

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang Penelitian**

Kompetensi guru di Indonesia menjadi suatu objek penelitian yang esensial dalam dekade terakhir ini. Diawali dengan adanya beberapa perubahan kebijakan mengenai standar nasional pendidikan, termasuk di dalamnya standar kompetensi dan sertifikasi pendidik, menyebabkan kualitas pembelajaran yang diimplementasikan oleh para guru menjadi informasi yang semakin dibutuhkan. Kompetensi guru tersebut esensial karena menjadi input yang mempengaruhi hasil belajar siswa dan berdampak lebih lanjut terhadap kualitas pendidikan di sekolah (Mccaffrey, 2010; Sanders & Rivers, 1996). Penelitian-penelitian mengenai kompetensi guru dan faktor-faktor yang mempengaruhinya semakin banyak dilakukan dan memberikan informasi terkini yang bermanfaat untuk pengembangan pendidikan secara komprehensif. Penelitian-penelitian tersebut dikaitkan pula dengan data-data lain seperti hasil belajar siswa, program pengembangan kompetensi, dan beberapa aspek lainnya.

Hasil belajar siswa sebagai tujuan akhir dalam pembelajaran lebih dipengaruhi oleh kualitas guru dibandingkan beberapa faktor lain, seperti dijelaskan oleh Darling-Hammond (2011), bahwa kualifikasi guru yang lebih baik akan berdampak pada kualitas hasil belajar yang lebih baik pada hasil pembelajaran di kelas. Lebih lanjut dijelaskan pula bahwa latar belakang pendidikan secara formal merupakan bagian yang tidak dapat dipisahkan dari pembentukan kompetensi guru secara keseluruhan, tetapi pendidikan tersebut belum mencukupi, karena itu kompetensi dibentuk lebih lanjut selama perjalanan tugasnya menjadi guru. Kualifikasi dalam konsep tersebut lebih cenderung berkaitan dengan kualitas atau kemampuan guru secara menyeluruh, tetapi latar belakang pendidikan adalah pondasi dasar yang menjadi kekuatan bagi kompetensi guru tersebut. Kualifikasi yang lebih baik berarti kemampuan yang lebih baik secara faktual dan terbentuk dari kemampuan konseptual yang lebih banyak diperoleh dari pendidikan pra tugas serta berkembang sesuai dengan kemampuan belajarnya atau pengembangan profesinya selama bertugas

(Jovanova-Mitkovska, 2010). Guru yang lebih kompeten umumnya merupakan guru-guru yang berpengalaman mengajar lebih lama, dimana kompetensi tersebut lebih banyak dibentuk oleh lingkungan sekolah yang bermutu daripada pendidikan guru (Siswandoko & Suryadi, 2013). Sementara kompetensi guru tersebut berhubungan dengan kinerja, karena kinerja dipengaruhi aspek-aspek kompetensi yang dimiliki guru antara lain pengetahuan, penguasaan keterampilan, pengalaman, keyakinan diri, dan motivasi untuk melakukan pembelajaran dengan baik. Kompetensi yang terbentuk dari pendidikan keguruan dan pengembangan profesi tersebut akan menjadi basis untuk seorang guru memiliki kinerja yang baik dalam melaksanakan tugasnya. Aspek-aspek tersebut sudah dimiliki oleh para guru yang telah bertugas, tetapi tingkatan atau kualitas dari faktor-faktor tersebut tidak seluruhnya sesuai dengan standar yang ditetapkan. Hal itu berdampak pada kinerja yang ditampilkan kemungkinan tidak akan sesuai harapan. Tingkat kinerja guru tersebut akan cenderung berbanding lurus dengan tingkat hasil belajar siswa dari pembelajaran yang dilakukan. Untuk mengetahui tingkat kompetensi, kinerja, ataupun aspek-aspek yang mengkonstruksinya umumnya merupakan tujuan dari suatu sistem asesmen atau evaluasi yang handal dan akurat (Kementerian Pendidikan dan Kebudayaan, 2012, 2015).

Pengetahuan konten, pengetahuan konten pedagogis, efikasi diri, dan motivasi mengajar adalah aspek-aspek yang telah diteliti pengaruhnya terhadap kinerja dan menunjukkan bahwa aspek-aspek tersebut memang mempengaruhi kinerja guru (Abós dkk., 2018; Kulgemeyer & Riese, 2018). Sementara beberapa faktor seperti ciri-ciri personal dan program sertifikasi guru tidak memberikan pengaruh yang signifikan terhadap performa mengajar guru, (Corcoran & O'Flaherty, 2018; Kusumawardhani, 2017). Hasil-hasil penelitian tersebut telah menunjukkan bahwa hubungan antar faktor dapat membuatnya saling menguatkan atau bisa melemahkan dalam penilaian, seperti seorang guru memiliki pengetahuan konten pedagogis yang cukup baik, tetapi ternyata performanya rendah karena terpengaruh oleh motivasi yang rendah. Di sisi lain, ada guru-guru pemula yang memiliki semangat tinggi tetapi mendapat penilaian kinerja rendah karena kemampuan dalam pembelajarannya kurang baik. Berdasarkan penelitian-penelitian tersebut, aspek pengetahuan adalah salah satu aspek yang memiliki

pengaruh signifikan terhadap kinerja guru dalam pelaksanaan pembelajaran (Kulgemeyer & Riese, 2018).

Penelitian untuk penerapan asesmen pengetahuan konten dan pedagogi serta pengetahuan konten pedagogi (*pedagogical content knowledge/PCK*) telah banyak dilakukan sampai saat ini. Penelitian terhadap uji kompetensi yang dilakukan di Indonesia juga memperlihatkan bahwa tes tersebut mengukur pengetahuan konten dan pedagogi yang terkait. Hasil Uji Kompetensi Guru UKG tahun 2015 menunjukkan rata-rata yang masih rendah baik untuk kompetensi pedagogik maupun profesional (<https://www.kemdikbud.go.id>), termasuk untuk guru IPA (<https://p4tkipa.kemdikbud.go.id>), yaitu dengan rerata di bawah 50. Pengetahuan konten guru akan berpengaruh terhadap konten yang diberikan kepada siswanya, sementara pengetahuan pedagogi dan pengetahuan konten pedagogis akan mempengaruhi pembelajaran yang diterapkan oleh guru tersebut (Nilsson, 2014). Pengukuran kemampuan guru yang lebih kontekstual adalah mengukur kemampuan tersebut secara terintegrasi, karena pada konteks pembelajaran sebenarnya kemampuan guru tersebut diterapkan secara terintegrasi atau tidak terpisah-pisah. Salah satu bentuk integratif adalah konsep PCK yang merupakan representasi dari pencampuran konten dan pedagogi menjadi suatu pemahaman tentang cara mengorganisir, adaptasi, dan merepresentasikan aspek-aspek tertentu dari materi pelajaran dalam pengajaran (Putra, Widodo, & Sopandi, 2018; Shulman, 2015). Begitu pula TPACK yang merupakan pengembangan dari PCK dengan mengintegrasikan pengetahuan teknologi, dapat menjadi kerangka dasar yang lebih sesuai untuk suatu asesmen bagi para guru bahkan dampaknya terhadap hasil belajar siswa (Nurina dkk., 2019).

Kerangka kerja dari suatu asesmen juga menjadi perhatian para peneliti sebagai bagian yang tidak terpisahkan dari instrumen itu sendiri. Standar untuk menguji kemampuan guru umumnya merupakan standar yang ditetapkan oleh pemerintah, seperti yang berlaku di Indonesia berupa UKG (Kementerian Pendidikan dan Kebudayaan, 2015, 2020) atau di beberapa negara lain berupa pengukuran kemampuan guru atau calon guru (Okhremtchouk, Newell, & Rosa, 2013; SEAMEO INNOTECH, 2010). Sementara berbagai penelitian untuk mengembangkan instrumen dan mengukur kemampuan guru menggunakan

kerangka kerja antara lain *pedagogical content knowledge* (PCK), *technological pedagogical content knowledge* (TPACK) atau bagian dari pengetahuan terintegrasi tersebut seperti pengetahuan konten, pengetahuan pedagogi, serta penerapan pengetahuan dan kinerja dalam pelaksanaan pembelajaran (Bertram & Loughran, 2012; Kleickmann dkk., 2017; Putra, Widodo, & Sopandi, 2018; Wahlen dkk, 2020; Yeh dkk., 2015).

TPACK menjadi salah satu kerangka kerja untuk asesmen yang dianggap sesuai dengan konteks perkembangan zaman saat ini dan dapat mempengaruhi perkembangan pembelajaran sekaligus kemampuan guru dalam melakukan tugasnya (Bilici, Guzey, & Yamak, 2016; Mishra, 2019). Hasil-hasil penelitian sebelumnya yang menunjukkan bahwa tingkat TPACK sebagai suatu pengetahuan dipengaruhi oleh faktor-faktor internal dari guru itu sendiri, seperti pengetahuan dasar untuk mengajar, pengetahuan konten, kemampuan dalam menggunakan perangkat teknologi, efikasi diri, persepsi, perilaku, serta motivasi, ataupun faktor eksternal seperti kurikulum, ketersediaan fasilitas berupa perangkat teknologi di sekolah, serta pelatihan-pelatihan yang diberikan memberikan gambaran bahwa konsep ini memiliki kompleksitas yang tinggi (Kartal, Kartal, & Uluay, 2016; Koh, Chai, & Tsai, 2013; Schmidt dkk., 2009; Yulisman dkk., 2019). Pengembangan kemampuan TPACK juga telah menjadi bagian dari reformasi kurikulum di beberapa negara dan penerapan mata kuliah yang terkait teknologi telah menunjukkan potensi untuk meningkatkan TPACK para calon guru (Dalal, Archambault, & Shelton, 2017; Kovalik, Kuo, & Karpinski, 2013). Oleh karena itu mengukur TPACK menjadi suatu tantangan tersendiri karena kompleksitasnya tersebut. Instrumen-instrumen asesmen TPACK yang telah ada saat ini umumnya merupakan bagian dari penelitian untuk pengembangan TPACK guru ataupun masih diterapkan secara terbatas, karena kebutuhan sumber daya yang banyak jika akan diterapkan dalam skala luas (Anwar dkk., 2014; Bertram & Loughran, 2012; Hadiyanti, Widodo, & Rochintaniawati, 2015; Williams & Lockley, 2012). Berdasarkan hasil-hasil sebelumnya, maka masih terbuka peluang untuk mengembangkan instrumen yang valid dan akurat serta efisien, untuk dapat diterapkan dalam praktik sistem evaluasi atau asesmen bagi guru-guru secara luas.

Penelitian-penelitian yang dilakukan untuk mengukur atau menilai pengetahuan TPACK guru telah berkembang sejak kerangka kerja tersebut dipublikasikan (Koehler & Mishra, 2006). Jenis-jenis instrumen yang digunakan untuk mengukur TPACK pada periode 2006 – 2011 meliputi pengukuran berdasarkan laporan/evaluasi diri, penilaian unjuk kerja, wawancara, kuisisioner pertanyaan terbuka, dan observasi (Koehler, Shin, & Mishra, 2012). Tingkat TPACK guru tentunya dapat diukur dengan penerapan instrumen CoRes dan PaP-eRs, seperti kerangka PCK yang menjadi dasar pengembangannya (Bertram & Loughran, 2012; Hadiyanti dkk., 2015). Sementara pengukuran TPACK menggunakan instrumen tes belum menjadi prosedur yang umum digunakan. Instrumen tes uraian dapat memberikan hasil yang lebih objektif dengan penggunaan panduan penskoran yang akurat, walaupun unsur subjektivitas masih ada tetapi dianggap lebih rendah daripada pengisian kuisisioner sebagai evaluasi mandiri yang nilainya ditentukan oleh responden ataupun pengukuran yang menggunakan observasi sebagai sumber data yang harus melibatkan sejumlah besar observer sebagai penilai. Selain itu, secara konseptual TPACK adalah pengetahuan yang menjadi bagian dari kompetensi kognitif, sehingga pengukuran menggunakan prosedur tes dapat dilakukan, seperti juga mengukur pengetahuan ataupun keterampilan berpikir seseorang dalam konsep yang lain (Großschedl, Welter, & Harms, 2018; Jüttner dkk., 2013; Kastner & Stangla, 2011). Penelitian ini, yang menjadikan instrumen soal bentuk uraian sebagai produk dan TPACK sebagai kerangka kerja dalam pengukuran pengetahuan, memiliki potensi untuk menjadi bagian yang baru dalam sistem asesmen atau evaluasi kompetensi guru.

Asesmen yang baik memiliki kriteria atau karakteristik yang terstandar. Instrumen yang digunakan dalam suatu penilaian harus valid, dapat diandalkan dalam hal kesimpulan dan skor yang dihasilkannya, serta reliabilitas atau memiliki konsistensi hasil penilaian yang dilakukan dalam waktu yang berbeda (Hung & Yang, 2015). Selain itu umumnya instrumen dalam suatu asesmen diharapkan memiliki tingkat kesukaran yang sesuai dengan tujuan penilaian dan daya beda yang tinggi dalam pengukuran kemampuan yang dimiliki subjek tes (Budiyono, 2018). Asesmen yang baik juga memerlukan instrumen yang lebih autentik karena penggunaan instrumen yang memiliki otensitas lebih tinggi akan menghasilkan

informasi yang semakin relevan dengan kondisi sebenarnya yang dimiliki subjek tes (Sevcikova, 2018; Nkhoma dkk., 2020). Efisiensi dalam penyusunan soal yang valid (Rios & Wang, 2018) dan implementasi asesmen untuk subjek dalam jumlah banyak (Williamson dkk., 2010) juga menjadi pertimbangan untuk pengembangan dan penerapan instrumen yang terstandar. Beberapa karakteristik pokok tersebut juga dapat menjadi dasar untuk mengkaji instrumen-instrumen yang telah digunakan atau untuk pengembangan instrumen alternatif yang lebih baik dan lebih akurat.

Darling-Hammond dkk. (2013) telah melakukan studi yang menggambarkan bahwa suatu penilaian dapat dijadikan prediktor untuk situasi yang akan datang dalam kinerja guru. *Performance Assessment for California Teacher* (PACT) adalah suatu bentuk asesmen untuk para guru baru yang telah dikembangkan dan diterapkan di Amerika Serikat. Skor PACT dari individu sasaran merupakan prediktor yang signifikan dari keefektifan mengajar selanjutnya, yang diukur melalui perolehan hasil belajar siswanya untuk mata pelajaran English Language Arts dan Matematika. Perangkat PACT berkontribusi terhadap proses belajar untuk mempelajari cara mengajar, berupa tambahan pengetahuan dan keterampilan mengajar. Hasil studi tersebut menunjukkan bahwa pengembangan instrumen penilaian yang tepat merupakan suatu faktor esensial dalam pencapaian tujuan penilaian.

Kendala dalam penilaian kinerja atau kemampuan guru dalam skala luas saat ini jika tidak menggunakan instrumen objektif antara lain adalah jumlah guru yang banyak, integritas, konsistensi, dan kompetensi penilai, serta efisiensi waktu dan sumber daya dalam pelaksanaan penilaian (Jonsson & Svingby, 2007; Nehm & Haertig, 2012). Permasalahan-permasalahan tersebut dapat berimplikasi terhadap proses dan hasil penilaian yang dilakukan. Jika kendala-kendala tersebut tidak diatasi dan ujian tetap dilakukan, maka dampak yang dapat terjadi antara lain bias dari hasil pengukuran, kualitas tidak optimal, dan pelaksanaan tidak tuntas menyeluruh untuk semua guru. Data yang dihasilkan dari pengukuran yang bias tidak dapat dijadikan dasar tindak lanjut yang relevan, khususnya untuk pengembangan profesi yang efektif (Sumaryanta dkk., 2018). Bias juga dapat terjadi jika hasil pengukuran melibatkan penilai, bukan komputer, walaupun

penilai tersebut telah terlatih ataupun seorang pakar. Faktor jumlah guru yang banyak juga akan membuat pelaksanaan penilaian lebih cenderung praktis daripada kualitasnya. Kebutuhan sumber daya juga akan menjadi menyulitkan dari segi non teknis, yang pada akhirnya dapat mempengaruhi kuantitas dan kualitas hasil penilaian.

Penerapan instrumen berupa uraian umumnya masih merupakan alternatif atau tambahan dari instrumen pilihan ganda pada pengukuran pengetahuan khususnya dalam konteks siswa. Kepraktisan dalam proses analisis butir soal dan skoring masih pertimbangan yang paling mempengaruhi. Pada soal berupa pilihan ganda biasa, peserta dapat menebak pilihan jika dia tidak mengetahui jawabannya. Tebakan tersebut bisa saja benar dan hal ini membuat bias hasil dari suatu tes. Jawaban yang salah dari peserta juga sulit dianalisis untuk mengetahui tingkat pemahamannya, khususnya jika soal dimanfaatkan untuk tujuan formatif. Menurut Kastner & Stangla (2011), soal pilihan ganda relatif dapat menggantikan soal uraian jika soal pilihan ganda berupa pilihan ganda biasa dengan multi respon. Permasalahannya adalah penyusunan soal pilihan ganda, baik yang berupa pilihan ganda biasa atau dengan multi respon lebih sulit daripada menyusun soal yang meminta jawaban uraian ataupun isian singkat. Berdasarkan penjelasan hasil penelitian Budiyo (2018), Funk & Dickson (2011), serta Sam dkk. (2016), hasil ujian siswa menggunakan instrumen berupa pilihan ganda menunjukkan hasil yang lebih tinggi daripada hasil tes dalam bentuk isian singkat untuk paket soal yang sama. Selain itu soal pilihan ganda juga memiliki keterbatasan dalam pengukuran untuk pengetahuan pada tingkatan yang kompleks secara objektif, misalnya untuk kemampuan analisis atau evaluasi, karena jawaban telah dibatasi dalam pilihan yang diberikan. Di sisi lain instrumen uraian dianggap lebih sesuai untuk mengukur tingkatan pengetahuan yang lebih kompleks tersebut. Oleh karena itu penelitian-penelitian mengenai penggunaan instrumen isian atau uraian singkat semakin diperlukan untuk menyediakan alternatif instrumen pengukuran yang lebih baik, termasuk untuk asesmen pengetahuan guru.

*Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report* menyarankan penggunaan tugas menulis karena dianggap dapat membantu siswa memperkuat pengetahuan statistika dan keterampilan

berkomunikasi secara praktis (Zimmerman dkk., 2018). Bagaimana jika tugas tersebut diterapkan dalam penilaian untuk guru? Pengukuran kemampuan guru selama ini juga umumnya menggunakan tes pilihan ganda. Pertanyaan respon terbuka, seperti uraian pendek, memberikan kesempatan kepada subyek untuk mempraktikkan komunikasi sesuai pelajarannya masing-masing. Instrumen tersebut juga memberikan data yang kaya terkait dengan pemahaman, memperkecil kemungkinan menebak dalam menjawab, dan tidak memberikan informasi yang salah kepada yang mengerjakannya. Penelitian tersebut menunjukkan tes uraian dapat menjadi alternatif yang lebih baik untuk mengukur pengetahuan daripada tes pilihan ganda. Penelitian yang dilakukan kepada siswa dapat diadaptasi untuk mengukur kemampuan guru, termasuk penggunaan bentuk tes uraian.

Soal berbentuk uraian umumnya terkendala dalam penskoran jika dilakukan secara manual karena penskoran untuk soal respon terbuka adalah suatu proses yang membutuhkan sumber daya secara intensif (Zhang, 2013). Jika jumlah peserta tes ribuan, maka diperlukan ratusan penskor orang sesuai dengan rasio dengan jumlah peserta. Seorang penilai atau *rater* akan kesulitan untuk konsisten menilai jawaban dari sasaran yang berjumlah banyak. Penilai yang berjumlah banyak juga akan memiliki kemungkinan perbedaan persepsi dalam melakukan penskoran. Hal tersebut akan berdampak pula terhadap bias yang akan ditimbulkan oleh masing-masing penskor orang. Setiap penskor orang memiliki batasan untuk pengukuran dalam aspek objektivitas, reliabilitas dan validitas (Liu dkk., 2016). Kemampuan rater yang berbeda-beda akan menghasilkan skor yang tidak konsisten (Powers, Escoffery, & Duchnowski, 2015). Umumnya jawaban uraian hasil tes diberi skor oleh beberapa rater, hal ini akan meningkatkan kebutuhan sumber daya dan dana. Kedua hal tersebut menjadi kendala dalam penerapan instrumen esai sebagai bagian dari asesmen untuk skala nasional. Pengembangan perangkat penilaian atau penskoran otomatis berbasis teknologi diperlukan untuk mengatasi kendala tersebut.

Penskoran otomatis dapat didefinisikan sebagai penggunaan teknologi komputer untuk mengevaluasi dan memberikan skor. Penskoran tersebut umumnya dilakukan dengan cara mengagregasi relevansi-konstruksi teks yang dapat diekstraksi dari respon tertulis dan dikombinasikan menjadi model



matematis untuk menghasilkan skor (Rios & Wang, 2018). Pemanfaatan sistem atau program *komputer-automated scoring* dapat mengatasi kendala terkait penskoran dari hasil jawaban soal uraian (Aji, Baizal, & Firdaus, 2011; Shermis dkk., 2010; Toranj & Ansari, 2012). Penggunaan aplikasi praktis dalam pendidikan guru dan penelitian, khususnya berupa penilaian otomatis berpotensi mendukung penilaian manusia, karena penilaian otomatis dapat mengurangi jumlah sumber daya yang diperlukan untuk penilaian tersebut (Wahlen dkk., 2020). Pemanfaatan program komputer ini menjadikan tes lebih efisien, hampir sama dengan penggunaan soal pilihan ganda yang memiliki kemudahan dalam penskoran. Penelitian untuk penerapan asesmen dengan penskoran berbantuan komputer membuka peluang tes yang dilakukan untuk sasaran berjumlah banyak. Metode berbantuan komputer lebih umum digunakan dalam tes menggunakan soal pilihan ganda, yang merupakan metode pada ujian nasional untuk siswa dan uji kompetensi guru. Uji kompetensi guru berfungsi sebagai tes formatif, hasil jawaban peserta diberikan kepada yang pihak yang berkepentingan, misalnya Dinas Pendidikan, kelompok kerja guru, atau pihak sekolah, selain Kementerian Pendidikan itu sendiri. Potensi instrumen soal berbentuk uraian yang jawabannya diskor menggunakan komputer tersebut menjadikannya tidak hanya bersifat efisien, tetapi juga meningkatkan kualitas data yang diperlukan untuk peningkatan kualitas guru.

Kecerdasan buatan yang dapat diterapkan dalam lingkup pendidikan ini dapat didefinisikan sebagai teknologi yang mensimulasikan kecerdasan manusia, untuk mendefinisikan dan mencoba menyelesaikan persoalan menggunakan komputer melalui peniruan cara manusia menyelesaikan suatu permasalahan (Russel & Norvig, 1995). Pemanfaatan program komputer seperti itu dapat menjadikan tes lebih efisien, sama dengan penskoran pada soal pilihan ganda (Aji dkk., 2011). Algoritma yang berbasis perbandingan sinonim atau struktur ontology merupakan kecerdasan artifisial yang dapat digunakan dalam penskoran untuk jawaban singkat atau uraian singkat (Firdausiah dkk., 2008). Program komputer dalam bentuk kecerdasan artifisial (*artificial intelligence/AI*) tidak hanya berupa program pengecekan similarisasi kalimat antara jawaban peserta tes dan kunci jawaban, tetapi dapat dilakukan penskoran dengan pendekatan rubrik

seperti pada beberapa perangkat penskor esai otomatis (Zhang, 2013). Penggunaan AI inilah yang memerlukan proses uji performa sebelum digunakan sebagai perangkat dalam asesmen. Instrumen asesmen yang dikembangkan tentunya memiliki karakter kontekstual tertentu, sehingga penggunaan perangkat pendukung berupa teknologi baru perlu diteliti kesesuaian dan efektifitasnya. Di sisi lain, perangkat teknologi yang digunakan belum pernah diuji performanya untuk instrumen yang sejenis dengan instrumen yang dikembangkan dalam penelitian ini. Oleh karena itu uji coba dan kajian yang berhubungan dengan penggunaan teknologi untuk asesmen ini masih diperlukan.

## 1.2. Rumusan Masalah Penelitian

Masalah yang akan diselesaikan dalam penelitian ini adalah “Bagaimana pengembangan instrumen asesmen berbentuk soal uraian untuk mengukur TPACK guru IPA dengan penskoran berbantuan komputer?”. Pertanyaan penelitian yang lebih rinci untuk rumusan masalah tersebut adalah:

1. Bagaimana karakteristik, validitas dan reliabilitas instrumen soal uraian yang dikembangkan untuk pengukuran TPACK guru Ilmu Pengetahuan Alam?
2. Bagaimana validitas dan reliabilitas rubrik yang digunakan sebagai panduan dalam penskoran terhadap jawaban soal uraian asesmen TPACK guru Ilmu Pengetahuan Alam?
3. Bagaimana performa penskoran berbantuan komputer yang digunakan untuk penskoran jawaban soal berbentuk uraian untuk pengukuran TPACK guru Ilmu Pengetahuan Alam?
4. Bagaimana hasil asesmen TPACK guru Ilmu Pengetahuan Alam menggunakan instrumen soal esai dengan penskoran jawaban berbantuan komputer?

## 1.3. Pembatasan Masalah

Penelitian ini mengkaji pengembangan instrumen soal uraian dan rubrik penskoran untuk asesmen pengetahuan guru. Instrumen asesmen yang dikembangkan meminta jawaban berupa uraian terbatas. Kerangka pikir yang menjadi landasan penyusunan instrumen adalah bagian dari pengetahuan konten pedagogis teknologis atau *technological pedagogical content knowledge*

(TPACK) dalam pandangan transformatif dan secara efisien. guru IPA serta uji coba sistem penskoran menggunakan bantuan komputer. Kajian dilakukan terhadap validitas, reliabilitas, serta efektifitas instrumen soal dan rubrik penskoran yang dikembangkan. Penelitian juga dilakukan untuk mengetahui performa perangkat lunak untuk penskoran jawaban uraian.

#### **1.4. Tujuan Penelitian**

Tujuan yang ingin dicapai dari penelitian ini adalah menghasilkan instrumen uraian yang dapat digunakan dalam pengukuran TPACK guru IPA SMP. Karakter instrumen yang diharapkan adalah autentik, kontekstual, valid, reliabel, dan efektif dengan sistem penskoran berbantuan komputer yang efisien dan reliabel untuk tes yang melibatkan subjek dalam jumlah besar.

#### **1.5. Manfaat Penelitian**

Penelitian ini diharapkan memberikan kontribusi yang pertama adalah menjadi model pengembangan instrumen asesmen pengetahuan (TPACK) guru yang lebih autentik sesuai dengan kebutuhan evaluasi pendidik di Indonesia. Kedua, hasil penelitian berupa instrumen soal dan rubrik dapat menjadi model instrumen yang efektif menjadi referensi untuk kebutuhan data tingkat kompetensi dalam sistem evaluasi kompetensi guru. Kontribusi ketiga adalah penggunaan teknologi dalam sistem evaluasi sebagai basis sesuai konteks dalam perkembangan teknologi pendidikan saat ini dan masa yang akan datang.

#### **1.6. Struktur Organisasi Disertasi**

Sistematika umum disertasi ini tersusun dari 5 bab utama serta bagian kelengkapan lainnya berupa abstrak, daftar pustaka, dan lampiran-lampiran. Masing-masing bab berisi penjelasan yang relevan, dengan garis besar seperti berikut ini:

##### 1) Bab I pendahuluan

Bab pertama sebagai pendahuluan disertasi tersusun oleh latar belakang penelitian, rumusan masalah penelitian, pembatasan masalah penelitian, tujuan penelitian, manfaat penelitian, serta struktur organisasi disertasi.

Subbab latar belakang, rumusan masalah, dan pembatasan masalah penelitian menjadi kerangka dasar untuk penulisan bab selanjutnya yang menggambarkan suatu proses penelitian yang utuh.

2) Bab II kajian pustaka dan kerangka pikir penelitian

Literatur yang menjadi rujukan untuk penelitian ini meliputi artikel atau laporan hasil-hasil penelitian sebelumnya serta pustaka ilmiah lainnya yang mengkaji tentang asesmen atau evaluasi kompetensi guru, instrumen pengukuran pengetahuan, kerangka kerja TPACK, serta sistem penskoran esai otomatis berbantuan komputer. Kajian pustaka ini menjadi dasar untuk pengembangan instrumen dan pembahasan temuan-temuan pada Bab IV.

3) Bab III metode penelitian

Bab III ini adalah penjelasan mengenai desain penelitian dan prosedur dalam pelaksanaan penelitian, termasuk partisipan dan instrumen yang dilibatkan dalam penelitian. Untuk penelitian ini prosedur yang dijelaskan sesuai dengan metode penelitian yang diterapkan, yaitu prosedur untuk penelitian dan pengembangan serta penyesuaiannya.

4) Bab IV temuan dan pembahasan

Penulisan untuk Bab IV sesuai dengan rumusan masalah penelitian pada Bab I, yaitu meliputi 4 bagian utama terkait dengan uji coba instrumen dan perangkat penskoran serta hasil implementasi instrumen yang telah disusun sebagai hasil pengembangan. Masing-masing pertanyaan penelitian dibahas sesuai dengan temuan dalam proses penelitian. Pembahasan dilakukan dengan merujuk berbagai literatur yang relevan dan telah dikaji sebelumnya sebagai penguat temuan-temuan dalam penelitian ini. Pembahasan dilakukan pula untuk kebaruan pada temuan-temuan dalam penelitian ini.

5) Bab V kesimpulan, implikasi, dan rekomendasi

Hasil kesimpulan dari pembahasan dituliskan dalam Bab V yang sesuai pula dengan perumusan masalah dan pertanyaan penelitian. Kesimpulan yang ditulis merupakan jawaban-jawaban ringkas dan jelas untuk permasalahan dan pertanyaan-pertanyaan penelitian yang sebelumnya dituliskan pada Bab I. Bab V juga berisi implikasi dan rekomendasi berdasarkan hasil penelitian yang dilakukan terhadap aspek-aspek terkait.