

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Penelitian ini dilaksanakan dengan menggunakan metode pengembangan. Hasil penelitian berupa pemaparan secara kualitatif dan kuantitatif. Desain penelitian yang digunakan adalah desain penelitian pengembangan (*Research and Development*) oleh Plomp (2013). Menurut Plomp (2013) penelitian pengembangan terdiri dari lima tahap, yaitu 1) investigasi awal (*preliminary investigation*); 2) tahap desain (*the desain phase*); 3) tahap realisasi/konstruksi (*the realization/construction*); 4) tahap evaluasi (*the stage of evaluation*); 5) tahap implementasi (*the stage of implementation*).

3.2 Partisipan Penelitian

Partisipan dalam penelitian ini adalah peserta didik SMP Program studi Ilmu Pengetahuan Alam (IPA) di kelas VIII pada mata pelajaran IPA. Penetapan partisipan dari peserta didik SMP IPA kelas VIII karena konsep yang akan diujikan masih berkaitan dengan kelas dibawahnya yaitu kelas VII. Partisipan penelitian ini diambil pada salah satu SMP Negeri yang berada pada kota Cilegon, provinsi Banten.

3.3 Populasi dan Sample

Penelitian ini dilaksanakan dengan menggunakan teknik *purposive sampling* dalam memilih sample penelitian. Menurut Sugiyono (2013), *purposive sampling* adalah teknik pengambilan sampel sumber data dengan pertimbangan tertentu agar nantinya data yang diperoleh dapat lebih representatif. Populasi merupakan wilayah generalisasi yang terdiri atas objek/subjek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian diambil kesimpulannya. Sedangkan sampel merupakan bagian dari populasi yang mewakili populasi tersebut. Populasi pada penelitian ini adalah seluruh peserta didik kelas VIII di salah satu SMPN yang berada di Cilegon, Provinsi Banten. Pemilihan peserta didik SMP kelas VIII sebagai subyek penelitian didasarkan pada pertimbangan peserta didik tersebut merupakan kelompok peserta didik yang dirasa siap untuk

menerima perlakuan penelitian ini baik secara waktu dan materi yang tersedia. SMPN tersebut dipilih karena berdasarkan kurikulum yang sedang berjalan yaitu kurikulum merdeka, peneliti ingin mengembangkan asesmen yang sesuai dengan kurikulum merdeka dan juga pemilihan partisipan ini didasarkan pada pertimbangan bahwa topik yang diambil yaitu seluruh materi semester 2 kelas 8 SMP pada mata pelajaran IPA. Karena studi ini berfokus pada pengembangan dan penggunaan asesmen diagnostik *three tier* sebagai alat penilaian pembelajaran untuk menganalisis tingkat pemahaman konsep peserta didik. Partisipan pada penelitian ini adalah peserta didik kelas 8 SMP di sekolah yang menggunakan kurikulum merdeka dalam pembelajaran IPA.

Sampel penelitian ditentukan berdasarkan *convenience sampling*. Pemilihan sampel didasarkan pada pertimbangan yang diperoleh dari guru dan kelas yang mendapatkan izin administratif dari pihak sekolah. Tujuan dilakukan pengambilan sampel seperti ini adalah agar penelitian dapat dilaksanakan secara efektif dan efisien terutama dalam hal pengawasan, kondisi subyek penelitian, waktu penelitian yang ditetapkan, kondisi tempat penelitian serta prosedur perizinan.

3.4 Instrumen Penelitian

Instrumen untuk mengumpulkan data pada penelitian ini disusun dalam bentuk tes tertulis berbentuk pilihan ganda bertingkat dan angket/kuisisioner. Instrumen tersebut terdiri atas: (a) tes diagnostik *three tier*, (b) kuisisioner *self efficacy*. Pengembangan instrumen ini dilakukan melalui beberapa tahap, yaitu tahap pembuatan, revisi, dan tahap uji coba instrumen. Uji coba instrumen dilakukan untuk melihat layak atau tidak suatu instrumen digunakan dalam penelitian. Hasil uji coba instrumen yang dilakukan dijelaskan dibawah ini:

3.4.1 Tes Asesmen Diagnostik *Three Tier*

Tes asesmen diagnostik yang digunakan untuk mengidentifikasi tingkat pemahaman konsep peserta didik disajikan dalam bentuk pilihan ganda bertingkat, yang terdiri atas 50 soal yang terdiri dari 5 materi IPA yang ada pada semester 2 kelas 8, satu materi diberikan 10 soal. Soal tes ini diberikan diawal pembelajaran sebelum peserta didik masuk ke bab

yang akan dipelajari, hal ini berguna untuk memberikan gambaran pada guru tentang pemahaman konsep awal peserta didik sudah sejauh mana, dan mengetahui berapa banyak peserta didik yang sudah paham konsep, paham sebagian, dan belum paham.

Tes diagnostik *three tier* yang dikembangkan mengacu pada aturan yang ada didalam kurikulum merdeka, dimana soal tersebut memiliki keterkaitan konsep dari satu tingkat dibawahnya sebanyak 6 soal, 2 tingkat dibawahnya sebanyak 2 soal, dan 2 soal tentang materi yang akan dipelajari. Pemilihan materi yang digunakan pada penelitian ini berdasarkan capaian pembelajaran dan tujuan pembelajaran yang ada didalam kurikulum merdeka sehingga materi yang digunakan sesuai dengan yang ada didalam kurikulum merdeka.

Tes diagnostik pilihan ganda *three tier* yang dikembangkan akan menghasilkan beberapa pola jawaban peserta didik, Kriteria penilaian pemahaman konsep peserta didik diadaptasi dari Peşman & Eryilmaz, (2010).

Tabel 3.1 Keputusan Terhadap pola Jawaban Peserta didik

Kategori	Tingkat Pertama	Tingkat Kedua	Tingkat Keyakinan
Tidak Paham Konsep	Salah	Salah	Tidak Yakin
Paham Konsep	Benar	Benar	Yakin
Paham Sebagian	Benar	Benar	Tidak Yakin
	Benar	Salah	Tidak Yakin
	Salah	Benar	Tidak Yakin
	Salah	Benar	Yakin
	Benar	Salah	Yakin

Soal tes yang baik menurut Arikunto (2013:72) memenuhi persyaratan: (a) Validitas; (b) Reliabilitas; (c) Objektivitas; (d) Praktikabilitas, dan (e) Ekonomis. Pada penelitian ini soal tes terlebih dahulu diuji reliabilitas, validitas, tingkat kesukaran, dan daya pembeda. Untuk mendapatkan hal tersebut, soal tes harus diujicobakan pada kelas

lain di sekolah pada tingkat yang sama. Analisis instrumen pada penelitian ini menggunakan *Rasch* model, yang merupakan salah satu teori respon butir dengan bantuan program *Winsteps*.

3.4.2 Reliabilitas Tes Asesmen Diagnostik *Three Tier*

Reliabilitas adalah tingkat atau derajat konsistensi dari suatu instrumen yang berkenaan dengan pertanyaan. Apakah suatu tes teliti dan dapat dipercaya sesuai dengan kriteria yang telah ditetapkan. Uji reliabilitas butir soal dilakukan dengan menggunakan *Rasch* model.

Pada *Rasch* model, reliabilitas digambarkan dengan nilai *person reliability*, *item reliability* dan *item separation*. Biasanya reliabilitas dianalisis dengan cara membandingkan nilai *person reliability* dan *item reliability* yang diperoleh dengan kriteria koefisien reliabilitas. Pada penelitian ini untuk mengukur tingkat reabilitas dari kumpulan soal, peneliti menggunakan aplikasi *winstep* ntuk mengujinya, dari hasil yang diperoleh dapat diinterpretasikan dengan menggunakan klasifikasi koefisien reabilitas. Pengujian reliabilitas soal dalam bentuk pilihan ganda di uji dengan menggunakan rumus *Alpha* sebagai berikut: (Arikunto, 2009).

$$r_{11} = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

Keterangan :

- r_{11} : Reliabilitas yang dicari
- $\sum \sigma_i^2$: Jumlah varians skor tiap-tiap item
- σ_t^2 : Varians total

Tabel 3.2 Koefisien Reabilitas Menurut *Rasch* Model

Nilai Reliability (Person/Item)	Interpretasi
> 0,94	Istimewa
0,91 – 0,94	Bagus sekali
0,81 – 0,90	Bagus
0,67 – 0,80	Cukup
< 0,67	Lemah

3.4.3 Validitas Tes Asesmen Diagnostik *Three Tier*

a. Validitas Isi dan Konstruk

Untuk mendapatkan soal yang memenuhi syarat validitas isi dan validitas konstruk, maka pembuatan soal dilakukan dengan meminta pertimbangan dan saran dari ahli, dosen pembimbing, guru IPA dan teman sejawat. Validitas konstruk disebut pula validitas bentuk soal atau validitas tampilan, yaitu keabsahan susunan kalimat atau kata-kata dalam soal sehingga jelas pengertiannya dan tidak menimbulkan tafsiran lain. Sedangkan validitas isi berarti ketepatan tes tersebut ditinjau dari segi materi yang diajukan, yaitu kesesuaian indikator dengan butir soal, kesesuaian soal dengan tingkat kemampuan peserta didik dan kesesuaian materi serta tujuan yang ingin dicapai. Berdasarkan hasil uji validitas muka dan isi, perbaikan instrumen terdapat pada susunan kalimat, baik pada instrumen tes diagnostik *three tier*, maupun pada kuisioner *self efficacy*. Setelah melalui proses perbaikan sesuai yang disarankan, instrumen tersebut digunakan pada penelitian.

b. Validitas Butir Soal

Untuk melihat validitas butir soal, digunakan model *Rasch*. Uji validitas ini dilakukan dengan bantuan program *winsteps*. Hal yang dilihat adalah berdasarkan nilai *Outfit Mean Square (MNSQ)*, *Outfit Z-Standard (ZSTD)*, dan *Point Measure Correlation (Pt Mean Corr)*. Dengan kriteria menurut Sumintono & Widhiarso (2013; 111) sebagai berikut.

Nilai *Outfit Mean Square (MNSQ)* yang diterima: $0,5 < MNSQ < 1,5$

Nilai *Outfit Z-Standard (ZSTD)* yang diterima: $- 2,0 < ZSTD < +2,0$

Nilai *Point Measure Correlation (Pt Mean Corr)*: $0,4 < Pt Mean Corr < 0,85$

Bila butir tes kemampuan pemecahan masalah matematis memenuhi setidaknya dua kriteria diatas, maka butir soal atau

pernyataan tersebut dapat digunakan, dengan kata lain butir tersebut valid. Hasil yang diperoleh dari uji validitas tes asesmen diagnostik *three tier* adalah sebagai berikut.

Berdasarkan kriteria validitas yang telah ditetapkan sebelumnya, *item measure* pada tabel 3.4 dianalisis. Hasilnya menunjukkan bahwa tiga butir soal tidak dapat digunakan, yaitu nomor 11, 4 dan 35 karena tidak masuk kedalam dua kriteria yang dapat diterima dan hanya masuk satu kategori yang dapat diterima. Sedangkan soal nomor 19 dan 24 soal masih dapat dipertahankan dan diperbaiki meskipun nilai *ZSTD* tidak memenuhi namun nilai *MNSQ* dan *Pt Measure Core* masih masuk kriteria, sehingga kedua butir soal tersebut masih dapat digunakan. karena masih memenuhi dua kriteria lainnya. Kesimpulan yang diperoleh dari uji validitas di atas adalah bahwa terdapat 47 soal yang dapat digunakan untuk penelitian.

3.4.4 Tingkat Kesukaran Tes Asesmen Diagnostik *Three Tier*

Tingkat kesukaran soal adalah besaran yang digunakan untuk menyatakan apakah suatu soal termasuk ke dalam kategori mudah, sedang atau sukar. Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Pada penelitian ini, uji tingkat kesukaran dilakukan menggunakan *Rasch* model dengan bantuan program *winsteps*. Pada program tersebut, disajikan urutan butir soal dari yang tersulit sampai pada butir soal yang termudah. Pada pemodelan Rasch, tingkat kesulitan butir soal dikategorikan berdasarkan Measure logit dan nilai Simpangan Baku (SD) logit item dan dibagi menjadi empat kategori sebagai berikut (Sumintono & Widhiarso, 2015):

Tabel 3.3 Koefisien Tingkat Kesukaran Menurut Rasch Model

Nilai Measure (logit)	Interpretasi Kesulitan Butir Soal
$Measure\ logit < -1,12$	Item sangat mudah
$-1,12 \leq Measure\ logit \leq 0,00$	Item mudah
$0,00 \leq Measure\ logit \leq 1,12$	Item sulit
$Measure\ logit > 1,12$	Item sangat sulit

Dalam konteks soal untuk menguji hasil belajar peserta didik, sebaran tingkat kesulitan yang beragam mulai dari yang paling sulit hingga yang mudah tidak menjadi persoalan. Hal ini tentu dikarenakan tingkat kemampuan dalam memahami setiap peserta didik juga berbeda-beda. Jika akan menggunakan soal untuk menyeleksi peserta didik pada kompetensi tertentu maka memilih butir soal yang memiliki kesulitan yang tinggi, sedang, dan mudah menjadi pilihan yang tepat, karena soal yang baik adalah soal yang memiliki tingkat kesulitan yang beragam.

3.4.5 Daya Pembeda Tes Asesmen Diagnostik *Three Tier*

Daya pembeda antara *item* dan *person* (*separation*) dapat dilakukan dengan menggunakan analisis model Rasch. Analisis ini terdapat pada tingkat abilitas individu. Dari hasil analisis yang didapat, menurut Sumintono dan Widhiarso, (2015) dapat diketahui bahwa semakin besar nilai separasi maka kualitas instrumen dalam hal keseluruhan responden dan butir makin bagus, karena dapat mengidentifikasi kelompok responden dan butir. Persamaan lain untuk melihat pengelompokan secara lebih teliti digunakan persamaan pemisahan strata (H):

$$H = \frac{(4 \times SEPARATION) + 1}{3}$$

Keterangan:

H : Nilai strata (Pengelompokan)

Separation : Nilai *separation* untuk responden yang dihasilkan

Adapun hasil pengujian daya pembeda untuk *item* dan *person* (*separation*) berdasarkan model Rasch menggunakan *software* Winstep adalah nilai separasi *item* sebesar 2,23 maka H = 3,30 dibulatkan menjadi 3. Artinya terdapat tiga kelompok butir soal yang dapat diidentifikasi, sedangkan untuk nilai separasi *person* sebesar 2,17 maka H = 3,16 dibulatkan menjadi 3.

3.4.6 Kuisiner *Self Efficacy*

Kuisiner *Self Efficacy* dalam penelitian ini diadaptasi dari Hsia et al., (2016), namun dilakukan beberapa perbaikan pada poin – poinnya

agar lebih sesuai dengan penelitian ini. Kuisisioner Ini terdiri dari sepuluh item menggunakan skala Likert empat poin. Peserta didik diminta untuk memberikan tanda (√) Pada jawaban yang dirasanya benar.

Tabel 3.4 Kuisisioner *Self Efficacy*

Dimensi	Pertanyaan	SS	S	TS	STS
Level (<i>Magnitude</i>)	Setiap tugas IPA kelas 8 semester 1 yang diberikan pasti saya kerjakan.				
	Saya selalu memiliki ide untuk bisa mengerjakan tugas IPA khususnya materi pada kelas 8 semester 1				
	Saya merasa tertantang ketika menghadapi soal IPA kelas 8 semester 1 yang sulit.				
	Saya yakin dapat mengikuti tes diagnostik materi IPA kelas 8 semester 1 dengan baik.				
	Saat besok diadakan tes diagnostik saya lebih memilih belajar daripada menonton acara TV kesukaan saya				
Kekuatan (<i>Strenght</i>)	Saya selalu mencoba menggunakan cara lain ketika gagal menyelesaikan soal materi IPA kelas 8 semester 1				
	Setiap kesulitan dalam tes diagnostik materi IPA kelas 8 semester 1 pasti bisa saya atasi dengan baik.				
	Saya tidak pernah menunda - nunda untuk mengerjakan tugas yang diberikan				
	Saya tidak pernah terlambat mengumpulkan tugas.				
	Saya memiliki kemampuan yang baik dalam melaksanakan tes diagnostik yang diberikan				
Generalisasi (<i>Generaality</i>)	Saya tidak akan menyerah sebelum mencoba mengerjakan soal IPA walau sesulit apapun.				
	Saat nilai saya baik saya lebih bersemangat untuk belajar agar nilai yang saya peroleh lebih baik lagi				
	Saya senang membaca berbagai buku IPA untuk memperoleh informasi baru.				

Saya menjadikan pengalaman untuk meningkatkan keyakinan dalam mencapai kesuksesan.

Saya tidak menjadi pesimis saat nilai IPA saya buruk

Validitas butir pernyataan kusioner *self efficacy* juga dilihat menggunakan model *rasch* dengan bantuan program *winsteps*. Sama halnya dengan butir soal tes asesmen diagnostik *three tier*, kriteria yang digunakan dalam menentukan validitas instrumen angket adalah berdasarkan nilai *Outfit Mean Square (MNSQ)*, *Outfit Z-Standard (ZSTD)*, dan *Point Measure Correlation (Pt Mean Corr)*. Dengan kriteria menurut Sumintono & Widhiarso (2013; 111) sebagai berikut.

Nilai *Outfit Mean Square (MNSQ)* yang diterima: $0,5 < MNSQ < 1,5$

Nilai *Outfit Z-Standard (ZSTD)* yang diterima: $- 2,0 < ZSTD < +2.0$

Nilai *Point Measure Correlation (Pt Mean Corr)*: $0,4 < Pt Mean Corr < 0,85$

3.5 Teknik Pengumpulan Data

Tabel 3.5 menunjukkan bentuk data yang diukur, instrumen yang digunakan untuk mendapatkan data tersebut, sumber dari data tersebut dan bagaimana cara mendapatkan data tersebut.

Tabel 3.5 Data, Instrumen, Sumber Data, dan Teknik Pengumpulan Data

Data	Instrumen	Sumber Data	Teknik Pengumpulan Data
Skor peserta didik untuk pemahaman konsep	Tes diagnostik three tier	Seluruh partisipan	Partisipan mengerjakan tes sebelum memasuki materi yang akan dipelajari
Skala <i>Self-efficacy</i>	Kuisisioner <i>Self-efficacy</i>	Seluruh partisipan	Sebelum partisipan mengerjakan tes diagnostik, mengisi kuisisioner <i>self-efficacy</i> terlebih dahulu

3.6 Teknik Analisis Data

Data yang dianalisis pada penelitian ini adalah kualitas butir soal, hasil penggunaan tes, dan kuisisioner *self efficacy*. Untuk menganalisis kualitas butir soal menggunakan aplikasi *winstep* dengan rasch model untuk melakukan uji validitas, reliabilitas, daya pembeda dan indeks kesukaran. Sedangkan untuk melihat hasil penggunaan tes dengan menggunakan analisis keputusan terhadap pola jawaban peserta didik, dan untuk *self efficacy* digunakan skala likert.

1. Analisis data hasil penggunaan asesmen diagnostik *three tier*

Analisis data hasil tes dianostik *three tier* yang dikembangkan dan digunakan untuk mengidentifikasi pemahaman konsep peserta didik diuji dengan beberapa langkah berikut ini:

a) Mengolah hasil asesmen diagnostik *three tier* yang diberikan

Besarnya nilai yang diperoleh peserta didik merupakan persentase dari skor maksimum ideal yang harus dicapai jika tes tersebut dikerjakan dengan hasil 100% benar (Purwanto, 2010). Penilaian dengan persen ini digunakan untuk menilai hasil kerja peserta didik pada tes diagnostik untuk mengetahui kesulitan belajar peserta didik. Berikut adalah rumus perhitungan nilai persentase peserta didik untuk menjaring pemahaman konsep peserta didik.

$$NP = \frac{R}{SM} \times 100$$

Keterangan :

NP : Nilai persen yang dicari

R : Skor mentah yang diperoleh peserta didik.

SM : Skor maksimum ideal tes yang bersangkutan.

100 : Bilangan tetap

b) Membagi peserta didik berdasarkan nilai menjadi kategori “Paham, Paham Sebagian, Tidak Paham, dan Miskonsepsi”

Tes diagnostik pilihan ganda *three tier* yang dikembangkan akan menghasilkan beberapa pola jawaban peserta didik, Kriteria penilaian pemahaman konsep peserta didik diadaptasi dari Peşman & Eryilmaz,

(2010).

c) Hitung rata – rata kelas

Menghitung nilai rata-rata keseluruhan yang diperoleh peserta didik

$$\text{Nilai rata – rata kelas} = \frac{\Sigma \text{Nilai seluruh siswa}}{\Sigma \text{siswa}}$$

Jika peserta didik mendapatkan nilai rata-rata kelas, maka mereka akan mengikuti pembelajaran sesuai fasenya. Peserta didik yang mendapat nilai di bawah rata-rata dikategorikan sebagai peserta didik yang mengalami kesulitan belajar dan akan mengikuti pembelajaran khusus atau pendampingan pada kompetensi yang belum terpenuhi. Sedangkan peserta didik dengan nilai di atas rata-rata akan mengikuti pembelajaran dengan pengayaan.

2. Analisis kuisisioner *self efficacy*

Kuisisioner penilaian data yang digunakan untuk menilai lembar *self efficacy* dari ahli menggunakan skala Likert Sugiyono (2013).

Tabel 3.6 Kategori Skor dalam Skala Likert

Keterangan	Skor
Sangat Setuju (SS)	4
Setuju (S)	3
Tidak Setuju (TS)	2
Sangat Tidak Setuju (STS)	1

Setelah memperoleh hasil penilaian dari para ahli, maka skor yang diperoleh dihitung dengan menggunakan rumus. Menurut (Arifin, 2010) sebagai berikut:

$$P = \frac{\Sigma R}{N} \times 100\%$$

Keterangan :

P = Persentase respon peserta didik

ΣR = Jumlah jawaban yang diberikan oleh validator

N = Jumlah skor maksimal atau ideal

2.1 Prosedur Penelitian

Berikut ini adalah tahapan-tahapan yang dilakukan dalam penelitian ini:

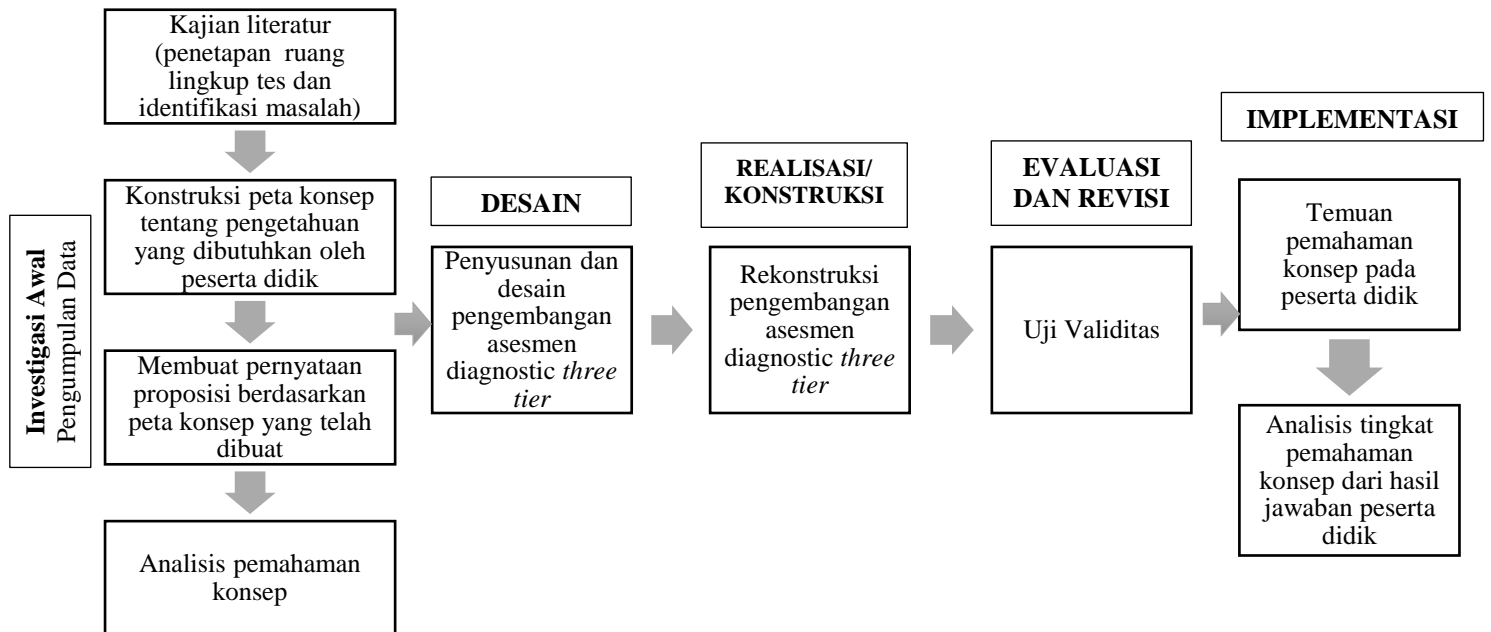


Diagram 3.1 Alur Pengembangan Penelitian

1) Investigasi Awal

Pada tahap ini dilakukan pengumpulan data melalui beberapa tahap:

- a) Kajian literatur untuk menetapkan ruang lingkup tes
Tahap awal penelitian ini adalah menentukan ruang lingkup dan tujuan tes melalui analisis CP dan TP SMP. Analisis ini bertujuan untuk menentukan indikator yang mewakili konsep pembelajaran yang nantinya akan dikembangkan menjadi butir soal untuk mengidentifikasi pemahaman konsep pada materi tersebut.
- b) Konstruksi peta konsep tentang pengetahuan yang dibutuhkan oleh peserta didik tentang materi IPA semester 2 Kelas 8.
- c) Membuat pernyataan proposisi berdasarkan peta konsep yang telah dibuat
- d) Analisis pemahaman konsep
Analisis pemahaman konsep bertujuan untuk memperoleh informasi tentang konsep alternatif peserta didik melalui *open ended two tier*

multiple choice. Soal-soal pada *open ended two tier multiple choice* merupakan soal pengembangan dari peta konsep dan pernyataan proposisi. Peneliti akan menggunakan jawaban peserta didik untuk pilihan jawaban pada *three tier test diagnostic* yang akan dikembangkan.

2) Desain Produk (*The Desain Phase*)

Desain produk berupa instrumen *three tier test diagnostic* yang berisi butir soal. Butir soal ini mengacu pada hasil analisis pemahaman konsep pada tahapan awal. Hasil dari identifikasi miskonsepsi berupa pola yang terdiri dari konsep benar dan konsep miskonsepsi. Berdasarkan pola tersebut maka peneliti akan mengembangkan *three tier test diagnostic*. Tes ini terdiri dari tiga tingkatan, tingkat pertama terdiri dari empat pilihan jawaban, tingkat kedua terdiri dari alasan pemilihan pada tingkatan pertama dan tingkatan ketiga terdiri dari dua pilihan tingkat keyakinan, yaitu yakin dan tidak yakin. Realisasi atau Konstruksi (*The Realization/Construction*)

Selanjutnya peneliti akan mengkonstruksi desain produk sebagai bentuk realisasi instrumen *three tier test diagnostic*. Konstruksi ini dipakai untuk mengidentifikasi pemahaman konsep peserta didik.

3) Evaluasi dan Revisi (*The Stage Of Evaluation*)

Peneliti melakukan uji kualitas asesmen diagnostik *three tier* yang telah dikembangkan. Uji kualitas instrumen dilakukan melalui 4 tahapan yaitu :

a) Uji Validitas

Validitas merupakan acuan yang menunjukkan tingkat kesahihan penelitian untuk mengklaim apa yang diukur sehingga kesimpulan bisa diperoleh dengan akurat (Fraenkel & Wallen, 2009). Uji validitas terdiri dari dua uji yakni uji validitas isi dan uji validitas empiris.

b) Reliabilitas

Reliabilitas adalah tingkat atau derajat konsistensi dari suatu instrumen yang berkenaan dengan pertanyaan. Apakah suatu tes teliti

dan dapat dipercaya sesuai dengan kriteria yang telah ditetapkan. Uji reliabilitas butir soal dilakukan dengan menggunakan Rasch model.

Pada Rasch model, reliabilitas digambarkan dengan nilai *person reliability*, *item reliability* dan *item separation*. Biasanya reliabilitas dianalisis dengan cara membandingkan nilai *person reliability* dan *item reliability* yang diperoleh dengan kriteria koefisien reliabilitas.

c) Daya Pembeda

Daya pembeda suatu soal merupakan kemampuan butir soal untuk membedakan peserta didik dengan kemampuan tinggi dan peserta didik dengan kemampuan rendah. Informasi mengenai daya pembeda sangat diperlukan untuk memperbaiki kualitas soal. Semakin tinggi daya pembeda soal berarti semakin mampu soal tersebut membedakan kemampuan peserta didik dalam memahami materi. Daya pembeda soal dilihat berdasarkan *nilai separation* untuk menghitung daya pembeda setiap butir soal dapat digunakan rumus sebagai berikut: (Arikunto, 2009).

$$DP = \frac{(WL - WH)}{n}$$

Keterangan:

- DP : daya pembeda
- WL : jumlah peserta didik yang gagal dari kelompok bawah
- WH : jumlah peserta didik yang gagal dari kelompok atas
- n : 27% X N (jumlah peserta didik)

Tabel 3.7 Koefisien Daya Pembeda

Koefisien D	Interpretasi
D < 0.00	Tidak Baik
0.0 < D ≤ 0.20	Jelek
0.20 < D ≤ 0.40	Cukup
0.40 < D ≤ 0.70	Baik
0.70 < D ≤ 1.00	Baik Sekali

d) Indeks Kesukaran

Indeks kesukaran adalah kemampuan tes dalam menjangar banyaknya subjek tes peserta tes dengan yang mengerjakan dengan

benar, maka tingkat kesukaran tes tersebut tinggi, sebaliknya jika hanya sedikit dari subjek yang tidak dapat menjawab dengan benar maka tingkat kesukarannya rendah (Sumintono & Widhiarso, 2015).

4) Implementasi

Instrumen *three tier test diagnostic* yang telah melalui tahapan evaluasi dan revisi di ujikan kepada peserta didik. *Three tier test diagnostic* akan mengungkapkan tingkat pemahaman konsep peserta didik pada materi yang diujikan.