

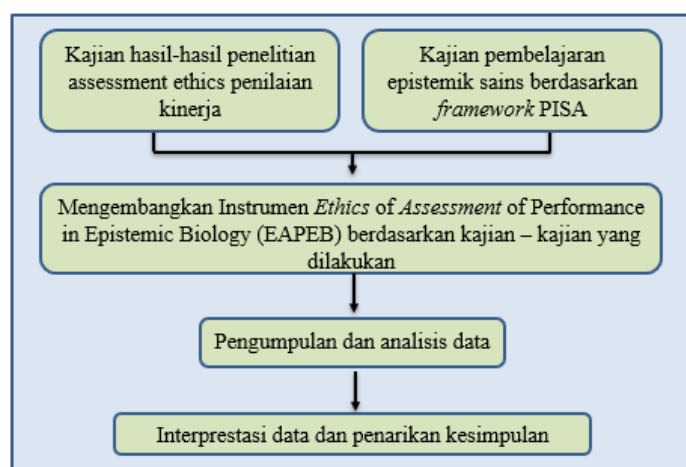
BAB III

METODE PENELITIAN

3.1. Metode dan Desain Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah pendekatan kuantitatif dengan metode penelitian deskriptif. Metode penelitian deskriptif merupakan alat yang sangat efektif dalam mengumpulkan dan menganalisis data untuk menggambarkan karakteristik dari fenomena atau situasi tertentu tanpa perlu memanipulasi variabel. Penggunaan survei, observasi, dan analisis data sekunder memungkinkan untuk secara akurat mengidentifikasi dan mendeskripsikan aspek-aspek dari sebuah kelompok atau fenomena. Keunggulan utama dari metode ini adalah kemampuannya untuk menggeneralisasi hasil dari sampel ke populasi yang lebih besar (Manjunatha, 2019).

Setelah dilakukan studi literatur untuk menemukan daftar pelanggaran etika asesmen kinerja pada kegiatan pembelajaran proses sains biologi dan kajian *framework* PISA 2025 mengenai pembelajaran epistemik, selanjutnya dilakukan pengembangan instrumen kuesioner dan pengujian kuesioner. Untuk mendukung data utama yang diperoleh maka diperlukan data tambahan untuk mendapatkan pemahaman mendalam tentang praktik *assessment ethics* di sekolah melalui metode wawancara (kualitatif). Adapun langkah-langkah utama yang dilakukan dalam penelitian ini dapat dilihat pada gambar 3.1.



Gambar 3. 1. Metode dan Desain Penelitian

Hasil kajian *framework* PISA 2025 mengenai pembelajaran epistemik dan asesmen kinerja dirumuskan ke dalam dua aspek utama, yaitu pelaksanaan pembelajaran epistemik biologi serta asesmen kinerja yang dikaitkan dengan pembelajaran epistemik. Hasil kajian ini kemudian digunakan sebagai dasar dalam penyusunan kisi-kisi untuk mengembangkan instrumen yang bertujuan mengidentifikasi pelaksanaan pembelajaran epistemik di sekolah, serta instrumen yang bertujuan mengidentifikasi pelaksanaan asesmen kinerja dalam pembelajaran epistemik (lihat Tabel 3.5.).

Kajian mengenai etika asesmen dilakukan dengan menganalisis pelanggaran yang terjadi dalam proses asesmen kinerja serta standar yang seharusnya diikuti oleh guru saat melaksanakan asesmen kinerja. Fokus dari kajian ini adalah untuk mengidentifikasi dan menganalisis kedua aspek tersebut agar sesuai dengan praktik asesmen, khususnya dalam konteks asesmen kinerja. Dengan memahami prinsip-prinsip etika yang mendasari proses asesmen kinerja, diharapkan dapat tercipta keadilan dan objektivitas dalam penilaian, sehingga hasil asesmen kinerja dapat mencerminkan kemampuan dan usaha siswa secara akurat. Hal ini juga penting untuk memastikan bahwa setiap siswa diberikan kesempatan yang sama untuk menunjukkan potensi mereka, tanpa terpengaruh oleh faktor eksternal yang tidak relevan.

Penelitian dan literatur yang mengkaji pelanggaran serta standar etika dalam asesmen kinerja masih tergolong sedikit. Hanya ada sejumlah kecil hasil penelitian dan sumber yang relevan dengan praktik etika dalam asesmen kinerja. Meskipun demikian, penelitian ini telah berhasil merangkum beberapa daftar standar etika yang berkaitan dengan asesmen kinerja. Tabel berikut menyajikan informasi mengenai daftar standar etika yang berkaitan dengan asesmen kinerja.

Tabel 3. 1. Daftar Standar *Assessment Ethics* pada Asesmen Kinerja

No.	Standar <i>Assessment Ethics</i>	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 asesmen kinerja yang sesuai dengan siswa dilengkapi dengan peninjauan dan uji coba sebelum digunakan	(Permendikbud-ristek No. 21 tahun 2022; AERA <i>et al.</i> , 2014)

No.	Standar Assessment Ethics	Sumber
4.	Spesifikasi tes harus menjelaskan tujuan tes, definisi konstruk atau domain yang diukur, populasi peserta asesmen yang dituju, dan interpretasi untuk tujuan penggunaan.	(Permendikbud-ristek No. 21 tahun 2022; AERA <i>et al.</i> , 2014)
5.	Instruksi yang diberikan kepada siswa harus berisi perincian yang cukup sehingga siswa dapat merespons tugas sesuai instruksi, jika diperlukan dapat memberikan contoh.	(AERA <i>et al.</i> , 2014)
6.	Membuat prosedur penilaian yang relevan asesmen kinerja	(AERA <i>et al.</i> , 2014)
7.	Memberikan informasi kepada siswa terlebih dahulu sebelum melakukan penilaian kinerja yang akan dilakukan, tujuan, keterbatasan dan fungsi penggunaannya.	(Schmeiser <i>et al.</i> , 1995)
8.	Melindungi kesehatan dan keselamatan semua peserta asesmen kinerja	(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 siswa secara menyeluruh.	(Schmeiser <i>et al.</i> , 1995)
11.	Memastikan perangkat asesmen kinerja 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 siswa	(Schmeiser <i>et al.</i> , 1995; Permendikbud-ristek No. 21 tahun 2022)
15.	Memberikan laporan kepada siswa 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 siswa 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)
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)

No.	Standar <i>Assessment Ethics</i>	Sumber
21.	Melindungi hak privasi siswa sebagai individu yang dinilai sebagai proses dari pengembangan nilai	(Schmeiser <i>et al.</i> , 1995)
22.	Mengevaluasi praktik penilaian yang tidak etis	(Wulan, 2018)

Data mengenai standar *assessment ethics* kemudian digunakan untuk mengembangkan instrumen yang bertujuan mengidentifikasi kegiatan pembelajaran epistemik yang dilakukan oleh siswa dan guru sebagai responden. Penelitian ini juga mempertimbangkan apakah praktik asesmen yang diterapkan telah sesuai dengan standar etika dan pedoman penilaian kinerja dalam konteks pembelajaran epistemik. Dengan demikian, diharapkan instrumen ini dapat memberikan wawasan yang lebih mendalam mengenai efektivitas dan keadilan proses penilaian kinerja pada pembelajaran sains di kelas.

3.2 Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan pada Tahun Ajaran 2023/2024 selama bulan Oktober - Desember 2024 dengan fokus terhadap pelaksanaan pembelajaran epistemik sains pada pembelajaran biologi dan pengukuran implementasi etika asesmen di sekolah-sekolah yang terletak di Kota Bandung dan Kota Cimahi. Sekolah-sekolah yang berpartisipasi dalam penelitian ini adalah yang telah terakreditasi oleh BAN-S/M Indonesia, dengan karakteristik akreditasi yang mencakup akreditasi A (unggul/sangat baik), akreditasi B (baik), dan akreditasi C (cukup baik/cukup).

3.3. Populasi dan Sampel

Populasi dalam penelitian ini terdiri dari seluruh SMA yang berada di Kota Bandung dan Cimahi. Sampel penelitian melibatkan siswa yang melakukan pembelajaran biologi dan guru yang mengajar biologi di sekolah-sekolah yang memenuhi kriteria akreditasi. Siswa yang menjadi responden adalah siswa kelas X, XI dan XII yang telah mengikuti serangkaian kegiatan pembelajaran biologi, termasuk asesmen kinerja. Responden dari kalangan guru adalah tenaga pendidik yang memfasilitasi proses belajar dan asesmen kinerja bagi siswa, dengan latar

belakang pengalaman mengajar yang bervariasi di sekolah-sekolah yang dijadikan sampel.

Sampel dalam penelitian ini ditentukan menggunakan prosedur *purposive sampling*, yaitu teknik pengambilan sampel non-probabilistik yang memilih partisipan secara sengaja berdasarkan kualifikasi tertentu dari sekolah. Kualifikasi yang dimaksud adalah akreditasi yang ditetapkan oleh BAN-S/M. Teknik ini diterapkan untuk mengidentifikasi dan memilih kelompok sampel yang memiliki karakteristik khusus terkait fenomena yang diteliti (Creswell & Guetterman, 2019).

Sebelum pengambilan sampel, dilakukan stratifikasi populasi berdasarkan karakteristik akreditasi sekolah (A, B, dan C) untuk menggambarkan variasi praktik pembelajaran epistemik dan asesmen yang berbeda, mulai dari yang sering diterapkan hingga yang jarang. Selain itu, pemilihan teknik ini juga didasarkan pada karakteristik sampel yang menjadi subjek penelitian, yaitu siswa dan guru yang telah melaksanakan pembelajaran hingga tahap penilaian. Ada 7 sekolah yang terletak di kota Bandung dan Cimahi yang dijadikan sampel dalam penelitian ini. Perincian mengenai jumlah sampel yang berkontribusi dalam penelitian ini dapat dilihat dalam berikut.

Tabel 3. 2. Sampel yang Digunakan dalam Penelitian

No.	Sekolah	Kategori	Jumlah Sampel Siswa	Jumlah Sampel Guru
1.	SMA-A-01	Unggul/Sangat Baik	65	2
2.	SMA-A-02	Unggul/Sangat Baik	60	2
3.	SMA-A-03	Unggul/Sangat Baik	98	2
4.	SMA-B-01	Baik	43	1
5.	SMA-B-02	Baik	28	1
6.	SMA-B-03	Baik	36	1
7.	SMA-C-01	Cukup Baik/Cukup	16	1
Total			346	10

3.4. Definisi Operasional

1. Pelaksanaan pembelajaran epistemik

Pelaksanaan pembelajaran epistemik adalah evaluasi terhadap bagaimana siswa membangun pengetahuan dan keterampilan ilmiah mereka dalam konteks pembelajaran Biologi. Pembelajaran epistemik yang dimaksud mencakup bagaimana siswa memperoleh, mengembangkan, menerapkan pengetahuan biologi yang diperolehnya serta mengevaluasi informasi secara kritis. Data pelaksanaan pembelajaran epistemik sains pada pembelajaran biologi didapat melalui instrumen berupa angket *Ethics of Assessment of Performance in Epistemic Biology* (EAPEB). Instrumen pelaksanaan pembelajaran epistemik mengacu pada *framework* PISA 2025 yang mencakup 3 indikator utama dan beberapa aspek pada masing-masing indikator. Jika dari hasil analisis data angket terdapat adanya temuan mengenai rendahnya pelaksanaan pembelajaran epistemik, maka dilakukan wawancara semi terstruktur kepada siswa untuk menelusuri lebih lanjut mengenai temuan tersebut.

2. Asesmen Kinerja Pembelajaran Epistemik

Asesmen kinerja adalah evaluasi sistematis untuk mengukur kemampuan individu dalam menerapkan pengetahuan, keterampilan, dan sikap dalam situasi praktis pada pelajaran biologi. Pelaksanaan asesmen kinerja yang dilakukan guru dilihat berdasarkan pengalaman siswa saat melakukan unjuk kerja pembelajaran biologi seperti, tugas proyek, diskusi, kegiatan praktikum, studi lapangan dan penggunaan rubrik penilaian asesmen kinerja oleh guru. Penelitian ini melihat apakah pelaksanaan asesmen kinerja sudah menilai epistemik siswa sesuai dengan *framework* PISA mengenai pembelajaran epistemik. Data pelaksanaan asesmen kinerja yang dilakukan guru didapat melalui instrumen berupa angket *Ethics of Assessment of Performance in Epistemic Biology* (EAPEB). Jika dari hasil analisis data angket terdapat adanya temuan mengenai rendahnya pelaksanaan asesmen kinerja pembelajaran epistemik, maka dilakukan

wawancara semi terstruktur kepada siswa untuk mengali lebih lanjut mengenai temuan tersebut

3. Etika Asesmen Kinerja

Etika dalam asesmen kinerja merujuk pada kumpulan aturan, sistem penilaian, komitmen terhadap nilai-nilai moral, dan norma-norma dalam asesmen pembelajaran epistemik. Daftar ini disusun melalui tinjauan literatur pada artikel-artikel, standar aturan khusus, dan buku-buku yang relevan. Hasil tinjauan literatur mencakup daftar pelanggaran etika asesmen dan daftar standar etika asesmen yang secara umum mencakup nilai-nilai moral atau aturan yang digunakan untuk mengatur pelaksanaan asesmen kinerja. Data pelaksanaan etika asesmen kinerja yang dilakukan guru didapat melalui instrumen berupa angket *Ethics of Assessment of Performance in Epistemic Biology* (EAPEB). Jika dari hasil analisis data angket terdapat adanya temuan pelanggaran etika pelaksanaan asesmen kinerja, maka dilakukan wawancara semi terstruktur kepada siswa untuk menelusuri lebih lanjut mengenai temuan tersebut.

3.5. Teknik Pengumpulan Data

Berikut adalah beberapa metode pengumpulan data yang digunakan dalam penelitian ini:

1. Melakukan studi literatur untuk mengumpulkan data yang diperlukan dalam mengembangkan instrumen *Ethics of Assessment of Performance in Epistemic Biology* (EAPEB). Instrumen EAPEB dirancang untuk mengumpulkan data yang relevan mengenai bagaimana pembelajaran epistemik dilaksanakan, serta untuk mengevaluasi sejauh mana praktik tersebut sesuai dengan standar etika asesmen kinerja yang telah ditetapkan (Lampiran 3).
2. Dalam penelitian ini, survei dilakukan dengan memberikan instrumen EAPEB kepada responden sampel untuk menilai pembelajaran epistemik sains, asesmen kinerja yang berkaitan dengan epistemik sains, serta etika asesmen kinerja di sekolah. Untuk meningkatkan kredibilitas data dan hasil penelitian,

proses survei dilaksanakan dengan mengikuti ketentuan operasional khusus yang dirinci pada tabel berikut.

Tabel 3. 3. 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 siswa dan responden guru.

- Melaksanakan wawancara semi-terstruktur berdasarkan analisis hasil survei untuk mengumpulkan informasi sekunder yang dapat memperkuat temuan penelitian. Wawancara ini dirancang untuk menggali lebih dalam perspektif responden dan mendapatkan wawasan tambahan yang relevan dengan isu-isu yang diidentifikasi dalam survei (Lampiran 13).

Berbagai instrumen penelitian digunakan dalam studi ini untuk mengumpulkan data yang diperlukan. Pemilihan instrumen dilakukan berdasarkan jenis data yang ingin direkam. Perincian lebih lanjut mengenai instrumen yang digunakan dapat dilihat dalam tabel berikut.

Tabel 3. 4. Instrumen Pengumpul Data Penelitian

No.	Instrumen	Teknik Pengumpulan Data	Data
1.	Lembar validasi instrumen siswa	Validasi ahli dan <i>Rasch Model</i>	Hasil validitas ahli dan data validitas aplikasi winstep
2.	Lembar validasi instrumen guru	Validasi ahli dan <i>Rasch Model</i>	Hasil validitas ahli dan data validitas aplikasi winstep
3.	Instrumen EAPEB siswa	Penyebaran angket kepada siswa	Data praktik pembelajaran epistemik, penilaian kinerja pembelajaran epistemik, dan assessment ethics penilaian kinerja
4.	Instrumen EAPEB guru	Penyebaran angket kepada guru	Data praktik pembelajaran epistemik, penilaian kinerja pembelajaran epistemik, dan assessment ethics penilaian kinerja
5.	Lembar wawancara	Wawancara semi terstruktur	Data wawancara praktik pembelajaran epistemik, penilaian kinerja pembelajaran epistemik, dan assessment ethics penilaian kinerja

3.5.1. Instrumen *Ethics of Assessment of Performance in Epistemic Biology* (EAPEB) untuk Guru dan Siswa

Instrumen EAPEB yang diberikan kepada guru dan siswa digunakan untuk mengidentifikasi isu-isu pelanggaran selama proses asesmen kinerja dan pelanggaran etika yang terjadi selama pelaksanaan asesmen serta mengidentifikasi pelaksanaan pembelajaran epistemik sains. Kisi-kisi dikembangkan berdasarkan hasil analisis tematik yang dilakukan pada *framework* sains PISA 2025 (OECD, 2023) dan daftar *assessment ethics* pada asesmen kinerja yang dikumpulkan dari berbagai sumber melalui studi literatur. Instrumen EAPEB ini memiliki empat pilihan jawaban, yaitu selalu, sering, kadang-kadang, dan tidak pernah.

Berdasarkan hasil analisis terhadap *framework* PISA 2025, terdapat tiga tema utama yang diidentifikasi: (1) menjelaskan fenomena secara ilmiah, (2) menyusun dan mengevaluasi desain penyelidikan ilmiah serta menginterpretasikan data secara kritis, serta (3) meneliti, mengevaluasi, dan menggunakan informasi ilmiah untuk pengambilan keputusan. Ketiga tema ini selanjutnya dikembangkan menjadi kisi-kisi yang disajikan dalam tabel berikut.

Tabel 3. 5. Kisi-kisi Instrumen Pembelajaran Epistemik dan Asesmen kinerja Guru dan Siswa

No.	Aspek	No.	Indikator
1.	Pelaksanaan pembelajaran epistemik	1.	Menjelaskan fenomena secara ilmiah
		2.	Menyusun dan mengevaluasi desain penyelidikan ilmiah serta menginterpretasikan data secara kritis.
		3.	Meneliti, mengevaluasi, dan menggunakan informasi ilmiah untuk pengambilan keputusan
2.	Asesmen kinerja pembelajaran epistemik	4.	Asesmen kinerja untuk menjelaskan fenomena secara ilmiah
		5.	Asesmen kinerja untuk menyusun dan mengevaluasi desain penyelidikan ilmiah serta menginterpretasikan data secara kritis.
		6.	Asesmen kinerja untuk meneliti, mengevaluasi, dan menggunakan informasi ilmiah untuk pengambilan keputusan

Pelanggaran etika asesmen dibagi menjadi tiga aspek utama: (1) persiapan dan pengembangan asesmen kinerja yang digunakan, (2) pelaksanaan asesmen kinerja, dan (3) pemrosesan serta tindak lanjut dari asesmen kinerja. perincian kisi-kisi instrumen EAPEB yang digunakan untuk mengumpulkan data dari siswa diadaptasi dari sumber-sumber berikut: Webber & Lupart (2012), Wulan (2018), dan Zlatkin-Troitschanskaia *et al.* (2019) dan Saputra (2024). Kisi-kisi instrumen *assessment ethics* guru dan siswa dapat dilihat pada tabel berikut.

Tabel 3. 6. Kisi-kisi Instrumen Assessment Ethics Guru dan Siswa

No.	Aspek	No.	Indikator
1.	Persiapan dan pengembangan instrumen asesmen kinerja biologi	1.	Metode penilaian sesuai dengan latar belakang dan pengalaman belajar siswa
		2.	Konten dan bahasa yang digunakan dalam tes tidak bersifat sensitif, seksis, atau menyinggung
		3.	Task pada asesmen kinerja mengandung unsur bias gender
		4.	Task pada asesmen kinerja bertentangan dengan norma masyarakat setempat
		5.	Task pada asesmen kinerja bertentangan dengan norma agama
		6.	Task pada asesmen kinerja bertentangan dengan budaya siswa tertentu
		7.	Task pada asesmen kinerja mengganggu keadaan psikis siswa tertentu
		8.	Rubrik yang digunakan dalam asesmen kinerja menguji “kompetensi asing” yang dapat memungkinkan terjadinya bias
2.	Pelaksanaan asesmen kinerja	9.	Asesmen kinerja dilaksanakan secara <i>fair</i> dan tidak merugikan siswa sebagai peserta
		10.	Pelaksanaan asesmen kinerja kurang mengakomodasi latar belakang sosial-ekonomi siswa
		11.	Pelaksanaan asesmen kinerja nyaman dikerjakan oleh siswa
		12.	Melaksanakan sistem hukuman melalui pengurangan nilai atas tindakan yang tidak relevan dalam pelaksanaan asesmen kinerja
		13.	Menggunakan rubrik dan instrumen asesmen kinerja yang tidak dapat diakses oleh semua siswa
3.	Pemrosesan dan tindak lanjut asesmen kinerja	14.	Menuntut siswa untuk meningkatkan kemampuan tanpa adanya umpan balik
		15.	Meluluskan siswa yang belum mencapai kriteria kelulusan
		16.	Mengambil keputusan atas dasar pertimbangan yang bersifat subjektif atau pertimbangan lain yang tidak relevan
		17.	Mengambil keputusan atas informasi yang tidak lengkap
		18.	Mengkomunikasikan hasil asesmen kinerja dengan cara menjatuhkan harga diri siswa
		19.	Membanding-bandingkan hasil asesmen kinerja siswa di depan kelas
		20.	Memberikan tanggung jawab kepada siswa dengan nilai baik untuk bekerja lebih dari siswa lainnya
		21.	Mempertontonkan hasil asesmen kinerja dengan kualitas rendah sebagai contoh buruk tanpa seizin pemiliknya
		22.	Tidak mengakomodir penafsiran skor asesmen kinerja secara individual secara detail

Indikator yang digunakan dalam pengembangan instrumen EAPEB ini seragam antara guru dan siswa, namun perbedaan terletak pada formulasi pertanyaannya. Pada instrumen untuk siswa, pertanyaan dirancang untuk menggali pengalaman mereka terkait pelaksanaan pembelajaran epistemik sains yang dilakukan oleh guru serta apakah penilaian kinerja telah dilaksanakan guru secara etis. Sementara itu, instrumen untuk guru dirancang untuk menyelidiki sejauh mana mereka telah melaksanakan pembelajaran epistemik sains kepada siswa dan menilai kinerja siswa secara etis dan sesuai standar yang berlaku. Dengan pendekatan ini, penelitian dapat memberikan gambaran holistik tentang pelaksanaan etika penilaian kinerja epistemik sains pada pembelajaran biologi.

3.6. Teknik Analisis Data

Analisis data dilakukan setelah semua data yang diperlukan dalam penelitian terkumpul, dan merupakan langkah penting yang menentukan kualitas serta keandalan temuan penelitian. Proses analisis ini terbagi menjadi dua tahap utama, yaitu uji coba kelayakan dan analisis data primer penelitian. Setiap tahap akan dibahas secara rinci pada sub-bagian berikut.

3.6.1. Analisis Uji Coba Kelayakan Instrumen

Uji coba kelayakan instrumen adalah langkah krusial yang harus dilakukan untuk memastikan bahwa instrumen yang dikembangkan dapat digunakan secara efektif dalam penelitian. Proses ini mencakup analisis validitas konstruk yang dilakukan oleh para ahli (validator), serta analisis validitas empiris yang menggunakan *Rasch Model* dalam uji coba penyebaran instrumen pada skala yang lebih kecil.

a. Analisis Uji Validitas Konstruk

Validitas konstruk bertujuan untuk menilai sejauh mana instrumen dapat mengukur konsep yang dimaksud, dengan melibatkan pendapat para ahli yang memiliki keahlian di bidang terkait. Melalui *expert judgment*, para ahli memberikan penilaian terhadap setiap item dalam instrumen, memastikan bahwa pertanyaan yang diajukan relevan dan sesuai dengan indikator yang ingin diukur. Validitas konstruk instrumen oleh ahli (validator) dilakukan sebelum instrumen disebarkan

kepada responden, baik dalam skala ukuran sampel kecil maupun skala ukuran sampel yang sebenarnya. Dalam penelitian ini, validitas ahli dilakukan oleh dua orang validator. Ahli asesmen dan ahli pembelajaran serta ahli konten (isi) dilakukan oleh kedua dosen pembimbing penulisan tesis ini.

Analisis konstruk dilakukan dengan mempresentasikan penilaian instrumen berdasarkan skor kelayakan yang diberikan oleh ahli konten (validator). Skor kelayakan tersebut kemudian dihitung persentasenya menggunakan rumus sebagai berikut:

$$\text{Persentase Kelayakan} = \left(\frac{\text{Jumlah Skor Kelayakan}}{\text{Jumlah Maksimal Skor}} \right) \times 100\%$$

Penggunaan rumus di atas dapat mengevaluasi sejauh mana instrumen tersebut memenuhi standar yang ditetapkan oleh para ahli. Proses ini penting untuk memastikan bahwa instrumen yang digunakan dalam penelitian tidak hanya valid secara teoritis, tetapi juga praktis dan dapat diterapkan dalam konteks penelitian yang lebih luas. Hasil dari analisis ini memberikan gambaran yang jelas mengenai kualitas instrumen dan membantu dalam pengambilan keputusan terkait penggunaan instrumen tersebut dalam pengumpulan data.

Pengembang instrumen dapat menentukan kriteria validitas berdasarkan jumlah item dan interval yang digunakan dalam skor validasi. Dalam konteks ini, validitas empiris diukur dengan menggunakan delapan item pernyataan yang diisi oleh validator ahli dengan skor berskala Likert dari 1 hingga 5. Menurut Yusup (2018), validitas mencakup berbagai aspek, termasuk validitas isi, konstruk, dan kriteria, yang semuanya berkontribusi pada kualitas data yang dihasilkan. Untuk memastikan bahwa setiap item dalam instrumen dapat dianggap valid, penting untuk melakukan uji validitas yang melibatkan evaluasi oleh para ahli (Dewi & Sudaryanto, 2020). Selain itu, lebar interval dapat ditentukan menggunakan rumus berikut:

$$\text{Lebar Interval} = \frac{\text{Skor Maksimal} - \text{Skor Minimal}}{\text{Jumlah Kategori}}$$

Rumus ini membantu dalam menetapkan batasan untuk setiap kategori skor validitas, sehingga memudahkan dalam interpretasi hasil dari penilaian yang

diberikan oleh validator. Dengan pendekatan ini, pengembang dapat memastikan bahwa instrumen yang digunakan memenuhi standar validitas yang telah ditentukan dan dapat memberikan data yang akurat dan reliabel dalam konteks penelitian.

Dengan menggunakan rumus tersebut, kriteria yang digunakan untuk menganalisis hasil validasi oleh para ahli dapat diperoleh sebagai berikut.

Tabel 3. 7. 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 EAPEB untuk siswa memperoleh skor sebesar 39 dengan persentase 97,5%. Dengan hasil tersebut, instrumen untuk siswa dinyatakan “sangat layak”. Namun, untuk meningkatkan kualitas instrumen tersebut, validator ahli memberikan catatan bahwa item pernyataan dalam instrumen EAPEB perlu disusun menggunakan bahasa yang sesuai dengan jenjang siswa di SMA.

Instrumen EAPEB 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 perlu direvisi pada beberapa bagian. Beberapa saran revisi tersebut meliputi meringkas kalimat agar tidak terlalu panjang, memperbaiki kalimat agar relevan dengan aspek atau indikator yang diukur, serta mencegah kemungkinan kalimat untuk dipilih seluruhnya oleh guru sebagai responden. Hal ini penting untuk memastikan bahwa instrumen tidak hanya valid secara teoritis, tetapi juga praktis dan efektif dalam mengukur kompetensi yang diharapkan dari para guru.

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

Pengujian kelayakan instrumen yang dikembangkan sangat penting untuk memastikan efektivitasnya. Dalam penelitian ini, validitas dan reliabilitas instrumen dianalisis menggunakan *Rasch Model* dengan bantuan aplikasi Winstep.

Rasch Model, yang diperkenalkan oleh ahli matematika Georg Rasch, berfungsi untuk menganalisis probabilitas dari pilihan yang tersedia dalam instrumen. Prinsip dasar dari model ini bersifat probabilistik, di mana individu dengan kemampuan lebih tinggi dibandingkan orang lain diharapkan memiliki peluang lebih besar untuk menjawab soal dengan benar.

Rasch Model tidak hanya menguji validitas dan reliabilitas, tetapi juga dapat mengevaluasi kesesuaian antara person dan item secara simultan, memberikan informasi yang lebih mendalam tentang kualitas instrumen yang digunakan. Item yang dianggap sulit cenderung menurunkan kemungkinan individu untuk memberikan jawaban yang benar, sehingga penting untuk mempertimbangkan aspek ini dalam pengembangan instrumen (Sumintono & Widhiarso, 2014). Dengan demikian, penerapan *Rasch Model* dalam analisis instrumen dapat meningkatkan kualitas hasil penelitian secara keseluruhan.

Uji coba dilaksanakan di dua kelas yang terletak di Kota Bandung dan Cimahi. Proses uji coba ini juga mempertimbangkan pembagian strata yang diterapkan dalam teknik sampling. Oleh karena itu, uji coba dilakukan pada kelas-kelas dengan akreditasi sekolah A dan B. Sekolah-sekolah yang memiliki akreditasi C tidak dilibatkan dalam uji coba karena terbatasnya jumlah responden yang memenuhi kriteria tersebut.

Data untuk validitas empiris dianalisis setelah uji coba terbatas dilakukan pada populasi yang telah ditentukan dalam penelitian. Uji coba melibatkan 38 siswa dan 7 guru mata pelajaran Biologi. Analisis validitas empiris dilakukan dengan menggunakan aplikasi *WinStep* yang menerapkan *Rasch Model*.

Pengukuran validitas empiris dilakukan dengan mempertimbangkan nilai *logit* pada *outfit MNSQ*, *outfit ZSTD*, dan *PT-Measure-Al Correlation*. *PT-Measure Correlation* umumnya digunakan untuk mengevaluasi daya beda suatu instrumen. Dalam hal ini, validitas empiris dari kuesioner yang menggunakan respons skala Likert hanya mengacu pada *outfit MNSQ* dan *outfit ZSTD*, dengan syarat nilai yang diterima adalah: $0,5 < MNSQ < 1,5$ dan $-2,0 < ZSTD < +2,0$. Apabila salah satu kriteria tersebut terpenuhi, maka item tersebut dinyatakan layak untuk digunakan dalam penelitian (Sumintono & Widhiarso, 2015).

Reliabilitas atau konsistensi butir pernyataan pada instrumen dianalisis menggunakan *Rasch Model* melalui aplikasi Winstep. Sebuah instrumen dinyatakan reliabel berdasarkan nilai *Item Reliability* dan *Person Reliability*. *Item reliability* mengacu pada tingkat keandalan butir soal dalam mengukur konsep yang dimaksud, sedangkan *person reliability* menunjukkan konsistensi dan keandalan skor yang diperoleh dari individu (responden). Secara umum, nilai reliabilitas item dan responden dianggap baik jika mencapai minimal 0.8 atau lebih.

Berikut adalah analisis validitas empiris dan reliabilitas untuk masing-masing responden, baik siswa maupun guru:

1. Hasil Analisis Validitas & Reliabilitas Instrumen EAPB Siswa

Tabel berikut menunjukkan *output item misfit order* dari hasil analisis data uji coba siswa yang dilakukan menggunakan aplikasi Winstep.

Tabel 3. 8. Ouput Item Misfit Order Instrumen Siswa

<i>Item</i>	<i>Outfit MNSQ</i>	<i>Interpretasi Outfit MNSQ</i>	<i>Outfit ZSTD</i>	<i>Interpretasi Outfit ZSTD</i>	Keputusan
E60	1.42	Terpenuhi	1.43	Terpenuhi	Digunakan
E56	1.27	Terpenuhi	1.05	Terpenuhi	Digunakan
C36	1.49	Terpenuhi	1.79	Terpenuhi	Digunakan
C31	1.51	Belum terpenuhi	1.9	Terpenuhi	Digunakan
C35	1.39	Terpenuhi	1.51	Terpenuhi	Digunakan
E59	1.09	Terpenuhi	0.44	Terpenuhi	Digunakan
C34	1.23	Terpenuhi	0.99	Terpenuhi	Digunakan
C37	1.52	Belum terpenuhi	2	Belum terpenuhi	Diperbaiki
C33	1.35	Terpenuhi	1.43	Terpenuhi	Digunakan
C38	0.91	Terpenuhi	-0.33	Terpenuhi	Digunakan
E62	1.44	Terpenuhi	1.75	Terpenuhi	Digunakan
C28	0.82	Terpenuhi	-0.76	Terpenuhi	Digunakan
D43	2.19	Belum terpenuhi	3.99	Belum terpenuhi	Diperbaiki
E63	1	Terpenuhi	0.08	Terpenuhi	Digunakan
C39	1.29	Terpenuhi	1.25	Terpenuhi	Digunakan
D44	0.96	Terpenuhi	-0.12	Terpenuhi	Digunakan
E58	1.23	Terpenuhi	1.02	Terpenuhi	Digunakan
E50	0.96	Terpenuhi	-0.09	Terpenuhi	Digunakan
D46	0.74	Terpenuhi	-1.23	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>
E49	1.21	Terpenuhi	0.95	Terpenuhi	Digunakan
D47	0.98	Terpenuhi	-0.04	Terpenuhi	Digunakan
D45	1.25	Terpenuhi	1.11	Terpenuhi	Digunakan
D48	2.45	Belum terpenuhi	4.81	Belum terpenuhi	Diperbaiki
E57	0.92	Terpenuhi	-0.3	Terpenuhi	Digunakan
C32	0.95	Terpenuhi	-0.17	Terpenuhi	Digunakan
E54	1.02	Terpenuhi	0.16	Terpenuhi	Digunakan
E55	1.02	Terpenuhi	0.18	Terpenuhi	Digunakan
E61	1.16	Terpenuhi	0.77	Terpenuhi	Digunakan
E53	1.23	Terpenuhi	1.06	Terpenuhi	Digunakan
B26	1.19	Terpenuhi	0.91	Terpenuhi	Digunakan
A5	0.68	Terpenuhi	-1.64	Terpenuhi	Digunakan
A18	0.78	Terpenuhi	-1.08	Terpenuhi	Digunakan
B21	0.67	Terpenuhi	-1.72	Terpenuhi	Digunakan
A8	0.8	Terpenuhi	-0.97	Terpenuhi	Digunakan
B25	0.67	Terpenuhi	-1.76	Terpenuhi	Digunakan
A3	0.72	Terpenuhi	-1.47	Terpenuhi	Digunakan
D41	0.62	Terpenuhi	-2.1	Belum terpenuhi	Digunakan
B20	0.71	Terpenuhi	-1.52	Terpenuhi	Digunakan
A17	1.18	Terpenuhi	0.89	Terpenuhi	Digunakan
B22	0.66	Terpenuhi	-1.86	Terpenuhi	Digunakan
A10	0.54	Terpenuhi	-2.69	Belum terpenuhi	Digunakan
A14	1.21	Terpenuhi	1.03	Terpenuhi	Digunakan
B27	0.81	Terpenuhi	-0.93	Terpenuhi	Digunakan
D42	1.54	Belum terpenuhi	2.38	Belum terpenuhi	Diperbaiki
E52	1.6	Terpenuhi	2.59	Belum terpenuhi	Digunakan
A4	0.79	Terpenuhi	-1.06	Terpenuhi	Digunakan
B23	0.61	Terpenuhi	-2.24	Belum terpenuhi	Digunakan
A9	0.58	Terpenuhi	-2.39	Belum terpenuhi	Digunakan
A7	0.78	Terpenuhi	-1.14	Terpenuhi	Digunakan
P19	0.76	Terpenuhi	-1.26	Terpenuhi	Digunakan
A12	0.73	Terpenuhi	-1.46	Terpenuhi	Digunakan
B24	0.69	Terpenuhi	-1.7	Terpenuhi	Digunakan
C29	1.17	Terpenuhi	0.87	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>
A1	0.69	Terpenuhi	-1.71	Terpenuhi	Digunakan
A2	0.75	Terpenuhi	-1.29	Terpenuhi	Digunakan
A13	0.75	Terpenuhi	-1.3	Terpenuhi	Digunakan
E51	1.07	Terpenuhi	0.4	Terpenuhi	Digunakan
A6	0.75	Terpenuhi	-1.32	Terpenuhi	Digunakan
D40	0.78	Terpenuhi	-1.12	Terpenuhi	Digunakan
A11	0.9	Terpenuhi	-0.44	Terpenuhi	Digunakan
A16	0.59	Terpenuhi	-2.35	Belum Terpenuhi	Digunakan
A15	0.82	Terpenuhi	-0.9	Terpenuhi	Digunakan
C30	0.82	Terpenuhi	-0.9	Terpenuhi	Digunakan

Pada Tabel 3.8, terlihat bahwa empat soal belum memenuhi kriteria yang ditetapkan dan perlu ditinjau kembali, yaitu C37, D43, D48, D42. soal D43, yang berbunyi " LKPD dalam kegiatan praktik hanya dapat dilaksanakan oleh siswa dengan kondisi sosial-ekonomi tertentu" dan soal D48, yang berbunyi "Terdapat siswa yang tidak memiliki gawai (*handphone*/laptop/komputer) untuk mengakses LKPD secara daring", tidak memenuhi salah satu dari dua kriteria yang ada. Nilai *MNSQ Outfit* dan *ZSTD Outfit* untuk soal ini sangat tinggi, yang menunjukkan bahwa soal D43 dan D48 kurang jelas dan berpotensi menimbulkan ambiguitas. Oleh karena itu, diperlukan perbaikan dalam penyusunan kalimat item agar lebih mudah dipahami sehingga dapat digunakan.

Pada soal C37 "LKPD dalam kegiatan praktik membuat saya merasa tertekan atau stres" dan soal D42 "Saat kegiatan praktik berlangsung, guru membimbing siswa dengan baik sehingga tidak ada siswa yang bermain-main" belum memenuhi nilai kriteria dimana nilai *MNSQ Outfit* dan *ZSTD Outfit* tinggi. Oleh karena itu, penting untuk menganalisis konten soal dan cara pemahaman responden terhadap soal. Keputusan dalam penelitian ini adalah untuk tetap menggunakan semua item pertanyaan setelah mempertimbangkan hasil analisis tersebut. Mungkin diperlukan revisi kecil pada pernyataan soal untuk meningkatkan kejelasan dan pemahaman.

Analisis reliabilitas pada instrumen yang dikembangkan dilakukan dengan menggunakan aplikasi Winstep. Hasil dari analisis ini, yang ditampilkan dalam Gambar 3.2, menunjukkan nilai reliabilitas tes Cronbach Alpha (KR-20) sebesar 0,92. Nilai Cronbach Alpha yang mencapai 0,92 ini mengindikasikan bahwa instrumen tes memiliki tingkat reliabilitas yang sangat baik (Sumintono & Widhiarso, 2015).

Menurut Budiastusi dan Agustinus (2018), nilai Cronbach Alpha di atas 0,80 menunjukkan bahwa instrumen tersebut dapat diandalkan dan konsisten dalam pengukuran. Selain itu, penelitian oleh Hayati dan Lailatussaadah (2016) juga mendukung pernyataan ini dengan menyatakan bahwa reliabilitas yang tinggi pada instrumen tes sangat penting untuk memastikan bahwa hasil pengukuran dapat dipercaya dan memberikan informasi yang akurat mengenai kemampuan peserta.

Nilai Person Reliability yang diperoleh adalah 0,90, yang menunjukkan bahwa instrumen yang dikembangkan akan tetap stabil dan konsisten jika diterapkan pada sampel lain dalam populasi. Sementara itu, nilai Item Reliability yang didapat cukup tinggi, yaitu 0,92. Hal ini mengindikasikan bahwa item-item dalam instrumen tersebut memiliki keajekan yang baik dan dapat diandalkan untuk mengukur konsep yang diinginkan. Dengan demikian, kedua nilai reliabilitas ini menegaskan bahwa instrumen yang dikembangkan memenuhi standar yang diperlukan untuk digunakan dalam penelitian lebih lanjut.

TABLE 3.1 UJI VALIDITAS.xlsx ZOU062WS.TXT Nov 13 2024 20:03
 INPUT: 38 PERSON 63 ITEM REPORTED: 38 PERSON 63 ITEM 4 CATS WINSTEPS 5.7.1.0

SUMMARY OF 38 MEASURED PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	142.6	63.0	-.38	.17	.99	-.37	1.02	-.18
SEM	3.4	.0	.10	.00	.08	.45	.08	.45
P.SD	20.9	.0	.59	.01	.49	2.76	.52	2.73
S.SD	21.2	.0	.60	.01	.50	2.80	.52	2.77
MAX.	208.0	63.0	1.50	.19	2.40	6.56	2.73	7.59
MIN.	112.0	63.0	-1.29	.16	.31	-5.20	.41	-4.06
REAL RMSE	.18	TRUE SD	.56	SEPARATION	3.02	PERSON RELIABILITY .90		
MODEL RMSE	.17	TRUE SD	.56	SEPARATION	3.30	PERSON RELIABILITY .92		
S.E. OF PERSON MEAN = .10								

PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00

CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .92 SEM = 5.94
 STANDARDIZED (50 ITEM) RELIABILITY = .90

SUMMARY OF 63 MEASURED ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	86.0	38.0	.00	.22	1.05	.01	1.02	-.07
SEM	2.3	.0	.11	.00	.05	.22	.05	.20
P.SD	18.0	.0	.87	.03	.41	1.71	.37	1.56
S.SD	18.2	.0	.88	.03	.42	1.73	.37	1.58
MAX.	112.0	38.0	1.94	.31	2.56	5.08	2.45	4.81
MIN.	52.0	38.0	-1.17	.20	.53	-2.79	.54	-2.69
REAL RMSE	.25	TRUE SD	.84	SEPARATION	3.39	ITEM	RELIABILITY	.92
MODEL RMSE	.22	TRUE SD	.85	SEPARATION	3.78	ITEM	RELIABILITY	.93
S.E. OF ITEM MEAN = .11								

ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00

Global statistics: please see Table 44.
 UMEAN=.0000 USCALE=1.0000

Gambar 3. 2. Hasil Reliabilitas Instrumen Siswa

2. Hasil Analisis Validitas & Reliabilitas Instrumen EAPEB Guru

Respons hasil uji coba pada guru dianalisis menggunakan aplikasi Winstep. Validitas empiris diperoleh dari hasil *item measure* yang ditampilkan pada gambar 3.3.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-AL CORR.	EXP.	EXACT MATCH OBS% EXP%	ITEM
62	7	7	6.81	1.85	MAXIMUM MEASURE				.00	.00	100.0 100.0	P62
43	9	7	3.95	.77	2.57	1.83	5.75	3.27	-.69	.27	85.7 74.0	P43
56	9	7	3.95	.77	2.57	1.83	5.75	3.27	-.69	.27	85.7 74.0	P56
58	10	7	3.44	.66	1.98	1.49	3.69	2.76	-.43	.31	71.4 63.3	P58
60	10	7	3.44	.66	1.78	1.28	2.23	1.64	-.05	.31	71.4 63.3	P60
47	11	7	3.05	.60	1.98	1.62	1.59	1.08	-.67	.35	57.1 54.7	P47
48	12	7	2.72	.56	.59	-.79	.70	-.47	-.02	.37	42.9 50.9	P48
59	13	7	2.42	.53	1.24	.62	1.64	1.26	-.40	.39	57.1 48.4	P59
46	14	7	2.15	.51	.78	-.34	.84	-.20	-.09	.41	57.1 47.3	P46
61	14	7	2.15	.51	2.01	1.88	2.08	1.89	-.15	.41	71.4 47.3	P61
45	16	7	1.65	.50	.49	-1.19	.49	-1.19	.15	.42	71.4 46.8	P45
49	17	7	1.40	.49	1.53	1.11	1.50	1.06	.08	.42	14.3 47.6	P49
55	17	7	1.40	.49	2.25	2.11	2.28	2.15	-.57	.42	14.3 47.6	P55
4	18	7	1.16	.50	.14	-2.86	.13	-2.91	.82	.42	100.0 47.5	P4
2	19	7	.91	.50	.41	-1.43	.42	-1.38	.77	.42	71.4 47.9	P2
57	19	7	.91	.50	1.93	1.69	1.86	1.59	.13	.42	14.3 47.9	P57
15	20	7	.65	.51	.86	-.13	.84	-.18	.59	.42	42.9 48.6	P15
51	20	7	.65	.51	2.09	1.87	2.09	1.88	.30	.42	28.6 48.6	P51
63	20	7	.65	.51	2.04	1.81	2.24	2.06	-.57	.42	57.1 48.6	P63
10	21	7	.39	.52	.43	-1.31	.46	-1.23	.93	.41	57.1 47.8	P10
20	21	7	.39	.52	.43	-1.31	.46	-1.23	.93	.41	57.1 47.8	P20
22	21	7	.39	.52	1.32	.76	1.39	.86	-.07	.41	57.1 47.8	P22
17	22	7	.10	.54	1.45	.95	1.35	.80	.64	.40	57.1 49.0	P17
21	22	7	.10	.54	.29	-1.85	.32	-1.73	.94	.40	71.4 49.0	P21
23	22	7	.10	.54	.77	-.34	.75	-.38	.82	.40	42.9 49.0	P23
25	22	7	.10	.54	.39	-1.42	.43	-1.30	.78	.40	71.4 49.0	P25
54	22	7	.10	.54	1.78	1.43	1.99	1.71	-.17	.40	42.9 49.0	P54
5	23	7	-.21	.57	1.45	.94	1.74	1.34	-.22	.38	57.1 52.7	P5
8	23	7	-.21	.57	.44	-1.21	.51	-1.00	.34	.38	71.4 52.7	P8
9	23	7	-.21	.57	.65	-.59	.64	-.62	.76	.38	71.4 52.7	P9
14	23	7	-.21	.57	.77	-.32	.79	-.27	.64	.38	71.4 52.7	P14
16	23	7	-.21	.57	.65	-.59	.64	-.62	.76	.38	71.4 52.7	P16
26	23	7	-.21	.57	.32	-1.64	.39	-1.37	.57	.38	71.4 52.7	P26
27	23	7	-.21	.57	.77	-.32	.79	-.27	.64	.38	71.4 52.7	P27
50	23	7	-.21	.57	1.23	.59	1.56	1.10	.03	.38	57.1 52.7	P50
52	23	7	-.21	.57	1.36	.80	1.40	.86	.51	.38	42.9 52.7	P52
53	23	7	-.21	.57	.44	-1.20	.45	-1.19	.96	.38	71.4 52.7	P53
7	24	7	-.55	.61	.28	-1.73	.31	-1.56	.86	.36	100.0 56.1	P7
12	24	7	-.55	.61	1.25	.61	1.06	.30	.40	.36	71.4 56.1	P12
18	24	7	-.55	.61	.81	-.20	.70	-.42	.78	.36	71.4 56.1	P18
30	24	7	-.55	.61	.69	-.45	.62	-.63	.89	.36	71.4 56.1	P30
11	25	7	-.97	.68	1.60	1.06	1.32	.67	.32	.33	57.1 62.9	P11
13	25	7	-.97	.68	.75	-.28	.64	-.47	.37	.33	57.1 62.9	P13
24	25	7	-.97	.68	.55	-.73	.59	-.59	.56	.33	85.7 62.9	P24
28	25	7	-.97	.68	.40	-1.16	.39	-1.09	.82	.33	85.7 62.9	P28
29	25	7	-.97	.68	1.01	.21	.78	-.18	.81	.33	85.7 62.9	P29
39	25	7	-.97	.68	.45	-1.01	.44	-.95	.75	.33	85.7 62.9	P39
41	25	7	-.97	.68	.64	-.52	.81	-.13	.35	.33	85.7 62.9	P41
1	26	7	-1.49	.79	.81	-.07	1.15	.44	.15	.28	85.7 72.5	P1
3	26	7	-1.49	.79	.88	-.04	.76	-.10	.28	.28	57.1 72.5	P3
6	26	7	-1.49	.79	.52	-.65	.43	-.72	.77	.28	85.7 72.5	P6
19	26	7	-1.49	.79	.73	-.22	.79	-.04	.38	.28	85.7 72.5	P19
32	26	7	-1.49	.79	.52	-.65	.43	-.72	.77	.28	85.7 72.5	P32
35	26	7	-1.49	.79	.52	-.65	.43	-.72	.77	.28	85.7 72.5	P35
38	26	7	-1.49	.79	.82	-.06	1.20	.50	.13	.28	85.7 72.5	P38
40	26	7	-1.49	.79	.82	-.06	1.20	.50	.13	.28	85.7 72.5	P40
44	26	7	-1.49	.79	.52	-.65	.43	-.72	.77	.28	85.7 72.5	P44
31	27	7	-2.31	1.06	.64	-.12	.38	-.38	.69	.21	85.7 86.2	P31
33	27	7	-2.31	1.06	.64	-.12	.38	-.38	.69	.21	85.7 86.2	P33
34	27	7	-2.31	1.06	.64	-.12	.38	-.38	.69	.21	85.7 86.2	P34
36	27	7	-2.31	1.06	.64	-.12	.38	-.38	.69	.21	85.7 86.2	P36
37	27	7	-2.31	1.06	.64	-.12	.38	-.38	.69	.21	85.7 86.2	P37
42	27	7	-2.31	1.06	.64	-.12	.38	-.38	.69	.21	85.7 86.2	P42
MEAN	21.3	7.0	.10	.68	.98	-.03	1.11	.07			68.4 59.8	
P. SD	5.3	.0	1.75	.22	.61	1.10	1.08	1.25			19.6 12.4	

Gambar 3. 3. Hasil Validasi Instrumen Guru

Beberapa item soal memiliki nilai *Outfit MNSQ* dan *Outfit ZSTD* yang tidak sesuai dengan kriteria ideal seperti ditampilkan pada Tabel 3.12.

Tabel 3. 9. Item Soal Tidak Sesuai Kriteria

Soal	<i>Outfit MNSQ</i>	<i>Outfit ZSTD</i>	Keterangan	Keputusan
D43	5.75	3.27	<i>Misfit</i> signifikan (<i>MNSQ</i> tinggi).	Diperbaiki
E56	5.75	3.27	<i>Misfit</i> signifikan (<i>MNSQ</i> tinggi).	Diperbaiki
E55	2.28	2.15	<i>Misfit</i> , variasi terlalu besar.	Diperbaiki
E62	0.00	0.00	<i>Overfit</i> , jawaban terlalu seragam.	Dibuang

Soal D43, E56, E55 menunjukkan *Misfit* yang signifikan yang terindikasi bahwa soal menimbulkan ambiguitas bagi responden sehingga soal tersebut perlu direvisi untuk memperjelas redaksinya. Soal E62 menunjukkan *Overfit* dimana pola jawaban responden terlalu seragam sehingga soal ini dibuang untuk memastikan instrumen yang digunakan sesuai dengan tujuan pengukuran dan layak digunakan.

Untuk mengevaluasi reliabilitas instrumen yang dikembangkan untuk guru, penelitian ini memeriksa nilai Cronbach Alpha (K-20) yang ditampilkan pada *output* aplikasi Winstep. Nilai Cronbach Alpha (K-20) yang diperoleh adalah 0,88, yang menunjukkan bahwa instrumen tersebut memiliki reliabilitas yang sangat baik. Hal ini didukung oleh nilai Person Reliability sebesar 0,87, yang dikategorikan baik, serta Item Reliability sebesar 0,78, yang dikategorikan cukup baik.

TABLE 3.1 Angket Penelitian-guru.xlsx ZOU734WS.TXT Dec 18 2024 21:24
 INPUT: 7 PERSON 63 ITEM REPORTED: 7 PERSON 63 ITEM 4 CATS WINSTEPS 5.7.1.0

SUMMARY OF 7 MEASURED PERSON								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	192.1	63.0	1.25	.21	1.02	.00	1.11	.26
SEM	5.9	.0	.25	.01	.12	.64	.24	1.03
P.SD	14.5	.0	.60	.01	.29	1.57	.59	2.53
S.SD	15.7	.0	.65	.02	.31	1.70	.64	2.73
MAX.	210.0	63.0	2.03	.23	1.47	2.34	2.40	5.38
MIN.	166.0	63.0	.23	.19	.60	-2.45	.57	-2.46

REAL RMSE	.22	TRUE SD	.56	SEPARATION	2.54	PERSON RELIABILITY		.87
MODEL RMSE	.21	TRUE SD	.57	SEPARATION	2.72	PERSON RELIABILITY		.88
S.E. OF PERSON MEAN = .25								

PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00								
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .88 SEM = 4.94								
STANDARDIZED (50 ITEM) RELIABILITY = .85								

SUMMARY OF 62 MEASURED ITEM								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	21.6	7.0	.00	.66	.98	-.03	1.11	.07
SEM	.6	.0	.20	.02	.08	.14	.14	.16
P.SD	5.0	.0	1.60	.16	.61	1.10	1.08	1.25
S.SD	5.1	.0	1.61	.16	.62	1.11	1.09	1.26
MAX.	27.0	7.0	3.95	1.06	2.57	2.11	5.75	3.27
MIN.	9.0	7.0	-2.31	.49	.14	-2.86	.13	-2.91

REAL RMSE	.74	TRUE SD	1.42	SEPARATION	1.91	ITEM RELIABILITY		.78
MODEL RMSE	.68	TRUE SD	1.45	SEPARATION	2.14	ITEM RELIABILITY		.82
S.E. OF ITEM MEAN = .20								

MINIMUM EXTREME SCORE: 1 ITEM 1.6%								

Gambar 3. 4. Hasil Reliabilitas Instrumen Guru

3.6.2. Analisis Data Penelitian

Data yang digunakan dalam penelitian ini dibagi menjadi dua kategori, yaitu data primer dan data sekunder. Pengategorian ini didasarkan pada jenis data dan metode analisis yang digunakan. Secara umum, data sekunder dalam penelitian ini mencakup hasil studi literatur yang dianalisis pada tahap awal penelitian, serta data wawancara yang dilakukan setelah survei. Sementara itu, data primer terdiri dari data numerik yang diperoleh melalui penyebaran instrumen EAPEB kepada responden. Dengan demikian, kedua jenis data ini saling melengkapi untuk memberikan pemahaman yang lebih komprehensif mengenai topik yang diteliti.

a. Analisis Data Sekunder

Data sekunder dalam penelitian ini mencakup hasil kajian literatur tentang pembelajaran epistemik berdasarkan tinjauan *Framework PISA 2025*, asesmen epistemik, serta standar etika asesmen dalam konteks asesmen kinerja, dan data wawancara semi terstruktur. Data tersebut dianalisis menggunakan metode analisis kualitatif tematik. Analisis tematik merupakan metode yang digunakan untuk mengidentifikasi pola atau tema dari data yang telah dikumpulkan oleh peneliti. Menurut Fereday dan Muir-Cochrane (2006), analisis tematik adalah pendekatan yang efektif untuk mengeksplorasi data kualitatif secara mendalam dan menemukan keterkaitan pola-pola dalam fenomena yang diteliti.

Berikut adalah tiga prosedur dalam melakukan analisis tematik berdasarkan Braune dan Clarke (2006):

1. Membiasakan Diri dengan Data

Langkah pertama dalam analisis tematik adalah membiasakan diri dengan data yang telah dikumpulkan. Menurut Braun dan Clarke (2006), peneliti perlu membaca dan meninjau data secara menyeluruh untuk mendapatkan pemahaman yang lebih baik tentang konten dan konteksnya. Ini melibatkan mendengarkan atau membaca data beberapa kali untuk mengidentifikasi konsep, ide, atau pola yang muncul. Pencatatan detail selama tahap ini sangat penting untuk membantu dalam identifikasi tema di langkah-langkah selanjutnya.

2. Membuat Kode Awal

Setelah peneliti merasa akrab dengan data, langkah berikutnya adalah membuat kode awal. Ini melibatkan penandaan teks dengan kata atau frasa yang relevan yang mencerminkan esensi dari konten. Kode-kode ini harus singkat dan deskriptif, serta berkaitan erat dengan isi data. Pada tahap ini, penting untuk memberi kode pada semua aspek data yang berhubungan dengan pertanyaan penelitian, sehingga semua informasi yang relevan dapat teridentifikasi.

3. Mencari dan Mengembangkan Tema

Setelah proses pengkodean, peneliti kemudian beralih untuk mencari tema dari kode-kode yang telah dibuat. Tema ini harus mencerminkan pola penting dalam

data yang relevan dengan pertanyaan penelitian. Peneliti perlu meninjau dan menyempurnakan tema-tema tersebut untuk memastikan bahwa mereka menangkap makna esensial dari data. Pada tahap ini, peneliti dapat mendefinisikan dan memberi nama pada tema-tema tersebut sebelum melanjutkan ke analisis dan pelaporan temuan.

b. Analisis Data Primer

Data utama atau primer yang diperoleh dari sampel berbentuk kuantitatif, sehingga memerlukan analisis statistik untuk diolah secara sistematis dan menghasilkan informasi yang bermakna. Pengolahan data dimulai dengan pemberian skor pada setiap butir pertanyaan yang terdapat dalam instrumen penelitian. Proses pemberian skor ini bertujuan untuk mengukur tingkat respons secara objektif sesuai dengan panduan yang telah ditetapkan. Panduan ini dirancang agar penilaian dapat dilakukan secara konsisten dan mencerminkan persepsi responden terhadap pertanyaan yang diajukan. Skor yang dihasilkan nantinya digunakan untuk mendukung proses analisis statistik lebih lanjut, serta penyajian hasil penelitian dalam bentuk yang mudah dipahami. Panduan pemberian skor secara rinci akan dijelaskan pada bagian berikut.

Tabel 3. 10. Panduan Pemberian Skor Instrumen EAPEB

Pilihan Jawaban	Skor
A	4
B	3
C	2
D	1

Penelitian ini bertujuan untuk mengeksplorasi fenomena terkait pelaksanaan asesmen kinerja dari perspektif siswa dan guru. Selanjutnya, analisis statistik deskriptif diterapkan untuk menganalisis praktik *Assessment Ethics* berdasarkan data yang diperoleh. Menurut Ridwan (2020), statistik deskriptif berfungsi untuk memberikan gambaran menyeluruh tentang data, termasuk informasi mengenai nilai rata-rata (*mean*), standar deviasi, varians, nilai maksimum dan minimum, total, serta rentang data.

Data kuantitatif dianalisis dengan menggunakan metode statistik deskriptif, mencakup perhitungan rata-rata (*mean*), persentase, serta interpretasi data berdasarkan kategori interval yang telah ditetapkan. Proses ini bertujuan untuk memberikan wawasan yang jelas dan mendetail mengenai pola distribusi data. Analisis ini digunakan untuk memahami secara menyeluruh pelaksanaan pembelajaran epistemik, strategi asesmen yang diterapkan dalam pembelajaran tersebut, serta penerapan etika asesmen (*assessment ethics*) dalam konteks pendidikan.

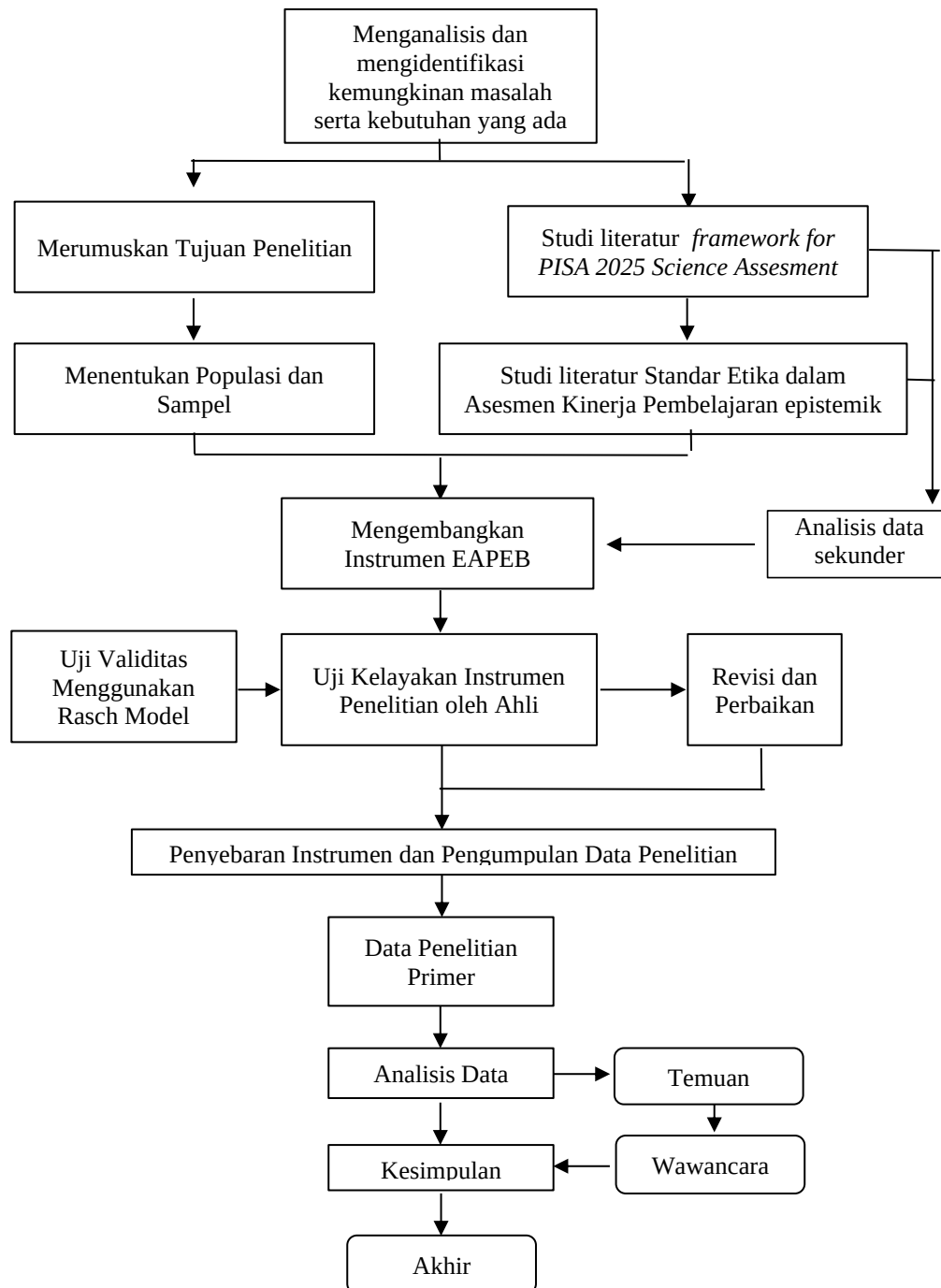
Pelaksanaan dinilai berdasarkan rata-rata skor yang disesuaikan dengan panduan dari Ridwan (2020). Hasil analisis kuantitatif terkait penilaian tersebut disajikan dalam Tabel 3.14.

Tabel 3. 11. Interpretasi Skala Penilaian

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

Berdasarkan informasi yang berhasil dikumpulkan, data kemudian diolah dan divisualisasikan untuk mempermudah pemahaman dan interpretasi hasil penelitian. Dengan menggunakan visualisasi ini, diharapkan pembaca dapat dengan mudah melihat pola dan tren yang muncul dari data yang ada. Hasil penelitian disajikan dalam bentuk nilai rata-rata dan persentase, yang kemudian dikelompokkan berdasarkan lebar interval dan rentang pada setiap aspek yang diteliti. Pendekatan ini memungkinkan pembaca untuk melihat distribusi data secara lebih jelas. Visualisasi ini tidak hanya membantu dalam menyajikan data secara lebih terstruktur dan sistematis, tetapi juga bertujuan memberikan gambaran yang lebih menyeluruh mengenai implementasi pembelajaran epistemik, asesmen kinerja dalam pembelajaran epistemik, dan penerapan etika asesmen (*assessment ethics*), yang menjadi temuan utama dalam penelitian ini. Dengan demikian, diharapkan hasil penelitian ini dapat dipahami dengan lebih mudah dan memberikan wawasan yang mendalam tentang praktik yang diteliti.

3.7. Alur Penelitian



Gambar 3. 5. Alur Penelitian