

BAB III

METODE PENELITIAN

A. Metode dan Desain Penelitian

Penelitian yang dilakukan adalah penelitian deskriptif-eksploratif dengan pendekatan non-eksperimen. Penelitian deskriptif adalah penelitian yang bertujuan untuk menyelidiki keadaan atau kondisi suatu hal yang hasilnya dipaparkan secara lugas dan apa adanya (Arikunto, 2010: 3). Penelitian eksploratif adalah penelitian yang permasalahannya belum pernah diteliti atau sedikit sekali informasi mengenai permasalahannya dan bertujuan untuk:

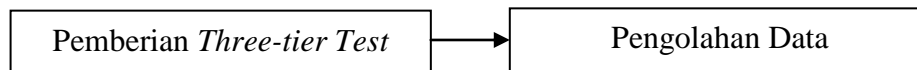
1. Memperoleh gambaran yang jelas mengenai permasalahan.
2. Menentukan alternatif tindakan yang dilakukan.
3. Menentukan variabel-variabel penelitian dan pengujian lebih lanjut (Masyhuri dan Zainudin, 2008: 48 dan 45).

Dengan demikian, penelitian deskriptif-eksploratif ini merupakan penelitian yang bertujuan untuk menyelidiki kondisi miskonsepsi siswa SMP pada materi kalor dengan menggunakan *Three-tier Test* agar diperoleh informasi atau gambaran yang jelas mengenai miskonsepsi siswa tersebut.

Peneliti tidak memberikan perlakuan dalam bentuk kegiatan pembelajaran terlebih dahulu kepada siswa, yang terpenting siswa sudah mempelajari materi kalor sebelum diberikan *Three-tier Test*. Berdasarkan pengertian dan sifat-sifat miskonsepsi, walaupun waktu antara pembelajaran mengenai kalor dengan pemberian tes cukup lama, miskonsepsi muncul atau hilang tidak disebabkan oleh waktu, melainkan karena faktor epistemologi, psikologis, dan pedagogik (Bal, 2011: 285) sebagaimana dijelaskan pada kajian teori.

Sebelum tes dilaksanakan, sekitar seminggu sebelum tes siswa diberikan informasi mengenai materi yang akan diteskan. Hal ini dilakukan agar siswa mempelajari kembali (*review*) materi yang akan diteskan, sehingga hasil yang diperoleh pada *Three-tier Test* sesuai dengan yang diharapkan peneliti. Desain yang digunakan dalam penelitian ini disesuaikan dengan bentuk pendekatan

penelitian non-eksperimen. Karena pengambilan data hanya dilakukan sekali, maka pola penelitiannya digambarkan seperti dalam Gambar 3.1.



Gambar 3.1 Pola Penelitian *One-Shot Design*

B. Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah siswa kelas VII dari 76 SMP Negeri di Kabupaten Kuningan, Jawa Barat. Sampel yang dibutuhkan dalam penelitian ini ditentukan dengan teknik *cluster random sampling*. Sugiyono (2011: 122) mengemukakan bahwa pengambilan sampel dengan teknik *cluster random sampling* dari populasi yang di dalamnya terdapat kelompok yang berstrata perlu menggunakan *stratified random sampling* pula. Pengambilan sampel dengan teknik *cluster random sampling* ini terdiri dari dua tahap, yaitu:

1. Menentukan sampel daerah.
2. Menentukan orang-orang yang ada pada daerah tersebut.

Berdasarkan tahap tersebut, langkah pertama yang peneliti lakukan untuk memilih sampel penelitian yaitu menentukan sampel sekolah yang dijadikan tempat penelitian. Peneliti mengambil tiga sekolah secara acak di mana masing-masing sekolah ini mewakili sekolah klaster atas, menengah, dan bawah. Langkah selanjutnya yaitu menentukan sampel individu dari ketiga sekolah tersebut, dalam hal ini adalah menentukan sampel siswa.

Peneliti menentukan jumlah sampel berdasarkan tingkat ketelitian atau kepercayaan hasil penelitian. Dengan mempertimbangkan dana, waktu, dan tenaga, peneliti mengambil jumlah sampel berdasarkan tingkat kepercayaan sebesar 90% atau taraf kesalahan sebesar 10%. Jumlah sampel ditentukan menggunakan rumus yang dikembangkan oleh *Isaac* dan *Michael*, atau dengan melihat tabel penentuan jumlah sampel berdasarkan rumus tersebut (Sugiono, 2011: 128). Populasi siswa kelas VII dari 76 SMP Negeri di Kabupaten Kuningan, Jawa Barat, diketahui sebesar 37.961 siswa. Dengan demikian, banyaknya sampel minimal yang harus digunakan dalam penelitian ini adalah 269 sampel. Jumlah

sampel penelitian dari setiap sekolah dicantumkan dalam Tabel 3.1. Peneliti memilih SMP Negeri karena kurikulum di sekolah negeri ini menggunakan urutan Standar Kompetensi yang sesuai dengan KTSP sehingga siswa kelas VII ini dipastikan sudah mempelajari materi mengenai kalor.

Tabel 3.1 Jumlah Sampel Penelitian

Nama Sekolah	Jumlah Siswa Kelas VII (N)	Jumlah Sampel (N_S)
SMPN 2 Kuningan	359	122
SMPN 7 Kuningan	372	126
SMPN 5 Kuningan	66	23
Jumlah Total (Σ)	797	271

Banyaknya jumlah minimal sampel ditentukan menggunakan rumus:

$$N_{tabel} = 269 \quad (3.1)$$

Banyaknya jumlah sampel (N_S) ditentukan menggunakan rumus:

$$N_S = \frac{N}{\Sigma N} \times N_{tabel} \quad (3.2)$$

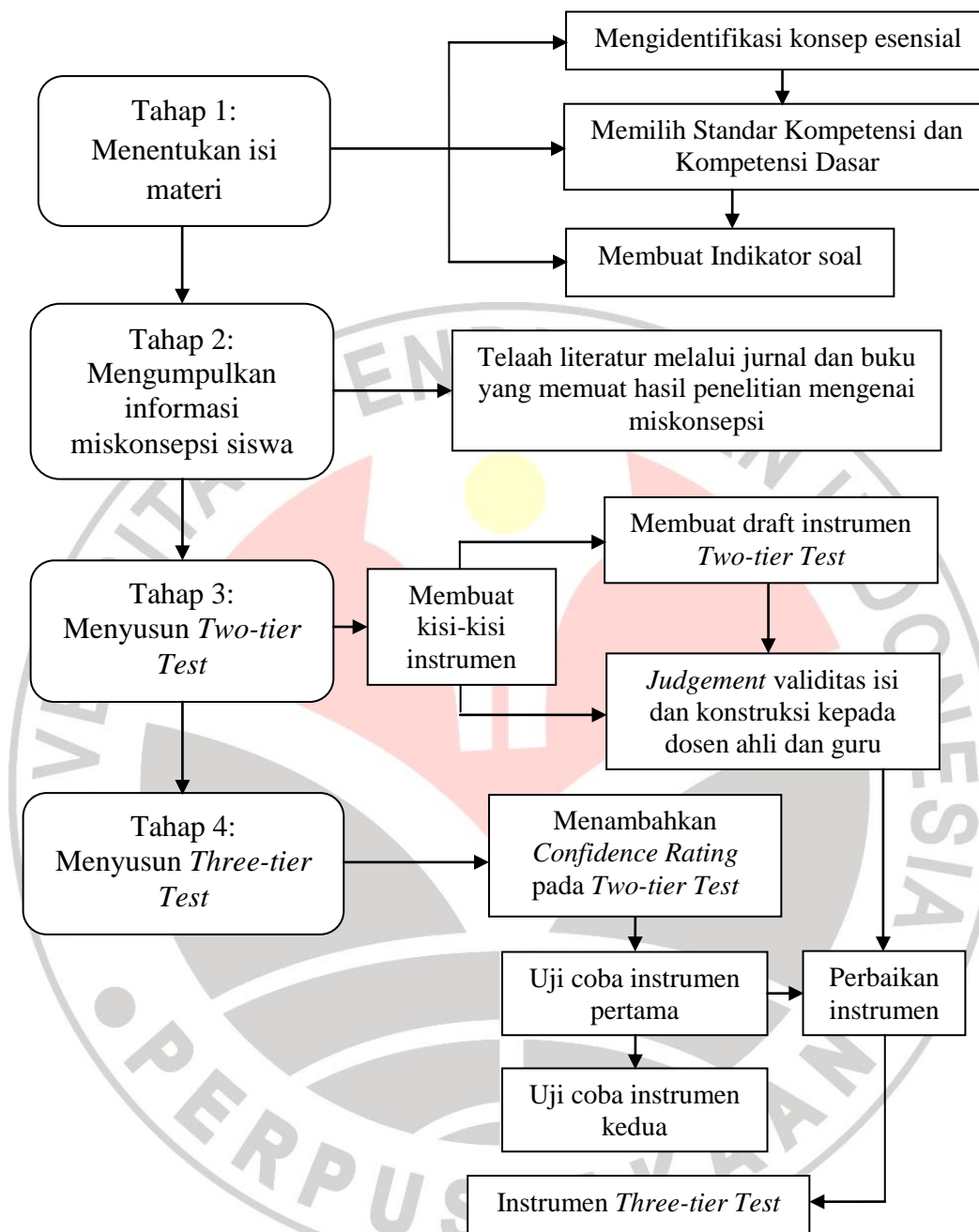
Dengan:

N_S = jumlah sampel yang dibutuhkan dari tiap sekolah

N = jumlah siswa kelas VII dari tiap sekolah

C. Instrumen Penelitian

Instrumen yang digunakan dalam penelitian ini adalah *Three-tier Test* yang berfungsi untuk mengevaluasi kemampuan memahami dan mengidentifikasi miskonsepsi siswa. Instrumen ini berbentuk pilihan ganda tiga tingkat dimana pada soal tingkat keduanya disisipkan opsi berbentuk isian kosong (*free response*). Hal ini bertujuan untuk mengetahui apakah siswa mengalami miskonsepsi baru atau miskonsepsi yang tidak terdapat dalam literatur sebelumnya. Instrumen ini mengukur kemampuan memahami (C2) berdasarkan taksonomi Anderson, yang meliputi proses kognitif menafsirkan, mencontohkan, mengklasifikasikan, menyimpulkan, membandingkan, dan menjelaskan. Peneliti mengadopsi dan mengadaptasi tahap pembuatan *Two-tier Test* yang dibuat oleh Treagust (2007: 394) yang digambarkan dalam Gambar 3.2.



Gambar 3.2 Bagan Tahap Pembuatan *Three-tier Test*

Tahap pembuatan *Three-tier Test* tersebut dijelaskan sebagai berikut.

1. Menentukan Konten atau Materi

Materi yang digunakan dalam tes ini adalah materi kalor. Setelah materi ditentukan, selanjutnya adalah mengidentifikasi konsep-konsep esensial yang ada dalam materi tersebut. Peneliti menggunakan dua Buku Sekolah Elektronik (BSE)

mata pelajaran IPA untuk SMP kelas VII untuk membantu mengidentifikasi konsep-konsep esensial. Selanjutnya peneliti membuat indikator soal mengacu pada standar kompetensi, kompetensi dasar, dan konsep-konsep esensial yang sudah diidentifikasi.

2. Mengumpulkan Informasi Miskonsepsi

Peneliti melakukan telaah literatur untuk mengetahui miskonsepsi mengenai kalor yang sering dialami oleh siswa berdasarkan penelitian terdahulu. Informasi miskonsepsi ini digunakan untuk membuat distraktor pada soal tingkat kedua (*Second Tier*).

3. Menyusun *Two-tier Test*

Soal pada *Two-tier Test* ini terdiri dari dua tingkat soal, yaitu soal tingkat pertama (*first tier*) yang berfungsi untuk menilai pengetahuan deskriptif siswa dan soal tingkat kedua (*second tier*) berbentuk butir-butir alasan atas jawaban pada soal tingkat pertama yang berfungsi untuk menilai pola pikir siswa. Soal tingkat kedua ini terdiri dari lima opsi jawaban, empat opsi berupa pernyataan tertulis, sedangkan satu opsi lainnya dalam bentuk isian kosong. Penggunaan opsi dalam bentuk isian kosong ini bertujuan untuk mengidentifikasi konsepsi siswa yang tidak sesuai dengan opsi-opsi yang dicantumkan dan untuk menghindari jawaban yang ragu-ragu dari siswa agar siswa benar-benar mengungkapkan konsep yang dipahaminya. *Two-tier Test* yang sudah selesai dibuat kemudian dikonsultasikan kepada dua dosen ahli dan satu guru mata pelajaran IPA (Fisika) untuk mengevaluasi validitas isi dan konstruksi soal-soal dalam instrumen tersebut.

4. Menyusun *Three-tier Test*

Setelah melalui tahap validasi oleh ahli, *Two-tier Test* ini kemudian ditambahkan dengan *Confidence Rating* dengan enam opsi jawaban, yaitu hanya menebak, sangat tidak yakin, tidak yakin, yakin, sangat yakin, dan sangat yakin sekali. *Two-tier Test* yang dikombinasikan dengan *Confidence Rating* ini selanjutnya dinamakan *Three-tier Test*. Setelah selesai dibuat, *Three-tier Test* ini kemudian diujicobakan kepada siswa. Hasil uji coba dan hasil validasi ahli pada akhirnya dijadikan acuan untuk menentukan soal mana saja yang layak digunakan dalam *Three-tier Test* berdasarkan nilai validitas, indeks kesukaran, dan daya

pembeda tiap butir soal. Uji coba diberikan sebanyak dua kali dengan menggunakan instrumen yang sama (*single test double trial*) untuk menentukan reliabilitas eksternal *Three-tier Test*. Uji coba instrumen diberikan pada siswa yang sama dan merupakan siswa dari tiga sekolah yang akan dijadikan tempat penelitian. Rentang waktu uji coba pertama dengan uji coba kedua adalah dua minggu.

Peneliti mengujikan soal dalam bentuk *Three-tier Test* karena peneliti menggunakan dua miskonsepsi yang tidak ada dalam literatur. Jika ternyata hasil uji coba menunjukkan bahwa siswa mengalami miskonsepsi pada distraktor tersebut, maka miskonsepsi ini dapat digunakan dalam instrumen pada saat penelitian. Selain itu, uji coba dalam bentuk *Three-tier Test* memungkinkan peneliti untuk menggali miskonsepsi siswa yang tidak ada dalam distraktor soal yang dapat dipertimbangkan untuk digunakan sebagai distraktor jika miskonsepsi ini banyak dipilih oleh siswa.

D. Prosedur Penelitian

Secara garis besar, penelitian ini dibagi menjadi tiga tahap, yaitu (1) tahap persiapan; (2) tahap pelaksanaan; (3) tahap akhir. Penjelasan lebih rinci mengenai tahap-tahap tersebut dijabarkan sebagai berikut.

1. Tahap Perencanaan

Pada tahap ini peneliti mengidentifikasi masalah dan merumuskan masalah, dilanjutkan dengan menyusun *Two-tier Test*. Instrumen *Two-tier Test* yang telah selesai dibuat kemudian divalidasi kepada ahli, ditambahkan *Confidence Rating*, lalu diujicobakan dalam bentuk *Three-tier Test* kepada 65 siswa. Selanjutnya peneliti menganalisis hasil uji coba yang meliputi analisis validitas, indeks kesukaran, dan daya pembeda butir soal serta reliabilitas instrumen, hingga dapat menentukan butir soal yang layak digunakan dalam instrumen *Three-tier Test* untuk penelitian.

2. Tahap Pelaksanaan

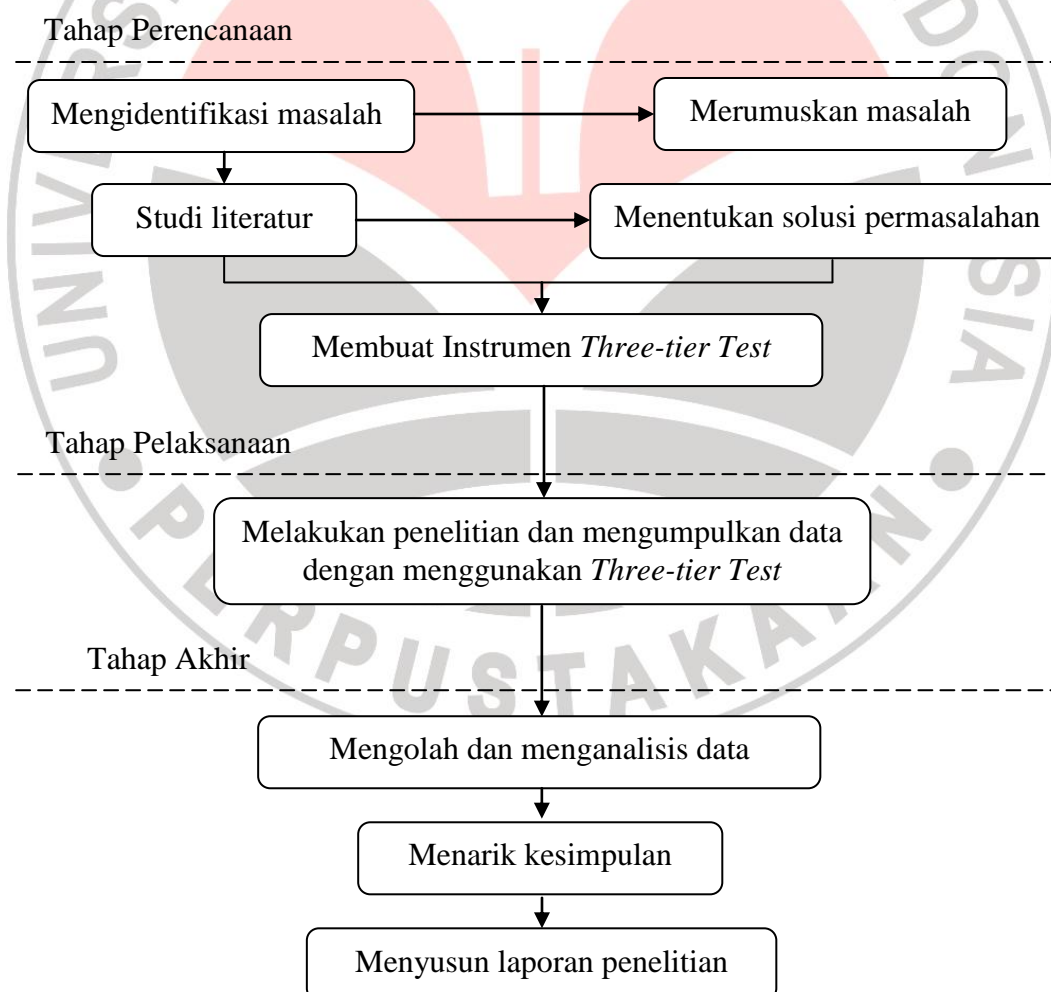
Setelah melalui tahap analisis hasil uji coba dan perbaikan instrumen, *Three-tier Test* kemudian diteskan kepada 271 siswa kelas VII yang berasal dari

tiga SMP Negeri dengan klaster yang berbeda-beda. Siswa yang dijadikan sampel penelitian ini berbeda dengan siswa yang mengikuti uji coba instrumen. Atau dengan kata lain, sampel penelitian ini tidak termasuk siswa yang sudah mendapatkan tes pada saat uji coba.

2. Tahap Akhir

Pada tahap ini peneliti mengolah data untuk mengetahui profil jawaban-jawaban siswa agar dapat mengidentifikasi miskonsepsi yang dialami siswa, lalu menganalisis apakah ditemukan miskonsepsi baru yang tidak ada dalam literatur, dan menganalisis bagaimana prevalensi miskonsepsinya. Setelah itu, peneliti menyimpulkan hasil temuannya dan membuat laporan penelitian.

Tahapan prosedur penelitian ini digambarkan dalam Gambar 3.3 berikut ini.



Gambar 3.3 Tahapan Prosedur Penelitian

E. Teknik Analisis Instrumen

Instrumen yang baik harus dapat mengukur apa yang hendak diukur (valid) dan dapat dipercaya untuk digunakan sebagai alat pengumpul data (reliabel) (Arikunto, 2010: 211). Sebab, sebuah tes mungkin saja reliabel tetapi tidak valid. Sebaliknya, tes yang valid biasanya reliabel (Arikunto, 2009: 87).

Peneliti menganalisis validitas, indeks kesukaran, dan daya pembeda butir soal pada *Three-tier Test*. Jika jawaban siswa pada soal tingkat pertama (*first tier*) benar, jawaban pada soal tingkat kedua (*second tier*) juga benar, dan siswa yakin terhadap jawaban pada kedua tingkat soal tersebut atau skala *Confidence Rating* yang dipilih siswa lebih dari 4 ($CR > 4$), maka siswa diberi skor 1 (Pesman dan Eryilmaz, 2010: 212). Jika selain jawaban tersebut, maka siswa diberi skor 0.

Validitas instrumen, indeks kesukaran butir soal, daya pembeda butir soal, dan reliabilitas instrumen ditentukan dengan cara sebagai berikut.

1. Validitas

Validitas adalah suatu ukuran yang menunjukkan tingkat-tingkat kevalidan atau kesahihan suatu instrumen. Sebuah instrumen dikatakan valid apabila mampu mengukur apa yang diinginkan (Arikunto, 2010: 211).

Validitas *Three-tier Test* ini terdiri dari validitas logis dan empiris. Uji validitas logis yang meliputi validitas isi dan konstruksi dilakukan dengan mengonsultasikan setiap butir soal dalam *Three-tier Test* kepada dua dosen ahli dan satu guru mata pelajaran IPA (Fisika). Uji validitas empiris dilakukan dengan teknik analisis validitas butir soal (Arikunto, 2010: 215) dan analisis daya pembeda untuk menguji validitas instrumen secara keseluruhan (Sugiyono, 2011: 180). Validitas butir soal ditentukan menggunakan rumus korelasi *Product Moment Pearson* (r_{xy}) dengan angka kasar (Arikunto, 2010: 213), yaitu:

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y^2 - (\Sigma Y)^2\}}} \quad (3.3)$$

Dengan:

r_{xy} = koefisien korelasi antara variabel X dan Y, dua variabel yang dikorelasikan

X = skor tiap butir soal

Y = skor total tiap butir soal

N = jumlah siswa

Penafsiran indeks validitas butir soal dilakukan dengan membandingkan indeks korelasi (r_{xy}) yang didapat terhadap indeks korelasi pada tabel harga kritik r *Product-Moment*. Jika indeks korelasi hitung lebih besar dari indeks korelasi (dengan N tertentu) pada tabel harga kritik r *Product-Moment*, maka butir soal dikatakan valid (Arikunto, 2010: 227). Jika sebaliknya, maka butir soal dikatakan tidak valid atau tidak dapat digunakan.

Sementara itu, validitas instrumen secara keseluruhan ditentukan menggunakan rumus *t-test* (Sugiyono, 2011: 181). Langkah awal yang perlu dilakukan adalah mengurutkan skor dari yang tertinggi hingga terendah. Lalu menentukan kelompok tinggi yaitu 27% sampel yang memiliki skor tertinggi dan kelompok rendah yaitu 27% sampel yang memiliki skor terendah. Rumus signifikansi daya pembeda yang digunakan untuk menentukan validitas instrumen secara keseluruhan adalah:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{gab} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (3.4)$$

$$S_{gab} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2) - 2}} \quad (3.5)$$

Dengan:

t = signifikansi daya pembeda

\bar{X}_1 = skor rata-rata kelompok tinggi

\bar{X}_2 = skor rata-rata kelompok rendah

n_1 = jumlah peserta kelompok tinggi

n_2 = jumlah peserta kelompok rendah

S_{gab} = standar deviasi gabungan

S_1 = standar deviasi kelompok tinggi

S_2 = standar deviasi kelompok rendah

Penafsiran validitas instrumen dilakukan dengan membandingkan harga t hitung terhadap tabel distribusi harga t , dengan $dk = (n_1 + n_2) - 2$. Jika harga t hitung lebih besar dari pada harga t tabel, maka instrumen dikatakan valid (Sugiyono, 2011: 182).

2. Indeks Kesukaran

Indeks kesukaran suatu butir soal adalah proporsi dari keseluruhan siswa yang menjawab benar pada butir soal tersebut (Arikunto, 2009: 208). Indeks kesukaran dihitung dengan rumus:

$$P = \frac{B}{JS} \quad (3.6)$$

Dengan:

P = indeks kesukaran soal

B = banyaknya siswa yang menjawab soal itu dengan benar

JS = jumlah seluruh siswa peserta tes

Arti nilai indeks kesukaran soal (P) dapat diketahui dengan cara membandingkan nilai tersebut dengan Tabel 3.2.

Tabel 3.2 Klasifikasi Indeks Kesukaran
(Arikunto, 2009: 210)

Nilai P	Kriteria
1,00	Terlalu mudah
0,70 – 1,00	Mudah
0,30 – 0,70	Sedang
0,00 – 0,30	Sukar
0,00	Terlalu sukar

3. Daya Pembeda

Daya pembeda suatu butir soal adalah kemampuan butir soal untuk membedakan siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah (Arikunto, 2009: 213-214). Daya pembeda dapat dihitung dengan menggunakan rumus:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B \quad (3.7)$$

Dengan:

D = indeks daya pembeda

J_A = jumlah peserta kelompok atas

J_B = jumlah peserta kelompok bawah

B_A = jumlah peserta kelompok atas yang menjawab benar

B_B = jumlah peserta kelompok bawah yang menjawab benar

P_A = proporsi peserta kelompok atas yang menjawab benar

P_B = proporsi peserta kelompok bawah yang menjawab benar

Untuk menghitung indeks daya pembeda, perlu dibedakan antara kelompok kecil dengan kelompok besar. Jika jumlah sampel kurang dari 100, maka kelompok ini dibagi dua menjadi sama besar, yaitu 50% sampel dengan skor teratas sebagai kelompok atas (J_A) dan 50% sampel dengan skor terbawah sebagai kelompok bawah (J_B). Sedangkan jika jumlah sampel lebih dari 100, maka hanya diambil 27% sampel dengan skor teratas sebagai kelompok atas (J_A) dan 27% sampel dengan skor terbawah sebagai kelompok bawah (J_B) (Arikunto, 2009: 212). Jika indeks daya pembeda bernilai negatif, sebaiknya soal diperbaiki atau dibuang saja karena artinya soal tersebut tidak baik (Arikunto, 2009: 218). Klasifikasi indeks daya pembeda ditunjukkan dalam Tabel 3.3.

Tabel 3.3 Klasifikasi Indeks Daya Pembeda
(Arikunto, 2009: 218)

Indeks D	Kriteria
0,70 – 1,00	Baik sekali
0,40 – 0,70	Baik
0,20 – 0,40	Cukup
0,00 – 0,20	Jelek

4. Reliabilitas

Reliabilitas menunjukkan bahwa suatu instrumen cukup dapat dipercaya untuk digunakan sebagai alat pengumpul data karena instrumen tersebut sudah baik (Arikunto, 2010: 221). Reliabilitas yang dicari untuk *Three-tier Test* ini adalah reliabilitas eksternal dan internal. Uji reliabilitas eksternal dilakukan dengan teknik *single test double trial*.

Uji reliabilitas eksternal dilakukan dengan mengorelasikan hasil uji coba instrumen pertama dengan hasil uji coba instrumen kedua. Rumus yang digunakan untuk menentukan reliabilitas eksternal *Three-tier Test* adalah rumus korelasi *Product Momen Pearson* sebagai berikut.

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y^2 - (\Sigma Y)^2\}}} \quad (3.8)$$

Dengan:

r_{xy} = koefisien korelasi antara tes pertama dengan tes kedua

X = skor siswa pada tes pertama

Y = skor siswa pada tes kedua

N = jumlah siswa

Selain uji reliabilitas eksternal, peneliti melakukan uji reliabilitas internal *Three-tier Test* yang ditentukan dengan menggunakan rumus *Cronbach Alpha*, yaitu:

$$r_{xy} = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right) \quad (3.9)$$

Dengan:

r_{xy} = reliabilitas instrumen

k = banyaknya butir pertanyaan atau banyaknya soal

$\sum \sigma_b^2$ = jumlah varians butir

σ_t = varians total

Penafsiran indeks reliabilitas instrumen dilakukan dengan membandingkan indeks reliabilitas (r_{xy}) hasil hitung terhadap indeks korelasi pada tabel harga kritik *r Product-Moment*. Jika indeks reliabilitas hitung lebih besar dari indeks korelasi (dengan N tertentu) pada tabel harga kritik *r Product-Moment*, maka instrumen dikatakan reliabel (Arikunto, 2010: 227). Jika sebaliknya, maka instrumen dikatakan tidak reliabel.

F. Hasil Uji Coba Instrumen

Uji coba instrumen dilakukan di tiga sekolah yang digunakan sebagai tempat penelitian. Instrumen yang diujicobakan diberikan dalam bentuk *Three-tier Test*. Siswa yang mengikuti uji coba instrumen ini berjumlah 65 orang. Hasil uji coba instrumen tersebut dijelaskan sebagai berikut.

1. Validitas, Daya Pembeda, Indeks Kesukaran, dan Reliabilitas *Three-tier Test*

Data yang digunakan untuk menentukan validitas, indeks kesukaran, dan daya pembeda butir soal adalah skor siswa pada *Three-tier Test* hasil uji coba pertama saja. Skor pada skala *Confidence Rating* digunakan untuk menentukan

apakah siswa mengalami dua miskonsepsi baru. Miskonsepsi baru ini diindikasikan berdasarkan adanya siswa yang memilih dua distraktor miskonsepsi yang tidak terdapat dalam literatur, yang dilampirkan pada Lampiran A.1.

Berdasarkan hasil uji coba instrumen, indeks validitas butir soal berkisar antara $DIV/0$ hingga 0,611. Tabel harga kritik r *Product-Moment* dengan $N = 65$, menunjukkan harga $r_{t(5\%)} = 0,244$ dan $r_{t(1\%)} = 0,317$. Berkaitan dengan validitas butir soal, peneliti membuat keputusan untuk menggunakan atau tidak menggunakan butir soal tertentu berdasarkan pada hasil uji validitas. Butir soal yang valid akan digunakan dalam instrumen dan butir soal yang tidak valid tidak akan digunakan dalam instrumen. Jika indeks validitas butir soal hasil hitung dibandingkan dengan tabel harga kritik r *Product-Moment* pada tingkat kepercayaan 95% atau tingkat kesalahan 5% ($r_{t(5\%)}$), maka diperoleh 11 butir soal yang valid atau dapat digunakan dalam *Three-tier Test*. Butir soal yang digunakan dalam *Three-tier Test* adalah butir soal nomor 2, 4, 5, 6, 7, 8, 10, 12, 14, 15, dan 16. Seluruh butir soal ini memuat seluruh konsep esensial yang sudah diidentifikasi sebelumnya, yaitu energi, kalor, suhu, massa, kalor jenis zat, tekanan, menguap, mendidih, titik didih zat, mencair, mengembun, membeku, dan titik beku zat. Tabel analisis butir soal pada *Three-tier Test* berdasarkan hasil uji coba instrumen ini dapat dilihat lebih rinci pada Lampiran B.3.

Sementara itu, dari 11 butir soal yang digunakan, hanya satu soal yang termasuk dalam indeks kesukaran dengan kriteria sedang, 10 butir soal lainnya termasuk dalam kriteria sukar. Indeks kesukaran rata-rata diketahui sebesar 0,12. Jika dibandingkan terhadap tabel indeks kesukaran (Tabel 3.2), indeks ini termasuk dalam kriteria sukar. Hal ini menunjukkan bahwa secara umum tingkat kesulitan soal dalam *Three-tier Test* termasuk dalam kriteria sukar.

Validitas *Three-tier Test* secara keseluruhan dapat diketahui dengan membandingkan harga signifikansi daya pembeda (t) hasil hitung terhadap tabel distribusi harga t . Berdasarkan pengolahan data diperoleh harga t sebesar 20,9 dengan $dk = 34$. Harga t dalam tabel dengan $dk = 30$ adalah $t_{(95\%)} = 1,697$. Harga t hitung ternyata lebih besar dari harga dalam tabel distribusi t . Dengan demikian dapat dikatakan bahwa *Three-tier Test* ini valid.

Reliabilitas instrumen yang ditentukan dalam penelitian ini adalah reliabilitas internal dan eksternal *Three-tier Test*. Indeks reliabilitas *Three-tier Test* hasil uji coba dirangkum dalam Tabel 3.4.

Tabel 3.4 Reliabilitas *Three-tier Test* berdasarkan Hasil Uji Coba

Reliabilitas		Indeks (r_{xy})
Reliabilitas Internal	• Uji coba pertama	0,419
	• Uji coba kedua	0,300
Reliabilitas Eksternal	<i>Test-retest</i>	0,593

Reliabilitas *Three-tier Test* dapat diketahui dengan membandingkan indeks reliabilitas hitung terhadap tabel harga kritik r *Product Moment*. Harga r pada tabel dengan $N = 65$, menunjukkan harga $r_{t(5\%)} = 0,244$ dan $r_{t(1\%)} = 0,317$. Tabel 3.4 menunjukkan bahwa indeks reliabilitas (r) hitung lebih besar dari harga r pada tabel harga kritik r *Product-Moment*. Dengan demikian dapat dikatakan bahwa *Three-tier Test* ini reliabel. Pengolahan data indeks kesukaran, validitas, reliabilitas internal, daya pembeda butir soal hasil uji coba instrumen pertama, reliabilitas internal instrumen hasil uji coba kedua, dan reliabilitas eksternal instrumen dapat dilihat pada Lampiran B.1, B.2, B.4, dan B.5.

2. Miskonsepsi Baru dan Analisis Kombinasi Jawaban pada *Three-tier Test*

Miskonsepsi yang digunakan sebagai distraktor dalam soal tingkat kedua (*second tier*) diperoleh dari telaah literatur dan dua diantaranya tidak ada dalam literatur. Peneliti menggunakan miskonsepsi tersebut karena kesulitan dalam membuat distraktor pada saat menyusun soal. Miskonsepsi tersebut adalah:

- (1) kalor yang diperlukan untuk menaikkan suhu benda hanya dipengaruhi oleh massa benda saja.
- (2) kalor yang diperlukan untuk menaikkan suhu benda dipengaruhi oleh volumenya.

Berdasarkan hasil uji coba instrumen pertama, diperoleh beberapa siswa yang mengalami miskonsepsi tersebut. Contoh lembar jawaban beberapa siswa yang mengalami dua miskonsepsi tersebut dapat dilihat pada Lampiran B.6.

Beberapa contoh jawaban siswa yang mengalami dua miskonsepsi di atas ditunjukkan pada Gambar 3.4 dan Gambar 3.5 berikut ini.

6.1 a b X d
6.2 a X c d
e. ...
6.3 1 2 3 X 5 6

Gambar 3.4 Contoh Jawaban Siswa yang Mengalami Miskonsepsi pada Topik Miskonsepsi Nomor (1)

7.1 a b c d
7.2 a b c d
e. ...
7.3 1 2 3 4 5 6

Gambar 3.5 Contoh Jawaban Siswa yang Mengalami Miskonsepsi pada Topik Miskonsepsi Nomor (2)

Peneliti mengadopsi dan mengadaptasi teknik menganalisis kombinasi jawaban untuk mengidentifikasi miskonsepsi siswa yang digunakan oleh Kaltakci dan Didiş (2007: 500). Mereka menggunakan *Three-tier Test* dengan dua opsi tingkat keyakinan, yaitu yakin dan tidak yakin yang telah dirangkum sebelumnya dalam Tabel 2.3.

Tabel 2.3 Analisis Kombinasi Jawaban pada *One-tier*, *Two-tier*, dan *Three-tier*

Analisis Tingkat Soal	Kategori	Tipe Jawaban
<i>One-tier</i>	Memahami Konsep	Jawaban benar
	Miskonsepsi	Jawaban salah
<i>Two-tier</i>	Memahami Konsep	Jawaban benar + alasan benar
	Error	Jawaban salah + alasan benar
	Miskonsepsi	Jawaban benar + alasan salah Jawaban salah + alasan salah
<i>Three-tier</i>	Memahami Konsep	Jawaban benar + alasan benar + yakin
	<i>Lack of Knowledge</i>	Jawaban benar + alasan benar + tidak yakin
		Jawaban salah + alasan benar + tidak yakin
		Jawaban benar + alasan salah + tidak yakin
		Jawaban salah + alasan salah + tidak yakin
Error	Jawaban salah + alasan benar + yakin	
Miskonsepsi	Jawaban benar + alasan salah + yakin Jawaban salah + alasan salah + yakin	

Berdasarkan Tabel 2.3 dapat diketahui bahwa miskonsepsi diidentifikasi dengan menganalisis kombinasi jawaban siswa pada soal *Two-tier* dengan skala tingkat keyakinan siswa. Pada penelitian ini, tingkat keyakinan yang digunakan adalah skala *Confidence Rating* seperti yang digunakan oleh Caleon dan Subramaniam (2010: 943) dalam penelitiannya.

Tabel 3.5 Skala dan Kriteria *Confidence Rating* (CR)

Skala <i>Confidence Rating</i> (CR)	Kriteria
1	hanya menebak
2	sangat tidak yakin
3	tidak yakin
4	yakin
5	sangat yakin
6	sangat yakin sekali

Teknik menganalisis jawaban dan mengidentifikasi miskonsepsi dalam penelitian ini merupakan gabungan dari teknik analisis pada Tabel 2.3 dan Tabel 3.5 yang secara sederhana dirangkum dalam Tabel 3.6.

Tabel 3.6 Teknik Analisis Kombinasi Jawaban pada *Three-tier Test*

Tingkatan Soal	Kategori	Tipe Jawaban	
<i>Three-tier</i>	Memahami Konsep	Jawaban benar, alasan benar, dan skala $CR \geq 4$	
	Error	Jawaban salah, alasan benar, dan skala $CR \geq 4$	
	<i>Lack of Knowledge</i>		Jawaban benar, alasan benar, dan skala $CR \leq 3$
			Jawaban benar, alasan salah, dan skala $CR \leq 3$
			Jawaban salah, alasan benar, dan skala $CR \leq 3$
			Jawaban salah, alasan salah, dan skala $CR \leq 3$
	Miskonsepsi		Jawaban benar, alasan salah, dan skala $CR \geq 4$
			Jawaban salah, alasan salah, dan skala $CR \geq 4$

Tabel 3.6 dan Kisi-kisi Instrumen *Three-tier Test* yang dapat dilihat pada Lampiran A.2 merupakan pedoman untuk mengidentifikasi kategori jawaban siswa dalam penelitian *Three-tier Test*. Dalam kisi-kisi instrumen *Three-tier Test* dijelaskan kunci jawaban dan kombinasi jawaban yang menunjukkan miskonsepsi dalam setiap butir soalnya. Akan tetapi, hasil uji instrumen menemukan bahwa

terdapat kombinasi jawaban yang tidak termasuk dalam kategori pada Tabel 3.6. Kombinasi jawaban tersebut adalah:

1. jawaban salah, alasan salah, skala $CR \geq 4$, tetapi jawaban dengan alasan tidak berhubungan.
2. jawaban benar, alasan salah, skala $CR \geq 4$, tetapi jawaban dengan alasan tidak berhubungan.

Agar lebih jelas, Gambar 3.6 menunjukkan contoh jawaban siswa pada *Three-tier Test* yang memiliki kombinasi jawaban salah, alasan salah, dan skala $CR \geq 4$ dimana jawaban dengan alasan tidak berhubungan.

- 16.1 Diantara peristiwa berikut ini yang termasuk dalam perubahan wujud zat yang memerlukan kalor adalah ...
- a. baju basah yang dijemur di tempat panas menjadi kering
 - b. uap air berubah wujud menjadi tetesan air di sekitar permukaan bawah tutup panci
 - c. air yang dimasukkan ke dalam lemari es berubah menjadi es
 - d. uap air berubah wujud menjadi tetesan air di permukaan dedaunan
- 16.2 Alasan yang paling tepat atas jawaban Anda pada pertanyaan 16.1 adalah ...
- a. uap air menerima sejumlah kalor yang terkandung dalam udara agar bisa mengembun menjadi tetesan air
 - b. uap air menerima sejumlah kalor yang terkandung dalam tutup panci sehingga mengembun di bawah permukaannya
 - c. kalor digunakan untuk menguapkan air dalam baju sehingga baju menjadi kering
 - d. udara di dalam lemari es menarik semua kalor yang terkandung dalam air hingga air berubah menjadi es
 - e.
- 16.3 Apakah Anda yakin terhadap jawaban pada pertanyaan 16.1 dan 16.2?
- | | |
|-----------------------|------------------------|
| 1. hanya menebak | 4. yakin |
| 2. sangat tidak yakin | 5. sangat yakin |
| 3. tidak yakin | 6. sangat yakin sekali |

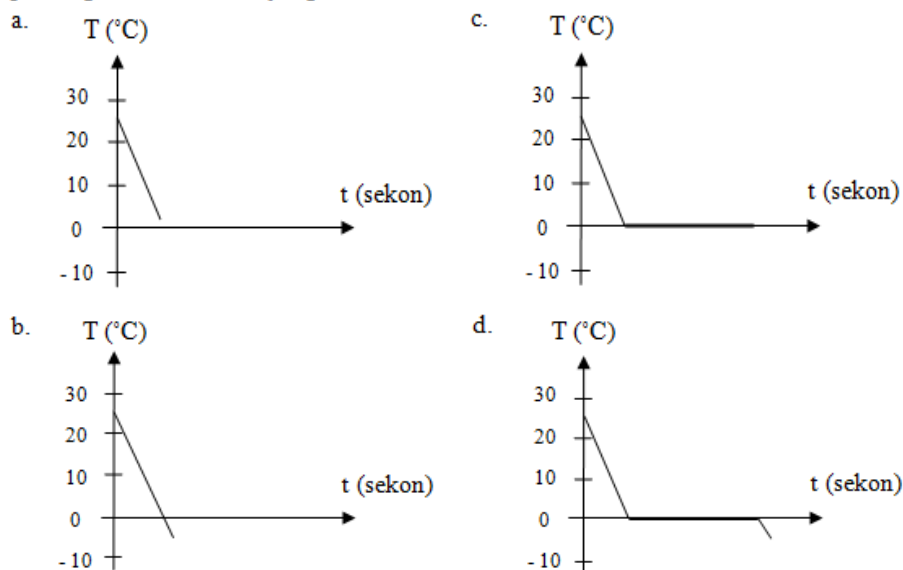
Gambar 3.6 Soal Nomor 16 dalam *Three-tier Test*

16.1	X	b	c	d		
16.2	a	X	c	d		
	e. ...					
16.3	1	2	3	X	5	6

Gambar 3.7 Contoh Jawaban Siswa pada Soal Nomor 16 dalam *Three-tier Test*

Selain contoh jawaban pada Gambar 3.7, Gambar 3.9 berikut ini menunjukkan jawaban siswa pada *Three-tier Test* yang memiliki kombinasi jawaban benar, alasan salah, dan skala $CR \geq 4$ dimana jawaban dengan alasan tidak berhubungan.

14.1 Air bersuhu 25°C didinginkan di dalam kulkas hingga mencapai suhu -5°C . Grafik pendinginan air tersebut yang benar adalah ...



14.2 Alasan yang paling tepat atas jawaban Anda pada pertanyaan 14.1 adalah ...

- suhu air turun hingga melewati titik beku dan suhunya mencapai -5°C
- suhu air turun hingga mendekati suhu 0°C , namun tidak bisa mencapai suhu 0°C karena air tidak bisa bersuhu 0°C
- suhu air turun hingga membeku pada suhu 0°C , selama membeku suhunya tidak berubah lagi, setelah itu tidak mengalami penurunan suhu lagi karena suhu es akan selalu 0°C atau tidak bisa lebih rendah dari 0°C
- suhu air turun hingga membeku pada suhu 0°C , selama membeku suhunya tidak berubah lagi hingga semua air berubah menjadi es, dan setelah itu suhu es turun hingga -5°C
-

14.3 Apakah Anda yakin terhadap jawaban pada pertanyaan 14.1 dan 14.2?

- hanya menebak
- sangat tidak yakin
- tidak yakin
- yakin
- sangat yakin
- sangat yakin sekali

Gambar 3.8 Soal Nomor 14 dalam *Three-tier Test*

14.1	a	b	c	d		
14.2	a	b	c	d		
	e.					
14.3	1	2	3	4	5	6

Gambar 3.9 Contoh Jawaban Siswa pada Soal Nomor 14 dalam *Three-tier Test*

Jika dilihat pada soal dan dibandingkan dengan kunci jawaban pada kisi-kisi instrumen, maka jawaban pada Gambar 3.7 dan 3.9 tidak menunjukkan miskonsepsi karena walaupun kombinasi jawaban dan alasannya salah tetapi tidak mengindikasikan miskonsepsi karena jawaban dan alasan tidak berhubungan.

Permasalahan dalam kombinasi jawaban siswa ini muncul karena bentuk soal yang disusun peneliti memungkinkan adanya siswa yang memilih jawaban yang tidak berhubungan. Jawaban yang tidak berhubungan ini dapat mengindikasikan adanya kecerobohan, adanya unsur menebak, atau kurangnya minat, perhatian siswa, atau pemahaman siswa mengenai cara mengerjakan soal. Kaltakci dan Didiş (2007: 500) mempertimbangkan jawaban seperti ini sebagai jawaban *error* (salah) dan bukan merupakan miskonsepsi. Oleh karena itu, peneliti memasukkan kombinasi jawaban dan alasan yang tidak berhubungan ini dalam kategori jawaban *Error* (salah) karena skala *Confidence Rating* ≥ 4 . Sedangkan, jika jawaban dan alasan yang tidak berhubungan ini berkombinasi dengan keyakinan siswa yang rendah terhadap jawabannya sendiri atau skala *Confidence Rating* ≤ 3 , maka kombinasi jawaban seperti ini dikategorikan sebagai *lack of knowledge* (Pesman dan Eryilmaz, 2010: 209). Berdasarkan penjelasan tersebut, teknik analisis kombinasi jawaban pada *Three-tier Test* yang digunakan dalam penelitian ini dirangkum dalam Tabel 3.7.

Tabel 3.7 Teknik Analisis Kombinasi Jawaban pada *Three-tier Test* Hasil Adaptasi dan Adopsi

Tingkatan Soal	Kategori	Tipe Jawaban
<i>Three-tier</i>	Memahami Konsep	Jawaban benar, alasan benar, dan skala CR ≥ 4
	Error	Jawaban salah, alasan benar, dan skala CR ≥ 4
		Jawaban benar, alasan salah, dan skala CR ≥ 4 (jawaban dan alasan tidak berhubungan)
		Jawaban salah, alasan salah, dan skala CR ≥ 4 (jawaban dan alasan tidak berhubungan)

Tabel 3.7 Teknik Analisis Kombinasi Jawaban pada *Three-tier Test* Hasil Adaptasi dan Adopsi (Lanjutan)

Tingkatan Soal	Kategori	Tipe Jawaban
<i>Three-tier</i>	<i>Lack of Knowledge</i>	Jawaban benar, alasan benar, dan skala $CR \leq 3$
		Jawaban benar, alasan salah, dan skala $CR \leq 3$
		Jawaban salah, alasan benar, dan skala $CR \leq 3$
		Jawaban salah, alasan salah, dan skala $CR \leq 3$
		Jawaban benar, alasan salah, dan skala $CR \leq 3$ (jawaban dan alasan tidak berhubungan)
		Jawaban salah, alasan salah, dan skala $CR \leq 3$ (jawaban dan alasan tidak berhubungan)
	Miskonsepsi	Jawaban benar, alasan salah, dan skala $CR \geq 4$
		Jawaban salah, alasan salah, dan skala $CR \geq 4$

G. Teknik Pengumpulan Data

Data yang dibutuhkan dalam penelitian ini adalah data miskonsepsi siswa mengenai materi kalor. Karena instrumen yang digunakan adalah *Three-tier Test*, maka pengumpulan data dilakukan dalam bentuk tes tertulis. Tes tulis ini diberikan kepada 271 siswa kelas VII yang berasal dari tiga SMP Negeri dengan klaster yang berbeda-beda.

H. Teknik Pengolahan dan Analisis Data

Sebelum melakukan teknik pengolahan data, jawaban siswa pada *Three-tier Test* dikelompokkan dalam variabel-variabel data berikut ini.

1. *Two-tier Test* (TT)

Pada variabel ini yang dinilai adalah jawaban siswa pada kedua tingkat soal. Jika jawaban pada soal tingkat pertama benar dan alasan yang dipilih pada soal tingkat kedua juga benar, maka siswa diberi skor 1. Jika selain jawaban tersebut, maka siswa diberi skor 0.

2. *Confidence Rating*

Pada variabel ini yang dinilai adalah jawaban siswa pada soal tingkat ketiga atau pada soal tingkat kepercayaan diri (*Confidence Rating*). Semakin besar skala

Confidence Rating yang dipilih siswa, semakin menunjukkan bahwa siswa yakin terhadap jawabannya.

3. *Three-tier Test*

Siswa yang benar menjawab pada *Two-tier Test* dan yakin atas jawabannya (skala *Confidence Rating* ≥ 4) dikatakan memahami konsep dan diberi skor 1. Selain dari jawaban itu, maka skornya adalah 0. Skor *Three-tier Test* ini digunakan untuk menentukan tingkat atau *rank* kemampuan memahami konsep siswa. Teknik yang digunakan adalah teknik Standar Deviasi dengan 3 *rank* (Arikunto, 2009: 263).

Langkah-langkah untuk menentukan tingkat kemampuan memahami konsep ini adalah:

1. Menjumlahkan skor semua siswa.
2. Menghitung nilai rata-rata (mean) dan simpangan baku atau standar deviasi (SD) menggunakan rumus berikut ini.

$$\bar{X} = \frac{\sum X}{N} \quad (3.10)$$

$$SD = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} \quad (3.11)$$

Dengan:

\bar{X} : skor rata-rata (mean)

X : skor atau jumlah kode miskonsepsi

N : jumlah siswa

SD : standar deviasi atau simpangan baku

3. Menentukan batas-batas kelompok, dengan:
 - a. Kelompok tinggi adalah semua siswa yang mempunyai skor sebanyak skor rata-rata +1 SD ke atas.
 - b. Kelompok sedang adalah semua siswa yang mempunyai skor antara -1 SD dan +1 SD.
 - c. Kelompok rendah adalah semua siswa yang mempunyai skor -1 SD dan yang kurang dari itu.

Diagnosis miskonsepsi dilakukan dengan cara menganalisis kombinasi jawaban siswa pada *Two-tier Test* dan skala *Confidence Rating* mengacu pada teknik analisis kombinasi jawaban yang ditunjukkan Tabel 3.7. Agar lebih mudah dalam mengolah data, setiap kategori jawaban diberi kode yang menunjukkan singkatan kategori tersebut. Secara sederhana, cara pemberian kode tersebut diringkas dalam Tabel 3.8 berikut ini.

Tabel 3.8 Kode Kategori Jawaban pada Pengolahan Data *Three-tier Test*

Kategori	Tipe Jawaban	Kode
Memahami Konsep	Jawaban benar, alasan benar, dan skala $CR \geq 4$	MK
Error	Jawaban salah, alasan benar, dan skala $CR \geq 4$	Error
	Jawaban benar, alasan salah, dan skala $CR \geq 4$ (jawaban dan alasan tidak berhubungan)	Error-TB
	Jawaban salah, alasan salah, dan skala $CR \geq 4$ (jawaban dan alasan tidak berhubungan)	Error-TB
<i>Lack of knowledge</i>	Jawaban benar, alasan benar, dan skala $CR \leq 3$	LoK-True
	Jawaban benar, alasan salah, dan skala $CR \leq 3$	LoK- ...
	Jawaban salah, alasan benar, dan skala $CR \leq 3$	LoK-Error
	Jawaban salah, alasan salah, dan skala $CR \leq 3$	LoK-...
	Jawaban benar, alasan salah, dan skala $CR \leq 3$ (jawaban dan alasan tidak berhubungan)	LoK-TB
	Jawaban salah, alasan salah, dan skala $CR \leq 3$ (jawaban dan alasan tidak berhubungan)	LoK-TB
Miskonsepsi	Jawaban benar, alasan salah, dan skala $CR \geq 4$	M- ...
	Jawaban salah, alasan salah, dan skala $CR \geq 4$	M- ...

Pernyataan miskonsepsi yang digunakan sebagai distraktor soal tingkat kedua (*Second Tier*) dalam *Three-tier Test* harus diurutkan terlebih dahulu. Pernyataan miskonsepsi ini seluruhnya berjumlah 22 pernyataan. Akan tetapi, miskonsepsi yang dapat digunakan dalam soal *Three-tier Test* untuk penelitian ini hanya berjumlah 17 pernyataan karena berdasarkan hasil uji coba instrumen telah ditentukan bahwa soal yang dapat digunakan hanya 11 butir soal. Hal tersebut mengakibatkan empat pernyataan miskonsepsi tidak dapat digunakan sebagai distraktor dan tidak dapat digali dari siswa. Empat miskonsepsi tersebut adalah:

1. jika kedua benda suhunya sama, maka kedua benda tersebut memiliki kalor yang sama (M-1).

2. suhu adalah ukuran banyaknya kalor (M-3).
3. terdapat dua jenis kalor, yaitu kalor dingin dan kalor panas (M-6).
4. suhu zat berubah selama mendidih (M-17).

Sehubungan dengan hal tersebut, saat mengolah data, titik-titik kosong pada kode kategori jawaban dalam Tabel 3.8 diisi dengan nomor urut miskonsepsi. Nomor urut miskonsepsi dan kombinasi jawaban siswa yang mengindikasikan miskonsepsi dapat dilihat dalam tabel Indikasi Miskonsepsi berdasarkan Kombinasi Jawaban Siswa pada *Three-tier Test* yang secara rinci dilampirkan dalam Lampiran A.3. Siswa dikatakan mengalami miskonsepsi jika jawaban pada *Three-tier Test* sesuai dengan kombinasi jawaban pada tabel indikasi miskonsepsi dalam Lampiran A.3 dan skala *Confidence Rating* yang dipilih adalah skala 4-6.

4. Kode Miskonsepsi

Kode miskonsepsi ini digunakan untuk menentukan *rank* atau tingkat miskonsepsi yang dialami siswa dan untuk menentukan persentase siswa yang mengalami miskonsepsi. Untuk menentukan tingkat miskonsepsi, peneliti mengubah kode miskonsepsi (M-...) menjadi kode angka. Semua kode jawaban siswa yang menunjukkan miskonsepsi (M-...) diubah menjadi angka 1. Selain kode jawaban miskonsepsi, maka kode diubah menjadi angka 0.

Penentuan ranking atau tingkat miskonsepsi siswa dalam penelitian ini menggunakan teknik Standar Deviasi dengan 3 *rank* (Arikunto, 2009: 263). Tiga tingkat miskonsepsi ini meliputi kelompok tinggi, sedang, dan rendah. Langkah-langkah untuk menentukan tingkat miskonsepsi ini sama seperti langkah-langkah untuk menentukan tingkat kemampuan memahami konsep.