

### BAB III METODE PENELITIAN

#### 3.1 Metode dan Desain Penelitian

Penelitian ini menggunakan desain penelitian *pre-experimental design* dengan *One Group Pretest-Posttest*. Desain ini menganalisis variabel terikat berpikir kritis dalam pembelajaran STEM-PBL pada materi suhu dan kalor. Dalam penelitian ini melibatkan satu kelompok yang sebelum perlakuan diberikan tes awal (*pretest*) kemudian dilanjutkan dengan *treatment* dan diakhir pembelajaran diberikan tes akhir (*posttest*). Alasan penggunaan desain ini adalah untuk mengetahui gambaran peningkatan berpikir kritis peserta didik setelah dilakukan pembelajaran berbasis STEM-PBL.

Tabel 3.1 Desain Penelitian *One Group (pretest - posttest) Design*

<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
O <sub>1</sub>	X	O <sub>2</sub>

Keterangan :

O<sub>1</sub> : Tes awal (*pretest*) berpikir kritis

O<sub>2</sub> : Tes akhir (*posttest*) berpikir kritis

X : Perlakuan menggunakan pembelajaran berbasis STEM-PBL

#### 3.2 Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh peserta didik kelas XI di salah satu SMA Negeri di kota Cimahi. Sedangkan sampel penelitian ini adalah terdiri dari satu kelas dengan jumlah 36 peserta didik. Pemilihan sampel dipilih secara *purposive sampling*, yaitu teknik penentuan sampel dengan pertimbangan tertentu agar dapat mendukung keterlaksanaan penelitian (Sugiono, 2013, hlm 124).

#### 3.3 Variabel Penelitian

Variabel bebas penelitian ini adalah pembelajaran berbasis STEM-PBL. Sedangkan variabel terikat penelitian ini adalah keterampilan berpikir kritis yang diukur dengan hasil *pretest* dan *posttest* peserta didik materi suhu dan kalor.

### 3.4 Prosedur Penelitian

Pada prosedur penelitian terdapat tiga tahapan, di antaranya yaitu tahap perencanaan, tahap pelaksanaan, dan tahap akhir. Tahapan tersebut dijelaskan sebagai berikut.

#### 3.4.1 Tahap Perencanaan

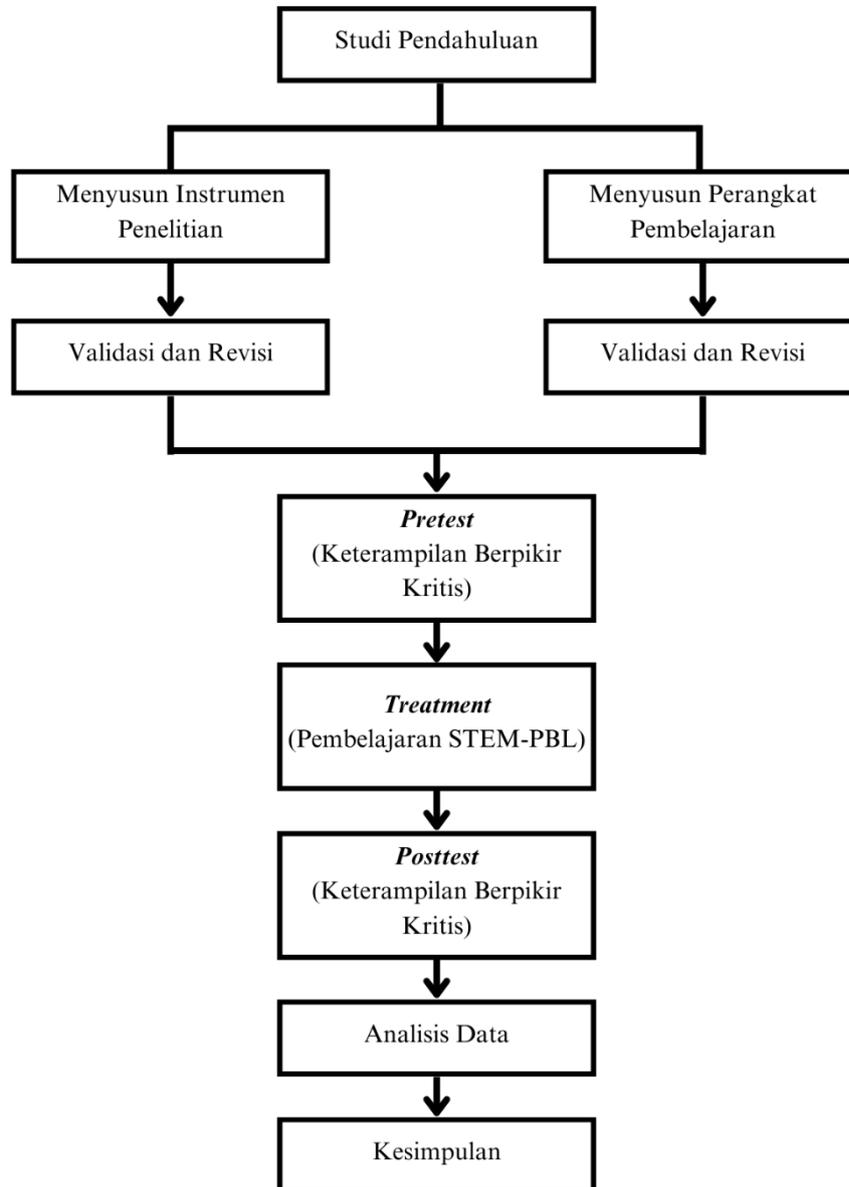
- 1) Melakukan studi pendahuluan berupa wawancara kepada guru dan pihak-pihak terkait, studi literatur jurnal, laporan penelitian dan buku mengenai penerapan pembelajaran berbasis STEM-PBL dan materi suhu dan kalor.
- 2) Menyusun perangkat pembelajaran yang meliputi modul ajar dan/atau RPP, bahan ajar materi suhu dan kalor serta LKPD biasa dan LKPD berbasis STEM.
- 3) Menyusun instrumen penilaian *pretest* dan *posttest* berbasis STEM berupa butir soal uraian, lembar observasi serta rubrik penilaiannya.
- 4) Melakukan validasi terhadap seluruh instrumen penelitian dengan bantuan dosen.
- 5) Melakukan uji coba dan analisis butir soal terhadap instrumen tes aspek pengetahuan untuk menentukan validitas dan reliabilitas soal yang akan digunakan.
- 6) Menentukan populasi dan sampel penelitian.

#### 3.4.2 Tahap Pelaksanaan

- 1) Sebelum diberikan perlakuan, peneliti memberikan *pretest* kepada 36 peserta didik untuk mengetahui kemampuan awal peserta didik terutama aspek berpikir kritis.
- 2) Memberikan perlakuan melalui kegiatan pembelajaran menggunakan pembelajaran berbasis STEM dengan model PBL Materi pokok pembelajaran yaitu materi suhu dan kalor.
- 3) Memberikan *posttest* untuk mengetahui keterampilan akhir siswa terkait dengan keterampilan berpikir kritis peserta didik saat setelah diberikan perlakuan.

### 3.4.3 Tahap Akhir

Tahap akhir penelitian merupakan proses pengolahan dan analisis data yang telah dikumpulkan. Mengolah dan menganalisis hasil *pretest* dan *posttest* sehingga didapatkan nilai N-gain. Jika seluruh data telah selesai dianalisis maka peneliti dapat menarik kesimpulan dari berdasarkan nilai N-gain yang diperoleh.



Gambar 3.1 Diagram Alir Prosedur Penelitian

### **3.5 Instrumen Penelitian**

#### **3.5.1 RPP (Rancangan Pelaksanaan Pembelajaran)**

RPP atau Rancangan Pelaksanaan Pembelajaran berisi KI dan KD, tujuan pembelajaran, sarana pembelajaran, materi ajar, LKPD berbasis STEM serta kegiatan pembelajaran STEM-PBL. RPP ini menjadi acuan bagi peneliti selama melakukan pembelajaran.

#### **3.5.2 Lembar Observasi Keterlaksanaan Pembelajaran**

Lembar observasi digunakan untuk memberi informasi tentang kesesuaian antara RPP yang telah disiapkan dengan kegiatan pembelajaran STEM-PBL. Observasi dilakukan oleh observer selama pembelajaran berlangsung.

#### **3.5.3 Soal Tes Keterampilan Berpikir Kritis**

Tes keterampilan berpikir kritis dilakukan sebelum dan setelah pembelajaran STEM-PBL. Soal dibuat dengan format uraian untuk melihat tingkat keterampilan berpikir kritis peserta didik berdasarkan kemampuan berpikir kritis yang dikembangkan oleh Ennis (2011).

### **3.6 Teknik Analisis Instrumen**

Pada penelitian ini, instrumen penelitian haruslah valid dan reliabel. Maka diperlukanlah uji validitas dan uji reliabel pada soal *pretest* dan *posttest* dengan menggunakan analisis data Rasch Model yang dibantu dengan *software* bernama *ministep*. Adapun penilaian yang dilakukan adalah dengan cara sebagai berikut.

#### **3.6.1 Validitas Ahli**

Sebelum melakukan uji coba, instrumen yang telah dirancang diuji validitasnya melalui penilaian dari penelaah atau ahli. Pembuktian validitas tipe ini menunjukkan kesesuaian isi tes dengan tujuan yang dibuat. Validasi ahli perlu mencakup tiga tipe validasi sebagai berikut.

- 1) Validitas isi instrumen terakit keterwakilan aspek dari keseluruhan yang diukur dan sejauh mana butir instrumen mencerminkan karakter yang diukur. Validasi isi mengandung beberapa aspek seperti kesesuaian butir soal terhadap indicator soal dan aspek yang ingin dicapai.
- 2) Validitas konstruk digunakan untuk mendapatkan informasi tentang kegunaan instrumen sebagai pengukur keterampilan berpikir kritis.

Instrumen dinilai oleh dua orang dosen ahli dan satu orang guru menggunakan lembar validitas instrumen.

- 3) Validasi bahasa memiliki kaitan dengan kerangka penyusunan butir soal berupa bentuk kalimat positif, penggunaan tata bahasa, tanda baca, dan pengejaan kata yang tepat.

Pada penelitian ini, penilaian validasi instrumen dilakukan oleh 3 validator, yaitu dua dosen ahli pembelajaran fisika dan satu guru fisika SMA. Adapun indikator validasi yang digunakan untuk menilai tiap butir soal, yaitu kemampuan butir soal dalam mengukur indikator keterampilan berpikir kritis, kebenaran konsep materi, kejelasan pertanyaan, bahasa pada butir soal sesuai Ejaan Bahasa Indonesia, dan rubrik penilaian dapat mengukur target yang diukur. Skala penilaian yang digunakan untuk menilai setiap indikator yaitu dengan rentang nilai 1-5. Peneliti melakukan perhitungan validitas isi menggunakan validasi Aiken. Berikut rumus untuk menghitung koefisien validasi Aikens's V (Azwar, 2013:134).

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

Keterangan:

$$S = r - l_0$$

r : Angka yang diberikan penilai

$l_0$  : Angka penilaian validitas yang terendah (misalnya 1)

c : Angka penilaian validitas tertinggi (misalnya 3)

n : Jumlah penilai

Setelah didapatkan indeks V berdasarkan perhitungan di atas, Aiken juga memberikan panduan untuk menentukan item tersebut diterima atau tidak. Nilai indeks V tersebut disajikan pada Tabel 3.2.

Tabel 3.2 Indeks V Aiken

No. of Items ( <i>m</i> ) or Raters ( <i>n</i> )	Number of Rating Categories ( <i>c</i> )													
	2		3		4		5		6		7			
	V	p	V	p	V	p	V	p	V	p	V	p		
2							1.00	.040	1.00	.028	1.00	.020		
3							1.00	.008	1.00	.005	1.00	.003		
3			1.00	.037	1.00	.016	.92	.032	.87	.046	.89	.029		
4					1.00	.004	.94	.008	.95	.004	.92	.006		
4			1.00	.012	.92	.020	.88	.024	.85	.027	.83	.029		
5			1.00	.004	.93	.006	.90	.007	.88	.007	.87	.007		
5	1.00	.031	.90	.025	.87	.021	.80	.040	.80	.032	.77	.047		
6			.92	.010	.89	.007	.88	.005	.83	.010	.83	.008		
6	1.00	.016	.83	.038	.78	.050	.79	.029	.77	.036	.75	.041		
7			.93	.004	.86	.007	.82	.010	.83	.006	.81	.008		
7	1.00	.008	.86	.016	.76	.045	.75	.041	.74	.038	.74	.036		
8	1.00	.004	.88	.007	.83	.007	.81	.008	.80	.007	.79	.007		
8	.88	.035	.81	.024	.75	.040	.75	.030	.72	.039	.71	.047		
9	1.00	.002	.89	.003	.81	.007	.81	.006	.78	.009	.78	.007		
9	.89	.020	.78	.032	.74	.036	.72	.038	.71	.039	.70	.040		
10	1.00	.001	.85	.005	.80	.007	.78	.008	.76	.009	.75	.010		
10	.90	.001	.75	.040	.73	.032	.70	.047	.70	.039	.68	.048		
11	.91	.006	.82	.007	.79	.007	.77	.006	.75	.010	.74	.009		
11	.82	.033	.73	.048	.73	.029	.70	.035	.69	.038	.68	.041		
12	.92	.003	.79	.010	.78	.006	.75	.009	.73	.010	.74	.008		
12	.83	.019	.75	.025	.69	.046	.69	.041	.68	.038	.67	.049		
13	.92	.002	.81	.005	.77	.006	.75	.006	.74	.007	.72	.010		
13	.77	.046	.73	.030	.69	.041	.67	.048	.68	.037	.67	.041		
14	.86	.006	.79	.006	.76	.005	.73	.008	.73	.007	.71	.009		
14	.79	.029	.71	.035	.69	.036	.68	.036	.66	.050	.66	.047		
15	.87	.004	.77	.008	.73	.010	.73	.006	.72	.007	.71	.008		
15	.80	.018	.70	.040	.69	.032	.67	.041	.65	.048	.66	.041		
16	.88	.002	.75	.010	.73	.009	.72	.008	.71	.007	.70	.010		
16	.75	.038	.69	.046	.67	.047	.66	.046	.65	.046	.65	.046		
17	.82	.006	.76	.005	.73	.008	.71	.010	.71	.007	.70	.009		
17	.76	.025	.71	.026	.67	.041	.66	.036	.65	.044	.65	.039		
18	.83	.004	.75	.006	.72	.007	.71	.007	.70	.007	.69	.010		
18	.72	.048	.69	.030	.67	.036	.65	.040	.64	.042	.64	.044		
19	.79	.010	.74	.008	.72	.006	.70	.009	.70	.007	.68	.009		
19	.74	.032	.68	.033	.65	.050	.64	.044	.64	.040	.63	.048		
20	.80	.006	.72	.009	.70	.010	.69	.010	.68	.010	.68	.008		
20	.75	.021	.68	.037	.65	.044	.64	.048	.64	.038	.63	.041		
21	.81	.004	.74	.005	.70	.010	.69	.008	.68	.010	.68	.009		
21	.71	.039	.67	.041	.65	.039	.64	.038	.63	.048	.63	.045		
22	.77	.008	.73	.006	.70	.008	.68	.009	.67	.010	.67	.008		
22	.73	.026	.66	.044	.65	.035	.64	.041	.63	.046	.62	.049		
23	.78	.005	.72	.007	.70	.007	.68	.007	.67	.010	.67	.009		
23	.70	.047	.65	.048	.64	.046	.63	.045	.63	.044	.62	.043		
24	.79	.003	.71	.008	.69	.006	.68	.008	.67	.010	.66	.010		
24	.71	.032	.67	.030	.64	.041	.64	.035	.62	.041	.62	.046		
25	.76	.007	.70	.009	.68	.010	.67	.009	.66	.009	.66	.009		
25	.72	.022	.66	.033	.64	.037	.63	.038	.62	.039	.61	.049		

(Aiken, 1985)

Kolom pertama pada tabel di atas menunjukkan jumlah *Raters*. Sedangkan *Number of Rating Categories* menunjukkan pilihan skala. Pada penelitian ini digunakan lima skala penentu validitas yaitu sangat tidak baik dengan skor 1, tidak baik dengan skor 2, cukup baik dengan skor 3, baik dengan skor 4, dan sangat baik dengan skor 5. Item instrumen pada penelitian ini dinilai oleh 3 penilai dengan 5 skala penentu validitas, maka nilai minimum indeks V pada tabel V adalah 0,92. Maka apabila nilai  $V_{hitung}$  lebih tinggi daripada nilai minimum  $V_{tabel}$ , maka butir validasi tersebut dinyatakan valid.

Hasil perhitungan indeks V yang diperoleh dari hasil validasi instrumen tes Kemampuan Berpikir Kritis oleh para ahli menggunakan formula Aiken disajikan pada Tabel 3.3.

Tabel 3.3 Hasil Analisis Data Validasi Instrumen Tes KBK

No Soal	Kemampuan butir soal dalam mengukur indikator KBK	Kebenaran konsep materi	Kejelasan pertanyaan	Bahasa pada butir soal sesuai dengan Ejaan Bahasa Indonesia	Rubrik Penilaian dapat mengukur target kemampuan yang diukur
1.	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
2.	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
3.	0,92 (Valid)	0,92 (Valid)	0,83 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
4.	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
5.	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)	0,83 (Tidak Valid)	0,83 (Tidak Valid)
6.	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,92	0,83 (Tidak Valid)
7.	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
8.	0,92 (Valid)	0,58 (Tidak Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
9.	0,92 (Valid)	0,58 (Tidak Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
10.	0,67 (Tidak Valid)	0,92 (Valid)	0,75 (Tidak Valid)	0,75 (Tidak Valid)	0,67 (Tidak Valid)
11.	0,67 (Tidak Valid)	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
12.	0,67 (Tidak Valid)	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
13.	0,67 (Tidak Valid)	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
14.	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
15.	0,92 (Valid)	0,75 (Tidak Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)
16.	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,92 (Valid)	0,83 (Tidak Valid)

Berdasarkan tabel hasil validitas Aiken di atas, nilai yang didapatkan berkisaran 0,58 – 0,92. Maka dari itu, terdapat beberapa soal yang dikategorikan valid dan tidak valid sesuai dengan keterangan yang sudah diberikan pada Tabel 3.3. Hasil analisis indeks Aiken pada 16 butir soal instrumen tes Kemampuan

Berpikir Kritis menunjukkan bahwa terdapat beberapa butir soal yang tidak valid karena nilai  $V$  aiken kurang dari nilai  $V_{\text{tabel}} = 0,92$ . Dalam hal ini validator menyarankan untuk dilakukan perbaikan terkait butir soal sesuai dengan mempertimbangkan saran dan masukan dari para validator yang terlampir.

### 3.6.2 Validitas Konstruk

Dalam pemodelan Rasch, uji validitas konstruk dapat dinilai berdasarkan uji unidimensionalitas instrumen. Unidimensionalitas merupakan ukuran yang penting untuk mengevaluasi apakah instrumen yang dikembangkan mampu mengukur apa yang seharusnya diukur. Berikut merupakan kategori nilai uji unidimensionalitas yang diterima (Sumintono & Widhiarso, 2015) disajikan pada Tabel 3.4.

Tabel 3.4 Kategori Nilai Uji Unidimensionalitas

Kategori	Nilai yang diterima
Raw-Variance	>20% (diterima)
	>40% (bagus)
	>60% (Istimewa)
Unexplained Variance	<15%
Unexplained Variance Eigen Value	<3%

Instrumen tes Kemampuan Berpikir Kritis diuji cobakan kepada peserta didik kelas XI di SMA Pasundan 1 Bandung dengan jumlah sampel uji coba 61 orang. Kemudian data tersebut dimasukkan ke *Microsoft Excel* kemudian diolah dengan *software ministep*. Hasil dari pengolahan uji validitas konstruk disajikan pada Gambar 3.2.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units			
	Eigenvalue	Observed	Expected
Total raw variance in observations =	28.2185	100.0%	100.0%
Raw variance explained by measures =	12.2185	43.3%	43.8%
Raw variance explained by persons =	5.3189	18.8%	19.0%
Raw Variance explained by items =	6.8997	24.5%	24.7%
Raw unexplained variance (total) =	16.0000	56.7%	56.2%
Unexplned variance in 1st contrast =	2.5009	8.9%	15.6%
Unexplned variance in 2nd contrast =	2.1759	7.7%	13.6%
Unexplned variance in 3rd contrast =	1.7331	6.1%	10.8%
Unexplned variance in 4th contrast =	1.6457	5.8%	10.3%
Unexplned variance in 5th contrast =	1.2162	4.3%	7.6%

Gambar 3.2 Hasil Analisis Uji Validitas Konstruk

Berdasarkan Gambar 3.2 data menunjukkan nilai *raw variance* > 20 % yaitu memiliki nilai 43,3% yang dikategorikan “bagus”. Sedangkan nilai *unexplained variance* menunjukkan tidak ada butir soal yang bermasalah, ditandai dengan nilai

*unexplained variance 1 st contrast* < 15% yaitu memiliki nilai 8,9%. Sehingga dapat disimpulkan bahwa butir soal mampu mengukur rentang kemampuan responden secara komprehensif.

### 3.6.3 Validitas Empiris

Validitas empiris digunakan untuk menguji kelayakan instrumen keterampilan berpikir kritis dengan cara membandingkan hasil uji coba instrumen dengan kriteria rentang yang dapat diterima. Validitas ini diolah menggunakan Model Rasch. Hal yang dapat terukur dari pemodelan Rasch ini adalah analisis pada butir, kemampuan peserta didik, dan instrumen secara keseluruhan. Model Rasch dapat memperlihatkan bagaimana hubungan kemampuan peserta didik terhadap tingkat kesukaran soal yang dibuat. Nilai kelayakan analisis validitas konten item dapat diinterpretasikan dengan kriteria rentang seperti pada Tabel 3.5.

Tabel 3.5 *Kategori Nilai Validitas Konstruk*

<b>Kriteria</b>	<b>Rentang yang diterima</b>
Outfit Mean Square (MNSQ)	$0,5 < \text{MNSQ} < 1,5$
Outfit Z Standard (ZSTD)	$-2 < \text{ZSTD} < 2$

MNSQ merupakan ukuran keacakan yaitu jumlah distorsi dalam sistem pengukuran. Nilai MNSQ merupakan nilai statistic chi-kuadrat yang dibagi dengan derajat bebas. Nilai MNSQ selalu positif bergerak dari nol hingga tak hingga. Nilai MNSQ digunakan untuk memantau kesesuaian data dengan model dengan harapan memiliki nilai 1 sebagai nilai ideal. Nilai MNSQ yang lebih besar daripada 1 memiliki arti bahwa data yang diobservasi memiliki variasi lebih banyak dari yang diprediksi oleh Rasch. Sedangkan nilai MNSQ kurang dari 1, mengindikasikan bahwa data yang diobservasi memiliki variasi lebih sedikit daripada yang di prediksi Rasch (Bond & Fox, 2015).

Nilai ZSTD adalah uji-t untuk hipotesis mengenai apakah data sesuai dengan model. Hasil ZSTD berupa nilai Z yaitu penyimpangan unit. Nilai Z dapat berupa negatif atau positif. Nilai negatif menunjukkan variasi yang sedikit dibandingkan pada model yang berarti seluruh responden yang berkemampuan tinggi mampu menjawab benar dan seluruh responden berkemampuan rendah menjawab salah pada butir tersebut. Sementara nilai positif memiliki arti bahwa

variasi jawaban lebih banyak dibandingkan pada model, sehingga respon jawaban tidak teratur dan tidak dapat di prediksi. Nilai ideal ZSTD ini adalah mendekati 0 (Bond & Fox, 2015).

Berikut merupakan data fit hasil analisis instrumen Kemampuan Berpikir Kritis dengan model Rasch disajikan pada Gambar 3.3.

SUMMARY OF 61 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	43.4	16.0	1.16	.30	1.03	.01	.99	-.04
SEM	1.0	.0	.09	.01	.05	.15	.05	.13
P.SD	7.7	.0	.68	.04	.40	1.16	.36	1.01
S.SD	7.7	.0	.69	.04	.40	1.17	.36	1.02
MAX.	60.0	16.0	3.11	.51	2.35	2.31	1.89	2.25
MIN.	24.0	16.0	-.44	.28	.25	-2.96	.35	-2.43
REAL RMSE	.33	TRUE SD	.60	SEPARATION	1.80	Person RELIABILITY	.76	
MODEL RMSE	.30	TRUE SD	.61	SEPARATION	2.02	Person RELIABILITY	.80	
S.E. OF Person MEAN = .09								
Person RAW SCORE-TO-MEASURE CORRELATION = .99								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .79 SEM = 3.56								
STANDARDIZED (50 ITEM) RELIABILITY = .93								
SUMMARY OF 16 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	165.4	61.0	.00	.15	1.02	-.18	.99	-.22
SEM	8.9	.0	.20	.01	.10	.60	.09	.52
P.SD	34.4	.0	.76	.02	.37	2.33	.35	2.02
S.SD	35.5	.0	.79	.02	.39	2.40	.36	2.08
MAX.	220.0	61.0	1.25	.21	1.62	3.44	1.75	3.83
MIN.	105.0	61.0	-1.42	.14	.37	-4.70	.42	-3.76
REAL RMSE	.17	TRUE SD	.74	SEPARATION	4.41	Item RELIABILITY	.95	
MODEL RMSE	.16	TRUE SD	.75	SEPARATION	4.80	Item RELIABILITY	.96	
S.E. OF Item MEAN = .20								
Item RAW SCORE-TO-MEASURE CORRELATION = -.99								
Global statistics: please see Table 44.								
UMEAN=.0000 USCALE=1.0000								

Gambar 3.3 Hasil Analisis Data Fit

Berdasarkan analisis data fit di atas, maka didapatkan nilai MNSQ dan ZSTD tiap *person* dan *item* yang menginterpretasikan kesesuaian data dengan model seperti disajikan pada Tabel 3.6.

Tabel 3.6 Pelaporan Uji Validitas

	MNSQ		ZSTD	
	Infit	Outfit	Infit	Outfit
<b>Person</b>	1,03	0,99	0,01	-0,04
<b>Item</b>	1,02	0,99	-0,18	-0,22

Berdasarkan data fit pada tabel di atas, nilai MNSQ dan ZSTD yang didapatkan sudah sesuai dengan kriteria yang digunakan untuk memeriksa konten

*person* dan *item* soal. Rentang nilai *infit* dan *outfit* MSNQ adalah 0,5 - 1,5, sedangkan nilai yang didapat untuk *infit* dan *outfit* MNSQ *person* dan *item* yaitu 1,03 dan 0,99 (*person*) serta 1,02 dan 0,99 (*item*). Nilai tersebut menginterpretasikan bahwa data yang diobservasi memiliki variasi yang sesuai dengan yang di prediksi oleh model Rasch (Bond & Fox, 2015).

Sedangkan nilai *infit* dan *outfit* ZSTD pada data fit menunjukkan hasil uji instrumen yang bernilai negatif. Hal ini berarti bahwa seluruh responden yang berkemampuan tinggi mampu menjawab benar dan seluruh responden berkemampuan rendah menjawab salah pada butir tersebut (Bond & Fox, 2015).

### 3.6.4 Uji Reliabilitas

Uji reliabilitas dilakukan untuk melihat tingkat ketetapan instrumen untuk mengukur keterampilan berpikir kritis peserta didik. Info mengenai reliabilitas instrume dapat dilihat pada hasil pengolahan data, kemudian menggunakan nilai *Alpha Cronbach* untuk mendefinisikan interkasi antara responden dengan butir soal secara keseluruhan. Kriteria nilai reliabilitas ini dapat dilihat pada Tabel 3.7.

Tabel 3.7 Kriteria Reliabilitas Instrumen

Skala	Kriteria Reliabilitas
$r_{11} < 0,20$	Sangat Rendah
$0,20 < r_{11} < 0,40$	Rendah
$0,40 < r_{11} < 0,70$	Sedang
$0,70 < r_{11} < 0,90$	Tinggi
$0,90 < r_{11} < 1,00$	Sangat Tinggi

Pada analisis Rasch Model, instrumen dapat meninjau reliabilitas kualitas butir soal serta konsistensi responden dalam menjawab soal. Hal ini ditinjau pada nilai *Person Reliability* dan *Item Reliability* (Sumintono dan Widhiarso, 2015). Kriteria nilai reliabilitas ini dapat dilihat pada Tabel 3.8.

Tabel 3.8 Kriteria Reliabilitas Person dan Item

Skala	Kriteria Reliabilitas
$< 0,67$	Lemah
0,67 - 0,80	Cukup
0,81 - 0,90	Bagus
0,91 - 0,94	Bagus Sekali
$> 0,94$	Istimewa

Hasil uji reliabilitas *person* dan *item* ditunjukkan seperti pada Gambar 3.4.

SUMMARY OF 61 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	43.4	16.0	1.16	.30	1.03	.01	.99	-.04
SEM	1.0	.0	.09	.01	.05	.15	.05	.13
P.SD	7.7	.0	.68	.04	.40	1.16	.36	1.01
S.SD	7.7	.0	.69	.04	.40	1.17	.36	1.02
MAX.	60.0	16.0	3.11	.51	2.35	2.31	1.89	2.25
MIN.	24.0	16.0	-.44	.28	.25	-2.96	.35	-2.43
REAL RMSE	.33	TRUE SD	.60	SEPARATION	1.80	Person	RELIABILITY	.76
MODEL RMSE	.30	TRUE SD	.61	SEPARATION	2.02	Person	RELIABILITY	.80
S.E. OF Person MEAN = .09								
Person RAW SCORE-TO-MEASURE CORRELATION = .99								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .79 SEM = 3.56								
STANDARDIZED (50 ITEM) RELIABILITY = .93								
SUMMARY OF 16 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	165.4	61.0	.00	.15	1.02	-.18	.99	-.22
SEM	8.9	.0	.20	.01	.10	.60	.09	.52
P.SD	34.4	.0	.76	.02	.37	2.33	.35	2.02
S.SD	35.5	.0	.79	.02	.39	2.40	.36	2.08
MAX.	220.0	61.0	1.25	.21	1.62	3.44	1.75	3.83
MIN.	105.0	61.0	-1.42	.14	.37	-4.70	.42	-3.76
REAL RMSE	.17	TRUE SD	.74	SEPARATION	4.41	Item	RELIABILITY	.95
MODEL RMSE	.16	TRUE SD	.75	SEPARATION	4.80	Item	RELIABILITY	.96
S.E. OF Item MEAN = .20								
Item RAW SCORE-TO-MEASURE CORRELATION = -.99								
Global statistics: please see Table 44.								
UMEAN=.0000 USCALE=1.0000								

Gambar 3.4 Hasil Analisis Uji Reliabilitas

Berdasarkan gambar hasil analisis di atas, didapatkan nilai *Alpha Cronbach*, *item reliability*, dan *person reliability* seperti disajikan pada Tabel 3.9.

Tabel 3.9 Pelaporan Uji Reliabilitas

	Rata-Rata Logit	Separation	Reliability	Alpha Cronbach
<i>Person</i>	1,16 (0,68)	1,80	0,76	0,79
<i>Item</i>	0,0 (0,76)	4,41	0,95	

Berdasarkan nilai pada tabel di atas, konsistensi jawaban dari peserta didik (*person reliability*) memiliki nilai 0,76 sehingga dapat dikategorikan “bagus” dengan kualitas butir soal instrument (*item reliability*) dengan nilai 0,95 adalah “istimewa”. Kemudian interpretasi nilai *alpha Cronbach* untuk instrumen tes Kemampuan Berpikir Kritis memiliki nilai 0,79 yang berarti memiliki “reliabilitas tinggi”. Ketiga nilai tersebut menunjukkan bahwa terdapat kesesuaian

antara butir soal dengan peserta didik yang baik sehingga dapat disimpulkan bahwa instrumen tes Kemampuan Berpikir Kritis dapat dikatakan reliabel.

### 3.6.5 Daya Diskriminasi Item

Daya diskriminasi item atau daya pembeda item berfungsi untuk mengukur kemampuan suatu butir soal untuk membedakan antara peserta didik berkemampuan tinggi dengan peserta didik berkemampuan rendah (Arikunto, 2015). *Pt Measure Corr* merupakan daya diskriminasi item yang memiliki prinsip yang sama dengan teori klasik korelasi poin biserial. Nilai *Pt Measure Corr* bernilai positif mengindikasikan bahwa seluruh responden dengan abilitas tinggi menjawab benar dan responden dengan abilitas rendah menjawab salah. Sedangkan nilai *Pt Measure Corr* negatif menunjukkan bahwa butir soal menyesatkan karena responden dengan kemampuan rendah mampu menjawab butir dengan benar dan responden dengan kemampuan tinggi justru menjawab salah (Smiley, 2015). Kriteria daya diskriminasi item berdasarkan *Pt Measure Corr* berkisar pada rentang  $0,4 < x < 0,85$ .

Hasil analisis uji coba instrumen tes Kemampuan Berpikir Kritis menunjukkan nilai *Pt Measure Corr* disajikan pada Gambar 3.5.

Item STATISTICS: MISFIT ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFINIT MNSQ	OUTFIT ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item		
16	137	61	.61	.14	1.62	3.44	1.75	3.83	A .32	.52	18.0	36.2	B16
6	207	61	-.92	.18	1.50	2.15	1.21	.91	B .50	.41	57.4	52.5	B6
14	157	61	.24	.14	1.44	2.57	1.38	2.14	C .42	.51	29.5	38.3	B14
4	123	61	.88	.14	1.42	2.36	1.43	2.30	D .31	.53	19.7	38.4	B4
7	119	61	.96	.14	1.41	2.31	1.27	1.51	E .54	.52	23.0	38.9	B7
1	220	61	-1.42	.21	1.22	.90	1.04	.25	F .39	.35	68.9	66.7	B1
11	214	61	-1.17	.19	1.10	.51	1.01	.11	G .47	.38	55.7	59.1	B11
5	132	61	.71	.14	1.04	.33	1.02	.16	H .44	.53	36.1	37.2	B5
10	202	61	-.77	.17	.98	-.02	.96	-.12	h .45	.43	45.9	48.4	B10
8	166	61	.06	.14	.78	-1.49	.88	-.70	g .49	.50	39.3	39.2	B8
15	105	61	1.25	.15	.71	-1.78	.82	-1.00	f .49	.52	47.5	42.6	B15
3	175	61	-.12	.14	.78	-1.37	.74	-1.53	e .68	.49	45.9	41.3	B3
2	182	61	-.27	.15	.70	-1.93	.68	-1.89	d .61	.48	45.9	42.4	B2
13	158	61	.22	.14	.63	-2.77	.61	-2.75	c .58	.51	62.3	38.3	B13
12	162	61	.14	.14	.56	-3.36	.57	-2.99	b .60	.51	59.0	38.1	B12
9	187	61	-.39	.15	.37	-4.70	.42	-3.76	a .68	.47	68.9	43.0	B9
MEAN	165.4	61.0	.00	.15	1.02	-.18	.99	-.22			45.2	43.8	
P.SD	34.4	.0	.76	.02	.37	2.33	.35	2.02			16.0	8.4	

Gambar 3.5 Hasil Analisis Daya Diskriminasi Item

Berdasarkan gambar di atas, terdapat beberapa soal yang dikategorikan tidak mampu mendeskriminasi karena memiliki nilai *Pt Measure Corr*  $< 0,4$  yaitu

pada butir soal 4 dengan nilai 0,31, butir soal 16 dengan nilai 0,32, dan butir soal 1 dengan nilai 0,39.

### 3.6.6 Tingkat Kesukaran

Uji tingkat kesukaran dilakukan untuk mengetahui tingkat kesukaran butir soal keterampilan berpikir kritis yang digunakan sebagai instrumen penelitian untuk mengukur keterampilan berpikir kritis peserta didik. Pemodelan Rasch menggunakan skor per orang maupun skor per butir soal untuk mengestimasi skor murni yang menunjukkan tingkat kemampuan individu serta tingkat kesukaran butir. Rasch menggunakan fungsi logaritma untuk mengonversikan nilai peluang probabilistik yang ada sehingga menghasilkan garis skala dengan interval yang sama, fungsi logaritma ini dinamakan skala logit (Sumintono dan Widhiarso, 2015).

Skala logit menunjukkan satuan yang dapat membandingkan yang disebut unit logit. Berdasarkan skala logit ini, peneliti bisa mendapatkan informasi mengenai tingkat kesukaran butir serta tingkat kemampuan peserta didik, serta pengelompokannya berdasarkan nilai standar deviasi.

Berdasarkan hasil analisis *measure* data menggunakan *software ministep*, didapatkan data tingkat kesukaran butir soal seperti pada Gambar 3.6.

Item STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	Item
15	105	61	1.25	.15	.71	-1.78	.82	-1.00	.49	.52	47.5	42.6	B15
7	119	61	.96	.14	1.41	2.31	1.27	1.51	.54	.52	23.0	38.9	B7
4	123	61	.88	.14	1.42	2.36	1.43	2.30	.31	.53	19.7	38.4	B4
5	132	61	.71	.14	1.04	.33	1.02	.16	.44	.53	36.1	37.2	B5
16	137	61	.61	.14	1.62	3.44	1.75	3.83	.32	.52	18.0	36.2	B16
14	157	61	.24	.14	1.44	2.57	1.38	2.14	.42	.51	29.5	38.3	B14
13	158	61	.22	.14	.63	-2.77	.61	-2.75	.58	.51	62.3	38.3	B13
12	162	61	.14	.14	.56	-3.36	.57	-2.99	.60	.51	59.0	38.1	B12
8	166	61	.06	.14	.78	-1.49	.88	-.70	.49	.50	39.3	39.2	B8
3	175	61	-.12	.14	.78	-1.37	.74	-1.53	.68	.49	45.9	41.3	B3
2	182	61	-.27	.15	.70	-1.93	.68	-1.89	.61	.48	45.9	42.4	B2
9	187	61	-.39	.15	.37	-4.70	.42	-3.76	.68	.47	68.9	43.0	B9
10	202	61	-.77	.17	.98	-.02	.96	-.12	.45	.43	45.9	48.4	B10
6	207	61	-.92	.18	1.50	2.15	1.21	.91	.50	.41	57.4	52.5	B6
11	214	61	-1.17	.19	1.10	.51	1.01	.11	.47	.38	55.7	59.1	B11
1	220	61	-1.42	.21	1.22	.90	1.04	.25	.39	.35	68.9	66.7	B1
MEAN	165.4	61.0	.00	.15	1.02	-.18	.99	-.22			45.2	43.8	
P.SD	34.4	.0	.76	.02	.37	2.33	.35	2.02			16.0	8.4	

Gambar 3.6 Hasil Analisis Tingkat Kesukaran Item

Berdasarkan pada gambar di atas, kolom terakhir menunjukkan urutan butir soal sesuai dengan tingkat kesukarannya dari yang tersulit ke termudah

berdasarkan pada nilai *measure* kolom ke-4. Urutan paling atas memiliki *logit* terbesar (butir soal ke-15) sampai urutan paling bawah yang memiliki *logit* terkecil (butir soal ke-1). *Logit* yang tinggi menunjukkan tingkat kesukaran soal yang tinggi pula.

### 3.6.7 Pengambilan Keputusan Butir Soal

Berdasarkan analisis data menggunakan *software ministep* diperoleh nilai penting mengenai *logit*, *SEM*, *outfit MNSQ*, *outfit ZSTD*, dan *Point Measure Correlation*. Nilai tersebut diperoleh dari data yang dtunjukkan pada Gambar 3.7.

Person: REAL SEP.: 1.80 REL.: .76 ... Item: REAL SEP.: 4.41 REL.: .95

Item STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-AL CORR.	EXP.	EXACT OBS%	MATCH EXP%	Item
1	220	61	-1.42	.21	1.22	.90	1.04	.25	.39	.35	68.9	66.7	B1
2	182	61	-.27	.15	.70	-1.93	.68	-1.89	.61	.48	45.9	42.4	B2
3	175	61	-.12	.14	.78	-1.37	.74	-1.53	.68	.49	45.9	41.3	B3
4	123	61	.88	.14	1.42	2.36	1.43	2.30	.31	.53	19.7	38.4	B4
5	132	61	.71	.14	1.04	.33	1.02	.16	.44	.53	36.1	37.2	B5
6	207	61	-.92	.18	1.50	2.15	1.21	.91	.50	.41	57.4	52.5	B6
7	119	61	.96	.14	1.41	2.31	1.27	1.51	.54	.52	23.0	38.9	B7
8	166	61	.06	.14	.78	-1.49	.88	-.70	.49	.50	39.3	39.2	B8
9	187	61	-.39	.15	.37	-4.70	.42	-3.76	.68	.47	68.9	43.0	B9
10	202	61	-.77	.17	.98	-.02	.96	-.12	.45	.43	45.9	48.4	B10
11	214	61	-1.17	.19	1.10	.51	1.01	.11	.47	.38	55.7	59.1	B11
12	162	61	.14	.14	.56	-3.36	.57	-2.99	.60	.51	59.0	38.1	B12
13	158	61	.22	.14	.63	-2.77	.61	-2.75	.58	.51	62.3	38.3	B13
14	157	61	.24	.14	1.44	2.57	1.38	2.14	.42	.51	29.5	38.3	B14
15	105	61	1.25	.15	.71	-1.78	.82	-1.00	.49	.52	47.5	42.6	B15
16	137	61	-.61	.14	1.62	3.44	1.75	3.83	.32	.52	18.0	36.2	B16
MEAN	165.4	61.0	.00	.15	1.02	-.18	.99	-.22			45.2	43.8	
P.SD	34.4	.0	.76	.02	.37	2.33	.35	2.02			16.0	8.4	

Gambar 3.7 Hasil Analisis Nilai Logit, outfit MNSQ, dan *Pt Measure Corr*

Berdasarkan Gambar 3.7, didapatkan nilai-nilai penting yang dijadikan sebagai bahan pertimbangan untuk pengambilan keputusan butir soal seperti disajikan pada Tabel 3.10.

Tabel 3.10 Pelaporan Parameter Butir

Kode Item	Measure (-2 < measure < +2)	SEM (SEM < 0.5 logit)	Outfit MNSQ (0.5 – 1.5)	Outfit ZSTD (-2,0 < ZSTD < +2,0)	Point Measure Correlation (0.4 – 0.85)	Keputusan
B1	-1,42	0,21	1,04	0,25	0,39	Terima
B2	-0,27	0,15	0,68	-1,89	0,61	Terima
B3	-0,12	0,14	0,74	-1,53	0,68	Terima
B4	0,88	0,14	1,43	-2,30	0,31	Tolak
B5	0,71	0,14	1,02	0,16	0,44	Terima
B6	-0,92	0,18	1,21	0,91	0,50	Terima
B7	0,96	0,14	1,27	1,51	0,54	Terima
B8	0,06	0,14	0,88	-0,70	0,49	Terima
B9	-0,39	0,15	0,42	-3,76	0,68	Tolak
B10	-0,77	0,17	0,96	-0,12	0,45	Terima
B11	-1,17	0,19	1,01	0,11	0,47	Terima
B12	0,14	0,14	0,57	-2,99	0,60	Tolak
B13	0,22	0,14	0,61	-2,75	0,58	Tolak
B14	0,24	0,14	1,38	2,14	0,42	Terima
B15	-1,25	0,15	0,82	-1,00	0,49	Terima
B16	0,61	0,14	1,75	3,83	0,32	Tolak

Berdasarkan tabel pelaporan parameter butir di atas, terdapat beberapa butir soal yang tidak sesuai dengan kriteria pada model Rasch. Pertama, terdapat soal yang memiliki nilai *Pt Measure Corr* yang lebih kecil dari batas minimum ( $0,4 < Pt Measure Corr < 0,85$ ), yaitu pada butir soal 1 dengan nilai 0,39, butir soal 4 dengan nilai 0,31, dan butir soal 16 dengan nilai 0,32. Nilai *Pt Measure Corr* yang rendah menandakan bahwa butir soal memiliki daya diskriminasi yang kecil.

Kedua, terdapat soal yang memiliki nilai *outfit ZSTD* yang lebih kecil dari batas minimum ( $-2,0 < ZSTD < 2,0$ ), yaitu pada butir soal 4 dengan nilai -2,30, butir soal 9 dengan nilai -3,76, butir soal 12 dengan nilai -2,99, dan butir soal 13 dengan nilai -2,75. Kemudian terdapat dua butir soal dengan nilai *outfit ZSTD* yang terlalu tinggi, yaitu pada butir soal 14 dengan nilai 2,14 dan butir soal 16 dengan nilai 3,83. Nilai *outfit ZSTD* yang terlalu besar dan terlalu kecil menunjukkan bahwa butir soal tidak kompatibel dengan model yang diharapkan (Bond & Fox, 2015). Ketiga, terdapat soal yang memiliki nilai *outfit MNSQ* yang terlalu rendah ( $MNSQ < 0,5$ ) yaitu pada butir soal 9 dengan nilai 0,42. Ketidakcocokan atau tidak fit data berdasarkan nilai MNSQ dan ZSTD merupakan indikasi bahwa terjadi miskonsepsi pada peserta didik terhadap butir soal (Boone, Staver & Yal, 2014).

Jika terdapat butir soal pada kedua kriteria yaitu MNSQ dan ZSTD tidak terpenuhi, itu artinya butir soal tersebut tidak bagus dan perlu direvisi atau diganti. Berbeda dengan tingkat kesukaran item yang sifatnya konsisten, tingkat kesesuaian item ini sangat dipengaruhi oleh besarnya ukuran sampel. Kesalahan kunci jawaban, banyaknya individu yang asal-asalan dalam mengerjakan soal, dan soal yang memiliki daya beda rendah dapat menurunkan nilai kesesuaian item (Suminto & Widhiarso, 2015).

Berdasarkan hasil analisis parameter butir soal di atas, maka dapat disimpulkan butir soal yang diteliti dan dinyatakan valid dapat digunakan untuk mengukur Kemampuan Berpikir Kritis peserta didik pada materi Suhu dan Kalor. Adapun beberapa butir soal yang tidak fit dengan Rasch model tetapi dapat diterima dengan pertimbangan hanya satu kriteria yang tidak terpenuhi yaitu pada butir soal 1, 12, 13, dan 14. Sedangkan untuk butir soal 4, 9, dan 16 adalah butir soal yang ditolak atau dibuang dengan pertimbangan terdapat dua kriteria data fit yang tidak fit dengan Rasch model.

### 3.7 Teknik Analisis Data

#### 3.7.1 Keterlaksanaan Pembelajaran STEM- PBL

Data keterlaksanaan pembelajaran STEM-PBL diperoleh berdasarkan hasil observasi pengawas atau guru selama pembelajaran berlangsung yang kemudian dicatat dalam lembar observasi keterlaksanaan pembelajaran. Skor keterlaksanaan pembelajaran STEM-PBL didapat menggunakan rumus :

$$\text{Skor keterlaksanaan} = \frac{\text{tahapan yang terlaksana}}{\text{jumlah keseluruhan tahapan}} \times 100\%$$

Skor keterlaksanaan yang didapat kemudian diinterpretasikan sesuai dengan kriteria yang disajikan pada Tabel 3.11.

Tabel 3.11 Interpretasi Skor Keterlaksanaan

Skor Keterlaksanaan	Kategori
$87,60 < x < 100,00\%$	Sangat Baik
$62,60 < x < 87,60\%$	Baik
$37,60 < x < 62,60\%$	Sedang
$25,00 < x < 37,60\%$	Kurang
$0,00 < x < 25,00\%$	Sangat Kurang

### 3.7.2 Peningkatan Keterampilan Berpikir Kritis

Data keterampilan berpikir kritis peserta didik didapat setelah melaksanakan *pretest* dan *posttest*. Peningkatan keterampilan berpikir kritis peserta didik dilihat dengan menghitung nilai N-gain yang dinormalisasi dengan rumus

$$\langle g \rangle = \frac{\% \text{ Skor Post test} - \% \text{ Skor Pre test}}{100 - \text{Skor Pre test}}$$

Nilai N-gain yang didapat kemudian diinterpretasikan sesuai dengan kriteria yang disajikan pada Tabel 3.12

Tabel 3.12 Interpretasi Nilai N-gain

Nilai N-gain	Kriteria
$(g) > 0,7$	Tinggi
$0,3 < (g) < 0,7$	Sedang
$g < 0,3$	Rendah

(Hake, 1999)

### 3.7.3 Pengaruh Pembelajaran STEM-PBL Terhadap Keterampilan Berpikir Kritis

#### 1) Uji Prasyarat Analisis

Uji prasyarat analisis diperlukan untuk mengetahui apakah analisis data untuk pengujian hipotesis dapat dilanjutkan atau tidak. Analisis varians mempersyaratkan bahwa data berasal dari populasi yang berdistribusi normal. Oleh karena itu, uji normalitas menjadi prasyarat sebelum dilakukan pengujian statistika parametrik.

Uji normalitas adalah uji statistik yang dilakukan untuk mengetahui sebaran distribusi sebuah data. Dengan demikian, uji normalitas ini mengasumsikan bahwa data di tiap variabel berasal dari populasi yang berdistribusi normal. Uji normalitas pada penelitian ini menggunakan metode uji Shapiro Wilk dengan bantuan aplikasi Statistika IBM SPSS. Uji normalitas Shapiro Wilk digunakan apabila sampel kurang dari 50 (Sarjono dan Julianita, 2011).

Dasar pengambilan keputusan dalam uji normalitas adalah sebagai berikut.

- Jika nilai signifikansi  $< 0,05$ , maka dikatakan bahwa data tidak terdistribusi normal.

- Jika nilai signifikansi  $> 0,05$ , maka dikatakan bahwa data terdistribusi normal.

## 2) Uji Hipotesis

Uji hipotesis dilakukan dengan tujuan untuk menjawab pertanyaan penelitian yang masih bersifat dugaan sementara. Pada penelitian ini, dibuat hipotesis berdasarkan rumusan masalah yaitu terdapat pengaruh penerapan pembelajaran STEM-PBL terhadap keterampilan berpikir kritis peserta didik. Untuk mengetahui ada tidaknya pengaruh dari penerapan pembelajaran STEM-PBL terhadap keterampilan berpikir kritis peserta didik digunakanlah uji *paired sample t test*. Metode uji *paired sample t test* digunakan untuk membandingkan rata-rata dua sampel yang diambil dari subjek yang sama (berpasangan) untuk dianalisis apakah terdapat pengaruh dan perbedaan yang signifikan dari kedua data (Syafriani, D., dkk, 2023). Analisis ini digunakan untuk pengujian membandingkan hasil *pretest* sebelum diberikan perlakuan dengan hasil *posttest* setelah diberikan perlakuan. Dasar pengambilan keputusan pada uji *paired sample t test* adalah sebagai berikut.

- Jika  $t_{hitung} < t_{tabel}$ , maka  $H_0$  diterima
- Jika  $t_{hitung} > t_{tabel}$ , maka  $H_a$  diterima

Untuk mengetahui besarnya nilai t hitung dapat dihitung menggunakan rumus tersebut.

$$t = \frac{\sum d}{\sqrt{\frac{N \sum d_i^2 - (\sum d_i)^2}{(N - 1)}}$$

Keterangan :

d = selisih nilai *pretest* dan *posttest*

N = banyak sampel

Selain menggunakan  $t_{tabel}$ , uji *paired sample t test* juga dapat dihitung menggunakan aplikasi Statistika IBM SPSS dengan pengambilan keputusan sebagai berikut.

- Jika nilai signifikansi  $< 0,05$ , maka  $H_0$  ditolak dan  $H_a$  diterima.
- Jika nilai signifikansi  $> 0,05$ , maka  $H_0$  diterima dan  $H_a$  ditolak.

Adapun maksud dari  $H_0$  dan  $H_a$  adalah sebagai berikut.

- $H_0$  : tidak terdapat pengaruh penerapan pembelajaran STEM-PBL terhadap keterampilan berpikir kritis peserta didik.
- $H_a$  : terdapat pengaruh penerapan pembelajaran STEM-PBL terhadap keterampilan berpikir kritis peserta didik.