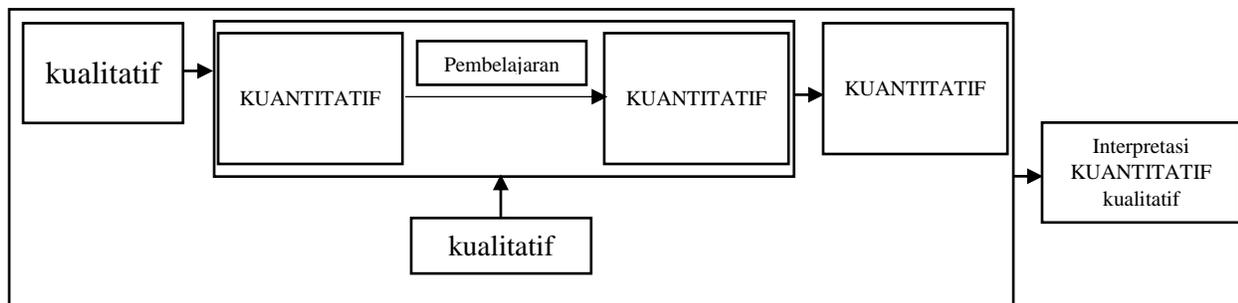


BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode penelitian campuran (*mixed methods*). Tashakkori dan Teddlie (2003), Tashakkori dan Creswell (2007), Johnson, Onwuegbuzie dan Turner (2007) mendefinisikan jenis penelitian campuran sebagai jenis penelitian dimana peneliti menggabungkan basis metode penelitian kuantitatif dan kualitatif. Desain penelitian yang digunakan adalah *embedded experimental model* (Creswell, 2006). Desain ini memiliki data kualitatif dalam sebuah desain eksperimental dimana waktunya menunjukkan tujuan untuk memasukkan data kualitatif (Creswell, 2014). Tahapan *embedded experimental model* dapat dilihat pada Gambar 3.1 berikut.



Gambar 3.1 *Embedded Experimental Model*

Penelitian ini melibatkan dua kelas yaitu terdiri dari kelas eksperimen dan kelas kontrol. Pada kelas eksperimen digunakan pembelajaran yang menerapkan GW-STM2I, sedangkan pada kelas kontrol dilakukan pembelajaran dengan tahapan pembelajaran yang sama dengan kelas eksperimen akan tetapi pada kelas kontrol tidak menggunakan GW-STM2I sebagai bahan ajar melainkan menggunakan buku konvensional/buku teks sebagai bahan ajarnya. Pada desain ini, tahap awal yang dilakukan adalah menggunakan metode kualitatif dengan

menganalisis karakteristik bahan ajar berbasis web berorientasi kecerdasan majemuk yang telah dikembangkan sebelumnya sehingga peneliti dapat melakukan perbaikan isi materi dan menyesuaikan GW-STM2I sesuai dengan tahapan pembelajaran, kemudian hasil datanya sebagai bahan kajian untuk masuk pada tahap berikutnya yaitu tahap penelitian menggunakan metode kuantitatif dan sebagainya. Tahap kedua merupakan pengumpulan serta analisis data kuantitatif. Data kuantitatif yang digunakan dalam penelitian ini diperoleh dari hasil pengolahan data instrumen *pretest* perubahan konsepsi sebelum melakukan pembelajaran kepada siswa baik kelas eksperimen maupun kelas kontrol. Selanjutnya, dilakukan pembelajaran dengan kelas eksperimen menggunakan GW-STM2I sebagai bahan ajarnya dan kelas kontrol menggunakan buku konvensional/buku teks sebagai bahan ajarnya. Setelah itu, pada tahap ketiga dilakukan pengambilan data kuantitatif kembali yaitu *posttest* perubahan konsepsi setelah melakukan pembelajaran. Selama proses tahap kedua dan ketiga berlangsung, dilakukan analisis pola perubahan konsepsi yang berlaku sebagai data kualitatif. Tahap selanjutnya merupakan tahap pengambilan data kualitatif dari hasil self-perceive aktivitas kecerdasan majemuk. Kemudian dari semua tahap tersebut diinterpretasikan dengan data Kuantitatif sebagai fokus utama penelitian sehingga memperoleh kesimpulan yang diharapkan.

3.2 Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah seluruh siswa kelas XI IPA di salah satu SMA di Bandung. Sampel diambil menggunakan teknik *convenience sampling* (Fraenkel, 2007). Sampel dari penelitian ini adalah dua kelas dari keseluruhan kelas populasi yaitu XI IPA. Kelas yang digunakan terdiri dari dua kelas yaitu satu kelas eksperimen berjumlah 29 siswa yang diberi perlakuan penerapan GW-STM2I dan satu kelas kontrol berjumlah 29 siswa yang diberi perlakuan penerapan buku teks yang tersedia di sekolah pada proses pembelajarannya. Namun, sehubungan dengan adanya pandemi Covid-19, penelitian ini dilakukan secara

online sehingga peneliti menyiapkan segala kebutuhan sesuai yang dibutuhkan oleh pembelajaran online.

3.3 Instrumen Penelitian

Instrumen yang digunakan pada penelitian ini terdiri dari dua data yakni data kualitatif dan data kuantitatif. Untuk jenisnya dibagi menjadi dua yakni instrumen tes dan non tes, yang akan dirinci sebagai berikut.

3.3.1 Angket kecerdasan majemuk

Angket kecerdasan majemuk yang digunakan berisi pertanyaan yang mewakili delapan tipe kecerdasan majemuk Gardner (2011). Siswa diharapkan untuk menjawab pertanyaan-pertanyaan dengan jawaban Sangat Setuju (SS), Setuju (S), Tidak Setuju (TS) dan Sangat Tidak Setuju (STS). Skor yang diberikan untuk setiap pertanyaan yang bersifat positif SS = 4, S = 3, TS = 2 dan STS = 1, kemudian dihitung persentase kecerdasan majemuk tersebut dari yang paling dominan.

3.3.2 Persepsi diri tentang aktivitas kecerdasan majemuk (*Self-perceive of multiple intelligences activities*)

Persepsi diri tentang aktivitas kecerdasan majemuk ini merupakan sebuah angket tentang persepsi diri terhadap aktivitas kecerdasan majemuk yang disajikan dalam setiap tahapan pembelajaran atau bisa disebut dengan *Self-perceive of multiple intelligences activities*. Angket ini digunakan untuk mengukur sejauh mana aktivitas siswa dalam pembelajaran, dimana siswa mengungkapkan apa yang siswa lakukan selama proses pembelajaran sesuai dengan kecerdasan majemuk yang diakomodasikan dalam pembelajaran dengan menjawab pertanyaan-pertanyaan dengan jawaban Sangat Setuju (SS), Setuju (S), Tidak Setuju (TS), dan Sangat Tidak Setuju (STS).

3.3.3 Instrumen tes Perubahan Konsepsi

Tes perubahan konsepsi diberikan sebanyak dua kali, yaitu di awal dan akhir pembelajaran. Tes ini bertujuan untuk mengukur konsepsi siswa sebelum dan sesudah diberikan perlakuan serta dapat melihat pola dari perubahan konsepsi tersebut. Tes awal digunakan untuk melihat kondisi awal subyek penelitian berkaitan dengan konsepsi siswa. Sedangkan tes akhir digunakan untuk melihat kondisi akhir perubahan konsepsi siswa. Hasil tes ini dihitung dengan *effect size* untuk melihat efektivitas dari penerapan GW-STM2I pada pembelajaran dalam merubah konsepsi siswa. Tes perubahan konsepsi ini berbentuk tes diagnostik dengan jenis pilihan ganda empat tingkat (*four tier test*) pada materi pemanasan global. Adapun bentuk tes perubahan konsepsi dapat dilihat pada Gambar 3.2.

Pertanyaan
Pilihan Jawaban (Tingkat 1)
A. _
B. _
C. _
D. _
Tingkat Keyakinan Jawaban (Tingkat 2)
A. _
B. _
C. _
D. _
Pilihan Alasan (Tingkat 3)
A. _
B. _
C. _
D. _
Tingkat Keyakinan Alasan (Tingkat 4)
A. _
B. _
C. _
D. _

Gambar 3.2 Instrumen Perubahan Konsepsi

Adapun kisi-kisi penyusunan tes perubahan konsepsi dapat dilihat pada lampiran 3. Agar tes dapat digunakan untuk pengambilan data maka harus dilakukan validitas, reliabilitas, tingkat kesukaran dan daya pembeda.

1. Uji Validitas

Validitas didefinisikan sebagai suatu kesesuaian, keberanian, dan kegunaan dari sebuah inferensi spesifik yang dibuat oleh peneliti berdasarkan data mereka kumpulan (Fraenkel dkk, 2011). Secara singkat, validitas adalah semua tentang penggambaran kesimpulan yang benar berdasarkan data yang diperoleh dari sebuah penilaian. Validasi dilakukan pada dua aspek yaitu *content-related evidence of validity* dan *criterion-related evidence of validity*. Validasi yang dilakukan menggunakan Rasch karena memperhatikan item dan respon serta korelasinya (Ardiyanti, 2016). Selain itu, pemodelan Rasch ini mampu melakukan prediksi terhadap data yang hilang (*missing data*) yang didasarkan kepada pola respon yang sistematis (Aziz, R., 2015).

Content-related evidence of validity ini mengacu pada konten dan format instrumen. Isi dan format tes harus sesuai dengan definisi variabel dan sampel subjek yang akan diukur. Oleh karena itu, validasi isi sebagian adalah masalah menentukan apakah isi instrumen adalah sampel yang memadai dari domain konten yang seharusnya diwakilinya. Aspek lain dari validasi konten adalah format instrumen, kejelasan pencetakan, kesesuaian bahasa, kejelasan arah, ukuran jenis dan lain sebagainya.

Criterion-related evidence of validity ini mengacu pada hubungan antar skor yang diperoleh dengan menggunakan instrumen dan skor yang diperoleh dengan menggunakan satu atau lebih instrumen lain. Kekuatan hubungan antara nilai tes dan kinerja kriteria memberikan kriteria terkait bukti validasi.

a. Validasi isi

Uji validitas yang dilakukan pada tahap ini adalah validasi isi. Validasi terhadap instrumen dengan validasi isi bertujuan untuk menilai kelayakan setiap butir soal yang dihasilkan. Validasi tes ditentukan dengan beberapa teknik secara kuantitatif dan kualitatif. Secara garis besar, indikator-indikator yang akan dinilai oleh para ahli adalah isi, bahasa dan tampilan. Pertama-tama item dari tes diagnostik ini diperiksa dan dinilai oleh para ahli berdasarkan isi dan format pada tes untuk memperoleh *content-related evidence of validity*. Tabel spesifikasi disiapkan untuk konteks dan kasus yang diukur dengan tes dan tabel untuk pilihan kesalahpahaman diperiksa oleh para ahli untuk kecukupan dan kompatibilitasnya dengan item uji. Termasuk didalamnya format, petunjuk arah, angka dan bahasa dinilai oleh para ahli.

Selanjutnya uji validitas dilakukan dengan menggunakan analisis uji *multi-rater (multi-faceted Rasch measurement)* pada *software* MINIFAC untuk dianalisis dengan Rasch. Langkah yang dilakukan untuk menghasilkan *output* uji *multi-rater*, adalah dengan membuat spesifikasi berkas data yang dipunyai dan jenis analisis yang dilakukan. File khusus tersebut tidak lain baris-baris perintah (*coding*) seperti halnya membuat satu program untuk tugas tertentu. *Coding* generik untuk analisis MINIFAC menggunakan program Notepad.

Coding generik untuk analisis MINIFAC menggunakan program Notepad dapat dilihat pada Lampiran 4. Hasil uji validitas isi instrumen perubahan konsepsi dapat dilihat seperti pada Gambar 3.3 berikut.

Measr	Soal	-Aspek	-Penilai	Scale
2	+ Soal 11 Soal 12 Soal 16 Soal 19	+	+	(4)
1	Soal 20	Keterkaitan		---
1	Soal 5	+ Seirama	+	+
	Soal 10 Soal 7	Mengungkapkan		3
0	+ Soal 1 Soal 17 Soal 21 Soal 22 Soal 4	Menjawab		---
	Soal 14 Soal 18	Jelas		2
	Soal 13 Soal 23 Soal 3	Sesuai		2
-1	+ Soal 2 Soal 6 Soal 8	+ EYD	Dosen A	---
-2				
-3			Dosen B Guru A	(1)
Measr	Soal	-Aspek	-Penilai	Scale

Gambar 3.3 Uji multi-rater (*multi-faceted Rasch measurement*)

Berdasarkan hasil yang didapat, secara garis besar menggambarkan kekuatan soal perubahan konsepsi yang sudah diisi oleh *judgment* dengan begitu instrumen soal tersebut sebagian dilakukan perbaikan sesuai dengan saran dan komentar yang diberikan *judgment*. Perbaikan dilakukan sebelum instrumen soal digunakan menjadi hal penting dalam pengembangan instrumen soal karena kesalahan bahasa atau kalimat merupakan faktor yang berpengaruh negatif terhadap performan tes pilihan ganda.

b. Validasi Empiris

Pada tahapan uji validasi ini, instrumen soal di uji cobakan. Setelah mendapatkan data hasil uji coba tersebut, selanjutnya uji validitas dilakukan dengan menggunakan menu *output 10. Item (column): fit order* pada *software* MINISTEP untuk di analisis dengan Rasch. Pengukuran uji validitas instrumen ini dilihat dari hasil skor pada menu *Outfit mean square (MNSQ)*, *Outfit Z-standard (ZSTD)*, dan *Point Measure Correlation (Pt Mean Corr)*. *Outfit* MNSQ berguna untuk melihat kesesuaian data dengan model yang digunakan.

Menurut Boone, dkk (2014), kriteria tersebut digunakan untuk mengetahui tingkat kesesuaian butir (*item fit*). Sedangkan penelitian yang dilakukan oleh Sabudin, dkk (2018) menyatakan bahwa *Point Measure Correlation (Pt Mean Corr)* hanya digunakan untuk mengetahui daya pembeda dari suatu instrumen. Maka untuk mengukur kevalidan atau kesahihan suatu instrumen hanya menggunakan skor pada menu *Outfit mean square (MNSQ)* dan *Outfit Z-standard (ZSTD)* pada *software* MINISTEP. Mengetahui kevalidan dari tiap soal dapat dicari dengan interpretasi menurut Boone, dkk (2014) pada Tabel 3.1 sebagai berikut.

Tabel 3.1 Interpretasi nilai *Output* MNSQ dan ZSTD

<i>Output item</i>	Skor	Keterangan
<i>Outfit MNSQ</i>	$0,5 < \text{MNSQ} < 1,5$	Diterima
<i>Outfit ZSTD</i>	$-2,0 < \text{ZSTD} < +2,0$	Diterima

Kriteria pengujian ini dapat dilihat dari Tabel 3.1, setiap soal dikatakan valid dan diterima jika 0,5 lebih kecil dari nilai MNSQ yang didapat dari perhitungan dan nilai MNSQ lebih kecil dari 1,5. Hasil perhitungan dapat dilihat pada Tabel 3.2.

Tabel 3.2 Interpretasi kevalidan item instrumen

No Soal	Skor MNSQ	Skor ZSTD	Keterangan
1	0.80	-0.53	Digunakan
2	1.07	0.89	Digunakan
3	0.82	-0.60	Digunakan
4	0.79	-0.53	Digunakan
5	1.22	0.77	Digunakan
6	1.00	0.06	Digunakan
7	1,08	0.45	Digunakan
8	0,97	-0,2	Digunakan
9	0.90	-0,41	Digunakan
10	0.75	-1.27	Digunakan
11	1.27	1,23	Digunakan
12	0,99	0,10	Digunakan

No Soal	Skor MNSQ	Skor ZSTD	Keterangan
13	0.89	-0.39	Digunakan
14	0.73	-0.97	Digunakan
15	1.43	1,94	Digunakan
16	0.92	-0.29	Digunakan
17	0.74	-0,62	Digunakan
18	1.01	0.14	Digunakan
19	0,97	-0,09	Digunakan
20	1,14	0.65	Digunakan
21	0.94	-0.10	Digunakan
22	0.96	-0.03	Digunakan
23	1,04	0.23	Digunakan

Berdasarkan Tabel 3.2 soal yang telah dilakukan pengujian validasi menunjukkan hasil yang valid dan diterima dengan begitu instrumen soal tersebut dapat digunakan sebagai instrumen perubahan konsepsi.

2. Uji Reliabilitas Instrumen

Reliabilitas instrumen mempunyai pengertian bahwa suatu instrumen cukup dapat dipercaya atau digunakan sebagai alat pengumpul data karena instrumen tersebut sudah baik dan tetap. Sebuah instrumen yang *reliabel* berapa kali digunakan dalam penelitian apapun hasilnya akan relatif sama sehingga hasil yang didapatkan dari instrumen tersebut dapat dipercaya. Uji reliabilitas pada penelitian ini menggunakan *software* MINISTEP 4.3.1. Perangkat lunak MINISTEP ini merupakan sebuah aplikasi pendukung untuk mengetahui kriteria dan kualitas suatu instrumen soal berdasarkan analisis Rasch. Salah satu menu *output* pada *software* ini yaitu *output 3.1 Summary Statistics* yang fungsinya yaitu untuk menampilkan beberapa nilai reliabilitas, diantaranya adalah *person reliability*, *item reliability*, dan *Cronbach alpha*. Nilai reliabilitas yang diambil dalam penelitian ini adalah *item reliability* menunjukkan kualitas item tes, artinya kita dapat mengetahui kualitas sebuah instrumen yang

digunakan dalam penelitian ini instrumen soal. Interpretasi untuk nilai *item reliability* dan *cronbach alpha* dapat dilihat pada Tabel 3.3 dan Tabel 3.4 berikut.

Tabel 3.3 Interpretasi item *reliability*

Nilai <i>item reliability</i>	Interpretasi
$0,91 \leq \text{Nilai}$	Istimewa
$0,91 \leq \text{Nilai} \leq 0,94$	Bagus Sekali
$0,80 \leq \text{Nilai} \leq 0,90$	Bagus
$0,67 \leq \text{Nilai} \leq 0,80$	Cukup
Nilai $< 0,67$	Lemah

Tabel 3.4 Interpretasi nilai *Cronbach alpha*

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

Nilai *person reliability* dan *item reliability* yang dipakai adalah *REAL RMSE* karena nilai ini merupakan kondisi terburuk reliabilitas batas bawah berdasarkan instrumen yang dipakai (Sumintono dan Widhiarso, 2015). Interpretasi mengenai besarnya nilai reliabilitas instrumen ditunjukkan pada Gambar 3.4.

SUMMARY OF 26 MEASURED (EXTREME AND NON-EXTREME) Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	7.8	23.0	-1.02	.60				
S.D.	4.5	.0	1.41	.28				
MAX.	20.0	23.0	2.28	1.83				
MIN.	.0	23.0	-4.92	.47	.70	-1.5	.45	-1.3
REAL RMSE	.69	TRUE SD	1.23	SEPARATION	1.79	Person	RELIABILITY	.76
MODEL RMSE	.66	TRUE SD	1.24	SEPARATION	1.88	Person	RELIABILITY	.78
S.E. OF Person MEAN = .28								

Person RAW SCORE-TO-MEASURE CORRELATION = .96
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .81

SUMMARY OF 23 MEASURED (NON-EXTREME) Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	8.8	26.0	.00	.52	.98	.0	1.06	.2
S.D.	4.6	.0	1.11	.09	.18	.8	.53	.9
MAX.	20.0	26.0	2.11	.79	1.37	1.7	2.89	2.9
MIN.	2.0	26.0	-2.59	.45	.66	-2.1	.36	-1.4
REAL RMSE	.54	TRUE SD	.97	SEPARATION	1.78	Item	RELIABILITY	.76
MODEL RMSE	.53	TRUE SD	.98	SEPARATION	1.85	Item	RELIABILITY	.77
S.E. OF Item MEAN = .24								

UMEAN=.0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = -.99
575 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 547.06 with 528 d.f. p=.2742
Global Root-Mean-Square Residual (excluding extreme scores): .3933

Gambar 3.4 Menu output *Summary Statistic*

Setelah dilakukan pengujian reliabilitas dengan menggunakan analisis Rasch, didapatkan hasil nilai *person reliability* sebesar 0,76 artinya konsistensi jawaban siswa berada dalam kategori cukup. Selanjutnya, untuk nilai *item reliability* yang muncul pada soal sebesar 0,76; artinya kualitas item tes pada instrumen tes yang digunakan berada dalam kategori cukup. Selanjutnya nilai *Cronbach Alpha* sebesar 0,81 yang berarti bahwa nilai yang mendekati angka 1 menunjukkan ukuran yang lebih konsisten (Boone dan Noltemeyer, 2017). Hasil perhitungan reliabilitas menunjukkan bahwa secara keseluruhan, instrumen tes diagnostik *four-tier test* yang diuji sudah reliabel dan dapat dipercaya sehingga sudah layak digunakan untuk mendiagnosis tingkat konsepsi siswa dalam penggunaan GW-STM2I pada materi pemanasan global.

3. Daya Pembeda

Daya pembeda digunakan untuk mengetahui kemampuan soal dalam membedakan peserta didik dengan kemampuan yang tinggi dan kemampuan yang rendah. Daya pembeda soal dapat diketahui menggunakan *software* MINISTEP 4.3.1 dengan menu *output Table 10 Item Fit Order* yang dianalisis dengan Rasch.

Korelasi *point-measure* mengacu pada hubungan antara tingkat kesulitan item tiap individu dan tingkat kesulitan soal secara keseluruhan, dimana nilai satu menunjukkan kemampuan peserta didik yang tinggi menjawab soal dengan benar dan sebaliknya yang menunjukkan korelasi sempurna menurut Rasch, nilai nol menunjukkan tidak ada hubungan antara respon item dan item secara keseluruhan, sedangkan nilai negatif menunjukkan masalah pada item soal karena sering mendapatkan skor rendah dibanding skor tinggi (Smiley, 2015).

Penggunaan *Point Measure Correlation (Pt Mean Corr)* seperti yang diungkapkan sebelumnya dapat memberikan informasi mengenai daya pembeda dari suatu instrumen soal yang digunakan untuk membedakan kemampuan peserta didik.

Untuk dapat mengetahui kategori daya pembeda tiap butir soal instrumen tes pada kolom *PT-MEASURE CORR.*, Smiley (2015) memberikan interpretasi untuk setiap nilai yang diberikan seperti yang terdapat pada Tabel 3.5 sebagai berikut.

Tabel 3.5 Interpretasi nilai Point Measure Correlation

<i>Pt Mean Corr</i>	Interpretasi
$0,40 < ID$	Sangat baik
$0,30 \leq ID \leq 0,40$	Baik
$0,20 \leq ID < 0,30$	Kurang baik
$ID < 0,20$	Jelek

Dari hasil perhitungan yang telah dilakukan, diperoleh nilai dan kriteria seperti pada Tabel 3.6 berikut.

Tabel 3.6 Daya pembeda butir soal

No. Soal	<i>Pt Mean Corr</i>	Interpretasi
S1	0,53	Sangat Baik
S2	0,38	Baik
S3	0,39	Baik
S4	0,56	Sangat Baik
S5	0,27	Kurang Baik
S6	0,40	Sangat Baik
S7	0,43	Sangat Baik
S8	0,65	Sangat Baik
S9	0,60	Sangat Baik
S10	0,67	Kurang Baik
S11	0,16	Jelek
S12	0,21	Kurang Baik
S13	0,42	Sangat Baik
S14	0,53	Sangat Baik
S15	0,20	Kurang Baik
S16	0,41	Sangat Baik
S17	0,53	Sangat Baik
S18	0,39	Baik
S19	0,48	Sangat Baik
S20	0,45	Sangat Baik
S21	0,42	Sangat Baik
S22	0,35	Baik
S23	0,24	Kurang Baik

4. Tingkat Kesukaran

Tingkat kesukaran bertujuan untuk membedakan soal yang terlalu mudah, mudah, sulit, dan sangat sulit. Hal ini diperlukan untuk menghasilkan distribusi soal yang baik. Tingkat kesukaran tes diagnostik STADT dapat diketahui dengan menggunakan *software* MINISTEP 4.3.1 pada menu *output Table 1 Variable (Wright) maps* dan *output Table 13 Item Measure* yang dianalisis dengan Rasch. *Output Variable (Wright) maps* pada *software* ini bertujuan untuk mendapat gambaran mengenai kekuatan setiap butir soal. Sesuai dengan penjelasan Boone dan Noltemeyer (2017) bahwa *wright maps* mendukung para peneliti untuk mengukur kekuatan dan kelemahan instrumen, mendokumentasikan penilaian butir tes, membandingkan teori dengan data eksperimen, dan memberikan panduan kepada para peneliti. Sedangkan *Item Measure* digunakan untuk mengklasifikasikan tingkat kesulitan tiap butir soal berdasarkan nilai logit dan nilai Standar Deviasi (SD) yang didapat (Sumintono dan Widhiarso, 2015).

Klasifikasi kategori tingkat kesukaran pada instrumen tes ini dapat diinterpretasikan seperti pada Tabel 3.7 sebagai berikut.

Tabel 3.7 Interpretasi tingkat kesukaran soal

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

Tingkat kesulitan butir dengan analisis Rasch pada dasarnya sama dengan taraf kesukaran teori tes klasik, yaitu perbandingan antara jumlah jawaban benar dengan jumlah soal yang diujikan. Perbedaannya, pada analisis Rasch nilai peluang diskalakan dengan memasukkan fungsi

logaritma, sehingga dapat menghasilkan suatu pengukuran dengan interval yang sama karena terjadi transformasi data *odd ratio* dengan logaritma (Sumintono dan Widhiarso, 2015).

Hasil untuk setiap soal dapat dilihat pada Tabel 3.8. Pada Tabel 3.8 dapat disimpulkan setiap aspek soal memiliki kriteria tingkat kesukaran soal termasuk dalam kategori sangat mudah sampai dengan sangat sukar.

Tabel 3.8 Hasil tingkat kesukaran tiap butir tes dengan analisis Rasch

No. Soal	Tingkat Kesukaran (TK)	Interpretasi
S1	0.85	Sukar
S2	-1.16	Sangat Mudah
S3	0.85	Sukar
S4	0.85	Sukar
S5	0.85	Sukar
S6	-0.76	Sangat Mudah
S7	-1.37	Sangat Mudah
S8	-2.59	Sangat Mudah
S9	-1.81	Sangat Mudah
S10	-0.76	Sangat Mudah
S11	0.08	Sukar
S12	2.11	Sangat Sukar
S13	0.31	Sukar
S14	0.31	Sukar
S15	0.31	Sukar
S16	0.08	Sukar
S17	1.58	Sangat Sukar
S18	0.08	Sukar
S19	-0.76	Sangat Mudah
S20	-0.76	Sangat Mudah
S21	-0.14	Mudah
S22	0.31	Sukar
S23	1.58	Sangat Sukar

3.4 Prosedur Penelitian

Prosedur penelitian pada penelitian ini disajikan pada Tabel 3.9 berikut.

Tabel 3.9 Prosedur Penelitian

Tahap	Fase	Prosedur
1	Kualitatif	<ul style="list-style-type: none"> - Melakukan redesain terhadap bahan ajar yang telah dikembangkan sebelumnya dengan menganalisis karakteristik bahar ajar tersebut - Memberikan angket tentang identifikasi tipe kecerdasan majemuk siswa
2	KUANTITATIF (kualitatif)	<ul style="list-style-type: none"> - Melakukan <i>pretest</i> tentang konsepsi - Melakukan pembelajaran yang menerapkan GW-STM2I pada kelas eksperimen dan menerapkan buku teks pada kelas kontrol - Melakukan <i>posttest</i> tentang konsep - Memberikan angket tentang self-perceive aktivitas kecerdasan majemuk - Melakukan analisis perubahan konsepsi dalam pembelajaran ditinjau dari tipe kecerdasan majemuk
3	KUANTITATIF (kualitatif)	<ul style="list-style-type: none"> - Melakukan analisis dan interpretasi data yang diperoleh dengan data kuantitatif sebagai fokus utama

A. Persiapan

Tahap awal penelitian ini adalah menelaah karakteristik bahan ajar berbasis web berorientasi yang telah dikembangkan sebelumnya sebagai dasar dalam mendesain GW-STM2I. Tahapan kegiatan yang dilakukan pada tahap ini dimulai dengan melakukan kajian terhadap karakteristik bahan ajar yang telah dikembangkan sebelumnya untuk mendapatkan informasi tentang

kelemahan dan kekurangan yang ada pada bahan ajar tersebut serta memberikan angket identifikasi kecerdasan majemuk untuk mendapatkan gambaran tentang keadaan kecerdasan yang dimiliki pada sampel penelitian, dilanjutkan dengan melakukan kajian literatur terhadap hasil penelitian mengenai bahan ajar berbasis web, kecerdasan majemuk, perubahan konsepsi serta sehingga penulis dapat mendesain ulang bahan ajar tersebut menjadi GW-STM2I.

A. Perencanaan

Tahap perencanaan ini yaitu menyusun desain GW-STM2I, membuat kisi-kisi instrumen perubahan konsepsi, dan perangkat pembelajaran yang diperlukan sesuai dengan data kuantitatif dan kualitatif. Penyusunan desain GW-STM2I ini dimulai dari menentukan batasan materi pembelajaran yang akan diajarkan maka dilakukan penelaahan kurikulum, membuat *storyboard* serta menginput bahan GW-STM2I pada website yang tersedia berdasarkan *storyboard*. Setelah itu, melakukan validasi GW-STM2I melalui judgement ahli dan melakukan perbaikan sesuai dengan saran dan komentar yang diberikan oleh judgement ahli, untuk instrumen soal dilakukan validasi instrumen melalui judgement ahli dan melakukan uji coba instrumen soal.

B. Pelaksanaan

Tahap ini dimulai melakukan pretest untuk mengukur perubahan konsepsi siswa sebelum pembelajaran berlangsung. Kemudian dilanjutkan dengan kegiatan pembelajaran dengan menerapkan GW-STM2I sebagai bahan ajar untuk tiga kali pertemuan pada kelas eksperimen dan menerapkan pembelajaran menggunakan buku konvensional/buku teks untuk tiga kali pertemuan pada kelas kontrol. Pada proses pembelajaran, siswa diberi LKPD yang mengakomodasi kecerdasan-kecerdasan untuk membantu siswa dalam memahami konsep materi pemanasan global. Setelah proses pembelajaran berakhir, lalu dilakukan posttest perubahan konsepsi dan dilakukan analisis pola perubahan konsepsi siswa berdasarkan hasil pre-post ditinjau dari tipe

kecerdasan majemuk. Selanjutnya, diberikan self-perceive aktivitas kecerdasan majemuk angket.

C. Tahap Akhir

Tahap akhir dilakukan analisis data yang telah diperoleh dan penyusunan laporan akhir. Hasil analisis data kemudian dibahas secara mendalam sehingga dapat menarik sebuah kesimpulan. Kesimpulan yang diambil merupakan jawaban atas permasalahan penelitian. Selain itu, penelitian ini menjadi rekomendasi-rekomendasi terhadap aspek penelitian yang kurang memadai.

3.5 Teknik Analisis Data

3.5.1 Teknik Analisis Angket Kecerdasan Majemuk

Angket ini diberikan untuk mengidentifikasi kecerdasan majemuk yang dimiliki siswa sebelum dan setelah pembelajaran dalam penerapan GW-STM2I. Angket kecerdasan majemuk ini diolah menggunakan skala likert dengan skala penilaian 1-4. Angket ini memiliki delapan tipe kecerdasan majemuk, setiap kecerdasan terdiri dari sepuluh buah pertanyaan sehingga seluruhnya terdiri dari 80 pernyataan. Skor tertinggi pada salah satu atau beberapa kolom kecerdasan majemuk menunjukkan siswa tersebut memiliki kecerdasan majemuk pada kolom tersebut. Selanjutnya, skor tertinggi setiap siswa dikelompokkan berdasarkan jenis kecerdasan majemuk untuk memperoleh jumlah siswa dan persentase setiap jenis kecerdasan majemuk.

3.5.2 Teknik Analisis Self-Perceive of Multiple Intelligences Activities

Analisis hasil angket self-perceive of Multiple Intelligences dilakukan berdasarkan hasil persepsi diri siswa terhadap aktivitas siswa dalam pembelajaran atau dengan kata lain siswa melakukan aktivitas sesuai dengan kecerdasan majemuk yang diakomodasikan dalam pembelajaran. Hasil analisisnya akan digunakan untuk melengkapi hasil kuantitatif dari efektivitas penerapan GW-STM2I dalam pembelajaran.

3.5.3 Teknik analisis efektivitas penerapan GW-STM2I terhadap kecerdasan majemuk dan perubahan konsepsi

3.5.3.1 Analisis *Effect Size*

Teknik analisis untuk menentukan uji keefektifan penggunaan GW-STM2I menggunakan uji dampak. Uji dampak diperoleh hasil bahwa implementasi GW-STM2I secara signifikan efektif dapat merubah konsepsi siswa, maka selanjutnya akan dicari ukuran pengaruhnya dengan menghitung *effect size* menggunakan *Effect Size Cohen's D* dengan persamaan sebagai berikut (Becker, 2000):

$$d_{cohen} = \frac{|M_A - M_E|}{SD_{pool}}$$

Dengan

$$SD_{pool} = \sqrt{\frac{(N_1 - 1)SD_A^2 + (N_2 - 1)SD_E^2}{N_1 + N_2 - 2}}$$

Keterangan:

d_{cohen} : *effect size*

M_A : nilai rata-rata *post test* perubahan konsepsi kelas GW-STM2I

M_E : nilai rata-rata *post test* perubahan konsepsi kelas buku teks

SD_{pool} : standar deviasi untuk kedua kelas

SD_A^2 : standar deviasi kelas GW-STM2I

SD_E^2 : standar deviasi kelas buku teks

N_1 : jumlah siswa kelas GW-STM2I

Hasil perhitungan d Cohen's selanjutnya diinterpretasikan berdasarkan Tabel 3.10 berikut.

Tabel 3.10 Tabel Interpretasi Cohen D

Cohen's Standard	Effect Size	Kedudukan Persentil
	2,0	97,7
	1,9	97,1
	1,8	96,4
	1,7	95,5
	1,6	94,5
	1,5	93,3
	1,4	91,9
	1,3	90,0
	1,2	88,0
	1,1	86,0
	1,0	84,0
	0,9	82,0
Large	0,8	79,0
	0,7	76,0
	0,6	73,0
Medium	0,5	69,0
	0,4	66,0
	0,3	62,0
Small	0,2	58,0
	0,1	54,0
	0,0	50,0

Becker, 2000

Interpretasi Cohen D disempurnakan oleh Becker (2000) dengan menggunakan kedudukan persentil. Jika *effect size* bernilai 0,0 maka rata-rata kelompok dengan perlakuan berada pada persentil ke-50 dari kelompok tanpa perlakuan dan demikian seterusnya.

Setelah mendapatkan nilai *effect size*, hasil pengukuran keterampilan berpikir kreatif dibandingkan menggunakan Analisis Peta *Wright (Person-Item Map)* pemodelan Rasch dengan program Ministep. Peta tersebut menggambarkan sebaran kemampuan siswa dan sebaran tingkat kesukaran soal.

3.5.3.2 Analisis Uji Beda

1. Uji Normalitas

Uji normalitas dilakukan untuk mengetahui apakah data penelitian berdistribusi normal atau tidak. Uji normalitas yang digunakan adalah uji *Shapiro-Wilk* menggunakan SPSS. Hasil uji normalitas akan menentukan langkah pengujian berikutnya. Jika data normal, maka langkah selanjutnya adalah melakukan uji homogenitas. Tetapi, jika data tidak normal, maka pengujian selanjutnya dilakukan dengan analisis statistik non parametrik menggunakan uji *Mann-Whitney U*.

2. Uji Homogenitas

Uji homogenitas bertujuan untuk melihat apakah suatu varians (keberagaman) data dari dua atau lebih kelompok bersifat homogen (sama) atau heterogen (tidak sama). Uji homogenitas yang digunakan adalah uji *Levene* menggunakan SPSS. Dalam penelitian ini, uji homogenitas digunakan untuk mengetahui apakah varians data *pre-post test* kelas eksperimen dan kelas kontrol bersifat homogen atau tidak. Jika varians data bersifat homogen, maka langkah uji hipotesis selanjutnya dilakukan dengan analisis statistik parametrik, menggunakan uji-t. Tetapi jika varians data tidak homogen, maka pengujian selanjutnya menggunakan uji-t'.

3. Uji-t

Pengujian hipotesis dilakukan menggunakan uji beda rata-rata atau Uji-t (*independent sample t test*) karena data bersifat homogen dan berdistribusi normal serta bersifat independen. Perhitungan koefisien t pada *independent sample t test* menggunakan SPSS. Uji-t bertujuan untuk mengetahui signifikansi perubahan konsepsi setelah penerapan GW-STM2I dan buku teks dalam pembelajaran menggunakan data *pre-post test*. Hipotesis dalam penelitian ini dirumuskan dalam bentuk hipotesis statistik sebagai berikut:

$H_0: \mu_1 = \mu_2$: (Tidak terdapat perbedaan rata-rata skor sebelum dan sesudah penerapan GW-STM2I dan penerapan bahan ajar konvensional)

$H_1: \mu_1 \neq \mu_2$: (Terdapat perbedaan rata-rata skor sebelum dan sesudah penerapan GW-STM2I dan penerapan bahan ajar konvensional)

Adapun kriteria pengambilan keputusan sebagai berikut:

H_0 diterima apabila : nilai probabilitas (Sig) > 0,05

H_0 ditolak apabila : nilai probabilitas (Sig) < 0,05

3.5.4 Analisis Pola Perubahan Konsepsi

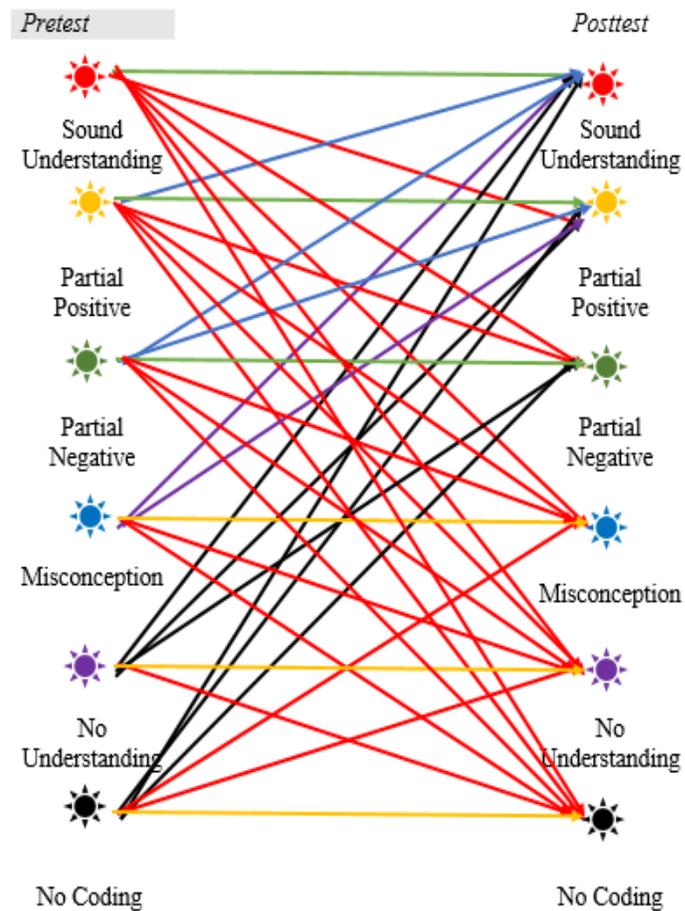
Analisis data pada perubahan konsepsi dilakukan dengan cara pengkodean data data *pretest* dan *posttest* setiap butir soal pada instrumen *four tier test*, dengan kode yang digunakan berdasarkan Tabel 3.11. Hasil pengkodean ini berupa tingkatan konsepsi yang dicapai siswa pada setiap butir soal setelah dilakukan implementasi GW-STM2I pada proses pembelajaran. Langkah selanjutnya, memberikan skor pada setiap butir soal sesuai dengan tingkatan konsepsi yang dicapai. Skor total dihitung dengan menjumlahkan skor yang diperoleh dari setiap butir soal. Tingkat konsepsi peserta didik

dikategorikan berdasarkan enam kategori konsepsi yang telah diadaptasi yaitu: *Sound Understanding* (SU); *Partial Positive* (PP); *Partial Negative* (PN); *Misconception* (MC); *No Understanding* (NU); dan *No Coding* (NC). Kategori ini adalah kombinasi dari tingkat pemahaman oleh Coştu (2008), penilaian oleh Kaltakçı, dkk (2017), dan kategori konsep oleh Amalia, dkk (2019).

Tabel 3.11 Kriteria Jawaban Instrumen Perubahan Konsepsi

Tingkat Konsepsi	Simbol	Tier-1	Tier-2	Tier-3	Tier-4	Skor
SU		Benar	Yakin	Benar	Yakin	2
PP		Benar	Tidak yakin	Benar	Yakin	1
		Benar	Yakin	Benar	Tidak yakin	
		Benar	Tidak yakin	Benar	Tidak yakin	
PN		Benar	Yakin	Salah	Yakin	1
		Benar	Tidak yakin	Salah	Yakin	
		Benar	Yakin	Salah	Tidak yakin	
		Benar	Tidak yakin	Salah	Tidak yakin	
		Salah	Yakin	Benar	Yakin	
		Salah	Tidak yakin	Benar	Yakin	
		Salah	Yakin	Benar	Tidak yakin	
Salah	Tidak yakin	Benar	Tidak yakin			
MC		Salah	Yakin	Salah	Yakin	0
NU		Salah	Tidak yakin	Salah	Yakin	0
		Salah	Yakin	Salah	Tidak yakin	
		Salah	Tidak yakin	Salah	Tidak yakin	
NC		Peserta didik tidak menjawab secara utuh tingkatan dalam butir soal				0

Analisis lebih lanjut yaitu menganalisis karakteristik perubahan tingkat konsepsi yang terjadi pada setiap siswa. Hal ini dilakukan dengan cara mengidentifikasi tingkatan konsepsi siswa yang dicapai pada saat *pretest* dan *posttest* untuk setiap butir soal. Skema setiap perubahan konsepsi yang terjadi untuk setiap konsep disajikan pada Gambar 3.5.



Gambar 3.5 Pola Perubahan Konsepsi