

BAB III

METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan di SMA KP Baleendah yang beralamat di Jl. Adipati Agung No. 32, Kecamatan Baleendah, Kabupaten Bandung, Jawa Barat. Pemilihan lokasi penelitian didasarkan pada hasil observasi awal dan diskusi dengan guru pamong Biologi. Hasil diskusi menunjukkan bahwa materi sistem reproduksi memiliki karakteristik yang kompleks, abstrak, dan sensitif sehingga memerlukan pendekatan pembelajaran yang mampu mendorong pemahaman konsep secara mendalam. Selain itu, meskipun SMA KP Baleendah telah mengimplementasikan pendekatan *deep learning* dalam Kurikulum Merdeka, proses pembelajaran Biologi pada materi sistem reproduksi masih cenderung menekankan penguasaan konsep secara faktual melalui hafalan sehingga keterlibatan kognitif peserta didik belum berkembang secara optimal. Oleh karena itu, sekolah ini dipilih sebagai lokasi penelitian untuk mengkaji penerapan model *Deep-Problem Based Learning (Deep-PBL)* dalam meningkatkan *Cognitive Engagement* peserta didik.

Pengumpulan data dilakukan pada semester genap tahun ajaran 2025/2026, yang disesuaikan dengan rencana pembelajaran materi Sistem Reproduksi di kelas XI. Proses penelitian berlangsung selama kurang lebih 2 minggu 4 kali pertemuan yang efektif, dengan rincian sebagai berikut:

Tabel 3. 1 Jadwal Pelaksanaan Penelitian

Tanggal	Pertemuan	Kelas	Status Kelas	Kegiatan / Materi
MINGGU PERTAMA				
27 April	Pertemuan 1	XI 5	Eksperimen	Penyampaian Materi Sistem Reproduksi Perempuan
28 April	Pertemuan 1	XI 7	Kontrol	Penyampaian Materi Sistem Reproduksi Perempuan
29 April	Pertemuan 2	XI 7	Kontrol	Penyampaian Materi Sistem Reproduksi Laki-laki + Pre-test
30 April	Pertemuan 2	XI 5	Eksperimen	Penyampaian Materi Sistem Reproduksi Laki-laki + Pre-test
MINGGU KEDUA				
4 Mei	Pertemuan 3	XI 5	Eksperimen	Eksperimen Deep PBL: Gangguan dan Teknologi Sistem Reproduksi
5 Mei	Pertemuan 3	XI 7	Kontrol	Eksperimen PBL: Gangguan dan Teknologi Sistem Reproduksi
6 Mei	Pertemuan 4	XI 7	Kontrol	Post-test + Pengisian Kuesioner
7 Mei	Pertemuan 4	XI 5	Eksperimen	Post-test + Pengisian Kuesioner

3.2 Desain Penelitian

Penelitian ini menggunakan metode kuantitatif dengan menggunakan pendekatan *quasi-experimental*. Soesana et al. (2023) menyatakan bahwa penelitian kuantitatif adalah jenis penelitian yang terstruktur dan ilmiah, yang fokus pada fenomena yang bersifat objektif dengan memanfaatkan angka-angka yang dianalisis secara statistik.

Selanjutnya, Soesana et al. (2023) mencatat bahwa desain *quasi-experimental* merupakan salah satu jenis desain eksperimen yang mencakup desain time series dan desain kelompok kontrol yang tidak setara, yang memungkinkan peneliti untuk meneliti hubungan sebab-akibat antara variabel independen dan variabel dependen tanpa harus mengubah kelompok kelas yang sudah ada. *Design* ini diambil karena dalam lingkungan pendidikan di sekolah, tidak mungkin untuk membentuk kelompok secara acak sepenuhnya tanpa mengganggu proses belajar yang sedang berlangsung. Metode *quasi-experimental* memberi kesempatan kepada peneliti untuk mengeksplorasi hubungan sebab akibat antara variabel independen dan variabel dependen sambil mempertahankan kelompok kelas yang sudah ada.

Desain yang dipilih adalah desain kelompok kontrol *pretest-posttest*, yang terdiri dari kelompok kontrol dan kelompok eksperimen, yang masing-masing mengikuti tes awal sebelum perlakuan dan tes akhir setelah perlakuan. Kelompok eksperimen mendapatkan pembelajaran dengan Model *Deep-Problem Based Learning (Deep-PBL)*, sementara kelompok kontrol belajar dengan model PBL konvensional. Sesuai dengan pernyataan Suwastini et al. yang menyatakan bahwa PBL merupakan pendekatan pendidikan yang menempatkan peserta didik dalam posisi pusat melalui kegiatan penyelesaian masalah terbuka, desain ini digunakan untuk menguji secara empiris efisiensi pengembangan model PBL yang terintegrasi dengan pendekatan *Deep learning*. Desain penelitian dirincikan sebagai berikut:

Tabel 3. 2 Desain Penelitian Pretest-Posttest Control Group Design

Kelompok	Pretest	Perlakuan	Posttest
Eksperimen	O ₁	X ₁	O ₂
Kontrol	O ₁	X ₂	O ₂

Keterangan:

O₁ = *Pretest* (tes awal sebelum perlakuan)

O₂ = *Posttest* (tes akhir setelah perlakuan)

X₁ = Perlakuan dengan Model *Deep-PBL* (kelompok eksperimen)

X₂ = Perlakuan dengan Model PBL Konvensional (kelompok kontrol)

3.3 Populasi dan Sampel

Menurut Sugiyono (2023), populasi merupakan wilayah generalisasi yang terdiri atas objek atau subjek yang memiliki karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari sehingga dapat ditarik suatu kesimpulan. Populasi dalam penelitian ini adalah seluruh peserta didik kelas XI SMA KP Baleendah pada semester genap Tahun Ajaran 2025/2026 yang mengikuti mata pelajaran Biologi. Penetapan kelas XI sebagai populasi didasarkan pada karakteristik subjek yang sesuai dengan tujuan penelitian, yaitu peserta didik yang sedang mempelajari materi

gangguan dan teknologi sistem reproduksi sesuai dengan Capaian Pembelajaran Biologi pada Kurikulum Merdeka.

Menurut Sugiyono (2022), sampel merupakan bagian dari jumlah dan karakteristik yang dimiliki oleh populasi. Teknik pengambilan sampel yang digunakan dalam penelitian ini adalah *purposive sampling*, yaitu teknik penentuan sampel berdasarkan pertimbangan tertentu. Berdasarkan hasil observasi awal dan diskusi dengan guru mata pelajaran Biologi, dipilih dua kelas yang memiliki kemampuan awal relatif setara berdasarkan hasil ulangan harian, telah mempelajari materi prasyarat yang sama, serta memiliki karakteristik yang sesuai dengan tujuan penelitian. Adapun kriteria pemilihan sampel dalam penelitian ini adalah sebagai berikut:

1. Peserta Didik yang terdaftar dan aktif di semester genap tahun ajaran 2025/2026
2. Peserta Didik yang telah mengikuti semua rangkaian pembelajaran, mulai dari pretest, perlakuan (eksperimen), hingga posttest
3. Dua kelas yang mempunyai kemampuan awal yang relatif setara berdasarkan hasil ulangan harian sebelumnya.

Dengan pertimbangan tersebut, dua kelas ditetapkan sebagai sampel, di mana satu kelas berfungsi sebagai kelompok eksperimen dan satu kelas lainnya sebagai kelompok kontrol. Berdasarkan kriteria tersebut, ditetapkan kelas XI-5 sebagai kelas eksperimen yang memperoleh pembelajaran menggunakan model *Deep-Problem Based Learning (Deep-PBL)* dan kelas XI-7 sebagai kelas kontrol yang memperoleh pembelajaran PBL konvensional. Jumlah sampel pada masing-masing kelas adalah 25 peserta didik, sehingga total sampel penelitian berjumlah 50 peserta didik. Penentuan kedua kelas tersebut dilakukan berdasarkan pertimbangan yang telah ditetapkan melalui teknik *purposive sampling* serta hasil diskusi dengan guru mata pelajaran Biologi.

3.4 Pengembangan Instrumen

Soesana et al. (2023) menekankan bahwa sebelum membuat instrumen penelitian, peneliti harus merumuskan definisi operasional untuk setiap variabel yang akan digunakan. Jika definisi konseptual variabel disusun berdasarkan sumber yang ada dalam tinjauan pustaka, maka bagian metodologi ini harus memperjelas definisi variabel dengan menggunakan instrumen yang diambil dari penelitian sebelumnya yang sudah terbukti valid.

1. Definisi Operasional Variabel

a. Efektivitas pembelajaran

Efektivitas *Deep-PBL* diukur melalui 5 indikator yang dikemukakan oleh Yusuf, (2018) dengan diukur melalui 3 instrumen yaitu:

- 1) Pengelolaan pelaksanaan pembelajaran, proses belajar mengajar komunikatif, dan aktivitas belajar diukur menggunakan lembar observasi.
- 2) Respons peserta didik diukur menggunakan angket yang mencakup tanggapan dan reaksi peserta didik terhadap pembelajaran meliputi antusias, perhatian, kepuasan, dan keingintahuan.
- 3) Hasil belajar diukur menggunakan soal *pretest* dan *posttest* pembelajaran dikatakan tuntas sesuai kriteria ketercapaian tujuan pembelajaran (KKTP).

b. Variabel Bebas: Model *Deep-PBL*

Model *Deep-PBL* dalam penelitian diartikan secara operasional sebagai model pembelajaran yang berfokus pada masalah yang diintegrasikan dengan prinsip-prinsip *Deep learning*, meliputi *mindful learning* (pembelajaran berkesadaran), *meaningful Learning* (pembelajaran bermakna), dan *joyful Learning* (pembelajaran menyenangkan). Model ini diimplementasikan melalui lima tahap sintaks, yaitu: (1) *Awareness*; (2) *Organization*; (3) *Investigation*; (4) *Presentation*; dan (5) *Presentation* (Putri et al., 2025). Keberlangsungan model ini diukur menggunakan lembar observasi yang diisi oleh observar selama proses pembelajaran berlangsung.

c. Variabel Bebas: Model PBL Konvensional

Model PBL konvensional dalam penelitian ini digambarkan sebagai pendekatan pembelajaran berbasis masalah yang dijalankan mengikuti urutan standar tanpa mengintegrasikan prinsip pembelajaran mendalam. Sintaks PBL konvensional yang digunakan dalam penelitian ini mengacu pada lima tahapan yang diadaptasi oleh Nurwidodo et al. (2025) berdasarkan Sezgin Selçuk (2015) dan Simamora et al. (2017), yaitu: (1) orientasi peserta didik terhadap masalah; (2) pengorganisasian proses belajar; (3) bimbingan dalam penyelidikan; (4) pengembangan dan penyajian hasil; serta (5) analisis dan evaluasi dari proses yang telah dilakukan.

d. Variabel Terikat: *Cognitive Engagement*

Dalam penelitian ini, *Cognitive Engagement* yang dimaksud mencakup kemampuan elaborasi, inferensi, dan metakognisi yang dianalisis berdasarkan kerangka ICAP (*Interactive, Constructive, Active, Passive*) (Chi & Wylie, 2014). Kerangka ICAP digunakan sebagai landasan untuk mengidentifikasi sejauh mana peserta didik terlibat secara kognitif dalam memproses informasi, mengonstruksi pemahaman, dan berinteraksi secara aktif, serta menerapkan strategi metakognitif melalui pemantauan dan refleksi terhadap proses berpikir selama pembelajaran berlangsung.

Cognitive Engagement dalam penelitian ini diukur melalui dua instrumen, yaitu:

- 1) Soal *pretest* dan *posttest*, yang digunakan untuk mengukur indikator elaborasi inferensi dan metakognisi. Soal disusun dalam bentuk uraian berbasis masalah yang menuntut peserta didik mengaitkan pengetahuan sebelumnya dengan konsep baru serta menarik kesimpulan secara logis dari kasus gangguan sistem reproduksi yang disajikan.

- 2) Angket *Cognitive Engagement*, yang diadaptasi dari instrumen *Survey of Cognitive and Collaborative Engagement (SCCE)*. Angket disusun dalam bentuk skala Likert yang memuat pernyataan terkait kesadaran belajar, pemantauan proses berpikir, dan evaluasi pemahaman peserta didik selama mengikuti pembelajaran.

e. Instrumen Penelitian

Menurut Sugiyono (2022), instrumen penelitian merupakan alat yang digunakan untuk mengukur fenomena alam maupun sosial yang diamati dalam suatu penelitian. Oleh karena itu, instrumen penelitian perlu dikembangkan secara sistematis agar sesuai dengan tujuan penelitian serta mampu menghasilkan data yang valid dan reliabel. Dalam penelitian ini, pengembangan instrumen dilakukan untuk mendukung pengumpulan data sesuai dengan variabel yang diteliti, yang meliputi tiga jenis instrumen, yaitu lembar observasi keterlaksanaan, instrumen *Cognitive Engagement*, pembelajaran, dan angket respons peserta didik..

1. Lembar Observasi *Cognitive Engagement*

Lembar observasi digunakan untuk mengamati keterlibatan kognitif peserta didik selama proses pembelajaran, khususnya pada indikator elaborasi, metakognisi, dan inferensi. Lembar observasi terdiri dari dua jenis, yaitu: Lembar observasi terdiri dari dua jenis, yaitu: (1) observasi pelaksanaan pembelajaran untuk memastikan kesesuaian dengan sintaks Model Deep-PBL, dan (2) observasi aktivitas peserta didik untuk menilai keterlibatan kognitif selama diskusi, penyelidikan, dan presentasi.. Lembar observasi ini digunakan untuk mengamati keterlibatan kognitif peserta didik selama proses pembelajaran, khususnya pada kegiatan diskusi, penyelidikan, presentasi, dan refleksi.. Observasi dilakukan menggunakan checklist dengan skala numerik (1–4) yang dilengkapi deskripsi pada setiap indikator (Soesana et al., 2023).

2. Tes *Cognitive Engagement* (Pretest dan Posttest)

Menurut Sugiyono (2022), teknik tes merupakan salah satu metode pengumpulan data yang digunakan untuk mengukur kemampuan atau pencapaian peserta didik melalui seperangkat butir soal. Dalam penelitian ini, instrumen tes disusun untuk mengukur *Cognitive Engagement* peserta didik pada materi sistem reproduksi manusia, khususnya pada indikator elaborasi inferensi dan metakognisi. Instrumen ini mengacu pada Capaian Pembelajaran (CP) dan Tujuan Pembelajaran (TP) Kurikulum Merdeka, terdiri dari 20 butir pilihan ganda dan 5 butir soal uraian berbasis HOTS. Soal dirancang khusus untuk mendorong peserta didik mengembangkan jawaban secara mandiri melalui proses elaborasi, sekaligus melatih kemampuan inferensi dalam menyelesaikan masalah kontekstual. Skala penilaian menggunakan rubrik penskoran (0–3) untuk soal uraian.

3. Angket Metakognisi dan Respons Peserta Didik

Instrumen angket digunakan untuk mengukur indikator metakognisi sebagai bagian dari *Cognitive Engagement*, serta respons peserta didik terhadap pembelajaran. Instrumen ini mencakup pernyataan tentang kesadaran belajar, strategi berpikir, dan refleksi pemahaman, serta diadaptasi dari dimensi respons peserta didik (Yusuf, 2018) yang meliputi antusias, perhatian, kepuasan, dan keingintahuan. Pengukuran menggunakan skala Likert empat poin (1 = Sangat Tidak Setuju hingga 4 = Sangat Setuju) tanpa pilihan netral.

3.5 Teknik Pengumpulan Data

Dua komponen utama yang menentukan kualitas data yang dihasilkan dari penelitian kuantitatif adalah teknik pengumpulan data dan instrumen penelitian. Teknik pengumpulan data mencakup pengumpulan data melalui berbagai metode seperti tes, kuesioner, observasi, dan dokumentasi (Soesana

et al., 2023). Dalam penelitian ini, teknik pengumpulan data yang digunakan adalah sebagai berikut.

a. Observasi

Observasi dilakukan selama proses pembelajaran oleh peneliti dan pendidik sebagai kolaborator untuk mengamati keterlibatan kognitif peserta didik. Soesana et al. (2023) menjelaskan bahwa observasi terstruktur menggunakan lembar checklist sebagai instrumen pengumpulan data. Dalam penelitian ini, observasi digunakan untuk menilai *Cognitive Engagement* peserta didik yang tampak melalui aktivitas elaborasi, metakognisi, dan inferensi selama diskusi, penyelidikan, dan presentasi. Observasi dilakukan menggunakan dua jenis lembar observasi. Lembar observasi keterlaksanaan pembelajaran menggunakan format checklist dengan pilihan "Ya" (skor 1) dan "Tidak" (skor 0), sedangkan lembar observasi keterlibatan kognitif menggunakan skala penilaian 1–4 berdasarkan indikator elaborasi, inferensi, dan metakognisi.

b. Tes

Teknik tes menggunakan seperangkat butir soal sebagai instrumen dan merupakan salah satu teknik utama dalam pengumpulan data kuantitatif (Soesana et al., 2023). Tes tertulis dalam penelitian ini digunakan untuk mengukur *Cognitive Engagement* peserta didik, khususnya pada indikator elaborasi dan inferensi. Tes dilaksanakan dalam dua tahap, yaitu *pretest* sebelum penelitian dan *posttest* setelah penelitian. Soal disusun dalam bentuk pilihan ganda dan uraian berbasis HOTS yang mendorong peserta didik mengembangkan jawaban sendiri serta menyelesaikan masalah kontekstual. Pelaksanaan tes dilakukan secara daring menggunakan Google Form yang dibagikan kepada peserta didik sesuai jadwal yang telah ditentukan.

c. Survei Kuesioner (Angket)

Angket digunakan untuk mengukur *Cognitive Engagement* pada indikator metakognisi serta respons peserta didik terhadap pembelajaran. Soesana et al. (2023) menyatakan bahwa teknik kuesioner merupakan metode pengumpulan data kuantitatif yang dapat dilakukan secara daring maupun cetak. Angket diberikan setelah proses pembelajaran (*posttest*) dalam bentuk Google Form. Pernyataan dalam angket mencerminkan kesadaran belajar, strategi berpikir, dan refleksi pemahaman peserta didik selama mengikuti pembelajaran. Pengisian angket dilakukan secara mandiri berdasarkan pengalaman belajar masing-masing peserta didik.

d. Dokumentasi

Menurut Sugiyono (2022), dokumentasi merupakan teknik pengumpulan data melalui dokumen, arsip, atau rekaman yang berkaitan dengan penelitian. Dalam penelitian ini, dokumentasi digunakan untuk merekam seluruh rangkaian kegiatan penelitian sebagai bukti pelaksanaan penelitian, meliputi pelaksanaan pretest, proses pembelajaran menggunakan model *Deep-Problem Based Learning* (Deep-PBL) dan PBL konvensional, pengerjaan LKPD oleh peserta didik, pelaksanaan posttest, serta dokumentasi berupa foto kegiatan pembelajaran dan dokumen pendukung lainnya..

3.6 Uji Instrumen Penelitian

1. Uji Validitas

Soesana et al. (2023) membedakan validitas instrumen menjadi tiga jenis, yaitu:

- a. *Content Validity* (validitas isi), yaitu ketepatan butir-butir pernyataan/soal yang memuat keseluruhan indikator yang akan diukur
- b. *Criterion Validity* (validitas kriteria), yaitu ketepatan alat ukur yang dibandingkan dengan instrumen yang telah digunakan secara luas.

c. *Construct Validity* (validitas konstruk), yaitu keterkaitan instrumen dengan kerangka konseptual keilmuan topik penelitian.

Dalam penelitian ini, uji validitas dilakukan melalui dua tahap: pertama, validitas isi melalui penilaian dua ahli (*expert judgment*); kedua, validitas empiris menggunakan rumus Korelasi *Product Moment Pearson*. Soesana et al. (2023) menegaskan bahwa butir soal dinyatakan valid apabila nilai rhitung > rtabel pada taraf signifikansi 5%.

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

Keterangan

r_{xy} = koefisien korelasi item-total

N = jumlah peserta uji coba

X = skor butir soal

Y = skor total

Kriteria keputusan:

Dikatakan valid jika $t_{hitung} > r_{tabel}$ pada taraf signifikan 5%

Tabel 3. 3 Tingkatan validitas (Sugiyono, 2016)

Nilai Koefisien Korelasi (r)	Tingkat Validitas
0,80 – 1,00	Sangat tinggi
0,60 – 0,79	Tinggi
0,40 – 0,59	Cukup
0,20 – 0,39	Rendah
0,00 – 0,19	Sangat rendah / tidak valid

Dalam penelitian ini, uji validitas dilakukan melalui dua tahap, yaitu validitas isi dan validitas empiris. Validitas isi dilakukan terlebih dahulu melalui *expert judgment* oleh dua orang validator, yaitu Marshanti Lisbania Gratia, M.Pd. dan Mamat Arohman, M.Pd. Penilaian dilakukan terhadap kesesuaian butir soal dengan tujuan pembelajaran, indikator soal, jenjang

kognitif (C2–C6), serta dimensi *Cognitive Engagement* yang diukur, yaitu elaborasi, metakognisi, dan inferensi. Hasil validasi menunjukkan bahwa seluruh butir soal telah sesuai dengan aspek yang diukur sehingga dinyatakan layak untuk dilanjutkan pada tahap uji coba empiris.

Setelah memperoleh persetujuan dari validator ahli, instrumen diujicobakan kepada peserta didik di luar sampel penelitian untuk mengetahui validitas empiris setiap butir soal. Uji coba instrumen dilaksanakan di SMA Negeri 1 Ciparay pada peserta didik kelas XI yang telah mempelajari materi Sistem Reproduksi dengan jumlah peserta uji coba sebanyak 30 peserta didik. Pemilihan peserta didik yang telah memperoleh materi tersebut bertujuan agar mereka memiliki pengetahuan yang memadai dalam menjawab soal, sehingga hasil pengujian validitas dan reliabilitas instrumen dapat menggambarkan kualitas instrumen secara objektif.

Kemudian validitas empiris dianalisis menggunakan korelasi *Product Moment Pearson* dengan bantuan program IBM SPSS Statistics. Hasil uji validitas instrumen soal esai disajikan pada tabel berikut.

Tabel 3. 4 Hasil Uji Validitas Soal

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Soal 1	22,77	239,151	0,597	0,935
Soal 2	24,07	243,099	0,749	0,929
Soal 3	24,07	219,513	0,800	0,925
Soal 4	24,57	252,392	0,600	0,935
Soal 5	23,97	241,895	0,781	0,928
Soal 6	25,60	253,559	0,623	0,934
Soal 7	23,57	224,875	0,796	0,925
Soal 8	24,23	211,013	0,845	0,923
Soal 9	23,67	216,368	0,844	0,922
Soal 10	23,80	213,614	0,865	0,921

Berdasarkan Tabel 3.4 hasil uji validitas, seluruh item soal esai yang digunakan dalam penelitian ini dinyatakan valid. Nilai Corrected Item-Total Correlation pada setiap butir soal berada dalam rentang 0,597 hingga 0,865,

yang seluruhnya telah melampaui batas minimum koefisien validitas, yaitu 0,30. Nilai validitas tertinggi diperoleh oleh Soal 10 dengan nilai 0,865, sedangkan nilai terendah terdapat pada Soal 1 dengan nilai 0,597. Selain itu, nilai Cronbach's Alpha if Item Deleted pada setiap butir soal berada pada rentang 0,921 hingga 0,935. Hal tersebut menunjukkan bahwa setiap butir soal memiliki kontribusi yang baik terhadap konsistensi internal instrumen secara keseluruhan. Dengan demikian, seluruh 10 butir soal esai dinyatakan memenuhi kriteria validitas dan layak digunakan sebagai instrumen pengumpulan data dalam penelitian ini.

Variasi tingkat validitas antarbutir, di mana sebagian item berada pada kategori sangat tinggi dan sebagian lain pada kategori cukup menurut kriteria Sugiyono (2016), juga dapat dijelaskan melalui kompleksitas proses kognitif yang dituntut oleh masing-masing soal. Butir soal yang menuntut proses berpikir lebih kompleks, seperti menganalisis kasus, mengevaluasi kredibilitas sumber, atau mensintesis informasi dari berbagai sumber, cenderung mengaktifkan lebih banyak dimensi *Cognitive Engagement* secara bersamaan (elaborasi, inferensi, dan metakognisi), sehingga korelasinya dengan skor total instrumen menjadi lebih kuat. Sebaliknya, butir soal yang mengukur aspek pengetahuan yang lebih spesifik dan tunggal secara alami memiliki korelasi yang sedikit lebih rendah dibanding butir lain, meskipun tetap berada di atas batas minimum validitas yang dipersyaratkan. Variasi semacam ini wajar terjadi pada instrumen yang mengukur konstruk multidimensi seperti *Cognitive Engagement*, dan justru menunjukkan bahwa instrumen mampu menangkap perbedaan kompleksitas kognitif antarbutir secara proporsional, bukan menjadi indikasi kelemahan instrumen.

2. Uji Reliabilitas

Soesana et al. (2023) menjelaskan bahwa reliabilitas merupakan nilai kepercayaan suatu hasil pengukuran. Uji reliabilitas dilakukan untuk mendapatkan instrumen yang dapat dipercaya keabsahannya sehingga

menghasilkan data yang benar-benar relevan dengan tujuan penelitian. Dalam penelitian ini, uji reliabilitas instrumen tes kognitif pilihan ganda menggunakan formula KR-20 (Kuder-Richardson), sedangkan untuk soal uraian dan angket menggunakan koefisien Alpha Cronbach. Berdasarkan kriteria Guilford (1956), instrumen dinyatakan reliabel apabila koefisien reliabilitas $\geq 0,70$.

Setelah seluruh butir soal dinyatakan valid, tahap selanjutnya adalah menguji reliabilitas instrumen. Uji reliabilitas dilakukan untuk mengetahui tingkat konsistensi instrumen dalam memberikan hasil pengukuran apabila digunakan pada kondisi yang relatif sama. Hasil uji reliabilitas instrumen disajikan pada tabel berikut;

Tabel 3. 5 Hasil Uji Realibitas Soal

Cronbach's Alpha	Jumlah Item (N of Items)	Kategori
0,935	10	Sangat Reliabel

Berdasarkan Tabel 4.3 hasil uji reliabilitas instrumen, diketahui bahwa seluruh data yang diuji berjumlah 30 responden dan seluruhnya dinyatakan valid (100%), sehingga tidak terdapat data yang dikeluarkan (*excluded*) dari proses analisis. Hasil pengujian reliabilitas menggunakan koefisien Cronbach's Alpha menunjukkan nilai sebesar 0,935 dengan jumlah item sebanyak 10 butir soal. Nilai tersebut berada pada rentang 0,90–1,00, sehingga termasuk dalam kategori sangat reliabel.

Tingginya koefisien reliabilitas instrumen juga erat kaitannya dengan dua faktor yang saling memperkuat, yaitu homogenitas konstruk yang diukur dan kualitas konstruksi rubrik penilaian. Soesana et al. (2023) menjelaskan bahwa reliabilitas pada dasarnya merupakan nilai kepercayaan terhadap hasil pengukuran, yang akan tinggi apabila instrumen secara konsisten mengukur satu konstruk yang sama pada setiap butirnya. Karena seluruh butir soal dalam penelitian ini dirancang untuk mengukur konstruk yang sama, yaitu

kemampuan kognitif peserta didik pada materi sistem reproduksi, maka secara teoretis instrumen ini memang berpotensi menghasilkan koefisien reliabilitas yang tinggi. Selain itu, kualitas rubrik penilaian turut menjadi faktor penentu, sebagaimana disoroti secara khusus oleh Marshanti Lisbania Gratia, M.Pd., yang menilai bahwa pembeda antarskor pada rubrik sangat jelas sehingga dapat meminimalkan subjektivitas penilai. Rubrik dengan kriteria penilaian yang jelas dan tidak ambigu akan menghasilkan skor yang lebih konsisten antarpenilai maupun antarwaktu pengukuran, sehingga varians kesalahan pengukuran (measurement error) dapat diminimalkan. Inilah yang menjelaskan mengapa koefisien Alpha Cronbach yang diperoleh jauh melampaui ambang batas minimum reliabilitas menurut Guilford (1956): tingginya nilai tersebut bukan semata hasil statistik yang berdiri sendiri, melainkan cerminan langsung dari kualitas desain rubrik yang telah divalidasi secara teoretis sebelum digunakan.

Secara keseluruhan, keselarasan antara hasil validasi konten oleh dua dosen ahli dan hasil uji validitas serta reliabilitas secara statistik memperlihatkan adanya proses triangulasi yang saling menguatkan. Hasil empiris yang diperoleh bukan sekadar angka yang berdiri sendiri, melainkan konfirmasi nyata atas apa yang secara teoretis telah diprediksi oleh kedua dosen ahli sejak tahap pengembangan instrumen. Keselarasan ini memberikan keyakinan yang lebih kuat bahwa instrumen soal esai dalam penelitian ini benar-benar mengukur konstruk yang dimaksud, bukan sekadar menghasilkan koefisien tinggi akibat kebetulan statistik, sehingga instrumen ini dapat dipertanggungjawabkan secara metodologis untuk digunakan sebagai alat pengumpulan data dalam penelitian ini.

3.7 Teknis Analisis Data

Analisis data adalah proses pengolahan data yang dilakukan untuk menjawab perumusan masalah dan menguji hipotesis penelitian. Menurut Soesana et al. (2023), analisis data dalam penelitian kuantitatif dilakukan secara sistematis melalui tahap deskripsi data dan pengujian statistik sesuai dengan tujuan penelitian. Analisis data dalam penelitian ini meliputi analisis data observasi, analisis data tes *Cognitive Engagement*, uji normalitas, uji homogenitas, uji hipotesis menggunakan Independent Sample t-test, perhitungan *N-Gain*, dan Effect Size, dan analisis data angket respons peserta didik

1. Analisis Lembar Observasi

Penelitian ini menggunakan lembar observasi yang dinilai oleh observer yaitu *pendidik* kelas. Lembar observasi digunakan untuk memperoleh data keterlaksanaan pelaksanaan pembelajaran dan aktivitas belajar peserta didik. Analisis lembar observasi menggunakan analisis deskriptif persentase dengan rumus:

$$P = (\sum \text{Skor Perolehan} / \sum \text{Skor Maksimal}) \times 100\%$$

Keterangan:

- P = persentase keterlaksanaan / aktivitas
- $\sum$ Skor Perolehan = jumlah skor yang diperoleh dari hasil observasi
- $\sum$ Skor Maksimal = jumlah skor maksimal seluruh indikator observasi

Hasil persentase kemudian diinterpretasikan berdasarkan kriteria berikut:

Tabel 3. 6 Kriteria Penilaian Lembar Observasi

Persentase (%)	Kriteria
81 – 100	Sangat Baik
61 – 80	Baik
41 – 60	Cukup
21 – 40	Kurang
0 – 20	Sangat Kurang

Data hasil observasi dideskripsikan secara naratif untuk menggambarkan tingkat keterlaksanaan pembelajaran dan aktivitas peserta didik selama proses pembelajaran berlangsung pada masing-masing pertemuan.

2. Uji Normalitas

Setelah instrumen dinyatakan valid dan reliabel serta data *pretest* dan *posttest* berhasil dikumpulkan dari kelompok eksperimen dan kelompok kontrol, tahap selanjutnya adalah melakukan uji prasyarat analisis. Fadillah dan Mahyudin (2026) menegaskan bahwa uji normalitas merupakan tahap fundamental dalam analisis data statistik, terutama ketika hendak menggunakan metode parametrik seperti uji t, ANOVA, maupun regresi linear. Metode-metode ini pada dasarnya mengasumsikan bahwa data yang dianalisis memiliki distribusi normal; ketika asumsi ini dilanggar, hasil analisis rentan terhadap bias dan dapat menurunkan validitas kesimpulan yang diambil.

Jika data tidak memenuhi asumsi ini, maka hasil analisis yang diperoleh dapat menjadi tidak valid. Dalam penelitian ini, uji normalitas dilakukan terhadap data *pretest* dan *posttest* pada kedua kelompok menggunakan uji Shapiro-Wilk. Pemilihan uji Shapiro-Wilk didasarkan pada pertimbangan bahwa uji ini dianggap sebagai metode paling sensitif dan efektif untuk mendeteksi ketidaknormalan dalam distribusi data, terutama pada sampel berukuran kecil (Razali & Wah, 2011, dalam Fadillah & Mahyudin, 2026).

Kriteria pengambilan keputusan dalam uji normalitas adalah apabila nilai signifikansi (p) $> 0,05$, maka data dinyatakan berdistribusi normal. Sebaliknya, apabila nilai signifikansi (p) $\leq 0,05$, maka data dinyatakan tidak berdistribusi normal (Fadillah & Mahyudin, 2026; Usmadi, 2020). Dalam penelitian ini, uji normalitas dilakukan terhadap data *pretest* dan *posttest* menggunakan uji Shapiro-Wilk. Pengujian normalitas dalam penelitian ini dilakukan dengan bantuan perangkat lunak SPSS, yang menurut Pallant (2020, dalam Fadillah & Mahyudin, 2026) mampu menampilkan hasil uji Kolmogorov-Smirnov dan Shapiro-Wilk

secara bersamaan, dilengkapi grafik boxplot dan normal Q-Q plot untuk evaluasi normalitas secara numerik maupun visual.

Tabel 3. 7 Kriteria Keputusan Uji Normalitas

Nilai Signifikansi (p)	Keputusan
$p > 0,05$	Data berdistribusi normal
$p \leq 0,05$	Data tidak berdistribusi normal

3. Uji Homogenitas

Setelah uji normalitas, dilakukan uji homogenitas varians sebagai prasyarat kedua sebelum penggunaan statistik parametrik. Fadillah dan Mahyudin (2026) menjelaskan bahwa uji homogenitas merupakan prosedur statistik yang bertujuan untuk menentukan apakah dua atau lebih kelompok data memiliki varians yang serupa atau homogen. Homogenitas varians merupakan prasyarat esensial dalam penggunaan analisis varians (ANOVA) ataupun uji t untuk dua sampel independen. Jika syarat ini tidak terpenuhi, maka risiko terjadinya kekeliruan dalam interpretasi hasil analisis menjadi lebih tinggi.

Menurut Fadillah dan Mahyudin (2026), urutan pelaksanaan uji sangat penting, umumnya uji normalitas dilakukan terlebih dahulu sebelum uji homogenitas. Hal ini disebabkan oleh fakta bahwa beberapa metode uji homogenitas, seperti Levene dan Brown-Forsythe, merupakan metode parametrik yang mengharuskan data berdistribusi normal. Oleh karena itu, dalam penelitian ini uji homogenitas baru dilaksanakan setelah data terbukti memenuhi asumsi normalitas.

Dalam penelitian ini, uji homogenitas dilakukan menggunakan *Levene's Test*. Pemilihan metode ini didasarkan pada keunggulannya sebagai uji yang dikenal luas karena ketahanannya, bahkan ketika data tidak sepenuhnya memenuhi asumsi normalitas, karena metode ini menghitung deviasi absolut nilai data dari rata-rata atau median kelompok kemudian membandingkan hasilnya antar kelompok (Fadillah & Mahyudin, 2026). Selain itu, Pallant (2020, dalam

Fadillah & Mahyudin, 2026) menyatakan bahwa apabila varians data tidak homogen, hasil yang diperoleh dari analisis komparasi berpotensi menyesatkan. Oleh sebab itu, jika nilai signifikansi dari uji Levene kurang dari 0,05, maka asumsi bahwa varians antar kelompok adalah sama tidak dapat dipertahankan, dan peneliti disarankan beralih ke metode alternatif seperti uji nonparametrik.

Kriteria pengambilan keputusan dalam uji homogenitas adalah apabila nilai signifikansi (p) $> 0,05$, maka varians antar kelompok dinyatakan homogen sehingga uji parametrik dapat dilanjutkan. Sebaliknya, apabila nilai signifikansi (p) $\leq 0,05$, maka varians tidak homogen dan peneliti perlu memilih metode analisis alternatif seperti Welch ANOVA atau uji nonparametrik seperti Mann-Whitney U (Fadillah & Mahyudin, 2026; Nuryadi et al., 2017). Pengujian homogenitas dalam penelitian ini dilakukan dengan bantuan perangkat lunak SPSS.

Tabel 3. 8 Kriteria Keputusan Uji Homogenitas

Nilai Signifikansi (p)	Keputusan
$p > 0,05$	Varians homogen (asumsi terpenuhi)
$p \leq 0,05$	Varians tidak homogen

4. Uji Hipotesis (Uji-t Dua Sampel Independen)

Uji hipotesis digunakan untuk mengetahui apakah terdapat perbedaan hasil belajar yang signifikan antara kelas eksperimen dan kelas kontrol. Uji hipotesis bertujuan untuk membuktikan kebenaran dari hipotesis yang diajukan dalam penelitian. Uji hipotesis yang digunakan adalah Uji-t 2 Sampel Independen, yaitu uji-t untuk dua sampel yang tidak saling terkait atau tidak berhubungan sama sekali (Sugiyono, 2017).

Dalam penelitian ini, uji hipotesis dilakukan menggunakan *Independent Sample t-test* untuk mengetahui perbedaan *Cognitive Engagement* antara kelas eksperimen dan kelas kontrol berdasarkan data yang telah memenuhi uji prasyarat. Rumus uji-t independen adalah sebagai berikut:

$$t = (\bar{X}_1 - \bar{X}_2) / \sqrt{[s^2p(1/n_1 + 1/n_2)]}$$

di mana: $s^2p = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2] / (n_1 + n_2 - 2)$

Keterangan:

- $\bar{X}_1$ = rata-rata skor kelas eksperimen
- $\bar{X}_2$ = rata-rata skor kelas kontrol
- s^2p = varians gabungan (pooled variance)
- n_1 = jumlah peserta didik kelas eksperimen
- n_2 = jumlah peserta didik kelas kontrol

Hipotesis yang diuji dalam penelitian ini adalah:

- H_0 : Tidak terdapat perbedaan signifikan *Cognitive Engagement* antara kelas eksperimen dan kelas kontrol
- H_1 : Terdapat perbedaan signifikan *Cognitive Engagement* antara kelas eksperimen dan kelas kontrol

Tabel 3. 9 Kriteria Keputusan Uji Hipotesis

Kondisi	Keputusan
$R_{hitung} > R_{tabel}$ atau $sig. < 0,05$	H_0 ditolak, H_1 diterima (terdapat perbedaan signifikan)
$R_{hitung} \leq R_{tabel}$ atau $sig. \geq 0,05$	H_0 diterima (tidak terdapat perbedaan signifikan)

5. Uji *N-Gain* (Peningkatan *Cognitive Engagement*)

N-Gain digunakan untuk mengukur besarnya peningkatan *Cognitive Engagement* dari *pretest* ke *posttest* pada masing-masing kelompok. Penggunaan *N-Gain* bertujuan untuk mengetahui seberapa besar efektivitas perlakuan yang diberikan dalam meningkatkan kemampuan peserta didik. Rumus normal gain adalah:

$$\text{Normal Gain} = \frac{\text{Skor Post Test} - \text{Skor Pre Test}}{\text{Skor Ideal} - \text{Skor Pre Test}}$$

Tabel 3. 10 Kriteria Peningkatan Hasil Belajar (Hake, 1988)

Rata-rata	Kriteria
$g > 0,7$	Tinggi
$0,3 \leq g \leq 0,7$	Sedang
$0 < g < 0,3$	Rendah

6. *Effect Size*

Untuk memperkuat hasil uji-t dan nilai *N-Gain*, dilakukan analisis *effect size* guna mengukur seberapa besar pengaruh penerapan model *Deep-PBL* terhadap peningkatan *Cognitive Engagement* peserta didik. *Effect size* merupakan ukuran kuantitatif yang menggambarkan besarnya pengaruh suatu perlakuan secara praktis, melampaui sekadar signifikansi statistik (Cohen, 1988).

Dalam penelitian ini, perhitungan *effect size* menggunakan formula Cohen's d berdasarkan data pretest-posttest dua kelompok (Becker & Park, 2011), sebagai berikut:

$$ES = [(\bar{X}_{postE} - \bar{X}_{preE}) - (\bar{X}_{postC} - \bar{X}_{preC})] / [(SD_{preE} + SD_{preC} + SD_{postC}) / 3]$$

Keterangan:

- $\bar{X}_{postE}$ = rata-rata posttest kelompok eksperimen
- $\bar{X}_{preE}$ = rata-rata pretest kelompok eksperimen
- $\bar{X}_{postC}$ = rata-rata posttest kelompok kontrol
- $\bar{X}_{preC}$ = rata-rata pretest kelompok kontrol
- SD_{preE} = standar deviasi pretest kelompok eksperimen
- SD_{preC} = standar deviasi pretest kelompok kontrol
- SD_{postC} = standar deviasi posttest kelompok kontrol

Tabel 3. 11 Kriteria Effect size (Cohen, 1988)

Nilai <i>Effect size</i> (ES)	Kategori
$0 < ES \leq 0,2$	Rendah
$0,2 < ES < 0,8$	Sedang
$ES \geq 0,8$	Tinggi

Analisis *effect size* melengkapi uji-t dengan memberikan informasi tentang seberapa besar perbedaan antara kelompok eksperimen dan kontrol dalam satuan standar deviasi. Nilai *effect size* yang tinggi ($ES \geq 0,8$) mengindikasikan bahwa perlakuan memberikan dampak yang besar dan bermakna secara praktis terhadap peningkatan *Cognitive Engagement* peserta didik (Ramadhani et al., 2021).

7. Analisis Kuesioner (Angket)

Analisis kuesioner/angket dalam penelitian ini meliputi angket metakognisi dan angket respons peserta didik. Analisis dilakukan dengan menghitung persentase skor setiap indikator menggunakan rumus deskriptif persentase sebagai berikut:

$$P = (\sum \text{Skor Responden} / \sum \text{Skor Ideal}) \times 100\%$$

Keterangan:

- P = persentase skor kuesioner
- $\sum$ Skor Responden = jumlah skor keseluruhan jawaban responden
- $\sum$ Skor Ideal = jumlah skor tertinggi (skor maksimum $\times$ jumlah responden)

Hasil persentase diinterpretasikan menggunakan kriteria yang sama dengan analisis lembar observasi (Tabel 3.6), kemudian dideskripsikan secara naratif untuk menggambarkan tingkat metakognisi peserta didik serta respons mereka terhadap penerapan model *Deep-PBL* dalam pembelajaran

3.8 Analisis Efektivitas Pembelajaran

Efektivitas model *Deep-PBL* diukur berdasarkan lima indikator menurut Yusuf (2018), yaitu pengelolaan pelaksanaan pembelajaran, proses pembelajaran yang komunikatif, respons peserta didik, aktivitas belajar, dan *Cognitive Engagement* peserta didik. Analisis efektivitas dilakukan dengan mengintegrasikan data dari lembar observasi, angket, dan tes. Kriteria keefektifan ditetapkan berdasarkan pencapaian Kriteria Ketercapaian Tujuan Pembelajaran (KKTP) $\geq 75\%$, dengan ketentuan bahwa nilai *N-Gain Cognitive Engagement* pada kelas eksperimen lebih tinggi secara signifikan dibandingkan kelas kontrol.

Tabel 3. 12 Ringkasan Variabel, Instrumen, dan Teknik Analisis Data

No.	Variabel	Instrumen	Sumber Data	Teknik Analisis
1	<i>Cognitive Engagement</i> (Elaborasi & Inferensi)	Soal Pretest–Posttest	Peserta didik	<i>N-Gain</i> , t-test / Mann-Whitney dan Effect Size
2	<i>Cognitive Engagement</i> (Metakognisi)	Angket Metakognisi	Peserta didik	Deskriptif Persentase
3	Keterlaksanaan Pembelajaran	Lembar Observasi Pembelajaran	<i>Pendidik</i> sebagai Observer	Deskriptif Persentase
4	Aktivitas Belajar (<i>Cognitive Engagement</i>)	Lembar Observasi Peserta Didik	Peserta didik	Deskriptif Persentase
5	Respons Peserta Didik	Angket Respons	Peserta didik	Deskriptif Persentase