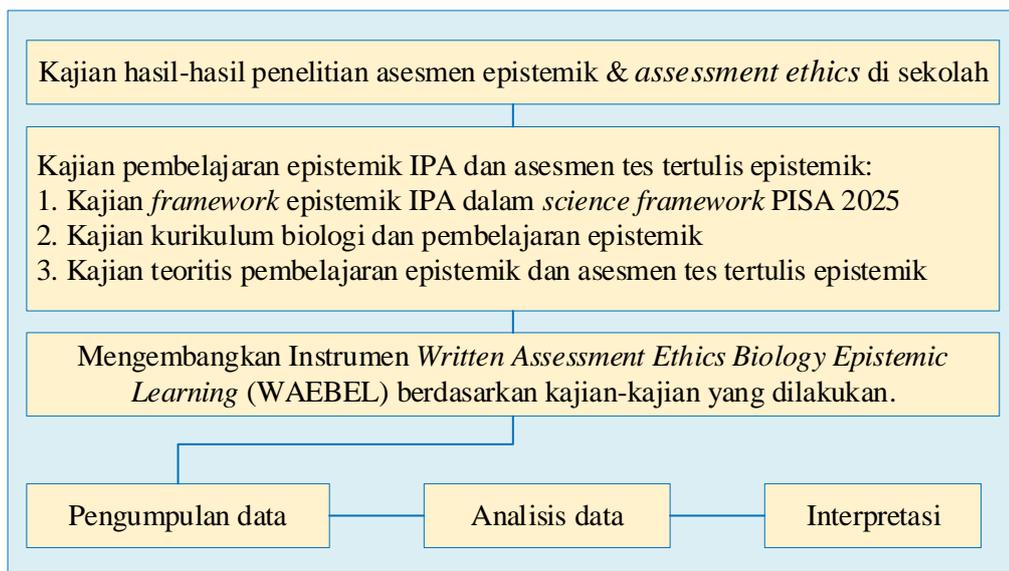


BAB III METODE PENELITIAN

3.1. Metode dan Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode penelitian deskriptif. Metode penelitian deskriptif dipilih karena kesesuaiannya dengan tujuan penelitian. Penelitian ini bertujuan untuk menganalisis praktik *assessment ethics* tes tertulis epistemik IPA pembelajaran biologi di SMA. Penelitian deskriptif tidak melakukan manipulasi dan intervensi terhadap variabel-variabel yang diteliti untuk mendapatkan gambaran suatu kondisi apa adanya (Mcmillan & Schumacher, 2001). Garis besar langkah yang dilakukan dapat dilihat pada Gambar 3.1.



Gambar 3.1. Implementasi Desain Penelitian Deskriptif dalam Penelitian

Keputusan untuk menggunakan metode penelitian deskriptif didasari oleh beberapa pertimbangan. Pertimbangan pertama, metode deskriptif dinilai sangat cocok untuk memberikan informasi otentik/apa adanya tentang praktik *assessment ethics* tes tertulis epistemik IPA di sekolah. Pertimbangan lainnya adalah metode deskriptif dengan pendekatan kuantitatif memungkinkan pengukuran indikator-indikator variabel penelitian, sehingga diperoleh gambaran yang jelas terkait variabel penelitian.

Penelitian ini mengembangkan instrumen WAEBEL berdasarkan studi *desktop research* untuk mengumpulkan data sekunder yang relevan dengan tujuan

penelitian. Data sekunder tersebut adalah hasil kajian *framework* epistemik dalam PISA 2025, hasil kajian kurikulum biologi, hasil kajian teoritis pembelajaran epistemik dan asesmen epistemik, serta hasil kajian *assessment ethics*.

Hasil-hasil kajian *framework* epistemik dalam PISA 2025, kajian kurikulum biologi, kajian pembelajaran epistemik dan asesmen epistemik digunakan untuk memetakan pembelajaran epistemik dan asesmen yang sesuai. Penelitian ini merangkum hasil-hasil kajian ke dalam dua aspek, yaitu pelaksanaan pembelajaran biologi dan asesmen tes tertulis dikaitkan dengan pembelajaran epistemik. Tabel 3.1. menunjukkan rincian masing-masing aspek.

Tabel 3.1. Hasil Kajian Kerangka Kerja IPA PISA 2025

No.	Aspek	No.	Cakupan Aspek
1.	Pelaksanaan pembelajaran epistemik	1.	Menerapkan model untuk memfasilitasi prediksi dan penjelasan
		2.	Memanfaatkan data dan bukti ilmiah dalam pembelajaran Biologi
		3.	Menggunakan pengetahuan ilmiah untuk pengambilan keputusan dan tindakan
		4.	Berkolaborasi dan bekerja sama dalam penyelidikan ilmiah
2.	Asesmen tes tertulis dikaitkan dengan pembelajaran epistemik	5.	Asesmen untuk menilai kemampuan menerapkan model untuk memfasilitasi prediksi dan penjelasan
		6.	Asesmen untuk menilai kemampuan memanfaatkan data dan bukti ilmiah dalam pembelajaran biologi
		7.	Asesmen untuk menilai kemampuan menggunakan pengetahuan ilmiah untuk pengambilan keputusan dan tindakan
		8.	Asesmen terhadap kemampuan kolaborasi dan bekerja sama dalam penyelidikan ilmiah

Kajian *assessment ethics* dilakukan dengan meninjau pelanggaran yang terjadi pada asesmen pembelajaran dan standar asesmen yang seharusnya dipatuhi oleh guru dalam menyelenggarakan asesmen di kelas. Kedua aspek tersebut diidentifikasi dan dianalisis agar sesuai dengan konteks praktik asesmen tes tertulis.

Penelitian dan literatur yang membahas tentang pelanggaran dan standar etika asesmen masih terbatas. Sangat sedikit jumlah hasil-hasil riset dan literatur yang relevan dengan praktik etika asesmen tes tertulis. Namun, penelitian ini telah merangkum beberapa pelanggaran dan standar etika asesmen tes tertulis. Tabel 3.2. menunjukkan pelanggaran *assessment ethics* tes tertulis.

Tabel 3.2. Hasil Kajian Pelanggaran Assessment Ethics Tes Tertulis

No.	Pelanggaran Assessment Ethics	Sumber
41.	Metode penilaian tidak sesuai dengan latar belakang dan pengalaman peserta didik sebelumnya.	(Webber & Lupart, 2012)

No.	Pelanggaran Assessment Ethics	Sumber
42.	Konten dan bahasa yang digunakan dalam tes dianggap sensitif, seksis, atau menyinggung	(Webber & Lupart, 2012)
43.	Soal yang digunakan dalam tes tertulis mengandung unsur bias gender	(Wulan, 2018; Zlatkin-Troitschanskaia <i>et al.</i> , 2019)
44.	Soal yang digunakan dalam tes tertulis bertentangan dengan norma masyarakat setempat	(Wulan, 2018)
45.	Soal yang digunakan dalam tes tertulis bertentangan dengan norma agama	(Wulan, 2018)
46.	Soal yang digunakan dalam tes tertulis bertentangan dengan budaya peserta didik tertentu	(Wulan, 2018)
47.	Soal yang digunakan dalam tes tertulis mengganggu keadaan psikis peserta didik tertentu	(Wulan, 2018)
48.	Soal yang digunakan dalam tes tertulis menguji “kompetensi asing” yang dapat memungkinkan terjadinya bias	(Wulan, 2018; Zlatkin-Troitschanskaia <i>et al.</i> , 2019)
49.	Tes tertulis dilaksanakan secara tidak <i>fair</i> dan merugikan peserta didik sebagai peserta tes.	(Wulan, 2018)
50.	Pelaksanaan tes tertulis kurang mengakomodasi latar belakang sosial-ekonomi peserta didik	(Wulan, 2018)
51.	Pelaksanaan tes tertulis tidak nyaman dikerjakan oleh peserta didik	(Wulan, 2018)
52.	Menerapkan sistem hukuman melalui pengurangan nilai atas tindakan yang tidak relevan dalam pelaksanaan tes tertulis	(Wulan, 2018)
53.	Menggunakan instrumen dan perangkat yang tidak dapat diakses oleh semua peserta didik	(Zlatkin-Troitschanskaia <i>et al.</i> , 2019)
54.	Menuntut peserta didik untuk meningkatkan kemampuan tanpa adanya umpan balik	(Wulan, 2018)
55.	Meluluskan peserta didik yang belum mencapai kriteria kelulusan	(Wulan, 2018)
56.	Mengambil keputusan atas dasar pertimbangan yang bersifat subjektif atau pertimbangan lain yang tidak relevan	(Wulan, 2018)
57.	Mengambil keputusan atas informasi yang tidak lengkap	(Wulan, 2018)
58.	Mengkomunikasikan hasil asesmen dengan cara menjatuhkan harga diri peserta didik	(Wulan, 2018)
59.	Membanding-bandingkan hasil tes tertulis peserta didik di depan kelas	(Wulan, 2018)
60.	Memberikan tanggung jawab kepada peserta didik dengan nilai baik untuk bekerja lebih dari peserta didik lainnya	(Wulan, 2018)
61.	Mempertontonkan hasil asesmen tes tertulis dengan kualitas rendah sebagai contoh buruk tanpa seizin pemiliknya	(Wulan, 2018)
62.	Tidak mengakomodir penafsiran skor tes secara individual secara detail	(Zlatkin-Troitschanskaia <i>et al.</i> , 2019)

Aspek standar *assessment ethics* juga diperlukan untuk menjawab pertanyaan penelitian yang dilakukan. Tabel 3.3. menunjukkan daftar standar *assessment ethics* tes tertulis yang sepatutnya dijadikan guru sebagai acuan pelaksanaan asesmen tes tertulis di kelas.

Tabel 3.3. Hasil Kajian Standar Assessment Ethics Tes Tertulis

No.	Standar Assessment Ethics	Sumber
1.	Menjelaskan cara untuk menghindari praktik penilaian yang tidak etis	(Wulan, 2018)
2.	Memberikan argumentasi terhadap praktik-praktik penilaian yang dinilai tidak etis	(Wulan, 2018)
3.	Mengembangkan/mengadaptasi tes tertulis yang sesuai dengan peserta didik dilengkapi dengan peninjauan dan uji coba sebelum digunakan	(Permendikbud-ristek No. 21 tahun 2022; AERA <i>et al.</i> , 2014)
4.	Spesifikasi tes harus menjelaskan tujuan tes, definisi konstruk atau domain yang diukur, populasi peserta ujian yang dituju, dan interpretasi untuk tujuan penggunaan.	(Permendikbud-ristek No. 21 tahun 2022; AERA <i>et al.</i> , 2014)
5.	Instruksi yang diberikan kepada peserta didik harus berisi rincian yang cukup sehingga peserta didik dapat merespons tugas sesuai instruksi, jika diperlukan dapat memberikan contoh.	(AERA <i>et al.</i> , 2014)
6.	Membuat prosedur penilaian yang relevan dengan tes tertulis	(AERA <i>et al.</i> , 2014)
7.	Memberikan informasi kepada peserta didik terlebih dahulu sebelum melakukan penilaian terkait tes yang akan dilakukan, tujuan, keterbatasan dan fungsi penggunaannya.	(Schmeiser <i>et al.</i> , 1995)
8.	Melindungi kesehatan dan keselamatan semua peserta asesmen tes tertulis	(Schmeiser <i>et al.</i> , 1995)
9.	Menerapkan prinsip objektif dan adil dalam pelaksanaan penilaian dan pengambilan keputusan	(Schmeiser <i>et al.</i> , 1995; Wulan, 2018; Permendikbud-ristek No. 21 tahun 2022)
10.	Menilai hanya pada bidang kompetensi tertentu yang telah ditetapkan sebagai tujuan pembelajaran dan mengungkapkan capaian peserta didik secara menyeluruh.	(Schmeiser <i>et al.</i> , 1995)
11.	Memastikan perangkat asesmen tes tertulis yang digunakan sebisa mungkin bebas dari bias	(Permendikbud-ristek No. 21 tahun 2022; Schmeiser <i>et al.</i> , 1995)
12.	Menggunakan hasil penilaian sesuai prosedur untuk pengambilan keputusan yang <i>fair</i>	(Wulan, 2018; Permendikbud-ristek No. 21 tahun 2022)
13.	Mencegah pihak lain untuk membuat laporan yang tidak akurat, klaim yang tidak berdasar, interpretasi yang tidak tepat atau pernyataan yang menyesatkan tentang hasil penilaian	(Schmeiser <i>et al.</i> , 1995)
14.	Menggunakan berbagai sumber dan jenis informasi yang relevan untuk menilai peserta didik	(Schmeiser <i>et al.</i> , 1995; Permendikbud-ristek No. 21 tahun 2022)
15.	Memberikan laporan kepada peserta didik yang berisi skor, deskripsi tertulis yang mudah untuk dimengerti	(Schmeiser <i>et al.</i> , 1995; Permendikbud-ristek No. 21 tahun 2022)
16.	Menghindari kesalahan prosedur dalam penilaian kelas seperti menggunakan penilaian sebagai hukuman atau menerapkan hukuman dalam sistem penilaian	(West, 1923; Wulan, 2018)
17.	Jika memungkinkan, memperbolehkan peserta didik untuk mengajukan banding apabila terindikasi kesalahan dalam proses penilaian dan menilai ulang penilaian tersebut	(Schmeiser <i>et al.</i> , 1995)
18.	Menganalisis praktik penilaian yang tidak etis yang dilakukan oleh guru	(Wulan, 2018)
19.	Menerapkan etika dan tanggung jawab dalam melaksanakan asesmen	(Wulan, 2018)

No.	Standar Assessment Ethics	Sumber
20.	Menunjukkan komitmen untuk menaati aturan dan hukum yang berlaku dalam melaksanakan penilaian sesuai dengan tanggung jawab profesional dengan kejujuran, integritas, kehati-hatian, dan keadilan	(Schmeiser <i>et al.</i> , 1995; Wulan, 2016)
21.	Melindungi hak privasi peserta didik sebagai individu yang dinilai sebagai proses dari pengembangan nilai	(Schmeiser <i>et al.</i> , 1995)
22.	Mengevaluasi praktik penilaian yang tidak etis	(Wulan, 2018)

Seluruh data sekunder yang didapatkan setelah melakukan *desktop research*, digunakan untuk mengembangkan instrumen WAEBEL yang dapat mengidentifikasi kegiatan pembelajaran epistemik melalui kegiatan ilmiah yang dilakukan oleh peserta didik sebagai responden, asesmen pembelajaran yang cocok untuk kegiatan tersebut, dan bagaimana implementasi asesmen dilakukan, apakah sudah sesuai dengan standar etika dan pedoman penilaian pembelajaran epistemik.

3.2. Lokasi dan Waktu Penelitian

Penelitian ini dilakukan pada Tahun Ajaran 2023/2024. Pengukuran implementasi *assessment ethics* dilakukan pada sekolah-sekolah yang berlokasi di Kota Bandung dan Kota Cimahi. Sekolah-sekolah yang terlibat dalam penelitian ini adalah sekolah yang sudah terakreditasi oleh BAN-S/M Indonesia. Karakteristik akreditasi yang digunakan adalah akreditasi A (unggul/sangat baik), akreditasi B (baik), dan akreditasi C (cukup baik/cukup).

3.3. Populasi dan Sampel

Populasi dalam penelitian ini adalah SMA di Kota Bandung dan Cimahi. Sampel dalam penelitian adalah peserta didik dan guru yang melakukan pembelajaran biologi di sekolah-sekolah yang memenuhi karakteristik akreditasi. Peserta didik adalah peserta didik kelas X dan XI yang telah melalui rangkaian kegiatan pembelajaran termasuk asesmen. Sedangkan responden guru adalah tenaga pendidik yang memfasilitasi kegiatan belajar dan asesmen peserta didik dengan pengalaman mengajar yang berbeda-beda pada sekolah sampel.

Sampel ditentukan menggunakan prosedur *purposive sampling*. *Purposive sampling* adalah teknik sampling non-probabilistik untuk memilih partisipan secara sengaja berdasarkan pertimbangan kualifikasi sekolah. Kualifikasi yang dimaksud adalah akreditasi yang ditentukan oleh BAN-S/M. Teknik ini dilakukan untuk

mengidentifikasi dan memilih kelompok sampel yang memiliki karakteristik khusus terkait fenomena yang diteliti dalam penelitian (Creswell & Guetterman, 2019).

Hal ini dilakukan karena peneliti membuat stratifikasi populasi berdasarkan karakteristik akreditasi sekolah (A, B, dan C) sebelum mengambil sampel. Stratifikasi tersebut dibuat untuk menggambarkan ragam praktik pembelajaran epistemik dan asesmen yang berbeda-beda, mulai dari yang terbiasa sampai jarang dalam konteks pelaksanaannya. Selain itu, pemilihan teknik ini juga didasari oleh karakteristik sampel yang menjadi subjek dalam penelitian yaitu peserta didik dan guru yang telah melaksanakan pembelajaran sampai tahap penilaian. Secara lebih rinci jumlah sampel yang berkontribusi dalam penelitian ini dapat dilihat dalam Tabel 3.4.

Tabel 3.4. Sampel yang Digunakan dalam Penelitian

No.	Sekolah	Kategori	Jumlah Sampel Peserta Didik	Jumlah Sampel Guru
1.	SMA-A-01	Unggul/Sangat Baik	67	1
2.	SMA-A-02	Unggul/Sangat Baik	69	2
3.	SMA-A-03	Unggul/Sangat Baik	98	2
4.	SMA-B-01	Baik	34	1
5.	SMA-B-02	Baik	36	1
6.	SMA-B-03	Baik	33	1
7.	SMA-C-02	Cukup Baik/Cukup	8	1
Total			345	9

3.4. Definisi Operasional

1. Praktik *Assessment Ethics* tes tertulis

Praktik *Assessment Ethics* tes tertulis merupakan penerapan etika asesmen dalam asesmen tes tertulis. *Assessment ethics* tes tertulis adalah kumpulan aturan, sistem nilai, komitmen moral, dan prinsip-prinsip yang mengatur pelaksanaan tes tertulis secara *fair* dan dapat dipertanggungjawabkan pada pembelajaran epistemik. Kumpulan aturan ini diperoleh melalui studi literatur mendalam yang mencakup artikel-artikel ilmiah, standar aturan khusus, dan buku-buku yang relevan dengan topik ini. Hasil kajian literatur berbentuk daftar pelanggaran *assessment ethics* tes tertulis (Tabel 3.2.) dan daftar standar *assessment ethics* tes tertulis (Tabel 3.3.) yang secara garis besar berisi nilai-nilai moral atau aturan-aturan yang digunakan untuk mengontrol pelaksanaan asesmen tes tertulis.

2. Tes tertulis pembelajaran epistemik

Asesmen pembelajaran epistemik merupakan penilaian terhadap proses peserta didik dalam membangun pengetahuan dan kompetensi sains pada pembelajaran biologi berdasarkan *framework* sains PISA 2025. Asesmen pembelajaran epistemik tersusun atas rangkaian proses pelaksanaan penilaian dalam pembelajaran yang berbentuk tes tertulis. Konteks yang terdapat pada tes tertulis yang dimaksud berkaitan dengan pengetahuan epistemik dalam pelajaran biologi di sekolah.

3.5. Teknik Pengumpulan Data

Beberapa teknik pengumpulan data yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan *desktop research* untuk mengumpulkan informasi sekunder yang diperlukan untuk mengembangkan instrumen *Written Assessment Ethics Biology Epistemic Learning* (WAEBEL).
2. Melakukan survei dengan memberikan instrumen WAEBEL kepada responden sampel untuk menilai pembelajaran epistemik IPA, asesmen tes tertulis epistemik IPA dan *assessment ethics* tes tertulis di sekolah. Survei dilakukan dengan ketentuan operasional khusus yang ditampilkan pada Tabel 3.5.

Tabel 3.5. Ketentuan Operasional Pelaksanaan Survei

No.	Aspek	Ketentuan Operasional
1.	Waktu	Disediakan waktu 90 menit untuk melakukan survei.
2.	Pembagian sesi	Instrumen diisi pada dua sesi, masing-masing sesi disediakan waktu 40 menit dengan jeda 10 menit di antara sesi I dan sesi II.
3.	Penyampaian	Item survei dibacakan secara terpusat di depan kelas untuk mengontrol waktu pengisian dan memberikan kesempatan bagi responden untuk bertanya jika ada item yang kurang jelas.
4.	Survei	Pelaksanaan survei dilakukan terpisah dan di waktu yang berbeda antara responden peserta didik dan responden guru.

Ketentuan khusus tersebut diterapkan untuk meningkatkan kredibilitas data dan hasil penelitian.

3. Melakukan wawancara semi terstruktur berdasarkan gambaran hasil survei untuk memperoleh informasi sekunder yang dapat mendukung temuan penelitian.

Beberapa instrumen penelitian digunakan dalam penelitian ini untuk merekam data-data yang dibutuhkan dalam penelitian. Instrumen digunakan tergantung data yang ditargetkan untuk direkam. Secara lebih rinci dilihat dalam Tabel 3.6.

Tabel 3.6. Instrumen Pengumpul Data Penelitian

No.	Data	Teknik Pengumpulan Data	Instrumen
1.	Kelayakan instrumen pembelajaran epistemik dan <i>assessment ethics</i> peserta didik oleh <i>validator</i>	Validasi ahli	Lembar validasi instrumen peserta didik
2.	Kelayakan pembelajaran epistemik dan <i>assessment ethics</i> guru oleh <i>validator</i>	Validasi ahli	Lembar validasi instrumen guru
3.	Tanggapan peserta didik	Survei	Instrumen WAEBEL untuk peserta didik
4.	Tanggapan guru	Survei	Instrumen WAEBEL untuk guru
5.	Tanggapan peserta didik	Wawancara semi terstruktur	Lembar wawancara yang disusun berdasarkan hasil survei

2.5.1. Lembar Validasi Instrumen

Lembar validasi dibuat berdasarkan instrumen yang digunakan, yaitu; instrumen *Written Assessment Ethics Biology Epistemic Learning* (WAEBEL) untuk peserta didik dan untuk guru. Instrumen ini digunakan untuk mendapatkan hasil analisis kelayakan berdasarkan pendapat ahli (*validator*) terkait isi dari instrumen yang disebarkan kepada responden. *Validator* ahli merupakan Dosen Program Studi Pendidikan Biologi, Universitas Pendidikan Indonesia.

Lembar validasi dibuat untuk menilai dua aspek pada instrumen, yaitu; aspek konstruksi dan aspek bahasa yang digunakan dalam instrumen. Kisi-kisi penilaian kelayakan pada instrumen WAEBEL yang dikembangkan dapat dilihat pada Tabel 3.7.

Tabel 3.7. Kisi-kisi Lembar Validasi

No.	Kriteria
I	Konstruksi
1.	Konten sesuai dengan indikator dan aspek yang dipersoalkan
2.	Pernyataan dibuat singkat dan jelas
3.	Kalimat bebas dari pernyataan yang bersifat negatif ganda
4.	Pernyataan bebas dari makna ganda
5.	Kalimat bebas dari kemungkinan disetujui/dikosongkan oleh semua responden
6.	Hanya berisi satu gagasan secara lengkap
II	Bahasa
1.	Menggunakan bahasa yang sesuai kaidah Bahasa Indonesia

No.	Kriteria
2.	Menggunakan bahasa yang komunikatif dan sesuai dengan responden

2.5.2. Instrumen *Written Assessment Ethics Biology Epistemic Learning* (WAEBEL) untuk Peserta Didik.

Instrumen tes untuk peserta didik dikembangkan berdasarkan daftar pelanggaran *assessment ethics* yang dikumpulkan dari berbagai sumber melalui studi literatur. Instrumen WAEBEL yang diberikan digunakan untuk mengetahui isu-isu pelanggaran hak peserta didik sebagai peserta asesmen tes tertulis dan pelanggaran etika selama proses pelaksanaan asesmen berlangsung. Instrumen WAEBEL yang digunakan memiliki lima pilihan jawaban, yaitu selalu, sering, kadang-kadang, jarang, dan tidak pernah. Kisi-kisi yang dikembangkan berdasarkan hasil analisis tematik yang dilakukan pada kerangka kerja IPA PISA 2025 (OECD, 2023) dan daftar pelanggaran *assessment ethics* tes tertulis.

Berdasarkan hasil analisis terhadap kerangka kerja, terdapat empat tema utama: (1) penerapan model untuk memfasilitasi prediksi, (2) pemanfaatan data dan bukti ilmiah, (3) penggunaan pengetahuan ilmiah dalam pengambilan keputusan dan tindakan, serta (4) kolaborasi dan kerja sama dalam penyelidikan ilmiah. Keempat tema tersebut kemudian dikembangkan untuk digunakan sebagai kisi-kisi yang ditampilkan pada Tabel 3.8.

Tabel 3.8. Kisi-kisi Instrumen Pembelajaran Epistemik dan Asesmen Tes Tertulis Peserta Didik

No.	Aspek	No.	Indikator
1.	Pelaksanaan pembelajaran epistemik	1.	Penerapan model untuk memfasilitasi prediksi dan penjelasan
		2.	Memanfaatkan data dan bukti ilmiah dalam pembelajaran Biologi
		3.	Menggunakan pengetahuan ilmiah untuk pengambilan keputusan dan tindakan
		4.	Berkolaborasi dan bekerja sama dalam penyelidikan ilmiah
2.	Asesmen tes tertulis dikaitkan dengan pembelajaran epistemik	5.	Asesmen untuk menilai kemampuan menerapkan model untuk memfasilitasi prediksi dan penjelasan
		6.	Asesmen untuk menilai kemampuan memanfaatkan data dan bukti ilmiah dalam pembelajaran Biologi
		7.	Asesmen untuk menilai kemampuan menggunakan pengetahuan ilmiah untuk pengambilan keputusan dan tindakan
		8.	Asesmen terhadap kemampuan kolaborasi dan bekerja sama dalam penyelidikan ilmiah

Sedangkan pelanggaran *assessment ethics* digolongkan berdasarkan empat aspek: (1) persiapan dan pengembangan asesmen tes tertulis yang digunakan, (2) pengembangan instrumen asesmen tes tertulis yang digunakan, (3) pelaksanaan asesmen tes tertulis, dan (4) pemrosesan dan tindak lanjut asesmen tes tertulis. Berikut merupakan rincian kisi-kisi instrumen WAEBEL yang digunakan untuk mengambil data dari peserta didik yang diadaptasi dari (Webber & Lupart, 2012; Wulan, 2018; Zlatkin-Troitschanskaia *et al.*, 2019).

Tabel 3.9. Kisi-kisi Instrumen *Assessment Ethics* Peserta Didik

No.	Aspek	No.	Indikator
1.	Persiapan dan pengembangan instrumen asesmen tes tertulis Biologi	1.	Metode penilaian sesuai dengan latar belakang dan pengalaman belajar peserta didik
		2.	Konten dan bahasa yang digunakan dalam tes tidak bersifat sensitif, seksis, atau menyinggung
		3.	Soal yang tes tertulis bebas dari unsur bias gender
		4.	Soal tes tertulis sesuai dengan norma masyarakat setempat.
		5.	Soal tes tertulis sesuai dengan norma agama.
		6.	Soal tes tertulis sesuai dengan kebudayaan peserta didik
		7.	Soal tes tertulis tidak mengganggu keadaan psikis peserta didik
		8.	Soal tes tertulis menguji “kompetensi asing” yang berpotensi menjadi bias.
2.	Pelaksanaan asesmen tes tertulis Biologi	9.	Tes tertulis dilaksanakan secara <i>fair</i> dan tidak merugikan peserta didik sebagai peserta tes.
		10.	Pelaksanaan tes tertulis mengakomodasi latar belakang sosial-ekonomi peserta didik
		11.	Pelaksanaan tes tertulis nyaman dikerjakan oleh peserta didik
		12.	Bebas dari sistem hukuman seperti mengurangi nilai atas tindakan yang tidak relevan dalam pelaksanaan tes tertulis
		13.	Instrumen dan perangkat tes tertulis dapat diakses oleh semua peserta didik
3.	Pemrosesan dan tindak lanjut asesmen tes tertulis	14.	Memberikan umpan balik setelah tes tertulis agar peserta didik dapat meningkatkan kemampuannya
		15.	Kelulusan peserta didik berdasarkan kriteria yang sudah ditetapkan sebelumnya
		16.	Mengambil keputusan secara objektif dan bebas dari pertimbangan lain yang tidak relevan
		17.	Mengambil keputusan atas dasar informasi yang lengkap
		18.	Mengkomunikasikan hasil asesmen tanpa menjatuhkan harga diri peserta didik
		19.	Tidak membanding-bandingkan hasil tes tertulis peserta didik di depan kelas
		20.	Tidak memberikan tanggung jawab lebih kepada peserta didik dengan nilai baik
		21.	Tidak mempertontonkan hasil asesmen tes tertulis dengan kualitas rendah sebagai contoh buruk tanpa seizin pemiliknya
		22.	Mengakomodir penafsiran skor tes secara individual secara detail

2.5.3. Instrumen *Written Assessment Ethics Biology Epistemic Learning* (WAEBEL) untuk guru

Instrumen WAEBEL untuk guru dikembangkan berdasarkan daftar standar *assessment ethics* tes tertulis. Daftar tersebut didapatkan melalui kajian literatur yang berisi prinsip-prinsip yang seharusnya dipahami dan dilakukan oleh guru sebagai asesor dalam kegiatan pembelajaran terutama yang berkaitan dengan asesmen tes tertulis. Instrumen akan memiliki lima pilihan jawaban sebagai berikut selalu, sering, kadang-kadang, jarang, dan tidak pernah. Kisi-kisi instrumen guru juga dikembangkan berdasarkan hasil analisis tematik yang dilakukan pada kerangka kerja IPA PISA 2025 (OECD, 2023).

Tabel 3.10. Kisi-kisi Instrumen Pembelajaran Epistemik dan Asesmen Tes Tertulis Guru

No.	Aspek	No.	Indikator
1.	Pelaksanaan pembelajaran epistemik	1.	Penerapan model untuk memfasilitasi prediksi dan penjelasan
		2.	Memanfaatkan data dan bukti ilmiah dalam pembelajaran Biologi
		3.	Menggunakan pengetahuan ilmiah untuk pengambilan keputusan dan tindakan
		4.	Berkolaborasi dan bekerja sama dalam penyelidikan ilmiah
2.	Asesmen tes tertulis dikaitkan dengan pembelajaran epistemik	5.	Asesmen untuk menilai kemampuan menerapkan model untuk memfasilitasi prediksi dan penjelasan
		6.	Asesmen untuk menilai kemampuan memanfaatkan data dan bukti ilmiah dalam pembelajaran Biologi
		7.	Asesmen untuk menilai kemampuan menggunakan pengetahuan ilmiah untuk pengambilan keputusan dan tindakan
		8.	Asesmen terhadap kemampuan kolaborasi dan bekerja sama dalam penyelidikan ilmiah

Daftar standar yang diperoleh dikategorikan berdasarkan hasil penelitian Wulan (2018) yang menyatakan bahwa daftar standar *assessment ethics* tes tertulis mencakup standar kompetensi etika sebagai berikut; 1) menjelaskan praktik-praktik yang tidak etis; 2) mempersiapkan tes tertulis yang akan digunakan; 3) menghindari kesalahan prosedur penilaian; dan 4) memiliki tanggung jawab legal dan etika dalam mengumpulkan informasi asesmen. Adapun kisi-kisi instrumen yang akan digunakan dapat dilihat pada Tabel 3.11. yang diadaptasi dari (AERA *et al.*, 2014; Schmeiser *et al.*, 1995; Wulan, 2018; Permendikbud-ristek No.21, 2022).

Tabel 3.11. Kisi-kisi Instrumen *Assessment Ethics* Guru

No.	Aspek	No.	Indikator
1.	Menjelaskan praktik-praktik yang tidak etis	1.	Menjelaskan cara untuk menghindari praktik penilaian yang tidak etis
		2.	Memberikan argumentasi terhadap praktik-praktik penilaian yang dinilai tidak etis
2.	Mempersiapkan tes tertulis yang akan digunakan	3.	Mengembangkan/mengadaptasi tes tertulis yang sesuai dengan peserta didik dilengkapi dengan peninjauan dan uji coba sebelum digunakan
		4.	Spesifikasi tes harus menjelaskan tujuan tes, definisi konstruk atau domain yang diukur, populasi peserta ujian yang dituju, dan interpretasi untuk tujuan penggunaan.
		5.	Instruksi yang diberikan kepada peserta didik harus berisi rincian yang cukup sehingga peserta didik dapat merespons tugas sesuai instruksi, jika diperlukan dapat memberikan contoh.
		6.	Membuat prosedur penilaian yang relevan dengan tes tertulis
3.	Menghindari kesalahan prosedur dalam penilaian	7.	Memberikan informasi kepada peserta didik terlebih dahulu sebelum melakukan penilaian terkait tes yang akan dilakukan, tujuan, keterbatasan dan fungsi penggunaannya.
		8.	Melindungi kesehatan dan keselamatan semua peserta asesmen tes tertulis
		9.	Menerapkan prinsip objektif dan adil dalam pelaksanaan penilaian dan pengambilan keputusan
		10.	Menilai hanya pada bidang kompetensi tertentu yang telah ditetapkan sebagai tujuan pembelajaran dan mengungkapkan capaian peserta didik secara menyeluruh.
		11.	Memastikan perangkat asesmen tes tertulis yang digunakan sebisa mungkin bebas dari bias
		12.	Menggunakan hasil penilaian sesuai prosedur untuk pengambilan keputusan yang <i>fair</i>
		13.	Mencegah pihak lain untuk membuat laporan yang tidak akurat, klaim yang tidak berdasar, interpretasi yang tidak tepat atau pernyataan yang menyesatkan tentang hasil penilaian
		14.	Menggunakan berbagai sumber dan jenis informasi yang relevan untuk menilai peserta didik
		15.	Memberikan laporan kepada peserta didik yang berisi skor, deskripsi tertulis yang mudah untuk dimengerti
		16.	Menghindari kesalahan prosedur dalam penilaian kelas seperti menggunakan penilaian sebagai hukuman atau menerapkan hukuman dalam sistem penilaian
		17.	Jika memungkinkan, memperbolehkan peserta didik untuk mengajukan banding apabila terindikasi kesalahan dalam proses penilaian dan menilai ulang penilaian tersebut
4.	Memiliki tanggung jawab legal dan etika dalam mengumpulkan informasi asesmen	19.	Menerapkan etika dan tanggung jawab dalam melaksanakan asesmen
		20.	Menunjukkan komitmen untuk menaati aturan dan hukum yang berlaku dalam melaksanakan penilaian sesuai dengan tanggung jawab profesional dengan kejujuran, integritas, kehati-hatian, dan keadilan

No.	Aspek	No.	Indikator
		21.	Melindungi hak privasi peserta didik sebagai individu yang dinilai sebagai proses dari pengembangan nilai
		22.	Mengevaluasi praktik penilaian yang tidak etis

3.6. Teknik Analisis Data

Analisis data dilakukan setelah data-data yang diperlukan dalam penelitian terkumpul. Analisis data dilakukan pada tahap uji coba kelayakan dan analisis data primer penelitian. Masing-masing akan dibahas secara rinci pada sub-bagian berikut.

2.6.1. Analisis Uji Coba Kelayakan Instrumen

Uji coba kelayakan instrumen perlu dilakukan untuk memastikan instrumen yang dikembangkan dapat digunakan dalam penelitian. Uji coba kelayakan dilakukan dengan analisis validitas konstruk oleh ahli (validator) dan analisis validitas empiris yang menerapkan *Rasch Model* pada uji coba penyebaran instrumen dalam skala yang lebih kecil. Secara sistematis bagian ini akan membahas analisis uji validitas konstruk dan validitas empiris.

a. Analisis Uji Validitas Konstruk

Penentuan validitas konstruk instrumen oleh ahli (validator) dilakukan sebelum instrumen disebarkan kepada responden dalam skala ukuran sampel kecil maupun skala ukuran sampel yang sebenarnya. Dalam penelitian ini validitas ahli dilakukan oleh tiga orang ahli. Ahli asesmen dan ahli pembelajaran dilakukan oleh kedua dosen pembimbing penulisan tesis ini, dan ahli konten (isi) yang dilakukan oleh dosen di luar penelitian ini.

Analisis konstruk dilakukan dengan mempresentasikan *rating* instrumen berdasarkan skor kelayakan yang diberikan oleh ahli konten (isi). Skor kelayakan kemudian dipersentasekan menggunakan rumus:

$$\% \text{ Kelayakan} = \frac{k}{Nk} 100\%$$

Keterangan:

% Kelayakan : persentase kelayakan
 Nk : skor capaian maksimum
 k : skor kelayakan yang diperoleh

Akbar (2013) menyatakan bahwa pengembang instrumen dapat menentukan sendiri kriteria validitas berdasarkan banyaknya item dan interval yang digunakan dalam skor validasi. Untuk validitas empiris, penelitian ini menggunakan delapan item pernyataan yang diisi oleh validator ahli dengan skor berskala Likert 1 sampai dengan 5. Berdasarkan hal tersebut maka dapat ditentukan lebar interval menggunakan rumus:

1. Penentuan skor maksimal

$$\text{Skor maks} = N \times L_{\text{maks}}$$

2. Penentuan skor minimal

$$\text{Skor min} = N \times L_{\text{min}}$$

3. Penentuan *range*

$$\text{Range} = \text{Skor maks} - \text{Skor min}$$

4. Rumus lebar interval

$$\text{Lebar interval} = \text{Range}/I$$

Keterangan:

N : banyaknya item instrumen validasi

L : skor perolehan skala Likert

I : besaran interval dalam skala Likert

Dengan menerapkan rumus tersebut, maka diperoleh kriteria yang digunakan dalam menganalisis hasil validasi oleh ahli sebagai berikut.

Tabel 3.12. Kriteria Skala Nilai Ahli

Kriteria Skor	Persentase (%)	Tingkat Validasi
33,7 – 40	84,25 – 100	Sangat layak
27,3 – 33,6	68,25 – 84	Layak
20,9 – 27,2	52,25 – 68	Cukup layak
14,5 – 20,8	36,25 – 52	Tidak layak
8 – 14,4	20 – 36	Sangat tidak layak

Berdasarkan tanggapan yang diberikan oleh validator ahli konten (isi), instrumen WAEBEL untuk peserta didik mendapatkan skor sebesar 39 dengan persentase 97,5%. Dengan demikian instrumen untuk peserta didik dinyatakan “sangat layak”. Untuk memperbaiki instrumen tersebut validator ahli menambahkan catatan agar item pernyataan pada instrumen WAEBEL dibuat menggunakan bahasa yang sesuai dengan jenjang peserta didik di SMA.

Instrumen WAEBEL yang ditujukan kepada guru mendapatkan skor 37 dengan persentase sebesar 92,5%. Ini berarti instrumen untuk guru juga dapat

dinyatakan “sangat layak”. Validator ahli memberikan penilaian umum bahwa instrumen dapat digunakan namun harus di revisi pada beberapa bagian; meringkas kalimat agar tidak terlampau panjang, perbaikan kalimat agar relevan dengan aspek/indikator yang diukur, dan mencegah kemungkinan kalimat untuk dipilih seluruhnya oleh guru sebagai responden.

b. Analisis Uji Validitas Empiris dan Reliabilitas *Rasch Model*

Instrumen yang dikembangkan harus diuji kelayakannya. Uji kelayakan validitas dan reliabilitas dalam penelitian ini akan menggunakan *Rasch Model* dengan bantuan aplikasi Winstep. Model dalam aplikasi tersebut adalah model yang dikembangkan pertama kali oleh ahli matematika Rasch. Model ini menganalisis probabilitas terhadap pilihan yang disediakan dalam instrumen. Prinsip dasar model Rasch adalah model probabilistik dimana “individu yang memiliki tingkat probabilitas yang lebih besar dibandingkan dengan individu lainnya seharusnya memiliki peluang yang lebih besar untuk menjawab soal instrumen dengan benar”. Pada prinsip tersebut, butir soal dengan kategori sulit akan menurunkan peluang individu untuk menjawab (Sumintono & Widhiarso, 2014).

Keunggulan *Rasch model* adalah mampu melakukan prediksi apabila terdapat data yang hilang (*missing data*). Prediksi didasari pada pola respons yang sistematis. Hal ini akan membuat hasil analisis statistik lebih akurat. *Rasch model* juga memiliki kemampuan untuk menghasilkan nilai pengukuran *standar error* pada instrumen yang digunakan (Sumintono & Widhiarso, 2014).

Uji coba dilakukan pada dua kelas yang tersebar di Kota Bandung dan Cimahi, Indonesia. Uji coba juga dilakukan dengan memperhatikan pembagian strata yang dilakukan pada teknik *sampling*. Oleh karena itu uji coba dilakukan pada kelas dengan akreditasi sekolah A dan B. Sekolah dengan akreditasi C tidak di uji coba karena keterbatasan jumlah responden dengan kriteria tersebut.

Data validitas empiris dianalisis setelah uji coba terbatas di dalam populasi yang ditetapkan dalam penelitian. Uji coba dilakukan pada 50 orang peserta didik dan 7 orang guru mata pelajaran Biologi. Analisis validasi empiris dilakukan menggunakan aplikasi WinStep Rasch Model.

Pengukuran validitas empiris dilihat berdasarkan nilai *logit* pada *outfit* MNSQ, *outfit* ZSTD, dan *PT-Measure-Al Coord. Pt-Measure Corr* biasanya

digunakan untuk melihat adanya daya beda dari suatu instrumen. Sehingga untuk melihat validitas empiris kuesioner dengan respons skala Likert yang dikembangkan hanya digunakan *oufit* MNSQ, *oufit* ZSTD dengan ketentuan nilai diterima : $0,5 < \text{MNSQ} < 1,5$, dan $-2,0 < \text{ZSTD} < +2,0$. Jika salah satu kriteria terpenuhi, maka item layak digunakan dalam penelitian (Sumirtono & Widhiarso, 2015).

Reliabilitas atau keajegan butir pernyataan pada instrumen juga diperoleh dari pengolahan *Rasch Model* pada aplikasi Winstep. Keputusan sebuah instrumen dinyatakan reliabel dilihat dari nilai *Item Reliability* dan *Person Reliability*. *Item reliability* mengukur sejauh mana item tes (pertanyaan) dapat diandalkan dalam mengukur konsep yang diinginkan. Sedangkan *person reliability* mengukur sejauh mana skor individu (responden) konsisten dan dapat diandalkan. Kriteria umum untuk menilai item dan *person reliability* yang baik adalah nilai minimal 0.8 atau lebih tinggi.

Analisis validitas empiris dan reliabilitas untuk masing-masing responden peserta didik dan guru adalah sebagai berikut:

1) Hasil Analisis Validitas & Reliabilitas Instrumen WAEBEL untuk Peserta Didik

Tabel 3.13. memperlihatkan *ouput item misfit order* hasil analisis data uji coba peserta didik yang didapatkan dengan bantuan aplikasi Winsep.

Tabel 3.13. Ouput Item Misfit Order Instrumen Peserta Didik

<i>Item</i>	<i>Outfit MNSQ</i>	Interpretasi <i>Outfit MNSQ</i>	<i>Outfit ZSTD</i>	Interpretasi <i>Outfit ZSTD</i>	Keputusan
D43	2.26	Belum Terpenuhi	5.32	Belum Terpenuhi	Diperbaiki
C61	1.57	Belum Terpenuhi	1.81	Terpenuhi	Digunakan
D53	1.78	Belum Terpenuhi	3.68	Belum Terpenuhi	Diperbaiki
D21	1.7	Belum Terpenuhi	3.05	Belum Terpenuhi	Diperbaiki
C71	1.65	Belum Terpenuhi	3.38	Belum Terpenuhi	Diperbaiki
D42	1.65	Belum Terpenuhi	3.19	Belum Terpenuhi	Diperbaiki
E61	1.37	Terpenuhi	1.54	Terpenuhi	Digunakan
B32	1.44	Terpenuhi	2.32	Belum Terpenuhi	Digunakan
E13	1.42	Terpenuhi	2.18	Belum Terpenuhi	Digunakan

<i>Item</i>	<i>Outfit MNSQ</i>	<i>Interpretasi Outfit MNSQ</i>	<i>Outfit ZSTD</i>	<i>Interpretasi Outfit ZSTD</i>	Keputusan
C41	1.14	Terpenuhi	0.54	Terpenuhi	Digunakan
E23	1.37	Terpenuhi	1.98	Terpenuhi	Digunakan
C21	1.21	Terpenuhi	0.84	Terpenuhi	Digunakan
D41	1.18	Terpenuhi	0.88	Terpenuhi	Digunakan
E31	1.17	Terpenuhi	0.77	Terpenuhi	Digunakan
E24	1.2	Terpenuhi	1.13	Terpenuhi	Digunakan
E12	1.17	Terpenuhi	1.02	Terpenuhi	Digunakan
C31	1.15	Terpenuhi	0.88	Terpenuhi	Digunakan
E51	1.05	Terpenuhi	0.32	Terpenuhi	Digunakan
E91	1.14	Terpenuhi	0.83	Terpenuhi	Digunakan
C22	0.97	Terpenuhi	-0.05	Terpenuhi	Digunakan
E81	1.03	Terpenuhi	0.19	Terpenuhi	Digunakan
D51	1.08	Terpenuhi	0.48	Terpenuhi	Digunakan
C62	0.86	Terpenuhi	-0.19	Terpenuhi	Digunakan
E11	1.08	Terpenuhi	0.54	Terpenuhi	Digunakan
A13	1.07	Terpenuhi	0.47	Terpenuhi	Digunakan
B23	1.07	Terpenuhi	0.42	Terpenuhi	Digunakan
C82	0.91	Terpenuhi	-0.49	Terpenuhi	Digunakan
C51	0.83	Terpenuhi	-0.56	Terpenuhi	Digunakan
E41	0.85	Terpenuhi	-0.74	Terpenuhi	Digunakan
E22	0.89	Terpenuhi	-0.67	Terpenuhi	Digunakan
D15	0.87	Terpenuhi	-0.8	Terpenuhi	Digunakan
D11	0.86	Terpenuhi	-0.79	Terpenuhi	Digunakan
A21	0.85	Terpenuhi	-0.91	Terpenuhi	Digunakan
B26	0.84	Terpenuhi	-0.93	Terpenuhi	Digunakan
B12	0.83	Terpenuhi	-0.99	Terpenuhi	Digunakan
D31	0.83	Terpenuhi	-0.88	Terpenuhi	Digunakan
C11	0.82	Terpenuhi	-1.04	Terpenuhi	Digunakan
D12	0.8	Terpenuhi	-1.3	Terpenuhi	Digunakan
A12	0.77	Terpenuhi	-1.44	Terpenuhi	Digunakan
C81	0.77	Terpenuhi	-1.46	Terpenuhi	Digunakan
B24	0.71	Terpenuhi	-1.62	Terpenuhi	Digunakan
B25	0.75	Terpenuhi	-1.56	Terpenuhi	Digunakan
B31	0.74	Terpenuhi	-1.71	Terpenuhi	Digunakan
B41	0.68	Terpenuhi	-2	Terpenuhi	Digunakan
A41	0.65	Terpenuhi	-2.38	Belum Terpenuhi	Digunakan
B22	0.65	Terpenuhi	-2.36	Belum Terpenuhi	Digunakan
B33	0.65	Terpenuhi	-2.36	Belum Terpenuhi	Digunakan
A11	0.58	Terpenuhi	-2.93	Belum Terpenuhi	Digunakan

<i>Item</i>	<i>Outfit MNSQ</i>	<i>Interpretasi Outfit MNSQ</i>	<i>Outfit ZSTD</i>	<i>Interpretasi Outfit ZSTD</i>	<i>Keputusan</i>
A14	0.56	Terpenuhi	-3.21	Belum Terpenuhi	Digunakan
A31	0.56	Terpenuhi	-3.15	Belum Terpenuhi	Digunakan
D14	0.56	Terpenuhi	-3.13	Belum Terpenuhi	Digunakan
D13	0.55	Terpenuhi	-3.23	Belum Terpenuhi	Digunakan

Pada Tabel 3.13. dapat dilihat bahwa lima item belum memenuhi kriteria dan perlu ditinjau; D43; C61; D53; D21; C71; D42. Item D43 "*Jika saya salah menjawab soal tes tertulis Biologi dalam tes, akan ada pengurangan skor*" tidak memenuhi salah satu dari dua kriteria. Nilai *MNSQ Outfit* dan *ZSTD Outfit* sangat tinggi. Ini menunjukkan bahwa item D43 kurang jelas dan berpotensi ambigu. Oleh karena itu, perlu dilakukan perbaikan redaksi pada item agar mudah dipahami.

Pada butir D53 "*Ujian Biologi Online diawasi oleh guru*", butir D21 "*Soal tes Biologi Tertulis mengabaikan kondisi sosial ekonomi peserta didik tertentu*", butir C71 "*Soal tes tertulis Biologi membuat mengganggu kesehatan psikis peserta didik*", dan item D24 "*Nilai ujian biologi mendapatkan proporsi besar dalam rapor*" belum memenuhi kriteria dimana nilai *MNSQ Outfit* dan *ZSTD Outfit* tinggi. Oleh karena itu, perlu untuk menganalisis isi item dan bagaimana peserta ujian memahaminya. Keputusan penelitian adalah tetap menggunakan semua item pertanyaan setelah melakukan pertimbangan. Revisi kecil pada pernyataan item mungkin diperlukan.

Analisis reliabilitas pada instrumen yang dikembangkan dilakukan dengan bantuan aplikasi Winstep. Hasil reliabilitas instrumen yang dikembangkan ditunjukkan dalam Gambar 3.2. diperoleh hasil yang menunjukkan nilai reliabilitas tes Cronbach Alpha (KR-20) yaitu 0,82. Nilai Cronbach Alpha sebesar 0,82 menunjukkan bahwa instrumen tes Anda memiliki reliabilitas yang bagus sekali (Sumirtono & Widhiarso, 2015).

Nilai *Person Reliability* yang didapatkan sebesar 0,80. Ini mengindikasikan bahwa instrumen yang dikembangkan akan stabil dan konsisten jika diberikan kepada sampel yang lain dalam populasi. Nilai *Item Reliability* yang diperoleh cukup tinggi yaitu 0,92. Ini berarti item-item dalam instrumen yang dikembangkan

memiliki keajekan yang baik dan dapat diandalkan dalam mengukur konsep yang diinginkan.

SUMMARY OF 50 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	175.1	60.0	.71	.16	1.02	-.17	1.02	-.18
SEM	2.1	.0	.06	.00	.07	.35	.07	.33
P.SD	14.6	.0	.39	.01	.47	2.45	.51	2.34
S.SD	14.7	.0	.40	.01	.48	2.48	.51	2.37
MAX.	206.0	60.0	1.63	.19	3.52	9.90	3.85	9.90
MIN.	140.0	60.0	-.19	.16	.38	-5.19	.38	-4.85
REAL RMSE	.18	TRUE SD	.35	SEPARATION	1.97	Person RELIABILITY	.80	
MODEL RMSE	.16	TRUE SD	.36	SEPARATION	2.16	Person RELIABILITY	.82	
S.E. OF Person MEAN = .06								
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .82 SEM = 6.17								
STANDARDIZED (50 ITEM) RELIABILITY = .80								

SUMMARY OF 60 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	145.9	50.0	.00	.19	1.04	-.01	1.02	-.07
SEM	2.9	.0	.10	.00	.05	.24	.04	.23
P.SD	22.1	.0	.75	.03	.35	1.84	.34	1.78
S.SD	22.2	.0	.75	.03	.35	1.86	.34	1.80
MAX.	193.0	50.0	1.78	.38	2.29	5.47	2.26	5.32
MIN.	89.0	50.0	-2.30	.17	.54	-3.38	.55	-3.23
REAL RMSE	.20	TRUE SD	.72	SEPARATION	3.53	Item RELIABILITY	.93	
MODEL RMSE	.19	TRUE SD	.72	SEPARATION	3.83	Item RELIABILITY	.94	
S.E. OF Item MEAN = .10								
Item RAW SCORE-TO-MEASURE CORRELATION = -.99								
Global statistics: please see Table 44.								
UMEAN=.0000 USCALE=1.0000								

Gambar 3.2. Hasil Reliabilitas Instrumen Peserta Didik

2) Hasil Analisis Validitas dan Reliabilitas Instrumen WAEBEL untuk Guru

Respons uji coba yang diberikan oleh guru juga di analisis menggunakan aplikasi Winstep. Validitas empiris dapat dilihat dari *ouput item misfit order* yang ditunjukkan pada Tabel 3.14.

Tabel 3.14. Ouput Item Misfit Order

Item	Outfit MNSQ	Interpretasi Outfit MNSQ	Outfit ZSTD	Interpretasi Outfit ZSTD	Keputusan
E71	3.47	Belum Terpenuhi	3.59	Belum Terpenuhi	Diperbaiki
E53	2.42	Belum Terpenuhi	2.17	Belum Terpenuhi	Diperbaiki
E81	2.29	Belum Terpenuhi	2.27	Belum Terpenuhi	Diperbaiki
E92	1.87	Belum Terpenuhi	1.53	Terpenuhi	Digunakan
E62	1.74	Belum Terpenuhi	1.44	Terpenuhi	Digunakan

<i>Item</i>	<i>Outfit MNSQ</i>	<i>Interpretasi Outfit MNSQ</i>	<i>Outfit ZSTD</i>	<i>Interpretasi Outfit ZSTD</i>	<i>Keputusan</i>
C11	1.61	Belum Terpenuhi	1.04	Terpenuhi	Digunakan
E33	1.46	Terpenuhi	1.05	Terpenuhi	Digunakan
E112	1.39	Terpenuhi	0.85	Terpenuhi	Digunakan
E11	1.38	Terpenuhi	0.88	Terpenuhi	Digunakan
E12	1.27	Terpenuhi	0.71	Terpenuhi	Digunakan
E32	1.34	Terpenuhi	0.70	Terpenuhi	Digunakan
A13	1.33	Terpenuhi	0.75	Terpenuhi	Digunakan
E35	1.23	Terpenuhi	0.61	Terpenuhi	Digunakan
E111	1.24	Terpenuhi	0.61	Terpenuhi	Digunakan
A14	1.22	Terpenuhi	0.56	Terpenuhi	Digunakan
F11	1.15	Terpenuhi	0.48	Terpenuhi	Digunakan
F32	1.20	Terpenuhi	0.53	Terpenuhi	Digunakan
A42	1.20	Terpenuhi	0.57	Terpenuhi	Digunakan
B23	1.17	Terpenuhi	0.48	Terpenuhi	Digunakan
E42	1.14	Terpenuhi	0.44	Terpenuhi	Digunakan
B33	1.11	Terpenuhi	0.37	Terpenuhi	Digunakan
D31	1.08	Terpenuhi	0.32	Terpenuhi	Digunakan
A12	1.07	Terpenuhi	0.30	Terpenuhi	Digunakan
E101	1.03	Terpenuhi	0.24	Terpenuhi	Digunakan
D12	0.94	Terpenuhi	0.07	Terpenuhi	Digunakan
B22	0.91	Terpenuhi	-0.04	Terpenuhi	Digunakan
E31	0.79	Terpenuhi	-0.16	Terpenuhi	Digunakan
B13	0.87	Terpenuhi	-0.09	Terpenuhi	Digunakan
B31	0.82	Terpenuhi	-0.20	Terpenuhi	Digunakan
C21	0.82	Terpenuhi	-0.26	Terpenuhi	Digunakan
D41	0.77	Terpenuhi	-0.41	Terpenuhi	Digunakan
A21	0.75	Terpenuhi	-0.46	Terpenuhi	Digunakan
A41	0.74	Terpenuhi	-0.50	Terpenuhi	Digunakan
B25	0.74	Terpenuhi	-0.50	Terpenuhi	Digunakan
B41	0.74	Terpenuhi	-0.50	Terpenuhi	Digunakan
E34	0.65	Terpenuhi	-0.44	Terpenuhi	Digunakan
B32	0.68	Terpenuhi	-0.61	Terpenuhi	Digunakan
E41	0.72	Terpenuhi	-0.42	Terpenuhi	Digunakan
E52	0.42	Terpenuhi	-0.40	Terpenuhi	Digunakan
D21	0.64	Terpenuhi	-0.78	Terpenuhi	Digunakan
E63	0.64	Terpenuhi	-0.78	Terpenuhi	Digunakan
A31	0.62	Terpenuhi	-0.69	Terpenuhi	Digunakan
B12	0.61	Terpenuhi	-0.69	Terpenuhi	Digunakan
D11	0.47	Terpenuhi	-0.87	Terpenuhi	Digunakan
E51	0.47	Terpenuhi	-0.87	Terpenuhi	Digunakan
E61	0.55	Terpenuhi	-0.85	Terpenuhi	Digunakan
F21	0.47	Terpenuhi	-0.87	Terpenuhi	Digunakan
F31	0.56	Terpenuhi	-1.02	Terpenuhi	Digunakan
A32	0.53	Terpenuhi	-1.05	Terpenuhi	Digunakan
B11	0.53	Terpenuhi	-0.96	Terpenuhi	Digunakan
E121	0.38	Belum Terpenuhi	-1.41	Terpenuhi	Digunakan
F61	0.38	Belum Terpenuhi	-1.41	Terpenuhi	Digunakan
A11	0.31	Belum Terpenuhi	-1.73	Terpenuhi	Digunakan
B21	0.24	Belum Terpenuhi	-2.00	Terpenuhi	Digunakan

Analisis yang dilakukan pada Tabel 3.13. menunjukkan bahwa item E71, E53, dan E81 masih perlu ditinjau kembali. Ketiga item tersebut belum memenuhi kriteria ZSTD (-2,0 hingga 2,0) atau MNSQ (0,5 hingga 1,5). Hal ini berarti ketiga item tersebut dianggap tidak mampu secara konsisten merekam sebuah konsep.

Item E71 yang berbunyi “Mencegah pihak lain untuk membuat laporan yang tidak akurat, klaim yang tidak berdasar, interpretasi yang tidak tepat atau pernyataan yang menyesatkan tentang hasil penilaian”. Item tersebut memperoleh nilai MNSQ dan ZSRTD yang tinggi. Asumsi yang ditarik adalah terjadi ambiguitas pada item ini, sehingga butir item ini perlu diperbaiki redaksinya agar bisa digunakan.

Hal yang sama juga terjadi pada item E53 “Mempertimbangkan faktor sosio-ekonomi peserta didik dalam mengembangkan soal dalam perangkat asesmen” dan item E81 “Menggunakan berbagai sumber dan jenis informasi yang relevan untuk menilai peserta didik”. Kedua item ini tetap digunakan namun direvisi terlebih dahulu.

Untuk meninjau reliabilitas instrumen yang dikembangkan untuk guru, penelitian ini melihat kembali nilai Cronbach Alpha (K-20) yang ditampilkan pada *output* aplikasi Winstep. Nilai Cronbach Alpha (K-20) yang diperoleh adalah 0,86. Ini mengindikasikan instrumen yang dikembangkan sudah dikategorikan bagus sekali dari segi reliabilitasnya. Hal ini juga dibuktikan dengan nilai *Person Reliability* yang dikategorikan bagus yaitu 0,84 dan *Item Reliability* yang dikategorikan cukup bagus yaitu 0.79.

TABLE 3.1 C:\Users\Widia\Desktop\1_RASCH BARU.pr ZOU750WS.TXT Jul 11 2024 11:42
 INPUT: 7 Person 58 Item REPORTED: 7 Person 58 Item 4 CATS WINSTEPS 5.4.2.0

SUMMARY OF 7 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	178.6	58.0	1.78	.24	1.01	.02	.99	-.07
SEM	4.6	.0	.26	.00	.08	.41	.08	.37
P.SD	11.2	.0	.63	.01	.19	.99	.18	.91
S.SD	12.1	.0	.68	.01	.21	1.07	.20	.98
MAX.	194.0	58.0	2.68	.26	1.40	1.93	1.38	1.69
MIN.	159.0	58.0	.71	.23	.83	-.96	.83	-.95
REAL RMSE	.25	TRUE SD	.58	SEPARATION	2.31	Person RELIABILITY		.84
MODEL RMSE	.24	TRUE SD	.58	SEPARATION	2.43	Person RELIABILITY		.86
S.E. OF Person MEAN = .26								
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00 (approximate due to missing data)								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .86 SEM = 4.22								
STANDARDIZED (50 ITEM) RELIABILITY = .84								
SUMMARY OF 58 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	21.6	7.0	.00	.70	.99	.01	.99	.02
SEM	.5	.0	.22	.01	.07	.13	.07	.13
P.SD	3.7	.0	1.65	.08	.54	.97	.55	.99
S.SD	3.7	.0	1.67	.08	.54	.97	.56	1.00
MAX.	27.0	7.0	5.09	1.09	3.65	3.71	3.47	3.59
MIN.	10.0	7.0	-3.02	.64	.24	-1.98	.24	-2.00
REAL RMSE	.76	TRUE SD	1.47	SEPARATION	1.93	Item RELIABILITY		.79
MODEL RMSE	.70	TRUE SD	1.50	SEPARATION	2.13	Item RELIABILITY		.82
S.E. OF Item MEAN = .22								
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00 (approximate due to missing data)								
Global statistics: please see Table 44.								
UMEAN=.0000 USCALE=1.0000								

Gambar 3.3. Hasil Reliabilitas Instrumen Guru

2.6.2. Analisis Data Penelitian

Data yang digunakan dalam penelitian ini dikategorikan ke dalam dua kategori, yaitu; data primer dan data sekunder. Pengkategorian tersebut didasari pada jenis data dan analisis data. Secara umum data sekunder yang digunakan dalam penelitian ini adalah data hasil studi literatur yang di analisis pada tahap awal penelitian, dan data wawancara yang dilakukan setelah survei dilakukan. Sedangkan data primer adalah data numerik yang didapatkan dari penyebaran instrumen WAEBEL kepada responden.

a. Analisis Data Sekunder

Data sekunder dalam penelitian ini terdiri atas hasil-hasil kajian literatur terkait pembelajaran epistemik berdasarkan tinjauan kurikulum dan *science framework* PISA 2025, asesmen epistemik, pelanggaran, standar *assessment ethics* pada konteks tes tertulis dan data wawancara semi terstruktur. Data-data tersebut akan dianalisis menggunakan metode analisis kualitatif tematik. Analisis tematik (*thematic analysis*) merupakan salah satu metode menganalisis data dengan tujuan mengidentifikasi pola atau menemukan tema pada data yang dikumpulkan oleh peneliti (Braun & Clarke, 2006).

Analisis tematik dianggap paling efektif untuk mengulas secara rinci hasil-hasil kajian literatur terkait pembelajaran epistemik, asesmen epistemik, dan *assessment ethics* pada konteks tes tertulis untuk menemukan keterkaitan pola-pola yang teridentifikasi dengan fenomena yang menjadi fokus penelitian (Fereday & Muir-Cochrane, 2006). Holloway dan Todres (2003) mengatakan bahwa *thematic analysis* ini merupakan dasar atau fondasi utama bagi peneliti yang akan menganalisis data yang bentuknya kualitatif.

Penelitian ini melewati tiga prosedur untuk melakukan analisis tematik, ditinjau berdasarkan (Heriyanto, 2018) yaitu;

1. Memahami data

Analisis dimulai dengan membaca dan memahami literatur yang relevan dengan penelitian secara menyeluruh. Literatur yang dimaksud yaitu *science framework* PISA 2025, standar kompetensi lulusan dalam kurikulum yang berlaku di sekolah yaitu kurikulum 2013 dan kurikulum merdeka, dan hasil-hasil penelitian, buku, kode etik terkait *assessment ethics*. Hal ini dilakukan untuk mendapatkan gambaran umum dan konteks relevansi dengan penelitian.

2. Menyusun kode

Identifikasi dan pemberian tanda dilakukan pada segmen-segmen data yang relevan dengan fokus penelitian. Penelitian ini berfokus pada pembelajaran epistemik, asesmen epistemik dan *assessment ethics*. Setelah penandaan selesai, peneliti mengelompokkan kode-kode tersebut berdasarkan kesamaan atau pola yang muncul.

3. Mencari tema

Peninjauan dan pengelompokkan kode-kode yang telah disusun dilakukan untuk menemukan tema-tema utama yang mencerminkan data yang dikumpulkan. Tema-tema ini kemudian dikaji dan diinterpretasikan untuk dijadikan sebagai instrumen yang digunakan menjawab pertanyaan penelitian serta menjelaskan fenomena yang diteliti.

b. Analisis Data Primer

Data primer yang diambil dari sampel adalah data kuantitatif, sehingga diperlukan statistika untuk mengolah data tersebut. Data dari instrumen akan diberi skor terlebih dahulu pada masing-masing butir pertanyaan yang diajukan. Adapun panduan pemberian skor adalah sebagai berikut.

Tabel 3.14. Panduan Pemberian Skor Instrumen WAEBEL

Pilihan Jawaban	Skor
A	4
B	3
C	2
D	1

Penelitian ini mengungkap fenomena terkait pelaksanaan asesmen tes tertulis melalui sudut pandang peserta didik dan sudut pandang guru. Pada tahap selanjutnya akan dilakukan analisis statistik deskriptif untuk menggambarkan praktik *Assessment Ethic* berdasarkan data yang diambil. Ghozali (2011) mengungkapkan bahwa statistik deskriptif akan memberikan gambaran atau deskripsi suatu data yang ditinjau dari nilai rata-rata (*mean*), standar deviasi varian, maksimum, minimum, total, dan *range*.

Analisis data kuantitatif menggunakan statistik deskriptif, analisis dilakukan dengan menghitung skor rata-rata (*mean*), menghitung persentase, dan menginterpretasikan temuan berdasarkan kategori interval untuk memperoleh gambaran tentang distribusi data. Hal tersebut dilakukan untuk memberikan gambaran implementasi pembelajaran epistemik, asesmen pembelajaran epistemik, dan *assessment ethics*.

Implementasi kemudian dinilai berdasarkan rata-rata skor yang diadaptasi dari Ridwan (2020). Penilaian menggunakan analisis kuantitatif dapat dilihat pada Tabel 3.15.

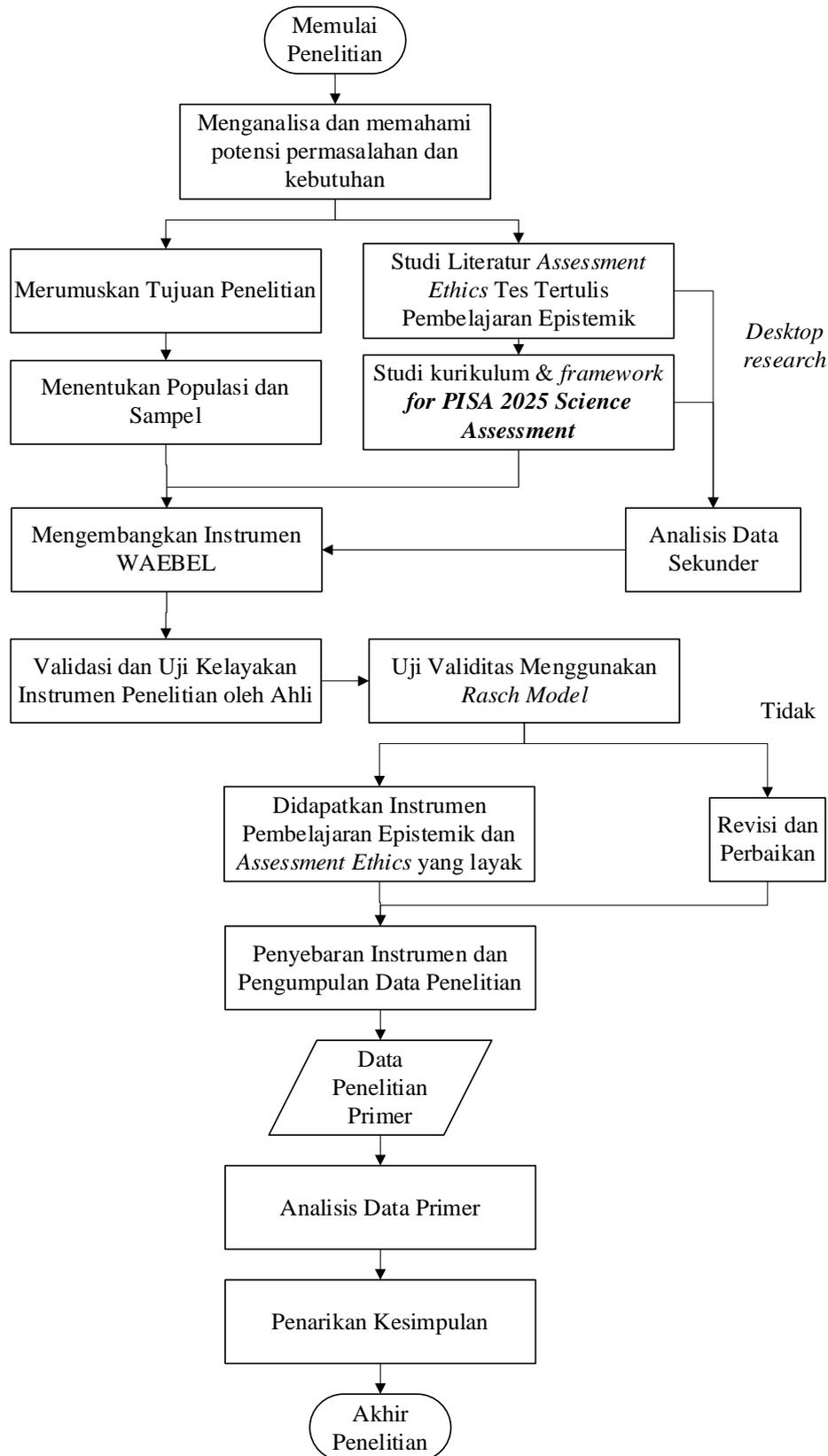
Tabel 3.15. Interpretasi Skala Penilaian

Pilihan Jawaban	Skor
Sangat Baik	4
Baik	3
Kurang	2
Sangat Kurang	1

(Ridwan, 2020)

Berdasarkan gambaran yang diperoleh, dibuat visualisasi data untuk memudahkan interpretasi. Temuan penelitian disajikan dalam bentuk nilai rata-rata dan persentase yang dikategorikan berdasarkan lebar interval dan *range* pada masing-masing aspek. Visualisasi ini bertujuan untuk memberikan gambaran yang lebih jelas dan mudah dipahami tentang implementasi pembelajaran epistemik, asesmen pembelajaran epistemik, dan *assessment ethics* sebagai temuan penelitian.

3.7. Alur Penelitian



Gambar 3.4. Skema Alur Penelitian