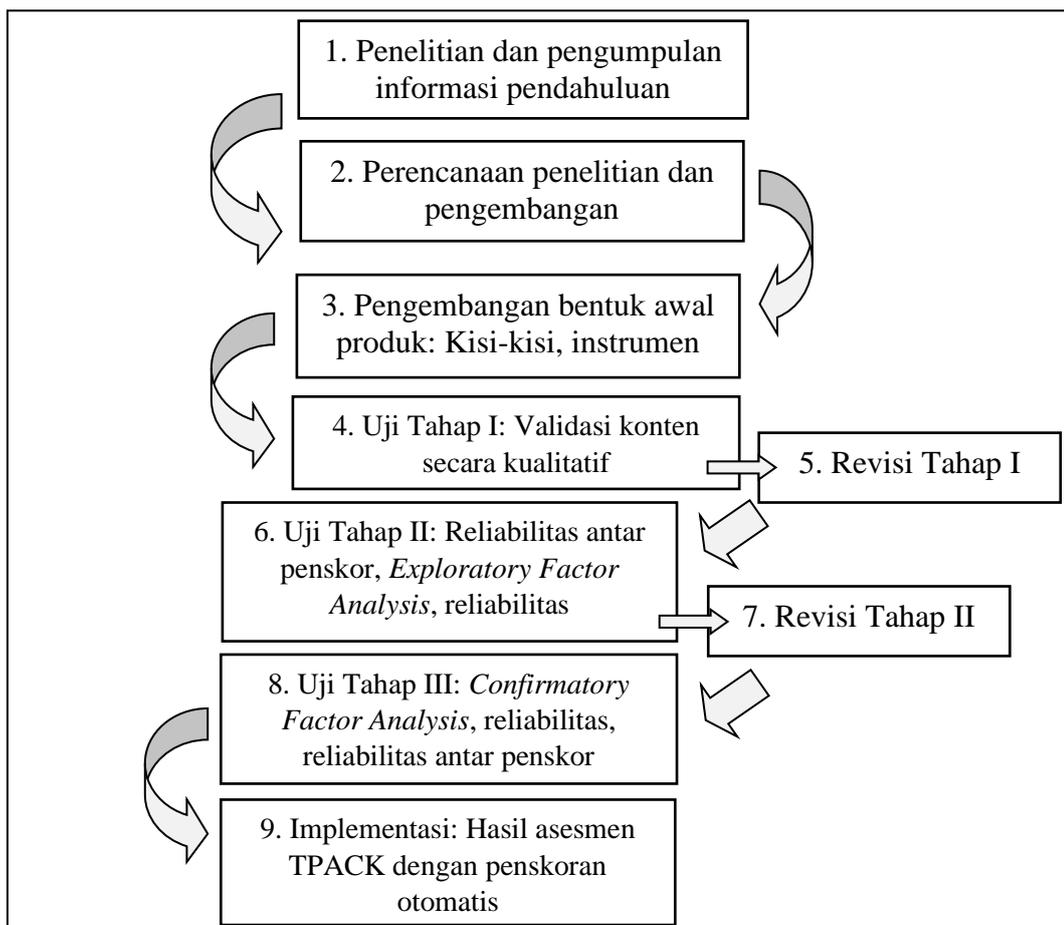


BAB III METODE PENELITIAN

3.1 Desain Penelitian

Penelitian ini dilakukan untuk menghasilkan instrumen yang dikembangkan melalui tahapan penelitian untuk pengujian kapabilitas dan kualitasnya. Berdasarkan tujuan tersebut, maka penelitian dilakukan dengan mengadaptasi metode penelitian dan pengembangan (*research and development/R and D*) dalam pendidikan menurut Borg dan Gall (1983). Metode ini telah dipilih karena komprehensif dan dapat diterapkan pada pengembangan produk secara umum, termasuk untuk pengembangan instrumen dan penggunaan suatu perangkat. Desain penelitian telah disesuaikan dengan kondisi dan konteks penelitian yang berbeda dari desain asli pada literatur. Desain penelitian yang telah dilakukan ditunjukkan pada Gambar 3.1 berikut ini.



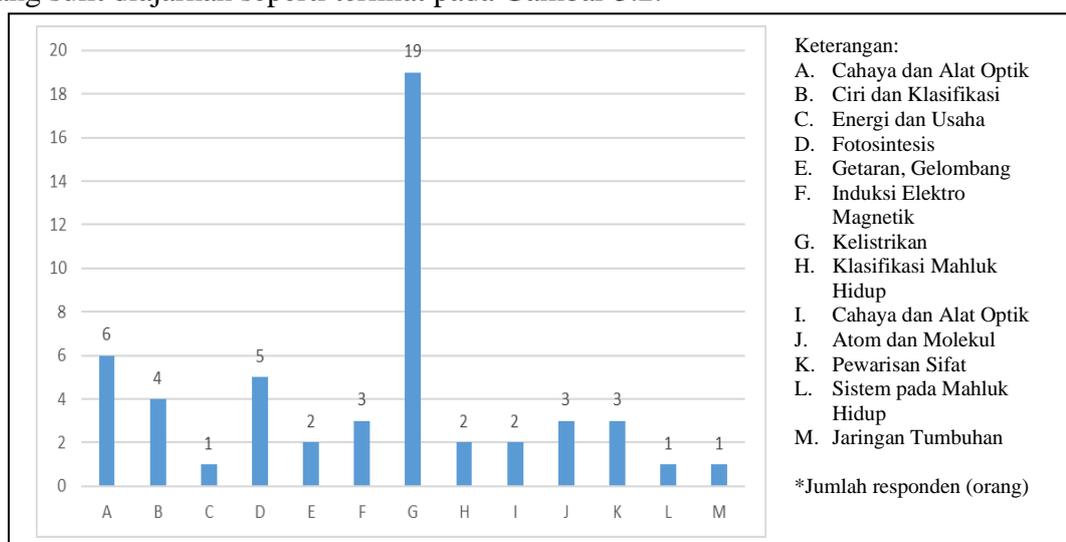
Gambar 3.1 Desain Penelitian dan Pengembangan Instrumen TPACK

3.2 Prosedur Pelaksanaan Penelitian

3.2.1 Penelitian dan Pengumpulan Informasi Pendahuluan

Penelitian pendahuluan yang dilakukan sebelum pengembangan instrumen adalah studi literatur yang terkait dengan asesmen untuk kompetensi guru. Studi literatur yang dilakukan bertujuan untuk mengetahui perkembangan asesmen kompetensi guru di Indonesia. Khususnya uji kompetensi guru (UKG) dan aspek-aspek lain yang terkait di dalam pelaksanaannya. Informasi dikumpulkan melalui proses pengkajian standar kompetensi guru dan asesmen yang pernah dilakukan sebelumnya kepada para guru, khususnya di Indonesia dan beberapa negara lain sebagai pembanding. Informasi diperoleh dari sumber-sumber berupa dokumen kebijakan, instrumen uji kompetensi guru tahun 2015, publikasi laporan program asesmen kompetensi guru IPA di Indonesia, serta artikel-artikel penelitian sebelumnya yang relevan. Studi literatur telah dilakukan secara semi sistematis dengan analisis secara kualitatif (Snyder, 2019). Hasil analisis terhadap informasi-informasi yang diperoleh kemudian menjadi dasar pengembangan instrumen dalam penelitian ini.

Penelitian pendahuluan juga dilakukan untuk menentukan dasar konten materi pelajaran IPA yang menjadi bagian dari instrumen. Penelitian awal ini dilakukan secara kuantitatif melalui pengisian kuisisioner oleh guru IPA dan kualitatif melalui kajian literatur. Hasil penelitian pendahuluan mengenai materi yang sulit diajarkan seperti terlihat pada Gambar 3.2.



Gambar 3.2 Grafik Hasil Kuisisioner Materi IPA Tersulit

Hasil penelitian pendahuluan seperti yang ditunjukkan pada Gambar 3.2 tersebut diperdalam dengan kajian mengenai penggunaan perangkat teknologi. Gabungan dari hasil tersebut disimpulkan bahwa materi Kelistrikan (rumpun Fisika) dan Fotosintesis (rumpun Biologi) merupakan materi-materi yang dianggap tersulit oleh para guru dan umumnya dapat diajarkan dengan pemanfaatan teknologi sebagai media atau sumber belajar. Penetapan materi untuk substansi instrumen ini cenderung lebih kualitatif karena keterbatasan jumlah responden secara kuantitatif (50 orang) sehingga kurang memadai untuk digeneralisir.

Penelitian pendahuluan sebagai pra pengembangan juga mengkaji mengenai penggunaan penskoran esai otomatis atau *Automated Essay Scoring (AES) system*. Studi ini bertujuan untuk mengetahui perkembangan penggunaan penskor esai berbantuan komputer dan kemungkinan yang dapat dilakukan terkait dengan penelitian ini. Informasi diperoleh dari literatur dan narasumber yang relevan. Narasumber yang dihubungi pada proses kajian adalah profesional yang berpengalaman dalam teknologi pendidikan dan pemangku kebijakan di Pusat Asesmen dan Pembelajaran (d/h Pusat Penilaian Pendidikan) Balai Penelitian dan Pengembangan, Kementerian Pendidikan dan Kebudayaan. Hasil studi ini kemudian menjadi dasar pemanfaatan program komputer (UKARA) yang telah dikembangkan oleh Pusat Asesmen dan Pembelajaran Kementerian Pendidikan dan Kebudayaan tersebut.

3.2.2 Perencanaan Penelitian dan Pengembangan Instrumen

Perencanaan yang disusun meliputi aspek akademik dan teknis pelaksanaan penelitian yang dilakukan. Aspek akademik yang direncanakan mencakup penetapan kerangka kerja instrumen, materi pelajaran yang menjadi bagian dari instrumen, tipe soal, dan standar-standar nilai atau koefisien untuk uji instrumen. Aspek teknis yang direncanakan mencakup kebutuhan-kebutuhan yang diperlukan untuk pengembangan instrumen, strategi pengambilan data atau pengujian, penjadwalan, validator dan penilai jawaban, rencana komunikasi dengan pihak di luar kampus, serta sumber daya lain yang terkait dengan penelitian dan pengembangan yang telah dilakukan.

Aspek-aspek akademik berdasarkan kajian literatur dan penelitian awal yang telah dilakukan sebelumnya. Aspek kerangka kerja instrumen yang tersebut adalah TPACK, khususnya aspek TPACK yang bersifat transformatif. Materi pelajaran IPA yang menjadi bagian dari soal adalah konsep Kelistrikan dan Fotosintesis. Tipe soal yang dianggap sesuai berupa soal-soal pemecahan masalah dan konseptual yang relevan dengan kurikulum. Sementara standar nilai untuk uji instrumen meliputi standar secara kualitatif dan kuantitatif. Standar kualitatif yang digunakan adalah triangulasi sumber untuk setiap uji yang dilakukan. Sementara standar untuk uji kuantitatif mencakup validitas, reliabilitas, statistik deskriptif, korelasi, dan reliabilitas antar penskor.

Aspek-aspek teknis direncanakan berdasarkan berbagai pertimbangan kontekstual dan kemampuan peneliti. Proses tahapan yang telah ditempuh untuk pengambilan data adalah pengajuan perijinan ke Dinas Pendidikan sesuai wilayah kabupaten/kota sasaran dan mengadakan pertemuan dengan partisipan untuk melakukan pengambilan data berupa tes. Pengambilan data tersebut sekaligus dengan penjadwalan dalam rentang waktu Januari – Maret 2020. Proses lainnya adalah menghubungi dosen-dosen validator dari UPI, dosen-dosen penskor dari UPI, Universitas Terbuka, dan STKIP 11 April, serta berkomunikasi dengan pengelola aplikasi UKARA di Pusat Asesmen dan Pembelajaran.

3.2.3 Pengembangan Produk Awal Instrumen TPACK dan Pendukung

Produk yang dikembangkan adalah instrumen soal dan rubrik penskoran sebagai produk utama dalam disertasi ini. Produk lainnya sebagai instrumen pendukung adalah format validasi dan kuisisioner. Semua produk dikembangkan berdasarkan hasil kajian atau kebutuhan berdasarkan literatur yang relevan dan persetujuan tim promotor dalam penelitian ini. Penjelasan untuk pengembangan instrumen-instrumen seperti berikut ini.

3.2.3.1 Pengembangan Desain Soal TPACK untuk Guru IPA

Instrumen soal didesain berdasarkan hasil kajian literatur dan penelitian pendahuluan. Desain mencakup tujuan khusus asesmen yang dikembangkan, kerangka kerja asesmen, jenis asesmen, tipe soal, bentuk jawaban, konten IPA

dalam soal, jumlah soal, dan alokasi waktu tes. Setiap aspek desain tersebut ditetapkan dengan pertimbangan akademis dan praktis sehingga kemudian instrumen dapat disusun dengan baik. Desain awal soal yang dikembangkan seperti pada Tabel 3.1.

Tabel 3.1 Desain Awal Instrumen Soal untuk Asesmen TPACK Guru IPA

No.	Aspek	Deskripsi
1.	Tujuan	Mengukur tingkat pengetahuan guru IPA
2.	Kerangka kerja	Pendekatan TPACK
3.	Jenis asesmen	Tes tertulis/uraian
4.	Tipe soal	Kasus
5.	Bentuk jawaban	Uraian singkat penyelesaian masalah
6.	Konten IPA pada soal	Kelistrikan dan Fotosintesis
7.	Jumlah soal	15 - 20 butir soal
8.	Alokasi waktu tes	60- 90 menit

3.2.3.2 Penyusunan Kisi-kisi Soal TPACK untuk Guru IPA

Penyusunan kisi-kisi dilakukan berdasarkan kerangka kerja TPACK sesuai dengan kajian pustaka dan hasil tahapan perencanaan penelitian. Kisi-kisi disusun memenuhi kaidah kisi-kisi (Thorndike & Thorndike-Christ, 2014) untuk instrumen soal yang mencakup komponen-komponen TPACK (Angeli dkk., 2016) sebagai standar dan indikator-indikator dengan pendekatan taksonomi tujuan pembelajaran (Anderson & Krathwohl, 2001). Kisi-kisi instrumen yang disusun mencakup 4 komponen TPACK yang diukur, standar dan indikator untuk masing-masing aspek, tingkatan kognitif (taksonomi), dan nomor soal untuk masing-masing indikator.

Indikator untuk masing-masing aspek pada format kisi-kisi tersebut merupakan sintesis dari hasil pengkajian literatur untuk asesmen TPACK guru IPA. Tingkatan ditetapkan sebagai batasan pengetahuan yang diukur, mulai dari “memahami” sampai “mengevaluasi” sesuai dengan pendekatan taksonomi tujuan pembelajaran yang dirujuk. Kisi-kisi tersebut kemudian diturunkan menjadi 15 butir soal berdasarkan pertimbangan dan perkiraan waktu yang memadai untuk mengerjakan soal uraian. Kisi-kisi telah direvisi sesuai dengan masukan dari validator pada pengujian instrumen tahap 1, tanpa mengubah keseluruhan format. Format kisi-kisi awal secara lengkap terdapat pada lampiran 3 laporan ini.

3.2.3.3 Penyusunan Instrumen Soal Esai

Instrumen utama yang dikembangkan dan diuji dalam penelitian ini telah disusun berdasarkan kisi-kisi yang telah disusun sebelumnya dan berdasarkan hasil kajian pustaka serta sesuai tujuan penelitian yang dilakukan. Instrumen awal berisi 15 soal esai dan sesuai dengan perencanaan merupakan soal-soal dengan tipe pemecahan masalah (kasus) dan konseptual dalam pembelajaran IPA. Soal merupakan pertanyaan atau permasalahan tunggal yang meminta jawaban dalam bentuk uraian singkat berupa penjelasan, poin-poin, ataupun langkah-langkah yang dapat menyelesaikan masalah atau jawaban yang diminta. Ada pula soal yang berupa pertanyaan untuk merumuskan jawaban yang relevan dengan kasus atau konsep yang diberikan. Semua butir soal dibentuk dari stimulus dan pertanyaan dalam kerangka TPACK serta terkait materi Kelistrikan atau Fotosintesis. Stimulus yang digunakan dapat berupa teks atau representasi (diagram, grafik) yang menjadi kasus dalam permasalahan yang harus dijawab. Istilah dan kalimat-kalimat yang digunakan dalam soal telah diupayakan sesuai dengan kaidah kebahasaan dan tingkat kemampuan berbahasa guru-guru yang mengerjakan soal secara umum. Instrumen soal dibuat dalam bentuk cetak dan digital (*Google Form*). Instrumen soal dilengkapi dengan data diri partisipan asesmen meliputi identitas, instansi, latar belakang pendidikan dan jenjang pendidikan terakhir, usia, masa kerja, sertifikasi, dan jenis kelamin. Instrumen soal terdapat pada lampiran 4.

3.2.3.4 Rubrik Penskoran Jawaban Esai

Konten rubrik atau pedoman penskoran disusun sesuai dengan kisi-kisi soal dan butir soal yang disusun sebelumnya. Sistematika format rubrik penskoran awal tersusun dari: 1) Penjelasan konsep; 2) Tingkat pengetahuan soal; 3) Tingkatan skor; dan 4) Kriteria untuk masing-masing skor. Konsep adalah penjelasan mengenai konsep pembelajaran yang relevan dengan soal sebagai konteks dan cakupan jawaban yang diharapkan. Tingkat pengetahuan soal adalah taksonomi pengetahuan yang diukur dalam soal. Tingkatan skor yang diberikan adalah 0 – 2 untuk setiap jawaban dan berlaku untuk semua soal. Kriteria jawaban adalah

pedoman untuk menetapkan skor yang diberikan, berupa batasan-batasan konsep dan penilaiannya secara kualitatif. Format rubrik penskoran pada lampiran 5.

3.2.3.5 Format Validasi Instrumen

Format validasi disusun sebagai panduan validasi isi instrumen soal dan rubrik penskoran. Validasi ini menjadi bagian dari pengujian tahap I, yang dilakukan secara kualitatif terhadap instrumen kisi-kisi soal, butir soal, dan rubrik penskoran. Validator diminta menyatakan butir soal sesuai atau tidak pada masing-masing aspek, untuk instrumen soal maupun rubrik penskoran. Catatan diminta sebagai penjelasan dan menjadi informasi dalam proses analisis hasil validasi dan serta dasar koreksi terhadap instrumen. Format instrumen validasi tercantum pada lampiran 6 dan hasil validasi sebagai bagian dari pengujian akan dijelaskan pada bagian selanjutnya.

3.2.3.6 Kuisisioner

Instrumen kuisisioner disusun untuk menjadi data dukung terhadap data jawaban soal oleh partisipan asesmen penelitian ini. Instrumen kuisisioner penelitian terdiri dari 2 macam, yaitu kuisisioner untuk uji keterbacaan instrumen soal dan instrumen kuisisioner data diri partisipan. Kuisisioner untuk uji keterbacaan digunakan pada uji tahap I, sementara kuisisioner data diri digunakan mulai pada uji tahap II.

a. Kuisisioner uji keterbacaan

Instrumen ini diberikan kepada partisipan untuk mengetahui waktu yang diperlukan untuk mengerjakan soal tes, apakah ada istilah yang tidak dipahami, kalimat yang kurang jelas atau rancu, tabel/diagram yang tidak dipahami, dan relevansi soal dengan mata pelajaran IPA yang diampu oleh partisipan. Kuisisioner merupakan pertanyaan “ya/tidak” dengan penjelasan yang bersifat terbuka atau dalam bentuk uraian (Lampiran 9).

b. Kuisisioner informasi kontekstual

Jawaban instrumen kuisisioner ini menjadi informasi tambahan yang dapat berhubungan dengan hasil tes. Pertanyaan yang diberikan adalah mengenai

pengalaman partisipan dalam mengajarkan materi Kelistrikan dan Fotosintesis kendala dalam mengajarkan materi-materi tersebut di sekolah masing-masing.

3.2.4 Pengujian Instrumen Tahap I

Pengujian tahap I merupakan uji instrumen untuk validitas isi secara kualitatif, yang dilakukan oleh tiga validator (pakar). Data berupa hasil isian format validasi untuk instrumen soal dan rubrik penskoran. Validitas isi juga mencakup aspek kontekstualitas konten soal, yaitu kesesuaian dengan kurikulum dan pembelajaran untuk mata pelajaran IPA SMP. Data berupa keputusan validator mengenai kesesuaian antara instrumen soal dengan kisi-kisi, konsep IPA, lingkup materi IPA, soal dengan rubrik, kriteria penskoran dalam rubrik. Catatan dari validator menjadi penjelasan terhadap keputusan yang diberikan. Untuk pengujian tahap I ini juga dilakukan uji coba keterbacaan soal menggunakan kuisioner oleh 100 orang partisipan sebagai informasi pendukung.

3.2.4.1 Pengumpulan Data Tahap I

Format validasi dan instrumen disampaikan kepada para validator dan kemudian dikembalikan melalui media internet sesuai dengan kondisi yang ada saat itu. Data hasil validasi oleh validator secara lengkap terdapat pada lampiran 7 dan lampiran 8. Kuisioner untuk uji keterbacaan dibagikan dibagikan bersama instrumen soal, yaitu dengan moda *Google Form*. Format kuisioner uji keterbacaan disebarikan melalui media grup *Whatsapp* Musyawarah Guru Mata Pelajaran IPA oleh koordinator kelompok masing-masing. Partisipan pengisian kuisioner ini adalah guru-guru IPA SMP yang bertugas di wilayah Provinsi Banten. Partisipan diambil secara acak dan bersifat sukarela. Data hasil pengisian kuisioner kemudian direkapitulasi menggunakan aplikasi Microsoft Excel. Data hasil uji keterbacaan terdapat pada lampiran 9.

3.2.4.2 Analisis Data untuk Pengujian Tahap I

Data yang telah digabungkan dari tiga validator kemudian diolah secara kualitatif dengan menerapkan metode triangulasi. Uji validitas kualitatif ini terutama untuk memvalidasi konten soal dan rubrik sebagai suatu instrumen untuk

asesmen guru IPA. Substansi mata pelajaran IPA dan pembelajaran IPA yang dilakukan menjadi fokus untuk validator, selain kesesuaian butir soal dengan indikator. Proses pengelompokan dari kesamaan atau perbedaan keputusan dilakukan untuk memilah penilaian dari validator. Kesamaan pendapat dari validator menjadi hasil akhir validitas suatu butir soal atau kriteria penskoran. Perbedaan pendapat dijadikan dasar untuk koreksi dan rekomendasi perbaikan soal/kriteria pada tahap pengembangan selanjutnya. Analisis secara kualitatif terhadap butir soal dan kriteria penskoran dilakukan berdasarkan validasi dan catatan dari validator. Kesimpulan yang diambil berdasarkan analisis adalah butir soal/kriteria “valid”, “perlu perbaikan”, atau tidak “valid”. Rangkuman dan triangulasi data untuk uji tahap I dapat dilihat pada lampiran 10 dan lampiran 11.

Untuk uji keterbacaan dilakukan analisis secara deskriptif sesuai dengan instrumen yang digunakan untuk pengumpulan data. Kesimpulan hasil analisis menjadi rekomendasi perlakuan yang diberikan terhadap setiap butir soal, yaitu “langsung dapat digunakan” atau “dapat digunakan dengan perbaikan”. Untuk soal-soal yang mendapat rekomendasi “dapat digunakan dengan perbaikan”, tindak lanjut dilakukan berdasarkan hasil jawaban partisipan pada kuisioner yang diberikan. Hasil analisis uji keterbacaan terdapat pada lampiran 12.

3.2.5 Revisi Instrumen Tahap I

Instrumen soal diperbaiki sesuai dengan hasil uji validitas isi dan catatan-catatan saran yang diberikan oleh validator. Soal yang tidak mendapat tanda silang dari validator dinyatakan telah valid secara isi dan kontekstual. Perbaikan soal difokuskan kepada validitas substansi soal dan kesesuaian dengan indikator, yaitu terkait dengan tingkat pengetahuan dan kerangka TPACK dalam pertanyaan dan stimulus soal. Perbaikan instrumen juga memperhatikan hasil uji keterbacaan yang diberikan oleh partisipan asesmen penelitian. Perbaikan dilakukan untuk meningkatkan keterbacaan soal sesuai masukan dari partisipan uji coba keterbacaan.

Revisi rubrik penskoran dilakukan sesuai hasil validasi dan saran yang diberikan oleh validator. Fokus perbaikan rubrik penskoran adalah pada kriteria untuk masing-masing skor. Saran perbaikan juga berupa penambahan komponen

rubrik penskoran, yaitu contoh jawaban ideal atau jawaban yang mendapat skor 2 sebagai skor tertinggi. Contoh jawaban diberikan sebanyak 3 macam yang berbeda secara pemilihan kata dan susunan kalimat, tetapi mempunyai maksud yang sama. Contoh jawaban diambil dari hasil isian partisipan asesmen, pada tahapan pengambilan data selanjutnya, dengan penyempurnaan sesuai konsep yang menjadi dasar penyusunan soal yang relevan.

3.2.6 Pengujian Instrumen Tahap II

Uji yang dilakukan pada tahap II adalah uji validitas konstruk terhadap instrumen soal secara kuantitatif. Dalam tahap II ini uji validitas dilakukan dengan prosedur analisis faktor eksploratori (*Exploratory Factor Analysis/EFA*). Data dalam uji tahap ini merupakan skor dari jawaban subjek. Penskoran dilakukan oleh peneliti menggunakan rubrik penskoran yang telah divalidasi pada tahap sebelumnya dan diuji reliabilitas antar penskor sebelum penskoran untuk EFA. Sebelum uji instrumen dalam tahap II ini, dilakukan pengujian secara kuantitatif terhadap rubrik penskoran terlebih dahulu. Uji yang dilakukan untuk mengetahui reliabilitas dan korelasi dalam penskoran sebagai indikator kelayakan rubrik yang digunakan. Tahapan pengumpulan data, proses uji rubrik, dan analisis data secara rinci dijelaskan selanjutnya.

3.2.6.1 Pengumpulan Data untuk Pengujian Tahap II

Data untuk pengujian tahap II ini terdiri dari data untuk uji rubrik penskoran dan data untuk instrumen soal secara kuantitatif. Partisipan untuk uji instrumen soal adalah 100 guru IPA SMP, yang menjawab soal dalam moda *Google Form* secara daring. Jumlah 100 orang subjek ditetapkan dengan pertimbangan telah dianggap memadai secara statistik (Pallant, 2011). Data berupa skor untuk jawaban subjek diperoleh dari 4 orang penskor, yaitu 3 orang dosen yang mengajar IPA/Biologi/Fisika dan peneliti sendiri. Penskoran dilakukan secara simultan, setelah fail jawaban subjek dan rubrik penskoran didistribusikan melalui surat elektronik kepada semua penskor. Hasil penskoran dikembalikan oleh para penskor kepada peneliti melalui moda yang sama. Peneliti kemudian menentukan skor akhir berdasarkan skor yang disepakati (persentase

kesepakatan) oleh minimal 2 dari 3 penskor atau 66,67%. Skor akhir tersebut kemudian menjadi data untuk uji kuantitatif instrumen. Seluruh hasil penskoran oleh 4 penskor menjadi lampiran 13 dalam laporan penelitian ini.

3.2.6.2 Uji Reliabilitas untuk Rubrik Penskoran

Uji reliabilitas untuk rubrik penskoran dilakukan 2 kali, yang pertama antara 2 orang penskor dan yang kedua melibatkan 3 orang penskor. Uji yang dilakukan 2 kali bertujuan untuk mengetahui persamaan dan perbedaan reliabilitas ketika jumlah penskor berbeda. Pada uji pertama analisis dilakukan terhadap hasil penskoran oleh Rater1 dan Rater3 (peneliti). Selanjutnya dilakukan uji reliabilitas kedua, tetapi sebelumnya dilakukan uji korelasi (Pearson) untuk mengetahui koefisien korelasi antar skor yang dihasilkan. Setelah melalui penghitungan korelasi dari 4 penskor tersebut, yang terlibat pada tahap kedua hanya 3 orang, yaitu Rater2, Rater3, dan Rater4. Hasil penskoran oleh Rater1 memiliki rata-rata korelasi paling rendah dan termasuk dalam kategori “cukup” (kurang dari 0,75) sehingga diputuskan tidak diikutsertakan dalam uji reliabilitas antar penskor untuk rubrik penskoran.

Uji secara statistik yang dilakukan untuk rubrik penskoran adalah menghitung:

- a. Alpha koefisien (*Cronbach's α coefficient*)
- b. Reliabilitas antar penskor *Intraclass Correlation Coefficient* (ICC)
- c. Persentase kesepakatan (*agreement*) antar penskor
- d. Koefisien korelasi Pearson (untuk menentukan jumlah penskor pada pengujian reliabilitas rubrik kedua)

Nilai Alpha yang dianggap memadai minimal 0,7 (Pallant, 2011; Wahlen dkk., 2020). Koefisien reliabilitas antar penskor yang dianggap memadai untuk uji ini minimal 0,75 (Koo & Li, 2016; Stemler, 2004; Van Lummel dkk., 2016). Persentase kesepakatan tidak menjadi dasar sebagai uji statistik, tetapi semakin tinggi persentase semakin memperkuat keyakinan terhadap rubrik yang digunakan (Harris dkk., 2010; Wahlen dkk., 2020). Penghitungan semua koefisien menggunakan menggunakan program SPSS Statistics 20, sementara nilai persentase dihitung menggunakan program Excel. Sebelum uji rubrik ini

dilakukan, pada data tersebut telah terlebih dahulu dilakukan uji normalitas dan linieritas. Hasil pengujian reliabilitas antar penskor terdapat pada lampiran 14.

Analisis reliabilitas dilakukan sesuai dengan kaidah analisis kuantitatif berdasarkan nilai atau koefisien yang diperoleh sesuai dengan fungsi masing-masing nilai tersebut. Hasil analisis menjadi dasar untuk keputusan penggunaan rubrik penskoran yang telah disusun.

3.2.6.3 Analisis Statistik Instrumen Soal pada Tahap II

Pengujian instrumen tahap II mengikuti prosedur EFA untuk mengetahui validitas konstruk dan koefisien Alpha *Cronbach's* untuk reliabilitas (Hair dkk., 2010). Nilai validitas yang diuji menggunakan analisis faktor melalui langkah-langkah penghitungan: 1) Uji Kaiser-Meyer-Olkin of Sampling Adequacy (KMO); 2) Uji Bartlett; 3) Nilai Measure of Adequacy (MSA); dan 4) Nilai Eigen. Nilai validitas dan reliabilitas dalam tahap II ini menjadi dasar revisi atau keputusan untuk penggunaan suatu butir soal selanjutnya. Interpretasi dan pembahasan dilakukan berdasarkan hasil EFA dan selanjutnya menjadi dasar untuk melakukan pengujian pada tahap III. Hasil uji secara statistik menggunakan SPSS untuk tahap II dapat dilihat dalam lampiran 15.

3.2.7 Revisi Instrumen Tahap II

Hasil uji tahap II ini merekomendasikan adanya soal yang tidak dapat dipakai karena tidak memenuhi standar uji. Perbaikan atau penggantian butir soal tidak dilakukan karena jumlah butir soal yang valid dan nilai reliabilitas hasil pengujian tahap II masih memadai. Selain itu, rekomendasi yang diberikan untuk pengujian tahap selanjutnya adalah penambahan jumlah partisipan uji. Penambahan jumlah partisipan dalam asesmen ini diharapkan meningkatkan kualitas uji statistik untuk instrumen soal.

3.2.8 Pengujian Instrumen Tahap III

Tahap III tahap akhir untuk proses pengujian instrumen soal yang dikembangkan dalam penelitian ini. Pengujian instrumen secara statistik menggunakan prosedur uji yang dilakukan sesuai prosedur analisis faktor

konfirmasi (*Confirmatory Factor Analysis/CFA*). Pada tahap III ini dilakukan pula pengujian untuk perangkat lunak penskor esai otomatis yaitu UKARA. Uji untuk penskor otomatis adalah reliabilitas antar penskor yang dibandingkan dengan penskor manusia (peneliti). Penjelasan lebih rinci seperti berikut ini.

3.2.8.1 Pengumpulan data untuk pengujian tahap III

Data untuk tahap ini dikumpulkan melalui moda *Google Form* seperti pada tahap sebelumnya. Tambahan subjek pada tahap III ini adalah 100 orang guru IPA yang berbeda dari subjek pada tahap sebelumnya, sehingga jumlah total data adalah 200 subjek. Penggabungan dianggap dapat dilakukan karena tidak ada perbedaan instrumen soal dan rubrik penskoran untuk tahap II dan III. Penggabungan dilakukan untuk menambah jumlah data yang diuji secara statistik dalam tahap III ini. Jawaban subjek kemudian diberikan skor oleh peneliti sesuai dengan rubrik penskoran. Data yang diuji dalam tahap III ini adalah skor dari 200 subjek.

Data berikutnya yang dikumpulkan pada tahap III ini adalah data hasil penskoran oleh UKARA, yaitu program komputer untuk penskoran jawaban esai secara otomatis. Proses penskoran atau komputasi yang dilakukan melalui tahapan seperti berikut ini:

- 1) Jawaban partisipan disiapkan sebagai bahan pelatihan AI pada UKARA disusun dalam bentuk tabel (fail Excel) untuk masing-masing butir soal yang telah diberi skor oleh manusia (peneliti).
- 2) Fail-fail tersebut kemudian diunggah ke dalam aplikasi UKARA yang bekerja dalam jaringan (*online*). Sistem penskoran otomatis diakses pada tautan <http://118.98.227.52:8080/>.
- 3) Fail Excel berisi jawaban diunggah ke dalam aplikasi satu-persatu, kemudian prosedur dilakukan untuk memproses penskoran
- 4) Sistem dapat diakses menggunakan user name dan kata kunci yang diperoleh dari dan dengan ijin dari Pusmenjar Kemdikbud, karena UKARA bukan suatu aplikasi untuk umum.
- 5) Salah satu prosedur yang menjadi ciri khas sistem AI ini adalah tahapan sistem membuat model.

- 6) Untuk mengakhiri proses, hasil penskoran dapat diunduh langsung setelah sistem melakukan semua tahapan.
- 7) Hasil akhir berupa tabel yang berisi jawaban, hasil penskor manusia, dan hasil penskor UKARA.

Proses penskoran oleh UKARA tersebut dilakukan 2 kali, yaitu untuk partisipan sejumlah 100 orang dan partisipan 200 orang. Untuk partisipan 200 orang tersebut merupakan gabungan 100 orang partisipan awal dan 100 orang partisipan tambahan. Hasil proses penskoran oleh UKARA terdapat pada lampiran 18 dan 19.

3.2.8.2 Analisis Statistik Instrumen Soal pada Tahap III

Pengujian secara kuantitatif untuk instrumen soal pada tahap III ini meliputi daya pembeda, tingkat kesukaran, uji validitas, dan reliabilitas. Khusus untuk uji validitas, pada tahap ini dilakukan lebih mendalam dibandingkan pada tahap sebelumnya. Uji secara statistik yang dilakukan sesuai prosedur analisis faktor konfirmatori (*Confirmatory Factor Analysis/CFA*). Dalam tahap III ini digunakan Model Persamaan Struktural atau *Structural Equation Model* (SEM) untuk CFA yang dilakukan (Green & Pearson, 2004; Hair dkk., 2010; Mirzaei dkk., 2019). Analisis dilakukan berdasarkan hasil secara kuantitatif yang diperoleh untuk menentukan kualitas instrumen soal. Koefisien yang dihitung berdasarkan SEM tersebut meliputi Chi-square relatif (CMIN/DF), Root mean square error of approximation (RMSEA), Comparative fit index (CFI), dan FMIN. Ambang batas yang dianggap memadai untuk kriteria model fit tersebut sebagai berikut: 1) $CMIN/DF < 5,00$; 2) $RMSEA \leq 0,05$; 3) $CFI > 0,9$; dan 4) FMIN semakin mendekati 0 semakin baik. Untuk analisis model struktur tersebut digunakan program komputer AMOS 22. Untuk daya pembeda dan tingkat kesukaran dianalisis dengan kerangka teori tes klasik (Aalaei, Ahmadi, & Aalaei, 2016; Quagrain & Arhin, 2017) menggunakan program SPSS 20 untuk menghitung indeks masing-masing butir soal yang telah memenuhi kriteria valid dan reliabel. Hasil uji secara statistik menggunakan program SPSS dan AMOS terdapat pada lampiran 16 dan 17.

3.2.8.3 Uji Reliabilitas untuk Penggunaan Penskor Berbantuan Komputer

Penggunaan penskor berbantuan komputer menjadi bagian dari penelitian dan pengembangan untuk disertasi ini. Pengujian penskor otomatis UKARA yang dikelola oleh Pusat Asesmen dan Pembelajaran menggunakan data berupa jawaban uraian peserta setelah hasil pengujian instrumen soal dan rubrik penskoran dianggap memadai. Untuk uji pada sistem penskoran berbantuan komputer ini yang dilakukan adalah uji reliabilitas antara hasil penskoran oleh peneliti atau sebagai penskor manusia (penskor 1) dan UKARA (penskor 2).

Tahap pertama adalah menguji sekaligus “melatih” sistem penskoran yang digunakan. Data yang diinput dan diuji untuk penggunaan UKARA adalah jawaban partisipan dengan jumlah yang sama seperti data untuk uji instrumen soal, yaitu 200 partisipan. Uji statistik yang dilakukan adalah menghitung Intraclass Correlation Coefficient (ICC) dan nilai korelasi Pearson (r) sebagai nilai reliabilitas antar penskor 1 dan penskor 2. Analisis yang dilakukan berdasarkan hasil uji kuantitatif tersebut. Selain itu analisis juga dilakukan untuk membandingkan hasil uji pada jumlah data yang berbeda dan komposisi skor yang berbeda. Perbandingan dengan jumlah partisipan yang berbeda dilakukan dengan jumlah 100 dan 200 partisipan. Perbandingan dengan komposisi skor yang berbeda adalah membandingkan nilai reliabilitas antar penskor untuk komposisi jawaban yang berimbang dan tidak berimbang pada masing-masing skor yang diberikan.

Tahap kedua pada uji untuk penskoran berbantuan komputer ini adalah uji penerapan sistem UKARA. Pengujian menggunakan jawaban dari 50 orang subjek tes atau partisipan dengan prosedur hampir sama dengan uji pada tahap 1, dengan perbedaan pada tahap kedua ini hasil penskoran oleh penskor 1 tidak diinput ke dalam sistem UKARA, sehingga sistem hanya menskor menggunakan hasil pelatihan pada tahap 1. Nilai reliabilitas yang dihitung adalah ICC dan r serta ditambah persentase kesepakatan antar penskor. Persentase kesepakatan antar penskor dihitung dan dianalisis untuk mengetahui kualitas hasil penskoran yang diberikan oleh UKARA secara mutlak.

3.2.9 Implementasi Asesmen TPACK dengan Penskoran Berbantuan Komputer

Tahap terakhir dari penelitian yang dilakukan adalah melakukan penilaian akhir terhadap hasil tes partisipan sebagai implementasi penggunaan instrumen. Penskoran dilakukan oleh UKARA dan kemudian hasil penskoran dikonversi menjadi nilai hasil asesmen untuk dianalisis secara deskriptif, perbandingan, dan korelasional. Analisis deskriptif untuk mengetahui hasil asesmen secara umum dan untuk masing-masing komponen TPACK yang diukur. Analisis deskriptif dilakukan pula terhadap hasil UKG 2015 khusus untuk subjek yang terlibat, hasilnya sebagai bahan analisis perbandingan dan korelasional di antara kedua asesmen yang telah dilakukan (UKG 2015 dan asesmen TPACK).

Analisis kuantitatif korelasional (Pearson) dan uji beda (uji homogenitas variansi) telah dilakukan pula untuk melihat adanya hubungan antara faktor-faktor yang dimiliki guru terhadap nilai TPACK yang diperoleh dan homogenitas nilai yang diperoleh dari kelompok-kelompok dalam setiap faktor. Faktor-faktor tersebut adalah latar belakang pendidikan, masa kerja, jenis kelamin, dan usia partisipan.

Hipotesis yang disusun untuk masing-masing variabel yang diuji secara statistik tersebut adalah sebagai berikut.

- 1) Hipotesis untuk variabel perbedaan latar belakang pendidikan
 - H₀: Ada perbedaan nilai TPACK untuk masing-masing kelompok berdasarkan latar belakang pendidikan yang berbeda
 - H₁: Tidak ada perbedaan nilai TPACK untuk masing-masing kelompok berdasarkan latar belakang pendidikan yang berbeda
- 2) Hipotesis untuk variabel perbedaan masa kerja
 - H₀: Ada perbedaan nilai TPACK untuk masing-masing kelompok berdasarkan masa kerja
 - H₁: Tidak ada perbedaan nilai TPACK untuk masing-masing kelompok berdasarkan masa kerja
- 3) Hipotesis untuk variabel perbedaan jenis kelamin/gender
 - H₀: Ada perbedaan nilai TPACK untuk masing-masing kelompok berdasarkan jenis kelamin

H₁: Tidak ada perbedaan nilai TPACK untuk masing kelompok berdasarkan jenis kelamin

4) Hipotesis untuk variabel perbedaan usia

H₀: Ada perbedaan nilai TPACK untuk masing-masing kelompok berdasarkan usia

H₁: Tidak ada perbedaan nilai TPACK untuk masing-masing kelompok berdasarkan usia

Pembahasan terhadap hasil implementasi mencakup hasil keseluruhan dan masing-masing komponen, autentisitas serta efektifitas asesmen yang dilakukan. Hasil keseluruhan atau total adalah nilai gabungan dari masing-masing komponen pengetahuan dalam kerangka kerja TPACK. Aspek autentisitas dibahas berdasarkan kajian teoritis dan jawaban subjek secara kualitatif. Sementara efektifitas didasarkan kepada hasil pengolahan data untuk masing-masing faktor yang dapat mempengaruhi dan kaitannya dengan pemanfaatan data sesuai tujuan penelitian.

3.3 Partisipan dan Konteks Penelitian

Partisipan yang dilibatkan dalam penelitian ini mencakup validator, penskor, dan subjek asesmen. Validator instrumen soal dan rubrik pada penelitian ini berjumlah 3 orang, yaitu ahli dalam pendidikan IPA dan berprofesi sebagai dosen pengampu mata kuliah Fisika/Biologi yang berasal dari FMIPA Universitas Pendidikan Indonesia (UPI) Bandung. Penskor jawaban esai, selain peneliti sendiri, adalah 3 orang dosen yang mengampu mata kuliah pendidikan Fisika/Biologi, masing-masing berasal dari UPI Bandung (Rater1), Universitas Terbuka Jakarta (Rater2), dan STKIP 11 April Sumedang (Rater4). Subjek asesmen dengan total berjumlah 250 orang, merupakan guru-guru IPA SMP dari 8 wilayah kabupaten/kota yang ada di Provinsi Banten. Penjaringan subjek dilakukan secara acak sesuai dengan kesediaan guru-guru tersebut. Penjaringan subjek asesmen juga telah melalui proses perijinan pada masing-masing dinas pendidikan dan pengawas pembina Musyawarah Guru Mata Pelajaran.

Pengajar IPA di SMP dalam wilayah provinsi Banten dianggap representatif sebagai konteks penelitian karena beragamnya kondisi sekolah dan

kemampuan guru yang ada berdasarkan literatur yang telah dikaji sebelumnya. Jumlah guru di Provinsi Banten juga termasuk cukup besar sehingga diasumsikan dapat memberikan informasi yang representatif untuk penelitian ini. Selain itu dalam wilayah provinsi Banten akses internet cukup memadai untuk pengambilan data melalui jaringan internet sebagai alternatif mengadakan pertemuan langsung.

3.4 Definisi Operasional

Aspek-aspek yang dikaji dalam penelitian ini adalah pengetahuan guru, instrumen soal uraian, dan sistem penskoran berbantuan komputer. Masing-masing variabel tersebut didefinisikan sebagai berikut:

- 1) Pengetahuan guru yang dimaksud dalam penelitian ini adalah pengetahuan konten pedagogis teknologi (TPACK) (Koehler & Mishra, 2009; Koehler dkk., 2012). Khusus pada penelitian ini, komponen-komponen TPACK yang diukur meliputi pengetahuan konten pedagogis (PCK), pengetahuan konten teknologis (TCK), pengetahuan pedagogis teknologi (TPK), dan pengetahuan konten pedagogis teknologi (TPCK) dengan pandangan transformatif (Angeli & Valanides, 2009; Angeli dkk., 2016). Pengetahuan guru diukur menggunakan instrumen soal uraian yang dikembangkan pada penelitian ini.
- 2) Instrumen soal uraian yang dikembangkan pada penelitian ini adalah soal yang meminta jawaban berupa uraian atau tanggapan terbatas dan rubrik penskorannya. Soal berupa pertanyaan yang meminta jawaban pada konten substantif dan bentuk tanggapan terfokus atau terbatas dengan struktur yang dibuat sendiri oleh subjek (Nitko & Brookhart, 2014; Yulyantar, 2017). Jawaban berupa penjelasan atau pemaparan singkat, berupa suatu kalimat atau lebih yang merupakan jawaban autentik subjek. Tipe soal berbasis konsep pemecahan masalah dari kasus dalam pembelajaran yang menguji kemampuan penalaran subjek untuk menyelesaikan masalah baru (tidak rutin) dan kontekstual menggunakan pengetahuan yang dimilikinya (Carson, 2007; Dostál, 2015; Reeff, Zabal, & Blech, 2006), sehingga soal terdiri dari stimulus berupa teks atau representasi informasi dan pokok soal atau pertanyaan. Asesmen juga menggunakan rubrik penskoran berupa panduan yang mencantumkan kriteria khusus untuk menilai tes akademik (Brookhart,

2013; Dawson, 2017). Kualitas instrumen yang memadai ditentukan oleh autentisitas, kontekstualitas, validitas, reliabilitas, dan efektifitasnya. Kualitas tersebut diukur melalui tahapan-tahapan uji coba secara kualitatif dan kuantitatif serta kajian literatur sesuai prosedur penelitian dan pengembangan yang diterapkan.

- 3) Penskor esai otomatis (*automated essay scoring*) adalah program komputer yang dikembangkan untuk memberikan nilai atau skor terhadap jawaban berupa kata, kalimat, atau uraian (Powers, Escoffery, & Duchnowski, 2015; Zimmerman dkk., 2018). Dalam penelitian ini digunakan penskor esai otomatis UKARA, yang berbasis kecerdasan buatan (*artificial intelligence*). UKARA tersebut adalah program komputer untuk penskoran jawaban yang dikembangkan dan dikelola oleh Pusat Asesmen dan Pembelajaran (d/h. Puspendik), Badan Penelitian dan Pengembangan, Kementerian Pendidikan dan Kebudayaan RI. Performa penskor otomatis yang baik ditentukan oleh reliabilitas dan kesamaanya dengan penskor manusia berupa nilai ICC, korelasi, dan persentase kesepakatan.