hal ini juga terlihat dari *fit-statistic*, yang menunjukkan bahwa terdapat soal yang mungkin kurang sesuai dengan standar yang diharapkan. Meskipun demikian, berdasarkan tiga kriteria tersebut instrumen

tes dapat dikatakan reliabel atau konsisten dalam pengumpulan data, karena ada pada kategori cukup, bagus, dan bagus sekali. Jika salah satu dari nilai tersebut rendah, perlu dilakukan peninjauan terhadap item-item dalam instrumen tersebut untuk mengidentifikasi mana yang mungkin perlu diperbaiki atau dihilangkan. (Sumintono & Widhiarso, 2015)

Instrumen non tes yang digunakan adalah lembar validitas *Physics Learning Media* berbasis *Rebuttal Texts* (PhyLeM-ReT), lembar validitas WonE-ForT, angket tanggapan peserta didik terhadap PhyLeM-ReT melalui *Predict Observe Explain Apply Write* (POEAW) pada topik usaha dan energi.

3.4 Prosedur Penelitian

Prosedur penelitian dilakukan secara bertahap sesuai dengan langkah-langkah yang terdapat pada tipe 1 *Product and Tool Research* dalam konsep DDR yang terbagi menjadi tahapan seperti berikut:

1) *Analysis* (Analisis)

Tahap pertama ini bertujuan untuk menganalisis dan merumuskan masalah yang akan diteliti. Dalam proses analisis ini memperhatikan kurikulum yang berlaku di sekolah, perkembangan peserta didik, serta kondisi yang ada di sekolah. Proses ini sering disebut sebagai pra-penelitian, yang mencakup analisis kurikulum, analisis peserta didik, dan analisis materi. Langkah selanjutnya adalah menganalisis capaian pembelajaran (CP) dan tujuan pembelajaran (TP). Selain itu, pada tahap ini, peneliti juga melakukan observasi atau studi lapangan untuk memperoleh informasi mengenai masalah-masalah atau hambatan yang dihadapi baik oleh peserta didik maupun guru dalam pembelajaran fisika. Studi lapangan dilakukan dengan memberikan tes *diagnostic* untuk mengetahui kategori pemahaman peserta didik, dimana terdapat 41% peserta didik mengalami miskonsepsi pada materi usaha dan energi. Selain itu, berdasarkan wawancara singkat terhadap guru dan peserta didik, ditemukan bahwa media pembelajaran yang digunakan hanya terbatas pada *power point* yang berisi materi saja.

2) *Design* (Perancangan)

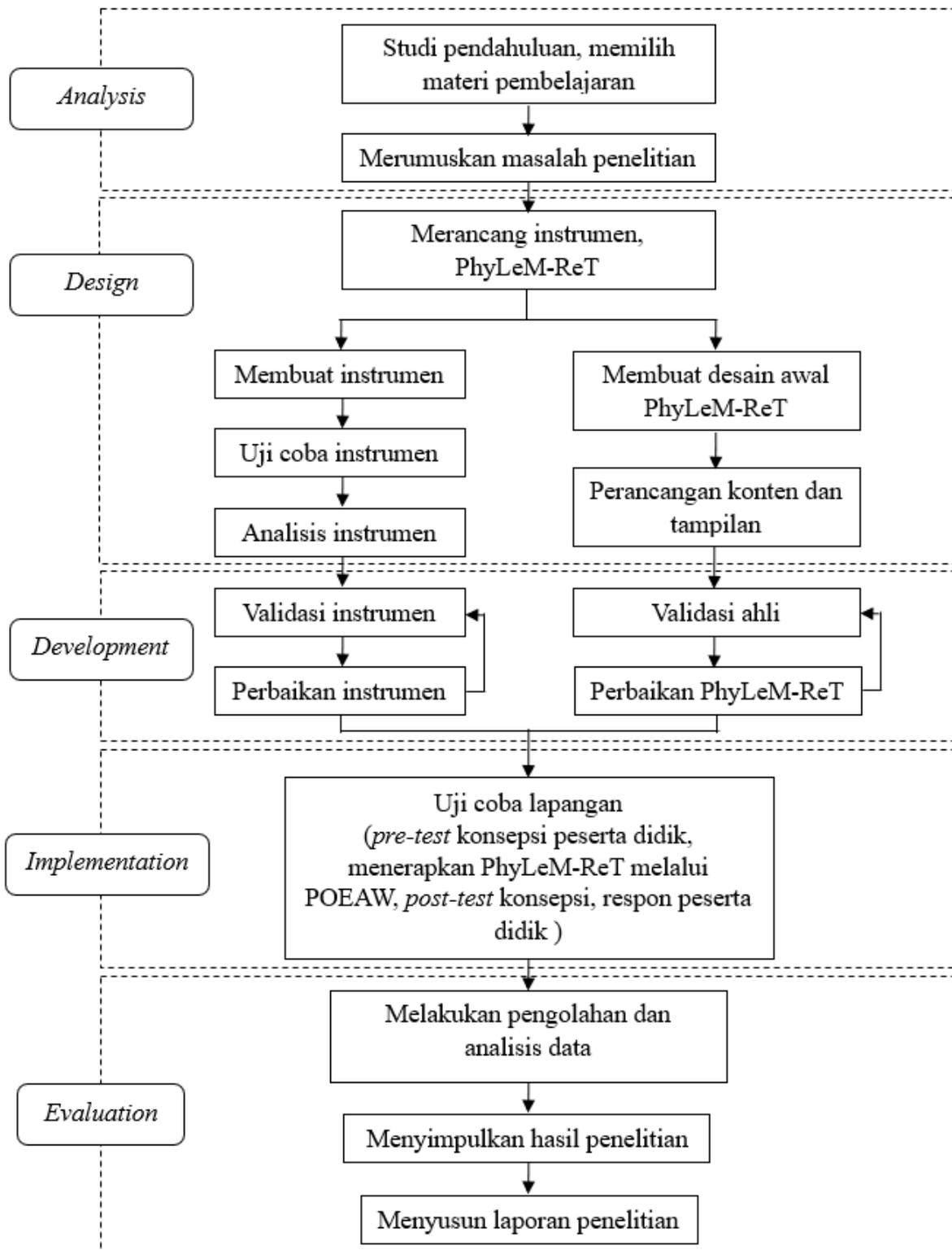
Tahap ini berfokus pada pengembangan desain awal *Physics Learning Media* berbasis *Rebuttal Texts* (PhyLeM-ReT), Rencana Pelaksanaan Pembelajaran (RPP), dan lembar validasi. Lembar validasi merupakan tahap dimana peneliti menyusun instrumen yang digunakan untuk menilai kelayakan PhyLeM-ReT yang telah dikembangkan (instrumen validasi). Selanjutnya, peneliti juga menyusun instrumen untuk menilai pemahaman konsep (instrumen tes) dan instrumen untuk mengevaluasi tanggapan peserta didik terhadap penggunaan PhyLeM-ReT yang telah dibuat.

3) *Development* (Pengembangan)

Pada tahap ini, pengembangan PhyLeM-ReT dilakukan menggunakan aplikasi Articulate Storyline. Setelah pengembangan selesai, produk tersebut akan diserahkan kepada validator untuk dilakukan pemeriksaan atau validasi agar dapat memastikan bahwa produk tersebut memenuhi kriteria kelayakan. Peneliti akan melakukan revisi berdasarkan masukan dan saran dari validator. Revisi ini bisa berlangsung selama beberapa kali sesuai kebutuhan, hingga produk akhir dianggap layak untuk digunakan. Hasilnya PhyLeM-ReT dikatakan valid dengan revisi menurut 4 ahli. Proses serupa juga diterapkan pada instrumen WonE-ForT sampai instrumen tersebut dinyatakan valid dan reliabel untuk mengukur pemahaman konsep peserta didik. Berdasarkan hasilnya diperoleh bahwa

4) *Implementation* (Implementasi)

Tahap ini menandai akhir dari proses penelitian pengembangan. Implementasi dilakukan dengan melakukan uji coba terhadap peserta didik, tanpa melakukan penyebaran secara luas, karena terbatas oleh waktu dan dana yang tersedia bagi peneliti. Sebelum pembelajaran menggunakan PhyLeM-ReT dimulai, peneliti akan memberikan pre-test untuk mengevaluasi pemahaman awal peserta didik. Selanjutnya, pembelajaran dilaksanakan dengan menerapkan PhyLeM-ReT yang diintegrasikan dengan model POEAW sesuai dengan rencana pembelajaran yang telah disusun pada fase analisis. Di akhir sesi pembelajaran, peneliti akan melaksanakan post-test untuk menilai pemahaman konsep yang diperoleh peserta didik. Selain itu, peneliti juga akan mengedarkan angket respons kepada peserta didik untuk menilai tanggapan mereka terhadap media yang dikembangkan. Data yang dikumpulkan dari uji coba lapangan akan dianalisis dan diolah, setelah itu peneliti akan menyimpulkan hasil dari penelitian yang telah dilakukan serta menyusun laporan penelitian. Prosedur penelitian ini juga digambarkan dalam skema yang terdapat pada Gambar 3.4.



Gambar 3.4. Prosedur Penelitian

3.5 Teknik Analisis Data

3.5.1 Analisis Karakteristik PhyLeM-ReT

Karakteristik pembelajaran menggunakan PhyLeM-ReT pada materi usaha dan energi dijelaskan mulai dari tahap awal pembuatan rancangan PhyLeM-ReT, yang mencakup proses mendesain hingga menjadikannya produk, validasi, dan deskripsi validasi yang telah memenuhi syarat serta layak digunakan untuk kegiatan pembelajaran dan pengumpulan data. Proses validasi ini dinilai dengan menggunakan skor dan kategori yang dapat dilihat pada Tabel 3.15 berikut:

Tabel 3.11. Kriteria Validasi

Skor Validasi	Kriteria Penilaian	Keterangan
3	Valid tanpa revisi	Dapat digunakan tanpa revisi
2	Valid dengan revisi	Dapat digunakan dengan revisi
1	Tidak valid	Belum dapat digunakan

Hasil validasi yang dilakukan oleh para ahli dianalisis menggunakan uji-rater dengan bantuan aplikasi Minifac. Setelah itu, hasil uji coba pada peserta didik juga dijelaskan untuk mengidentifikasi kekurangan yang berkaitan dengan manfaat praktis. Deskripsi mengenai karakteristik ini didukung oleh data yang diperoleh dari lembar keterlaksanaan aktivitas. Lembar observasi keterlaksanaan bertujuan untuk mengevaluasi penerapan PhyLeM-ReT melalui model POEAW dalam konteks kelas penelitian.

Lembar observasi dibagi menjadi dua jenis, yaitu observasi untuk guru dan observasi untuk peserta didik. Proses penilaian keterlaksanaan observasi dilakukan oleh dua orang guru menggunakan lembar observasi yang telah disediakan. Penilaian dilaksanakan setiap pertemuan sesuai dengan kegiatan pembelajaran yang tercantum dalam Modul Ajar. Untuk penilaian keterlaksanaan pembelajaran bagi peserta didik, diberikan skor mulai dari 1 sampai 4, dengan kriteria yang ditunjukkan pada Tabel 3.12.

Tabel 3.12. Kriteria Keterlaksanaan Pembelajaran

Skor	Deskripsi	Keterangan
4	Jika peserta didik selalu menunjukkan aktivitas seperti yang dituliskan dalam pernyataan pelaksanaan pembelajaran.	Sangat Baik
3	Jika peserta didik selalu menunjukkan aktivitas seperti yang dituliskan dalam pernyataan pelaksanaan pembelajaran tapi belum sepenuhnya baik.	Baik
2	Jika peserta didik memiliki kecenderungan menunjukkan aktivitas seperti yang dituliskan dalam pernyataan pelaksanaan pembelajaran.	Cukup Baik
1	Jika peserta didik kurang menunjukkan aktivitas seperti yang dituliskan dalam pernyataan pelaksanaan pembelajaran.	Kurang Baik

(Sumintono & Widhiarso, 2014)

Berdasarkan penilaian keterlaksanaan yang dilakukan oleh para observator, hasil yang diperoleh kemudian dianalisis dengan mengacu pada persentase yang dihasilkan. Selain itu, hasil penilaian keterlaksanaan ini juga dijelaskan secara deskriptif dengan merujuk pada catatan saran dan masukan yang diberikan oleh para observator di setiap tahapan model untuk setiap pertemuan yang berlangsung. Hal ini bertujuan untuk memberikan gambaran yang lebih rinci mengenai pelaksanaan dan untuk meningkatkan kualitas pelaksanaan di pertemuan berikutnya.

3.5.2 Analisis Kuantitas dan Kualitas Perubahan Konsepsi

Distribusi pemahaman peserta didik secara keseluruhan dapat dianalisis melalui *variable (wright) maps* dalam analisis Rasch yang dihasilkan dari hasil *pre-test* dan *post-test*. Berdasarkan jawaban peserta didik pada *pre-test* dan *post-test*, mereka dapat dikelompokkan kedalam beberapa kategori konsepsi. Kategori-kategori konsepsi tersebut disajikan dalam Tabel 3.13. berikut:

Tabel 3.13. Kategori Konsepsi

Kategori konsepsi	Tingkat 1 (Opsi)	Tingkat 2 (Tingkat Keyakinan)	Tingkat 3 (Alasan)	Tingkat 4 (Tingkat Keyakinan)	Skor
<i>Sound</i>					
<i>Understanding</i> (SU)	Benar	Yakin	Benar	Yakin	5
<hr/>					
<i>Partial Understanding</i> (PUP)	Benar	Yakin	Benar	Tidak yakin	4
	Benar	Tidak yakin	Benar	Yakin	
	Benar	Tidak yakin	Benar	Tidak yakin	
<hr/>					
<i>Partial Understanding</i> (PUT)	Benar	Yakin	Salah	Yakin	3
	Benar	Yakin	Salah	Tidak yakin	
	Benar	Tidak yakin	Salah	Yakin	
	Benar	Tidak yakin	Salah	Tidak yakin	
<hr/>					
<i>Partial Understanding</i> (PUN)	Salah	Yakin	Benar	Yakin	2
	Salah	Yakin	Benar	Tidak yakin	
	Salah	Tidak yakin	Benar	Yakin	
	Salah	Tidak yakin	Benar	Tidak yakin	
<hr/>					
<i>Misconception</i> (MC)	Salah	Yakin	Salah	Yakin	1
<hr/>					
<i>No Understanding</i> (NU)	Salah	Yakin	Salah	Tidak yakin	0
	Salah	Tidak yakin	Salah	Yakin	
	Salah	Tidak yakin	Salah	Tidak yakin	
<i>No Coding</i> (NC)	Jika tidak mengisi satu atau lebih item (tingkat)				-

1) Perhitungan konsepsi peserta didik

Berdasarkan informasi yang terdapat pada Tabel 3.13, terdapat tujuh kategori konsepsi yang dapat dikelompokkan dari peserta didik. Kategori-kategori tersebut meliputi SU, PUP, PUT, PUN, MC, NU, dan NC. Untuk mengevaluasi kuantitas

perubahan konsepsi setiap peserta didik dalam masing-masing kategori, dapat merujuk pada nilai $N\text{-change} < c >$. Nilai $< c >$ diperoleh dari analisis hasil *pre-test* dan *post-test* yang dilakukan terhadap setiap peserta didik. Selanjutnya untuk menghitung nilai tersebut, digunakan persamaan 3.1:

$$< c > = \begin{cases} \frac{\text{Skor posttest} - \text{Skor pretest}}{\text{Skor maksimum} - \text{Skor pretest}}, \text{posttes} > \text{pretest} \\ \text{Drop, posttest} = \text{pretest} = 100 \text{ or } 0 \\ 0, \text{posttest} = \text{pretest} \\ \frac{\text{Skor posttest} - \text{Skor pretest}}{\text{Skor pretest}}, \text{posttes} < \text{pretest} \end{cases} \quad (3.1)$$

(Marx and Karen, 2007)

Nilai $< c >$ yang diperoleh kemudian diinterpretasikan untuk mengetahui perubahan konsepsi peserta didik dalam kategori tinggi, sedang, rendah, atau negatif. Interpretasi nilai tersebut sesuai dengan yang tertera dalam Tabel 3.14 berikut.

Tabel 3.14. Interpretasi Nilai $N\text{-change}$

Nilai $N\text{-change}$	Interpretasi
$0,7 < <c> \leq 1$	Tinggi
$0,3 < <c> \leq 0,7$	Sedang
$0 \leq <c> \leq 0,3$	Rendah
$-1 \leq <c> < 0$	Negatif

(Marx and Karen, 2007)

Perhitungan konsepsi peserta didik setiap butir soal baik *pre-test*, maupun *post-test* dapat disajikan dalam bentuk persentase dengan menggunakan persamaan (3.2) sebagai berikut.

$$\text{Kategori Konsepsi (\%)} = \frac{\sum \text{siswa pada kategori konsepsi tertentu}}{\sum \text{seluruh siswa}} \times 100\% \quad (3.2)$$

2) Perubahan konsepsi peserta didik

Konsepsi peserta didik pada setiap butir soal dapat dianalisis perubahannya dengan persamaan (3.3).

$$\text{Perubahan Konsepsi (\%)} = \pm (\text{Kakhir}(\%) - \text{Kawal}(\%)) \quad (3.3)$$

Tanda (\pm) digunakan untuk menggambarkan perubahan konsepsi peserta didik sesuai dengan kriteria yang relevan untuk setiap soal. Penggunaan tanda (+)

mengindikasikan adanya perubahan yang diharapkan, yang mencakup kategori seperti SU, PUP, dan PUT. Disisi lain tanda (-) menunjukkan adanya perubahan yang tidak diinginkan, yang meliputi kategori PUN, MC, NU, maupun NC. Selain itu, tipe-tipe perubahan persentase konsepsi untuk setiap sub-materi telah dikelompokkan sebagaimana tercantum pada Tabel 3.15. berikut:

Tabel 3.15. Tipe Perubahan Persentase Konsepsi

Perubahan konsepsi	Tipe perubahan konsepsi
(+)	<i>Positive Change (PoC)</i>
(-)	<i>Negative Change (NeC)</i>
(0)	<i>Unchange (UC)</i>

3) Kategori perubahan konsepsi peserta didik

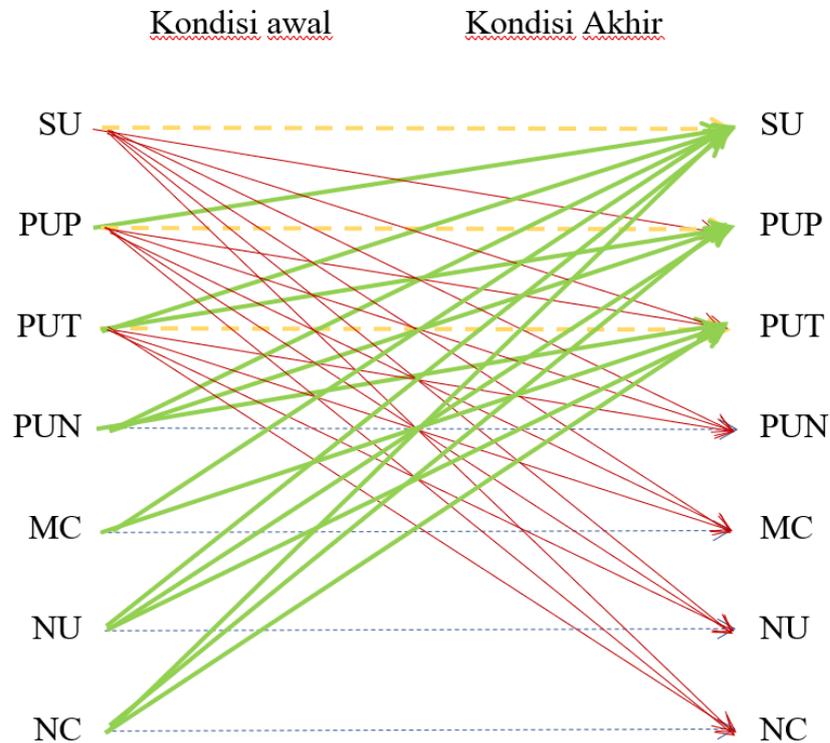
Karakteristik dari kategori perubahan konsepsi untuk setiap sub-materi yang dimiliki oleh peserta didik dapat dikelompokkan menjadi empat kategori, yaitu *Acceptable Change (ACh)*, *Not Acceptable (NA)*, *No change (NCh) (+)*, dan *No Change (NCh) (-)*. Kategori perubahan konsepsi ini diadaptasi dari penelitian yang dilakukan oleh Samsudin, *et al.*, (2016) dan rincian mengenai karakteristik tersebut dapat dilihat pada Tabel 3.16.

Tabel 3.16. Kategori Perubahan Konsepsi Peserta didik

Kategori perubahan konsepsi	Simbol perubahan konsepsi awal ke konsepsi akhir
<i>Acceptable Change (ACh)</i>	
<i>Not Acceptable (NA)</i>	
<i>No Change (NCh) (+)</i>	
<i>No Change (NCh) (-)</i>	

Kemungkinan kategori perubahan konsepsi peserta didik dari sebelum diberikan perlakuan (*pre-test*) hingga setelah perlakuan (*post-test*) diberikan simbol dan pengkodean seperti yang ditunjukkan dalam Gambar 3.5. *Acceptable Change (ACh)* adalah kategori perubahan konsepsi yang dapat diterima, dimana konsepsi yang sebelumnya berada pada kategori PUN, MC, NU, atau NC mengalami peningkatan menjadi SU, PUP, atau PUT. Disisi lain, *Not Acceptable (NA)* adalah

kategori perubahan konsepsi yang tidak dapat diterima, dimana konsepsi yang awalnya pada kategori positif seperti SU, PUP, atau PUT mengalami penurunan menjadi PUN, MC, NU, atau NC.



Gambar 3.5. Kemungkinan Kategori Perubahan Konsepsi

Kategori *No Change* (NCh) (+) mengacu pada kondisi di mana tidak ada perubahan kategori konsepsi peserta didik antara *pre-test* dengan *post-test*, namun konsepsi tersebut berada pada kategori positif seperti SU, PUP, dan PUN. Sebaliknya kategori *No Change* (NCh) (-) menunjukkan bahwa tidak ada perubahan kategori konsepsi dari *pre-test* ke *post-test*, tetapi konsepsi tersebut tetap pada kategori negatif seperti PUN, MC, NU dan NC. Proses perubahan atau ketidakberubahan konsepsi ini akan dijelaskan secara terperinci untuk setiap sub-materi.

3.5.3 Analisis Keterampilan Argumentasi

Untuk mengukur keterampilan argumentasi, peneliti menggunakan tes uraian. Tes ini dibangun berdasarkan komponen utama yang terdapat dalam argumentasi, yaitu *claim*, *data*, *warrant*, *backing*, *qualifier*, dan *rebuttal*. Komponen-komponen ini merupakan elemen kunci yang mendasari struktur argumentasi yang efektif.

Data yang diperoleh dari tes tersebut kemudian dianalisis menggunakan indikator keterampilan argumentasi sesuai dengan *Toulmin Argumentation Pattern*. Melalui analisis ini, hasil kemampuan argumentasi peserta didik dapat diukur dengan lebih akurat. Selanjutnya, Data yang diperoleh dianalisis menggunakan indikator keterampilan argumentasi *Toulmin's Argumentation Pattern*. Hasil analisis kemampuan argumentasi kemudian dikonversi berdasarkan kriteria level menurut *Toulmin's Argumentation Pattern* yang ditampilkan dalam Tabel 3.17.

Tabel 3.17. *Level Keterampilan Argumentasi*

Level	Kriteria	Keterangan
1	Argumentasi yang terdiri dari klaim yang sederhana	Buruk
2	Argumentasi yang memuat <i>claim</i> dengan <i>data</i> , namun tidak ada penjelasan eksplisit (<i>warrant</i>) tentang bagaimana data tersebut mendukung klaim	Kurang Baik
3	Argumentasi berisi klaim dengan data, serta terdapat penalaran atau penjelasan (<i>warrant</i>) yang menghubungkan data dengan klaim	Cukup Baik
4	Argumentasi selain berisi <i>claim</i> , <i>data</i> , <i>warrant</i> , juga terdapat <i>backing</i> atau pernyataan tambahan untuk memperkuat <i>warrant</i>	Baik
5	Argumentasi yang lebih kompleks dan panjang yang berisi <i>claim</i> , <i>data</i> , <i>warrant</i> , <i>backing</i> , dan <i>rebuttal</i>	Sangat Baik

(Toulmin., 2003)

Keterampilan argumentasi dapat dianalisis melalui berbagai indikator, salah satunya indikator yang dikembangkan oleh Toulmin. Toulmin menggambarkan argumen sebagai pernyataan yang didukung dengan alasan dan bukti yang logis. Menurutny sebuah argumen terdiri dari enam aspek utama, yaitu klaim (*claim*), data (*data*), pembenaran (*warrant*), syarat (*qualifier*), dukungan (*backing*), dan sanggahan (*rebuttal*) (Toulmin, 2003). Pemahaman dan penerapan komponen-komponen ini sangat penting dalam membangun argumen yang kuat dan meyakinkan. Indikator-indikator yang terkait dengan komponen-komponen tersebut dapat dilihat secara lebih rinci dalam Tabel 3.18 berikut.

Tabel 3.18. Indikator Keterampilan Argumentasi

No	Indikator	Deskripsi
1	Klaim (<i>claim</i>)	Peserta didik mampu menyampaikan pendapat atau pernyataan dan diterima oleh audien
2	Data atau bukti (<i>data</i>)	Peserta didik mampu menyampaikan fakta atau bukti yang digunakan untuk mendukung klaim
3	Penalaran (<i>warrant</i>)	Peserta didik mampu menyampaikan pernyataan logis yang umum dan bersifat hipotetis yang digunakan untuk menghubungkan klaim dengan data/bukti.
4	Dukungan tambahan (<i>backing</i>)	Peserta didik mampu menyampaikan pernyataan lebih lanjut yang mendukung penalaran/ <i>warrant</i> .
5	Syarat atau kualifikasi (<i>qualifier</i>)	Peserta didik mampu menyampaikan kalimat tambahan yang memperkuat klaim tertentu
6	Sanggahan (<i>rebuttal</i>)	Peserta didik mampu menyampaikan pernyataan yang berlawanan dengan klaim.

(Toulmin, 2003)

3.5.4 Analisis Respons Peserta Didik

Hasil respon peserta didik digunakan untuk menganalisis tanggapan peserta didik setelah melakukan pembelajaran dengan menggunakan model POEAW berbantuan PhyLeM-ReT. Angket respons diberikan dalam bentuk *google-form*. Respons berisikan pernyataan terkait dengan peran media PhyLeM-ReT untuk mengubah konsepsi dan melatih keterampilan argumentasi pada usaha dan energi. Instrumen yang digunakan berupa tanggapan peserta didik yang terdiri dari sebelas pernyataan positif dan sembilan pernyataan negatif seperti tertera pada (Lampiran A7).

Selanjutnya dianalisis menggunakan model Rasch yang ditinjau dari *variable (Wright) map*. Hasil distribusi *item respons* dan tanggapan peserta didik dijelaskan berdasarkan pernyataan yang sukar serta mudah untuk disetujui, baik untuk pernyataan positif maupun negatif. Tanggapan persetujuan dikategorikan

menjadi dua kategori yaitu, “sangat setuju” (SS), “setuju” (S) dan tanggapan pertidaksetujuan dibagi dengan dua kategori yaitu “tidak setuju” (TS), “sangat tidak setuju” (STS). Hal tersebut berguna untuk menentukan tanggapan peserta didik terhadap peranan PhyLeM-ReT dalam implementasi model POEAW.