

## **BAB III**

### **METODOLOGI PENELITIAN**

Bab ini terdiri dari empat bagian yaitu: paradigma penelitian; desain penelitian; lokasi dan partisipan penelitian; dan uji kualitas instrumen. Pada bagian paradigma penelitian diuraikan tentang bagaimana pandangan penulis melakukan asesmen *PCK* dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik. Selanjutnya bagian desain penelitian menjelaskan bagaimana prosedur penelitian yang dilakukan. Berikutnya dijelaskan lokasi dan partisipan penelitian, kemudian bagian akhir bab ini menjelaskan bagaimana menguji kualitas instrumen yang dikembangkan. Uji kualitas instrumen meliputi uji validitas, ujicoba awal untuk mendapatkan informasi tentang kepraktisan instrumen serta ujicoba utama untuk menguji reliabilitas (konsistensi) instrumen.

#### **3.1 Paradigma Penelitian**

Peningkatan kualitas pendidikan, termasuk instrumen pembelajaran fisika pada jenjang sekolah seharusnya dimulai dari usaha meningkatkan kualitas penyiapan calon guru di Perguruan Tinggi. Kualitas guru pertama-tama ditentukan oleh pendidikan calon guru di Lembaga Pendidikan Tenaga Kependidikan atau LPTK (Jalal & Supriadi, 2001). Semakin baik kualitas lulusan LPTK, semakin besar peluang untuk meningkatkan kualitas pendidikan. Ketercapaian mutu pendidikan guru bergantung pada bagaimana para calon guru belajar pengetahuan, keterampilan, dan karakteristik kepribadian yang dibutuhkan. Salah satu usaha penyiapan calon guru yang profesional dilakukan melalui pembekalan tentang *Pedagogical Content Knowledge (PCK)*.

*PCK* dibangun oleh sebuah paradigma yang menggambarkan pemahaman profesional guru secara khusus. Paradigma ini pertama kali diperkenalkan oleh Shulman (1986). *PCK* terus menjadi ide yang baik, yang masuk akal dan berguna baik bagi penelitian pendidikan sains maupun pendidik guru sains dalam rangka meningkatkan profesionalisme guru atau calon guru terkait pengetahuan konten dan pengetahuan pedagogi.

Guru harus memiliki kompetensi yang baik untuk melaksanakan tugas keprofesionalannya sebagai seorang pendidik, sebagaimana diamanatkan dalam

Peraturan Menteri Pendidikan Nasional nomor 16 Tahun 2007 tentang Standar Kualifikasi Akademik dan Kompetensi Guru. *PCK* merupakan pengetahuan profesional yang seyogyanya dikuasai oleh guru, konten pembelajaran (materi subjek) tidak semata-mata hanya menyangkut konsep, tetapi termasuk di dalamnya berbagai macam keterampilan, antara lain keterampilan komunikasi dan penyelesaian masalah saintifik sebagai bekal yang harus dimiliki siswa untuk dapat bersaing di era abad 21. Oleh karena itu *PCK* calon guru dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik harus diases untuk memastikan bahwa calon guru memiliki kemampuan *PCK* yang baik dalam pembelajaran.

Pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik dilakukan secara terintegrasi dengan materi subjek yang sedang dibelajarkan kepada siswa. Integrasi tersebut diases dengan instrumen yang dikembangkan khusus. Instrumen yang dirancang mengases komponen keterampilan komunikasi dan penyelesaian masalah yang dimunculkan dalam matriks *PCK* (matriks rencana pembelajaran). Matriks *PCK* diadaptasi dari *Content Representation (CoRe)* yang dikembangkan Loughran, Pamela dan Amanda (2004). Paradigma integrasi asesmen keterampilan komunikasi dan penyelesaian masalah dalam *CoRe* ditampilkan pada Gambar 3.1.



Gambar 3.1 Paradigma Integrasi Asesmen Pengembangan Keterampilan Komunikasi dan Penyelesaian Masalah dengan *CoRe*

Erwin, 2019

**PENGEMBANGAN INSTRUMEN ASESMEN PEDAGOGICAL CONTENT KNOWLEDGE CALON GURU FISIKA DALAM KONTEKS PEMBELAJARAN KETERAMPILAN KOMUNIKASI DAN PENYELESAIAN MASALAH SAINTIFIK**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

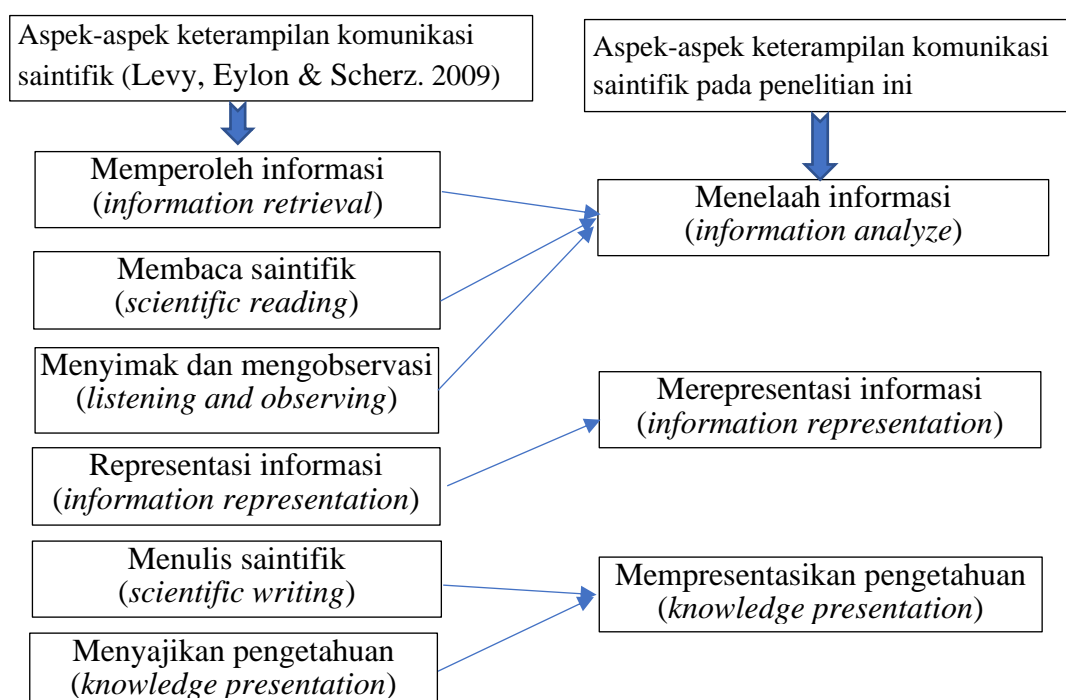
Gambar 3.1 memperlihatkan bagaimana pertanyaan-pertanyaan pedagogi yang difokuskan untuk mengungkap *PCK* guru dalam membelajarkan konsep (materi subjek), diperluas secara bersamaan untuk mengungkap *PCK* dalam membelajarkan keterampilan komunikasi dan penyelesaian masalah saintifik. Paradigma ini menunjukkan bahwa membelajarkan keterampilan komunikasi dan problem solving saintifik dilakukan terintegrasi dengan konsep. Dengan demikian pembelajaran keterampilan tersebut bukan keterampilan yang dibelajarkan secara terpisah dengan konsep. Misalnya pertanyaan pedagogi yang berbunyi “apa yang Anda inginkan dipelajari siswa tentang ide ini?”, pertanyaan ini harus dijawab guru tidak hanya tentang apa yang ingin dipelajari siswa tentang konsep, tetapi sekaligus menjawab apa yang diinginkan guru untuk dipelajari siswa tentang keterampilan komunikasi saintifik dan apa yang diinginkan guru untuk dipelajari siswa tentang penyelesaian masalah saintifik. Demikian seterusnya, semua pertanyaan pedagogi harus dijawab untuk bagian konsep, keterampilan komunikasi dan penyelesaian masalah saintifik.

Amalgamasi antara pengetahuan konten dan pengetahuan pedagogi merupakan wujud kemampuan guru membentuk *PCK* yang meliputi bagaimana guru menyesuaikan kurikulum atau rencana pembelajaran yang dikhususkan pada konten dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik, merumuskan tujuan pembelajaran sesuai dengan kurikulum yang disusun dan memilih strategi pembelajaran yang efektif pada materi khusus untuk mencapai tujuan yang telah dirumuskan.

Asesmen *PCK* dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik, dilakukan pada rencana pembelajaran yang disusun dalam bentuk matriks rencana pembelajaran oleh calon guru. Asesmen *PCK* dilakukan sebagai upaya mewujudkan guru profesional yang memiliki kompetensi sesuai dengan standar pendidik dan tenaga kependidikan, spesifik pada konten dalam konteks khusus komunikasi dan penyelesaian masalah saintifik.

Aspek-aspek keterampilan komunikasi saintifik meliputi: memperoleh informasi (*information retrieval*), membaca saintifik (*scientific reading*), mendengarkan dan mengobservasi (*listening and observing*), menulis saintifik (*scientific writing*), representasi informasi (*information representation*) dan

menyajikan pengetahuan (*knowledge presentation*) (Levy, Eylon & Scherz, 2009). Dalam penelitian ini aspek-aspek keterampilan komunikasi saintifik tersebut disederhanakan menjadi tiga aspek yang meliputi: menelaah informasi (*information Analyze*) (i), merepresentasi informasi (*information representation*) (ii), dan mempresentasikan pengetahuan (*knowledge presentation*) (iii). Aspek-aspek keterampilan komunikasi saintifik dimodifikasi dari Levy, Eylon dan Scherz (2009), dan disederhanakan yang terdapat pada Gambar 3.2.

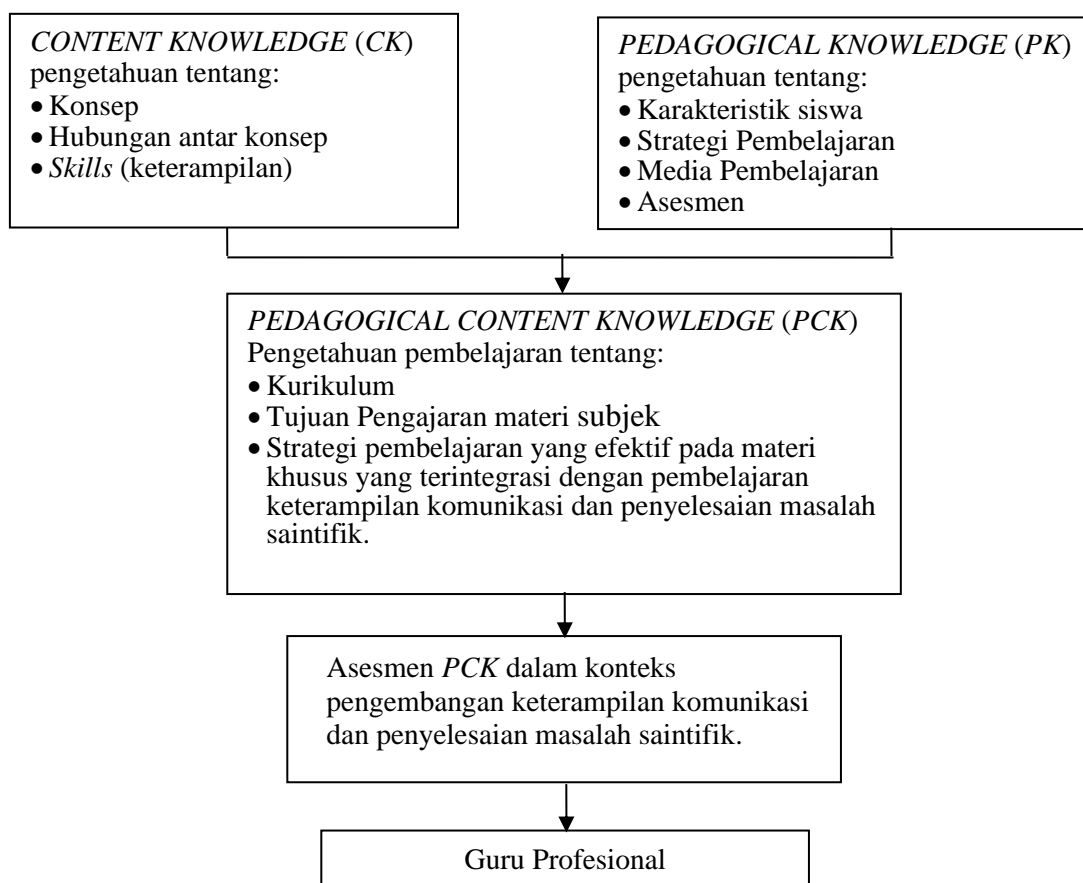


Gambar 3.2. Adaptasi Aspek-Aspek Keterampilan Komunikasi Saintifik dalam Penelitian

Sementara itu aspek-aspek keterampilan penyelesaian masalah diadaptasi dari langkah-langkah penyelesaian masalah yang digunakan Selcuk, Çaliskan dan Erol (2008) serta Sardi, Rizal dan Mansyur (2018). Aspek-aspek tersebut meliputi: memahami masalah (*understanding the problem*), merencanakan solusi (*devising a plan*), melaksanakan rencana (*carrying out the plan*) dan memeriksa kembali (*looking back*). Aspek-aspek keterampilan komunikasi dan penyelesaian masalah saintifik menjadi bagian yang harus dikembangkan dalam pembelajaran. Dengan demikian calon guru harus merencanakan bagaimana mengembangkannya dan mengintegrasikannya dengan pembelajaran materi subjek tertentu.

Rencana pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik secara eksplisit dituangkan dalam matriks *PCK* yang disusun oleh calon guru. Untuk mengungkap bagaimana *PCK* calon guru, matriks *PCK* yang disusun dinilai menggunakan rubrik yang dikembangkan khusus pada penelitian ini. Alur berpikir asesmen *PCK* dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik terdapat pada Gambar 3.3.

Empat bidang kebijakan utama dalam pengembangan *PCK* yaitu: pendidikan guru pra-jabatan, reformasi kurikulum, pengembangan profesional, dan evaluasi guru (Sickel, Carlson & Driel, 2015). Pengembangan awal *PCK* dilakukan pada saat calon guru sedang menempuh pendidikan di lembaga pendidikan guru tempat belajarnya.



Gambar 3.3. Alur Berpikir Asesmen *PCK* dalam Konteks Pengembangan Keterampilan Komunikasi dan Penyelesaian Masalah Saintifik

Pentingnya *PCK* bagi calon guru menghendaki adanya penelitian tentang bagaimana mengases *PCK* calon guru dalam konteks pengembangan keterampilan

komunikasi dan penyelesaian masalah saintifik, Asesmen tersebut dimaksudkan untuk dapat mengukur kemampuan *PCK* calon guru dalam membelajarkan keterampilan komunikasi dan penyelesaian masalah saintifik, sebelum calon guru lulus menempuh pendidikan profesi keguruan pada LPTK khususnya calon guru fisika.

### 3.2 Desain Penelitian

Secara sistematis penelitian ini dilakukan menggunakan *design and development research* dengan tiga tahapan besar yaitu perencanaan, produksi dan evaluasi (Richey & Klein, 2007) melalui tahapan-tahapan seperti yang disajikan pada Gambar 3.4. Tahap-tahap penelitian sebagaimana terdapat pada Gambar 3.4 dapat dijabarkan secara terperinci sebagai berikut.

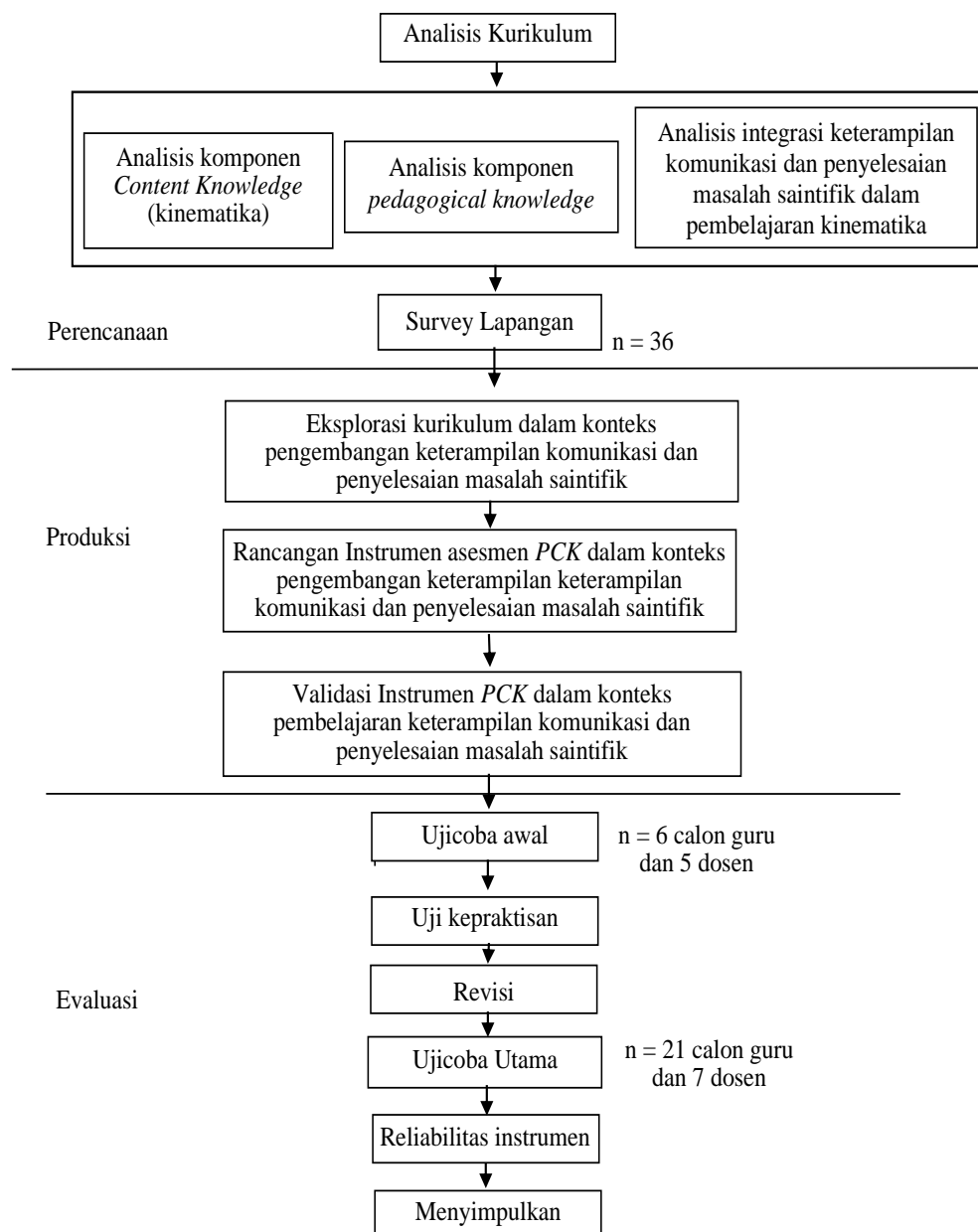
#### 1. Analisis kurikulum

Pada tahap awal dilakukan analisis kurikulum yang diberlakukan saat ini pada pembelajaran di sekolah. Analisis kurikulum dilakukan untuk mengkaji kompetensi inti dan kompetensi dasar yang harus dikuasai siswa dalam mempelajari materi kinematika. Analisis juga dilakukan untuk memahami kedalaman dan keluasan konten kinematika yang harus dipelajari siswa. Analisis kurikulum selanjutnya bertujuan untuk mengkaji strategi atau metode pembelajaran apa yang dianggap tepat untuk membelajarkan materi kinematika sehingga dapat diintegrasikan dengan konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik. Disamping analisis kurikulum, dilakukan analisis literatur terutama hasil-hasil penelitian yang berkaitan dengan *PCK* dan asesmen yang digunakan untuk menilainya.

#### 2. Persiapan/Survei Lapangan

Pada tahap persiapan dan survey lapangan dilakukan perencanaan untuk menyusun langkah-langkah yang akan ditempuh dalam rangka perumusan dan penyusunan instrumen asesmen *PCK* calon guru fisika yang dikhususkan pada konten dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik. Langkah-langkah yang disusun adalah sebagai berikut. Pertama, Lembaga Pendidikan Tenaga Kependidikan (LPTK)

ditentukan sebagai tempat belajar mahasiswa calon guru yang akan digunakan sebagai tempat penelitian. Kedua, studi lapangan dilakukan dengan berbagai kegiatan untuk mengungkap data awal tentang persepsi calon guru fisika tentang *PCK*, keterampilan komunikasi dan penyelesaian masalah saintifik. Langkah yang ditempuh untuk keperluan tersebut melalui penyebaran angket dan wawancara.



Gambar 3.4 Tahapan-tahapan Penelitian

Ketiga, deskripsi temuan dilakukan untuk mendeskripsikan hasil yang diperoleh pada saat studi lapangan, selanjutnya data studi lapangan dianalisis untuk mengetahui bagaimana persepsi mahasiswa calon guru fisika terhadap PCK yang dikhususkan pada konten dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik. Keempat, penyusunan draf awal instrumen asesmen PCK calon guru fisika yang dikhususkan pada konten dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik.

### 3. Penyusunan Draft Instrumen Asesmen *PCK*

Matriks pembelajaran yang digunakan diadaptasi dari Content Representation (*CoRe*) yang dikembangkan oleh Loughran, Pamela dan Amanda (2004) dengan melakukan penyesuaian seperlunya. Adaptasi dan penyesuaian *CoRe* dilakukan setelah ada komunikasi saintifik antara penulis dengan Loughran lewat email dan mendapat persetujuan untuk dikembangkan menjadi matriks pembelajaran untuk mengungkap *PCK* calon guru fisika dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik. *CoRe* yang dikembangkan oleh Loughran, Pamela dan Amanda (2004) berisi 8 pertanyaan pedagogi yang harus diisi oleh guru, pada penelitian ini direduksi menjadi enam pertanyaan, reduksi dilakukan karena instrumen yang dikembangkan ini diperuntukkan bagi calon guru yang baru memiliki sedikit pengalaman. Dua pertanyaan yang direduksi merupakan pertanyaan lanjutan dari pertanyaan sebelumnya. Kedua pertanyaan itu lebih bersifat mendalami dan lebih cocok diberikan pada guru yang sudah memiliki pengalaman yang banyak. Delapan pertanyaan pedagogi yang dikembangkan Loughran, Pamela dan Amanda (2004) dibandingkan dengan enam pertanyaan pedagogi yang digunakan dalam penelitian ini terdapat pada Tabel 3.1.

Pertanyaan nomor tiga yaitu: tentang hal-hal lain dari ide pokok yang diketahui guru tetapi belum saatnya diketahui siswa, merupakan pertanyaan lanjutan dan pendalaman dari pertanyaan nomor satu dan dua sehingga pertanyaan ini sudah tercakup dalam pertanyaan satu dan dua tersebut, pertanyaan nomor tiga tidak digunakan dalam penelitian ini. Pertanyaan



nomor enam (faktor lain yang mempengaruhi cara guru mengajarkan ide pokok), juga merupakan lanjutan atau pendalaman dari pertanyaan nomor lima, sehingga pertanyaan nomor enam ini juga tidak digunakan dalam penelitian ini.

Tabel 3.1  
*Adaptasi Pertanyaan Pedagogi CoRe Loughran, Pamela dan Amanda (2004) dalam Penelitian*

No	Pertanyaan Pedagogi CoRe	Pertanyaan Pedagogi yang digunakan dalam penelitian ini	Keterangan
1	Apa yang anda inginkan dipelajari siswa pada ide pokok ini?	Apa yang Anda inginkan untuk dipelajari siswa pada konsep/sub pokok bahasan /keterampilan ini?	digunakan
2	Mengapa penting bagi siswa untuk belajar ide pokok ini?	Mengapa penting bagi siswa mempelajari konsep (sub pokok bahasan)/keterampilan ini?	digunakan
3	Apa hal lain dari ide pokok ini yang anda ketahui tetapi belum saatnya diketahui siswa?	-	Tidak digunakan
4	Apa kesulitan/keterbatasan yang terkait dengan cara mengajarkan ide pokok ini?	Apa kesulitan/keterbatasan yang terkait dengan cara mengajarkan konsep (sub pokok bahasan)/keterampilan ini?	digunakan
5	Apa pengetahuan tentang pemikiran siswa yang mempengaruhi megajarkan ide pokok ini?	Bagaimana pemikiran (konsepsi) yang dimiliki siswa yang dapat mempengaruhi pengajaran konsep/ sub pokok bahasan /keterampilan ini?	digunakan
6	Faktor lain yang mempengaruhi cara anda mengajarkan konten ini?	-	Tidak digunakan
7	Bagaimana prosedur pengajaran yang Anda gunakan untuk mengembangkan ide pokok ini?	Bagaimana prosedur pengajaran yang Anda gunakan untuk mengembangkan konsep (sub pokok bahasan)/keterampilan ini?	digunakan
8	Bagaimana cara spesifik memastikan pemahaman atau kebingungan siswa mengenai ide pokok ini?	Bagaimana cara memastikan pemahaman atau kebingungan siswa secara spesifik mengenai konsep/sub pokok bahasan (atau memastikan keterampilan yang telah dikuasai siswa secara spesifik)?	digunakan

Sebanyak enam pertanyaan dari delapan pertanyaan dalam *CoRe* yang dikembangkan Loughran, Pamela dan Amanda (2004) dimodifikasi pada

matriks pembelajaran yang dikembangkan sebagai “tugas kinerja”, untuk mengungkap *PCK* calon guru fisika dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik dengan penyesuaian seperlunya. Matriks pembelajaran yang dikembangkan diminta diisi oleh calon guru fisika, matriks tersebut berfungsi sebagai tugas kinerja untuk mengungkap bagaimana *PCK* yang dimiliki oleh calon guru dalam rangka pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik.

Langkah yang ditempuh pada penyusunan draf produk instrumen asesmen *PCK* dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik adalah merancang matriks *PCK* sebagai perangkat untuk mengungkap *PCK* calon guru fisika dalam membelajarkan keterampilan komunikasi dan penyelesaian masalah saintifik. Selanjutnya dirancang rubrik yang berfungsi sebagai acuan dalam menilai matriks *PCK* yang telah diisi oleh calon guru fisika.

#### 4. Validasi Draft Instrumen Asesmen *PCK*

Draf instrumen asesmen *PCK* dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik yang telah dirancang kemudian divalidasi dengan meminta lima orang pakar yaitu pakar asesmen, *PCK* dan pembelajaran fisika. Proses validasi pakar dilakukan menggunakan perangkat lembar daftar cek (*check list*). Setelah dilakukan telaah dan *judgement* oleh para pakar, instrumen direvisi dan diperbaiki sesuai dengan saran para pakar.

#### 5. Ujicoba Awal Instrumen Asesmen *PCK*

Draf produk yang telah divalidasi dan diperbaiki kemudian diujicobakan, tujuan ujicoba untuk mendapatkan masukan apakah produk yang telah dibuat layak digunakan atau tidak. Melalui ujicoba awal ini diharapkan kualitas produk yang telah dikembangkan betul-betul teruji secara empiris. Detail kegiatan yang akan dilakukan pada ujicoba awal adalah menentukan partisipan untuk implementasi, memberikan arahan tentang penggunaan instrumen asesmen yang dikembangkan. Selanjutnya hasil pekerjaan calon guru mengisi matriks *PCK* dalam membelajarkan keterampilan komunikasi

dan penyelesaian masalah saintifik, dinilai menggunakan rubrik yang telah dirancang. Pada ujicoba awal partisipan yang terlibat sebanyak enam calon guru fisika dan lima dosen pada program studi Pendidikan Fisika.

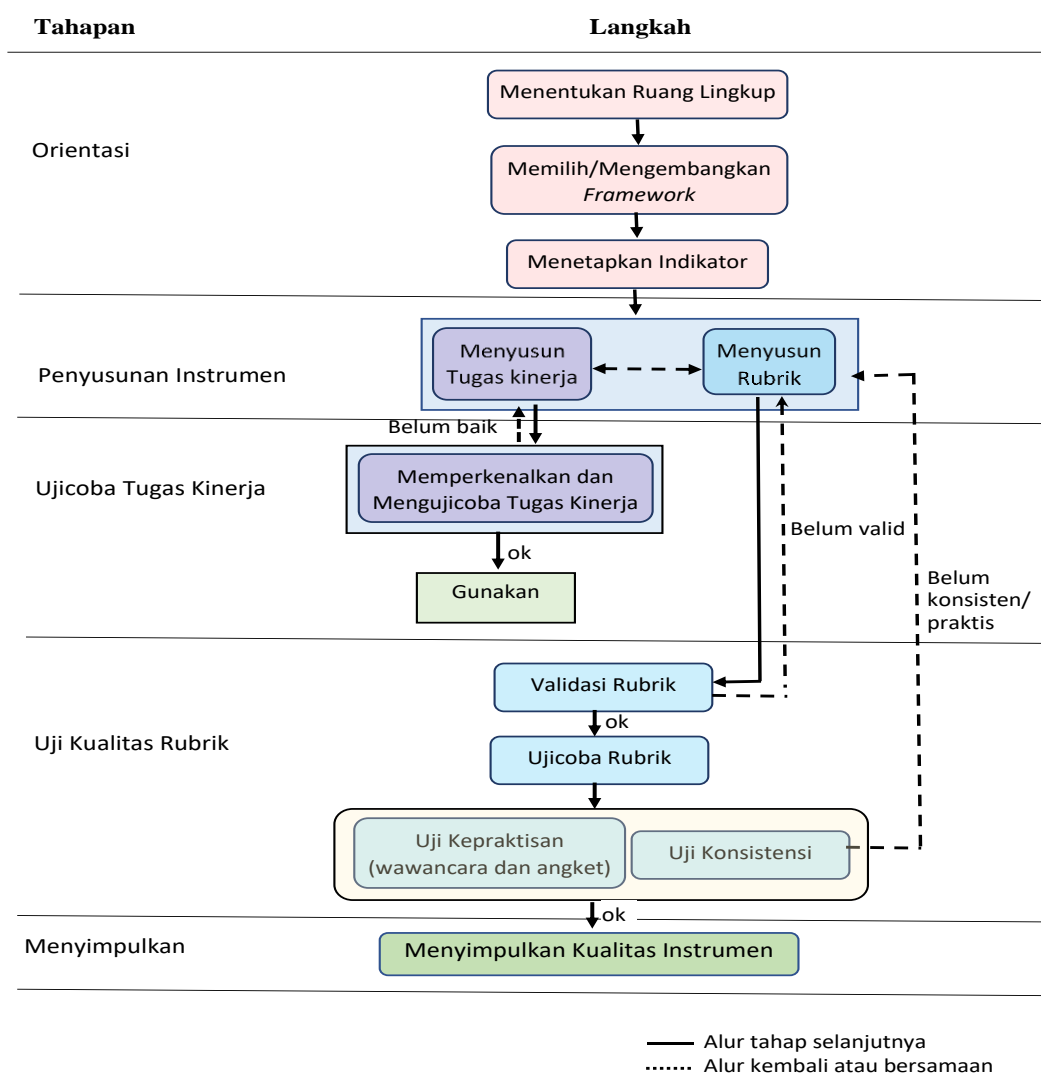
### 3.3 Lokasi dan Partisipan Penelitian

Penelitian ini dilakukan di sebuah Lembaga Pendidik dan Tenaga Kependidikan (LPTK) swasta di Sumatera Selatan yang menyelenggarakan program studi Pendidikan Fisika. Calon guru yang menjadi partisipan penelitian adalah mahasiswa yang telah atau sedang mengikuti perkuliahan pedagogi dan mata kuliah keahlian program studi (materi subjek) dan sedang mengikuti mata kuliah Program Pengenalan Lapangan (PPL) I dan PPL II. Perguruan tinggi lokasi penelitian sejak diberlakukan kurikulum tahun 2013 yang direvisi tahun 2016 mulai memasukkan aspek keterampilan abad 21 dalam penyusunan rencana pembelajaran serta mengaplikasikannya pada praktek pembelajaran yang dilakukan oleh mahasiswa calon guru. LPTK lokasi penelitian berada dalam naungan pondok pesantren sehingga suasana pembelajaran dilakukan dalam nuansa keagamaan, dengan demikian nilai-nilai spiritual dan ahlak (moral) menjadi perhatian utama dalam pembelajaran. Kurikulum program studi pada LPTK ditambah kurikulum lokal untuk pembinaan ahlak dan ibadah keagamaan.

Partisipan penelitian pada ujicoba awal berjumlah enam calon guru fisika, dan lima orang dosen pada program studi pendidikan fisika. Sedangkan pada ujicoba utama, partisipan berjumlah 21 calon guru fisika dan tujuh dosen pendidikan fisika, lima dosen berasal dari LPTK lokasi penelitian, sedangkan dua dosen lainnya berasal dari sebuah perguruan tinggi negeri kependidikan di kota Bandung. Calon guru yang menjadi partisipan penelitian sedang mengikuti PPL II di berbagai sekolah, calon guru ditugaskan melengkapi matriks *PCK* dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik. Matriks *PCK* yang disusun calon guru merupakan bentuk rencana pembelajaran yang akan dijadikan acuan dalam praktek pembelajaran yang mereka lakukan. Dosen yang menjadi partisipan penelitian bertugas menilai matriks *PCK* yang dilengkapi calon guru berpedoman kepada rubrik *PCK* yang dikembangkan.

### 3.4 Prosedur Penelitian

Prosedur penelitian yang diusulkan dalam rangka pengembangan instrumen asesmen PCK calon guru fisika dalam konteks pembelajaran keterampilan komunikasi dan penyelesaian masalah saintifik, dilakukan dengan tahapan-tahapan orientasi, penyusunan instrumen, ujicoba tugas kinerja, uji kualitas instrumen, dan menyimpulkan. Setiap tahapan terdiri dari langkah-langkah tertentu, setiap langkah tidak selalu merupakan urutan yang bersifat mutlak. Satu langkah dapat kembali ke langkah sebelumnya apabila diperlukan. Bagan model pengembangan instrumen asesmen PCK dalam konteks pembelajaran keterampilan komunikasi dan penyelesaian masalah saintifik yang diusulkan dalam penelitian ini ditampilkan pada Gambar 3.5.



Gambar 3.5 Model Pengembangan Instrumen Asesmen PCK

Erwin, 2019

**PENGEMBANGAN INSTRUMEN ASESMEN PEDAGOGICAL CONTENT KNOWLEDGE CALON GURU FISIKA DALAM KONTEKS PEMBELAJARAN KETERAMPILAN KOMUNIKASI DAN PENYELESAIAN MASALAH SAINTIFIK**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Gambar 3.5 memperlihatkan tahapan-tahapan dan langkah-langkah yang dilakukan dalam mengembangkan instrumen asesmen PCK dalam konteks pembelajaran keterampilan komunikasi dan penyelesaian masalah saintifik. Pada tahap orientasi, langkah pertama yang dilakukan menentukan ruang lingkup asesmen yang dikembangkan, dilanjutkan dengan pemilihan atau pengembangan *framework* yang akan digunakan dalam pengembangan instrumen. Menyusun indikator instrumen dilakukan setelah pemilihan atau pengembangan *framework* dilaksanakan.

Tahap penyusunan instrumen yang dilakukan berupa penyusunan rubrik, penyusunan rubrik dilakukan berdasarkan tugas kinerja yang direncanakan, karena rubrik yang disusun digunakan untuk menilai tugas kinerja yang dikerjakan calon guru. Penyusunan rubrik dapat dilakukan setelah tugas kinerja selesai disusun (Brown & Mevs, 2012; Metin, 2013), namun ada juga pendapat yang menyatakan bahwa berdasarkan hasil penelitian, menyusun rubrik dapat dilakukan terlebih dahulu kemudian dilanjutkan dengan penyusunan tugas kinerja.

Tahap ujicoba dilakukan setelah tugas kinerja selesai disusun, tahap ujicoba ini diawali dengan memperkenalkan matriks *PCK* kepada calon guru, hal ini perlu dilakukan karena calon guru belum mengenal Matriks *PCK* sebelumnya. Selanjutnya calon guru diminta mencoba melengkapi matriks *PCK* dengan memilih satu konsep (sub pokok bahasan) pada materi kinematika, merencanakan pembelajarannya terintegrasi dengan pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik. Ujicoba ini bertujuan untuk memastikan bahwa calon guru memahami tugas kinerja mereka yaitu cara melengkapi matriks *PCK*. Revisi tugas kinerja dilakukan pada saat ada tugas kinerja yang tidak dipahami calon guru.

Tahap uji kualitas rubrik dilakukan setelah rubrik selesai disusun. Uji kualitas rubrik meliputi uji validitas, uji reliabilitas (konsistensi), dan uji kepraktisan. Uji validitas dilakukan dengan meminta pertimbangan para pakar untuk memvalidasi rubrik, revisi rubrik dilakukan sesuai dengan saran-saran yang berikan para pakar, sampai akhirnya para pakar menilai rubrik tersebut baik. Uji reliabilitas dilakukan dengan uji konsistensi penilaian antar penilai, menggunakan uji statistik *ICC* dan *W Kendall*. Revisi rubrik kembali dilakukan pada saat hasil uji konsistensi

penilaian antar penilai belum memenuhi syarat minimum, sama dengan interpretasi koefisien korelasi yaitu kategori sedang (Gisev *et al.*, 2010). Selanjutnya pengujian konsistensi penilaian antar penilai kembali dilakukan setelah diadakan perbaikan pada rubrik, sampai diperoleh indeks konsistensi penilain yang memadai. Langkah selanjutnya pada tahap uji kualitas rubrik melakukan uji kepraktisan, kepraktisan rubrik dilihat dari sisi kebermanfaatan rubrik, kemudahan menggunakan rubrik, dan pertimbangan waktu yang dibutuhkan menggunakan rubrik. Tahap akhir pengembangan instrumen menyimpulkan kualitas instrumen berdasarkan uji kualitas yang dilakukan.

### 3.5 Instrumen Penelitian

Penelitian ini mengembangkan instrumen PCK dalam konteks pembelajaran keterampilan komunikasi dan penyelesaian masalah saintifik berbentuk rubrik *PCK*. Untuk menguji kualitas rubrik *PCK* dilakukan uji validitas, reliabilitas (konsistensi) penialain antar penilai dan kepraktisan instrumen. Instrumen yang digunakan untuk menguji validitas rubrik menggunakan lembar *judgement*. Lembar *judgement* diberikan kepada pakar yang diminta bantuan melakukan validasi terhadap rubrik yang dikembangkan. Pakar tinggal memberikan tanda centang pada kolom yang disediakan, berdasarkan pertimbangannya apakah deskriptor rubrik sesuai dengan pertanyaan pedagogi, apakah gradasi/tingkatan deskriptor jelas, dan apakah bahasa deskriptor jelas. *Form* lembar *judgement* yang digunakan untuk memvalidasi rubrik ditampilkan pada Gambar 3. 6.

RUBRIK PCK PEMBELAJARAN .....						Pertimbangan Validator (4)			
No	Pertanyaan Pedagogi	Skala dan Deskriptor (3)				Deskriptor sesuai dengan pertanyaan pedagogi	Deskriptor menunjukkan gradasi/tingkatan yang jelas	Bahasa deskriptor jelas dan mudah dipahami	Masukan / saran
		1	2	3	4	Ya	Tidak	Ya	
(1)	(2)	1	2	3	4	Ya	Tidak	Ya	Tidak
1	Pertanyaan 1	Deskriptor	Deskriptor	Deskriptor	Deskriptor				
2	Pertanyaan 2	Deskriptor	Deskriptor	Deskriptor	Deskriptor				
3	Pertanyaan 3	Deskriptor	Deskriptor	Deskriptor	Deskriptor				
4	Pertanyaan 4	Deskriptor	Deskriptor	Deskriptor	Deskriptor				
5	Pertanyaan 5	Deskriptor	Deskriptor	Deskriptor	Deskriptor				
6	Pertanyaan 6	Deskriptor	Deskriptor	Deskriptor	Deskriptor				

Gambar 3.6 *Form* Lembar *Judgement* Validasi Rubrik

Erwin, 2019

**PENGEMBANGAN INSTRUMEN ASESMEN PEDAGOGICAL CONTENT KNOWLEDGE CALON GURU FISIKA DALAM KONTEKS PEMBELAJARAN KETERAMPILAN KOMUNIKASI DAN PENYELESAIAN MASALAH SAINTIFIK**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Instrumen untuk menguji kepraktisan rubrik menggunakan lembar pedoman wawancara. Pedoman wawancara berisi beberapa pertanyaan untuk mengungkap kebermanfaatan yang dirasakan pengguna rubrik, kemudahan menggunakan rubrik, dan alokasi waktu yang dibutuhkan dalam menggunakan rubrik. Lembar pedoman wawancara ditampilkan pada Gambar 3.7.

PEDOMAN WAWANCARA
<p>Pedoman wawancara ini digunakan untuk menilai praktikabilitas rubrik PCK.</p> <p>Pertanyaan Penuntun Wawancara:</p> <ol style="list-style-type: none"> <li>1. Apakah Saudara pernah menilai menggunakan rubrik?</li> <li>2. Apakah rubrik ini menuntun anda untuk memberikan nilai?</li> <li>3. Apakah ada keraguan dalam memberikan skor antara 1, 2, 3 dan 4 pada rubrik ini?</li> <li>4. Apa kendala yang saudara rasakan saat menggunakan rubrik ini?</li> <li>5. Menurut pengalaman menggunakan rubrik yang baru saja saudara alami, berapa jumlah matriks yang bisa dinilai agar penilaian masih objektif?</li> <li>6. Adakah saran untuk perbaikan rubrik ini?</li> </ol>
<b>Gambar 3.7</b> Lembar Pedoman Wawancara

### 3.6 Analisis Data Pengujian Kualitas Rubrik

Instrumen yang dikembangkan dalam penelitian ini, merupakan instrumen asesmen *PCK* dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik berbentuk dan rubrik *PCK* untuk menilai tugas kinerja yang dikerjakan oleh calon guru fisika. Tugas kinerja yang dikembangkan berupa *form* matriks rencana pembelajaran (matriks *PCK*). Adapun rubrik *PCK* yang dikembangkan digunakan untuk menilai matriks rencana pembelajaran yang telah dilengkapi oleh calon guru fisika. Rubrik *PCK* yang dikembangkan ada tiga macam: rubrik untuk menilai matriks rencana pembelajaran konsep/sub pokok bahasan, rubrik untuk menilai matriks rencana pembelajaran pengembangan keterampilan komunikasi saintifik, dan rubrik untuk menilai matriks rencana pembelajaran pengembangan keterampilan penyelesaian masalah.

Calon guru fisika diminta untuk mengisi *form* matriks rencana pembelajaran. Matriks berisi enam pertanyaan pedagogi yang dimodifikasi dari *CoRe* Loughran, Pamela dan Amanda (2004), seperti terdapat pada Tabel 3.1. *Form*

matriks rencana pembelajaran merupakan sebuah format untuk memotret kemampuan *PCK* calon guru membelajarkan sebuah konsep (sub pokok bahasan) khusus, diintegrasikan dengan pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik peserta didik. *Form* matriks rencana pembelajaran diisi dengan menguraikan apa yang diinginkan untuk dipelajari siswa pada konsep (sub pokok bahasan)/keterampilan yang akan dikembangkan, mengapa penting bagi siswa mempelajari konsep (sub pokok bahasan)/keterampilan tersebut, kesulitan/keterbatasan yang terkait dengan cara mengajarkan konsep (sub pokok bahasan)/keterampilan tersebut, pemikiran (konsepsi) yang dimiliki siswa yang dapat mempengaruhi pengajaran konsep (sub pokok bahasan)/keterampilan itu, prosedur pengajaran yang digunakan untuk mengembangkan konsep (sub pokok bahasan)/keterampilan itu, dan cara memastikan pemahaman atau kebingungan siswa secara spesifik mengenai konsep/sub pokok bahasan (atau memastikan keterampilan yang telah dikuasai siswa secara spesifik).

### 3.6.1 Validitas Instrumen

Validitas adalah suatu ukuran kesahihan atau kevalidan sebuah instrumen. Sebuah instrumen dikatakan valid apabila instrumen tersebut dapat mengukur apa yang akan diukur (Azwar, 2012; Sugiyono, 2011; DeVon, *et al.*, 2007). Faktor validitas merupakan hal penting dalam memilih dan menerapkan suatu instrumen. Menetapkan validitas isi merupakan tugas kunci dalam membuat instrumen pengukuran (Haynes, Richard, & Kubany, 1995).

Validitas suatu instrumen sangat erat kaitannya dengan tujuan pembelajaran yang telah dirumuskan. Sebuah instrumen dapat dikatakan memiliki validitas yang baik apabila instrumen tersebut mampu menjalankan fungsinya dengan baik, yaitu mengukur atribut yang seharusnya diukur, jika fungsi instrumen berjalan baik maka akan dihasilkan data yang sesuai dengan tujuan pengukuran. Kevalidan sebuah instrumen juga dikatakan baik apabila mampu memberikan gambaran perbedaan yang sekecil-kecilnya antara satu subjek dengan subjek lainnya.

Instrumen yang layak digunakan sebagai alat ukur yang terpercaya, perlu dipastikan instrumen tersebut telah memenuhi persyaratan substansi, konstruksi, bahasa dan bukti validitas empirik. Instrumen yang disusun untuk kepentingan



penelitian maupun evaluasi perlu memperhitungkan beberapa pertimbangan, misalnya apa yang hendak diukur, apakah data yang akan dikumpulkan relevan dengan karakteristik penelitian dan sejauhmana skor yang terkumpul menggambarkan karakteristik yang menjadi target ukur. Berdasarkan sifat dan fungsinya, validitas instrumen digolongkan dalam tiga kategori besar, yaitu: 1) validitas isi (*Content validity*), 2) validitas konstruk (*Construct validity*), dan 3) validitas berdasarkan kriteria (*Criterion-related validity*) (Azwar, 2012).

Validitas isi merupakan validitas yang diestimasi dengan cara menganalisis isi instrumen. Validitas isi menggambarkan sejauhmana item-item penyusun instrumen mencerminkan atau mewakili keseluruhan isi objek yang akan diukur. Validitas isi suatu instrumen dikatakan valid apabila item-item penyusun instrumen mencerminkan keseluruhan konten atau aspek materi diujikan atau yang seharusnya dikuasai peserta didik secara proporsional. Untuk menimbang apakah suatu instrumen validitas isinya sudah baik atau belum, harus dilakukan penelaahan kisi-kisi instrumen untuk memastikan bahwa item-item penyusun instrumen sudah mewakili atau mencerminkan keseluruhan konten atau materi yang seharusnya dikuasai secara proporsional. Validitas isi suatu instrumen tidak ditentukan oleh suatu besaran yang dihitung menggunakan statistik, tetapi kesahihan isi suatu instrumen diperoleh berdasarkan telaah kisi-kisi instrumen. Oleh karena itu, validitas isi dilakukan dengan analisis logika, bukan dengan suatu koefisien validitas yang dihitung secara statistik. Berdasarkan uraian tersebut dapat dipahami bahwa estimasi validitas rubrik dilakukan melalui pertimbangan (*judgment*) dari para pakar yang relevan dengan isi objek yang akan diukur. Untuk memastikan validitas isi, instrumen yang dikembangkan perlu dinilai oleh pakar (Downing, 2006; Quellmalz *et al.*, 2012; Haladyana & Rodriguez, 2013). Berdasarkan hasil pertimbangan yang dilakukan oleh pakar, berapa besar derajat kesepakatan yang dicapai para pakar dapat dihitung dengan indeks *Content Validity Ratio (CVR)*. CVR merupakan salah satu metode yang digunakan untuk mengukur validitas isi berdasarkan hasil pertimbangan para pakar. Metode ini pada dasarnya untuk mengukur kesepakatan di antara penilai/*subject matter experts* tentang pentingnya item tertentu. Setiap penilai menjawab pertanyaan untuk setiap item dengan tiga pilihan jawaban, yaitu 1) esensial; 2) berguna tapi tidak esensial; dan 3) tidak

diperlukan. Jika lebih dari separoh jumlah panelis menyatakan bahwa item esensial, maka item tersebut setidaknya memiliki validitas isi.

Validitas isi dibedakan menjadi dua, yaitu validitas muka (*face validity*) dan validitas logis (*logical validity*). Validitas muka hanya menilai kesesuaian tampilan format (*appearance*) instrumen, sehingga jenis validitas ini dianggap tipe validitas yang paling rendah signifikansinya. Validitas muka dikatakan terpenuhi apabila instrumen mengukur karakteristik (*trait*) yang relevan, artinya item-item penyusun instrumen tidak menyimpang dari tujuan dilakukannya pengukuran. Validitas logis atau validitas sampling menunjukkan sejauhmana item-item instrumen mampu mewakili ciri-ciri atribut yang hendak diukur sebagaimana telah ditetapkan dalam lingkup (*domain*) ukurnya (Azwar, 1999).

Validitas konstruksi menunjukkan sejauhmana instrumen mengungkap karakteristik atau konstruk teoritis subjek yang hendak diukur. Analisis teoritik terhadap atribut yang hendak digali informasinya terlebih dahulu dilakukan untuk untuk memperoleh validitas konstruk yang baik. Validasi konstruk sebuah instrumen dilakukan melalui proses penelaahan atau justifikasi pakar atau melalui penilaian sekelompok panel yang merupakan orang-orang yang menguasai substansi atau konten dari variabel yang akan diukur.

Validitas kriteria disebut juga validitas empiris, artinya validitas ini ditentukan berdasarkan kriteria, baik kriteria internal maupun kriteria eksternal. Validitas kriteria diperoleh melalui hasil ujicoba instrumen kepada responden yang setara dengan responden yang akan diteliti. Pada kriteria internal, instrumen itu sendiri yang dijadikan sebagai kriteria, sedangkan pada kriteria eksternal, skor dari instrumen lain yang sudah teruji validitasnya yang menjadi kriteria (Djaali dan Muljono, 2008). Validitas yang diperoleh berdasarkan kriteria internal disebut validitas internal, sedangkan validitas yang diperoleh berdasarkan kriteria eksternal disebut kriteria eksternal. Validitas kriteria diperoleh dengan cara mengkorelasikan skor hasil tes dengan suatu kriteria tertentu. Kriteria dalam pengertian ini adalah variabel perilaku yang diprediksi oleh skor yang diperoleh melalui instrumen yang relevan. Namun demikian tidak semua jenis instrumen dapat menghasilkan sebuah indeks koefisien korelasi sebagai sebuah ukuran validitas kriteria. Validitas yang menjadi perhatian para pengembang instrumen pengukuran meliputi ketiga jenis

validitas, yaitu: validitas konten, validitas konstruk dan validitas kriteria. Ketiga validitas tersebut kemudian disatukan sebagai aspek fundamental dari teori yang lebih komprehensif mengenai validitas konstruk. Konsepsi validitas unifikasi ini memandang validitas konten dan validitas kriteria sebagai kasus khusus dari validitas konstruk (Liu, 2010). Sejalan dengan pendapat tersebut, Sugiyono (2011) menyatakan bahwa pengembangan instrumen non tes validitasnya cukup dilakukan dengan menguji validitas konstraknya saja. Rubrik sebagai salah satu instrumen non tes validitasnya ditentukan berdasarkan konstraknya. Konstruk rubrik ditentukan oleh deskriptor yang terkandung dalamnya. Konstruk rubrik divalidasi berdasarkan kesesuaian deskriptor dengan tugas kinerja dan tingkatan/gradasi kriteria deskriptor (Moskal & Mines, 2003).

Validitas internal disebut juga validitas butir, yaitu validitas yang ditentukan menggunakan besaran instrumen sebagai satu kesatuan (keseluruhan item penyusun instrumen) sebagai kriteria untuk menghitung validitas item-item penyusun instrumen. Artinya validitas tiap item dihitung menggunakan skor hasil ukur instrumen tersebut sebagai satu kesatuan sebagai kriteria, sehingga disebut juga sebagai validitas butir. Validitas internal ditentukan oleh sejauhmana skor setiap butir konsisten dengan hasil ukur secara keseluruhan. Hal tersebut mengandung arti bahwa validitas ini ditunjukkan oleh indeks koefisien korelasi setiap item (butir) instrumen dengan skor total instrumen. Pada saat indeks koefisien korelasi antara suatu butir dengan skor total menunjukkan angka yang positif dan signifikan maka butir tersebut dinyatakan valid berdasarkan kriteria validitas internal. Semakin besar koefisien korelasi antara skor suatu butir dengan skor total instrumen, maka dapat dikatakan bahwa butir tersebut memiliki konsistensi yang tinggi dengan skor total. Hal itu menunjukkan bahwa butir tersebut konvergen dengan butir-butir lain dalam mengukur konsep atau konstruk yang hendak diukur.

Hasil ukur sebuah instrumen yang sudah baku atau dianggap baku, dapat pula berupa hasil ukur instrumen lain yang sudah tersedia dan dapat dipercaya dijadikan kriteria pada validitas eksternal. Validitas eksternal ditunjukkan oleh suatu indeks koefisien korelasi yang merupakan hasil perhitungan statistik. Validitas eksternal ada dua macam, yaitu validitas bandingan (*concurrent validity*) dan validitas ramalan (*predictive validity*) (Azwar, 1999). Validitas bandingan

merupakan tingkat kejituan suatu instrumen. Hal ini dilihat dari korelasinya terhadap kecakapan yang telah dimiliki saat ini secara riil. Validitas bandingan diperoleh dengan cara mengkorelasikan hasil yang dicapai dalam tes yang dikembangkan dengan hasil tes yang dicapai dalam tes sejenis yang diketahui mempunyai validitas tinggi (misalnya tes standar). Validitas prediksi menunjukkan kemampuan instrumen meramalkan kemampuan yang akan dicapai seseorang di kemudian hari. Validitas ramalan diperoleh dengan mengkorelasikan hasil tes yang dicapai saat ini dengan hasil tes yang dicapainya kemudian. Jika hasil ukur instrumen yang sudah baku dijadikan sebagai kriteria eksternal, maka indeks koefisien korelasi diperoleh dengan mengkorelasikan hasil ukur instrumen yang sudah baku dengan hasil ukur instrumen yang sedang dikembangkan. Makin besar indeks koefisien korelasi yang diperoleh, maka validitas instrumen yang dikembangkan semakin baik (Sudijono, 2003). Untuk menguji signifikansi indeks koefisien korelasinya digunakan nilai  $r_{\text{tabel}}$  pada taraf signifikansi yang ditentukan.

Validasi instrumen asesmen *PCK* pada penelitian ini dilakukan melalui validasi pakar. Validasi pakar dilakukan untuk melihat kesesuaian tugas kinerja (pertanyaan dalam form matrikas rencana pembelajaran) dengan rubrik yang dikembangkan. Validasi dilakukan untuk menilai apakah deskriptor dalam rubrik sesuai dengan pertanyaan pedagogi dalam tugas kinerja, menilai apakah deskriptor rubrik memiliki gradasi/tingkatan yang jelas dan menilai apakah bahasa deskriptor dalam rubrik jelas dan mudah dipahami. Pakar yang melakukan validasi terhadap instrumen yang dikembangkan terdiri dari pakar asesmen, pakar *PCK* dan pakar pembelajaran fisika yang berjumlah lima pakar.

Instrumen direvisi sesuai dengan saran dan masukan dari para pakar. Selanjutnya dilakukan validasi kembali sampai pakar menyatakan bahwa instrumen yang dikembangkan valid. Validasi instrumen dilanjutkan secara kuantitatif dengan cara menghitung Indeks *Content Validity Ratio (CVR)*, menggunakan persamaan yang dikemukakan oleh Lawshe (1975).

$$CVR = \frac{n_e - (N/2)}{N/2} \quad \dots 3.1$$

Keterangan:

*CVR* = indeks *Content Validity Ratio*

$n_e$  = jumlah validator yang menyatakan setuju

$N$  = Jumlah total validator.

Berdasarkan kriteria yang dikemukakan Lawshe (1975), item dinyatakan valid untuk jumlah pakar (panelis) lima orang, apabila indeks *CVR* minimum 0,99. Kriteria tersebut dilakukan pada uji satu pihak, dengan taraf signifikansi 0,05. Mencermati kriteria tersebut, maka *instrumen* akan valid apabila semua pakar setuju dengan item instrumen. Kriteria tersebut kemudian direvisi berdasarkan kajian statistik, hasil revisi merekomendasi nilai kritis *CVR* untuk jumlah panelis lima orang dan taraf signifikansi 0,05 pada uji satu pihak sebesar 0,736 (Wilson, Pan & Schumsky, 2012). Apabila berdasarkan hasil validasi pakar dan perhitungan indeks *CVR* telah memenuhi syarat sebagai *instrumen* yang valid, selanjutnya dilakukan ujicoba awal penggunaan instrumen, ujicoba awal dilakukan bertujuan untuk mendapatkan masukan dari responden pengguna instrumen, guna perbaikan selanjutnya.

### 3.6.2 Ujicoba Awal dan Kepraktisan Instrumen

Tugas kinerja berupa penyusunan matriks *PCK* (matriks rencana pembelajaran) dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik, diperkenalkan kepada calon guru fisika saat mereka mengikuti mata kuliah Program Pengalaman Lapangan 1 (PPL 1). Setelah calon guru memahami dan dapat menyusun matriks *PCK*, mereka ditugaskan memilih satu konsep atau sub pokok bahasan pada materi kinematika, kemudian menyusun matriks *PCK* dalam konteks pengembangan keterampilan komunikasi dan penyelesaian masalah saintifik sesuai dengan konsep/sub pokok bahasan yang dipilih. Ujicoba awal dilakukan pada responden sebanyak enam calon guru fisika. Matriks *PCK* yang telah disusun oleh calon guru selanjutnya dinilai oleh lima dosen program studi Pendidikan Fisika, menggunakan rubrik *PCK* yang dikembangkan. Kelima dosen yang terlibat berasal dari perguruan tinggi tempat calon guru menempuh Pendidikan.

Setelah calon guru selesai melengkapi matriks *PCK*, calon guru diminta untuk merespon kuesioner yang berisi beberapa pertanyaan. Pertanyaan dalam

Erwin, 2019

**PENGEMBANGAN INSTRUMEN ASESMEN PEDAGOGICAL CONTENT KNOWLEDGE CALON GURU FISIKA DALAM KONTEKS PEMBELAJARAN KETERAMPILAN KOMUNIKASI DAN PENYELESAIAN MASALAH SAINTIFIK**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

kuesioner mengungkap kejelasan matriks PCK yang telah mereka lengkapi. Apakah pertanyaan dalam matriks dapat dipahami maksudnya. Untuk mendapatkan masukan terhadap rubrik yang dikembangkan, dilakukan wawancara dengan dosen pengguna rubrik, setelah mereka selesai menilai matriks PCK yang dibuat oleh calon guru. Wawancara juga dilakukan untuk mengungkap bagaimana kepraktisan instrumen yang meliputi kebermanfaatan rubrik dalam menilai matriks PCK, apakah ada keraguan memberi skor yang berpedoman pada rubrik, kesulitan yang dijumpai saat menggunakan rubrik dan alokasi waktu yang memadai menggunakan rubrik dalam menilai matriks secara objektif.

Kepraktisan (kepraktisan) termasuk salah satu ukuran yang menentukan apakah suatu instrumen baik atau tidak. Kepraktisan mengindikasikan suatu produk dapat digunakan sesuai dengan kondisi (Nieven & Folmer, 2013). Instrumen yang praktis merupakan instrumen yang dapat digunakan, berbiaya murah dalam kondisi normal dan penggunaannya bebas dari intervensi (Akker *et al.*, 2013). Seorang guru yang mengukur tanggapan peserta didik yang berjumlah 200 orang terhadap suatu produk pembelajaran dengan menggunakan tes essay akan merasakan cara ini kurang praktis. Diperlukan cara lain untuk mengukur tanggapan peserta didik tersebut, misalnya menggunakan angket tertutup. Kepraktisan dapat juga diartikan sebagai kemudahan dalam membuat, menggunakan dan pemeriksaan atau pengambilan keputusan yang objektif, sehingga keputusan yang diambil berdasarkan hasil pengukuran yang dilakukan tidak bias dan tidak meragukan.

Instrumen yang praktis merupakan instrumen khusus yang disesuaikan dengan ketersediaan waktu dan sumberdaya (Bachman & Palmer, 2009), faktor ekonomis dan kemudahan dalam mengadministrasikan dan proses memberi skor merupakan factor yang harus dipertimbangkan dalam mengembangkan instrumen, artinya bahwa pertimbangan biaya memperbanyak instrumen, berapa banyak sumberdaya administrator dan biaya yang dibutuhkan selama proses penilaian berlangsung, juga merupakan indikator kepraktisan instrumen (Utomo, Samsudi & Djuniadi, 2015). Sebuah tugas kinerja tergolong praktis apabila tugas tersebut sesuai dengan ketersediaan biaya, dapat diselesaikan peserta didik sesuai dengan waktu yang tersedia, kejelasan mengadministrasikan tugas, tugas juga harus sesuai dengan ketersediaan sumberdaya manusia yang ada, tugas tidak menuntut

sumberdaya material melebihi sumberdaya yang tersedia, dan mempertimbangkan waktu dan tenaga yang dibutuhkan dalam merancang dan menskor (Brown & Abeywickrama, 2010). Instrumen yang memerlukan biaya besar, sulit dalam menerapkannya, peserta didik membutuhkan waktu yang lebih banyak dibandingkan dengan waktu yang tersedia bukanlah instrumen yang praktis.

Dalam penelitian, kepraktisan instrumen sering diuji menggunakan kuesioner, wawancara dan observasi (Lutfiyah & Supardi, 2019; Kadir, Zaim & Refnaldi, 2018; Trianita, Zaim & Refnaldi, 2018; O'Malley *et al.*, 2013), karena kepraktisan dinilai berdasarkan penerapan instrumen yang dikembangkan di lapangan. Uji kepraktisan memberikan informasi bagaimana penggunaan instrumen tersebut pada kondisi penelitian yang dilakukan.

### **3.6.3 Ujicoba Utama dan Reliabilitas Instrumen (*Interrater Reliability*)**

Ujicoba utama dilakukan setelah dilakukan perbaikan instrumen berdasarkan masukan para dosen pengguna rubrik. Ujicoba utama bertujuan untuk menilai tingkat reliabilitas (konsistensi) rubrik. Ujicoba utama melibatkan 21 calon guru fisika yang sedang mengikuti Program Pengalaman Lapangan 2 (PPL 2), di berbagai sekolah menengah atas dan sekolah menengah kejuruan di Kabupaten Ogan Komering Ulu Timur.

Sebuah alat ukur yang baik harus mengukur dengan tepat apa yang akan diukur (valid) dan konsisten (andal, reliabel). Reliabilitas alat ukur ada dua aspek yaitu: konsistensi internal dan stabilitas (Murti, 2011). Jika sebuah instrumen terdiri dari beberapa item pertanyaan, maka skor dari setiap item pertanyaan seharusnya berkorelasi dengan skor total yang diperoleh, inilah yang dimaksud dengan konsistensi internal. Jika sebuah pengukur massa mengukur berulang kali  $10 \pm 0$  kg beras yang terdapat dalam kemasan, sedangkan sebuah alat pengukur massa lainnya mengukur beras dalam kemasan yang sama dengan hasil ukur  $10 \pm 2$  kg, maka alat pengukur massa pertama lebih stabil dibandingkan dengan yang kedua.

Pada umumnya indeks kesepakatan antar *rater* (*interrater reliability*) cocok untuk evaluasi variabel-variabel tertentu, perlu terlebih dahulu mengetahui jenis data yang akan diolah untuk menentukan teknik mana yang akan digunakan menganalisis indeks kesepakatan antar penilai. Ada empat jenis data yaitu: 1) data

nominal/kategori, yaitu data yang bentuknya kategori, data nominal diperoleh bukan melalui pengukuran tetapi diperoleh dengan cara membilang, angka dalam data nominal hanya berfungsi sebagai label atau identitas yang membedakan satu kategori dengan kategori lainnya, contohnya laki-laki dan perempuan; 2) data ordinal/rangking, yaitu data penjenjangan atau peringkat, angka ordinal menunjukkan peringkat yang berurutan, pada data ordinal tidak berlaku operasi hitung penjumlahan/pengurangan dan perkalian/pembagian. Misalnya rangking satu, dua, tiga dan seterusnya; 3) data interval, yaitu data hasil pengukuran yang memiliki jarak antar jenjang yang tetap, namun data interval tidak memiliki angka nol mutlak, sehingga data yang besarnya sepuluh tidak dapat dikatakan dua kali lima. Misalnya suhu lima puluh derajat Celcius tidak dapat dikatakan dua kali lebih panas dari dua puluh lima derajat Celcius; dan 4) data rasio, yaitu data hasil pengukuran yang sama dengan data interval, namun data rasio memiliki angka nol mutlak, pada data ini berlaku operasi hitung penjumlahan/pengurangan dan perkalian/pembagian, sehingga data yang besarnya seratus bisa dikatakan dua kali dari data lima puluh. Misalnya berat beban, beban yang massanya delapan kilogram dapat dikatakan memiliki massa empat kali lipat dari beban yang massanya dua kilogram (Wibowo, 2012),

Indeks konsistensi antar penilai untuk data berbentuk nominal dapat menggunakan indeks Cohen's kappa, *Intraclass Correlation Coefficient (ICC)* dan *Weighted kappa* jika penilainya (*rater*) hanya dua orang, jika penilai lebih dari dua orang dapat menggunakan *Fleiss' kappa* dan *ICC*. Data berbentuk ordinal dengan jumlah penilai dua orang dapat menggunakan *Weighted kappa* dan *ICC*, jika penilainya lebih dari dua orang, indeks konsistensi antar penilainya dapat menggunakan *Kendall Coefficient Concordance W* dan *ICC*. Sedangkan data berbentuk interval maupun rasio dapat menggunakan *Bland-Altman plots* dan *ICC* apabila jumlah penilai dua orang, jika jumlah penilai lebih dari dua orang dapat menggunakan *ICC* (Gisev, Bell & Chen, 2013). Jenis indeks konsistensi antar penilai yang dapat digunakan berdasarkan jenis data dan jumlah penilai ditampilkan pada Tabel 3.2.



Tabel 3.2  
*Contoh Indeks Konsistensi Antar Penilai yang Cocok digunakan berdasarkan Jenis Data*

	Jenis Data					
	Nominal		Ordinal		Interval dan Rasio	
	2 penilai	> 2 penilai	2 penilai	> 2 penilai	2 penilai	> 2 penilai
Indeks konsistensi antar penilai	Cohen's kappa	Fleiss' kappa	Weighted kappa	Kendall's <i>coefficient of concordance</i>	Bland-Altman plots	ICC
	ICC	ICC	ICC	ICC	ICC	ICC
	Weighted kappa					

Diadaptasi dari Gisev, Bell dan Chen, (2013)

Jenis indeks konsistensi antar penilai yang ditampilkan pada Tabel 3.2 bukanlah jenis indeks konsistensi yang lengkap, hanya menampilkan kemungkinan yang dapat digunakan. Sesuai dengan yang ditampilkan pada Tabel 3.2, statistik yang digunakan untuk menghitung indeks konsistensi antar penilai tergantung pada jenis datanya. Sebagai contoh, tiga orang pakar diminta untuk menilai produk-produk yang akan digunakan dalam sebuah kegiatan, penilaian dilakukan terhadap kecocokan produk dengan kegiatan. Setiap pakar secara independen menilai setiap produk menggunakan skala Likert-tipe 5 poin dengan skor 1 dinilai sebagai "tidak cocok" hingga 5 menjadi "sangat cocok". Tujuan dari penelitian ini adalah untuk menentukan tingkat persetujuan di antara tiga pakar dan seberapa mirip skor yang diberikan satu sama lain. Skala Likert merupakan jenis variabel ordinal, sesuai dengan Tabel 3.2, indeks konsistensi antar penilai yang sesuai digunakan adalah *Kendall's coefficient of concordance* atau *ICC*.

Pada penelitian ini, uji reliabilitas rubrik dilakukan menggunakan uji konsistensi indeks *Intraclass Correlation Coefficients (ICC)* dan indeks *Kendall's Coefficient Concordance W*. Kedua jenis uji ini digunakan untuk melihat kesesuaian skor antar individu yang diberikan penilai (*rater*). Disamping itu, data penelitian yang berbentuk ordinal dan penilaian dilakukan lebih dari 2 orang, sehingga digunakan indeks ICC dan W kendall. Data yang diuji konsistensinya berupa skor matriks *PCK* calon guru yang diberikan penilai berpedoman pada rubrik *PCK*.

### 3.6.3.1 *Intraclass Coefficient Correlation (ICC)*

*ICC* sudah digunakan secara luas untuk mengukur konsistensi antar penilai untuk data yang berbentuk peringkat atau ketika data dalam bentuk skala kontinu (McGraw & Wong, 1996; Shrout & Fleiss, 1979). *ICC* pertama kali diperkenalkan oleh Fisher (1950) untuk mempelajari kemiripan keluarga antara saudara kandung. Jika pengukuran dilakukan pada lebih dari dua kesempatan, atau lebih dari dua pengukur, menyangkut variabel berskala kontinu, maka reliabilitas dinilai dengan menggunakan *Intraclass Coefficient Correlation (ICC)* (Widiharso, 2005; Murti, 2011; Matthew, Milanowski & Miller, 2012). Pada *self-report*, reliabilitas ditentukan oleh tingkat konsistensi antar butir-butir penyusun instrumen, sedangkan pada reliabilitas *interrater* yang diuji konsistensinya adalah *raternya*. Dengan demikian disini posisi butir digantikan oleh *rater*. Dua orang *rater* dikatakan memiliki kesepakatan yang tinggi apabila kedua *rater* memberikan urutan skor yang sama atau hampir sama pada suatu subjek (Ebel & Frisbie, 1991).

Praktisi, peneliti, dan pembuat kebijakan sering menggunakan reliabilitas *interrater* sebagai istilah yang sering digunakan untuk konsistensi antar penilai. Beberapa pakar pengukuran mendefinisikan konsistensi antar penilai (*interrater reliability*) sebagai konsistensi pengukuran antara evaluator dalam urutan atau kedudukan relatif dari peringkat kinerja, terlepas dari nilai absolut dari peringkat masing-masing evaluator. Kesepakatan antar penilai (*interrater agreement*) adalah sejauh mana dua atau lebih evaluator yang menggunakan skala peringkat yang sama memberikan peringkat yang sama terhadap situasi identik yang diobservasi. Berbeda dengan konsistensi antar penilai, kesepakatan antar penilai adalah konsistensi pengukuran antara nilai absolut dari peringkat yang diberikan evaluator (Matthew, Milanowski & Miller, 2012).

Reliabilitas yang melibatkan *rater* ada dua macam yaitu kesepakatan antar *rater* (*interrater agreement*) dan reliabilitas antar *rater* (*interrater reliability*). Jika beberapa orang sedang diobservasi oleh dua orang *rater*, *rater* pertama menilai peserta didik A paling tinggi dibandingkan peserta didik yang lain, kemudian *rater* kedua juga memberikan skor paling tinggi pada peserta didik A, maka penilaian yang dilakukan kedua *rater* dapat dikatakan konsisten (Widiharso, 2012).

*ICC* dapat digunakan untuk menilai konsistensi, kesesuaian dan kesamaan skor yang diberikan oleh panelis untuk sampel yang sama. Oleh karena itu, korelasi *intraclass* merupakan ukuran homogenitas di antara para panelis (Bi & Kuesten, 2012). Indeks koefisien korelasi *intraclass* memiliki rentang teoritis dari nol hingga satu. Angka indeks nol menunjukkan tidak reliabel dan angka indeks satu menunjukkan reliabilitas sempurna (Denegar & Ball, 1993). *ICC* dapat mengukur kesamaan antara panelis dan sensitivitas panel dan panelis. Oleh karena itu, *ICC* dapat menangani masalah reliabilitas dan validitas. *ICC* dapat diperoleh dari berbagai jenis data untuk beragam eksperimen (Bi & Kuesten, 2012).

*ICC* merupakan estimasi reliabilitas yang lebih tepat daripada koefisien korelasi antar kelas seperti korelasi *product moment pearson* (Denegar & Ball, 1993). Oleh karena itu, banyak penelitian yang menggunakan *ICC* untuk mengestimasi reliabilitas. Beberapa penelitian yang menggunakan *ICC* untuk mengestimasi reliabilitas diantaranya penelitian yang dilakukan Schneider *et al.* (2013) tentang *intrarater* dan *interrater reliability* dari tes klinis terpilih pada pasien yang dirujuk untuk blok sendi diagnostik di tulang belakang leher, bertujuan mengukur reliabilitas *intra* dan *interrater* uji klinis terstandarisasi terpilih yang digunakan untuk penilaian pasien nyeri leher aksial yang dirujuk untuk blok sendi diagnostik. Uji reliabilitas menggunakan *intraclass correlation coefficients (ICC)*, koefisien kappa, dengan taraf kepercayaan 95% untuk rentang gerak serviks (ROM; 6 arah), ekstensi-rotasi (*ER*) tes, pemeriksaan tulang belakang manual (MSE), dan palpasi untuk nyeri paraspinal (PST) dari C2 hingga C7.

Penelitian yang dilakukan Bobak, Barr dan O'Malley (2018), juga menerapkan *ICC* untuk mengevaluasi reliabilitas *Observer OPTION*. Sebuah instrumen yang menggunakan penilai terlatih untuk mengevaluasi tingkat pengambilan keputusan bersama antara dokter dan pasien. Penelitian ini melibatkan dua penilai untuk mengevaluasi 311 rekaman pertemuan klinis pada tiga studi, untuk mengevaluasi dampak penggunaan alat bantu keputusan pribadi atas perawatan biasa. Disamping itu, Agarwal, Awasthi dan Walter, (2005) juga menggunakan *ICC* pada penelitiannya yang berjudul estimasi *intra-class correlation* untuk asesmen asupan vitamin a pada anak-anak. Menggunakan data penilaian kekurangan vitamin A klinis dan asupan makanan kaya vitamin A pada

anak-anak di sebuah distrik di India. Survei dilakukan terhadap 16 rumah tangga sebagai sampel dari 200 desa yang berada dalam delapan blok kabupaten yang dipilih secara acak. Empat tim terlatih - masing-masing terdiri dari petugas medis dan asisten peneliti non-medis - melakukan survei. Petugas medis memeriksa anak untuk mengetahui *bitot spot* dan mewawancarai orang tua untuk masalah kebutaan, imunisasi, dan dosis vitamin A yang dikonsumsi anak selama setahun terakhir. Asisten peneliti menggunakan kuesioner untuk mendapatkan data dari orang tua mengenai kesadaran mereka tentang kebiasaan tidur malam dan kebiasaan diet vitamin A, dengan fokus pada estimasi kuantitatif asupan makanan kaya vitamin A. *ICC* digunakan untuk mengukur ketergantungan antara dua anggota cluster yang sama, seperti blok (Faktor A) atau desa (Faktor B). hal Ini merupakan rasio komponen varians karena blok atau desa dengan total varians untuk masing-masing anak. Untuk masing-masing variabel, *ICC* dihitung pada tingkat blok dan desa.

Indeks *ICC* diperoleh dari rasio antara varians antar kelompok dengan varians total. Varians total pada suatu pengukuran berasal dari 3 sumber, yaitu: 1) varians yang dinilai (subjek); 2) varian penilai (*rater*); dan varians *random error*. Jika variasi penilai diasumsikan *random*, maka persamaan untuk menghitung indeks *ICC* (Murti, 2011) adalah:

$$ICC = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_o^2 + \sigma_e^2} \quad \dots\dots 3.2$$

Persamaan 3.1) digunakan untuk menghitung indeks kesepakatan antar penilai (*interrater agreement*). Kesepakatan antar penilai merupakan tingkat kecocokan antara penilai dalam memberikan skor pada objek yang dinilai. *interrater agreement* disebut juga reliabilitas absolut (*absolute reliability*). Apabila variasi penilai diasumsikan *fixed* pada persamaan 3.2), maka variasi penilai tidak diperhitungkan, sehingga persamaan *ICC* menjadi:

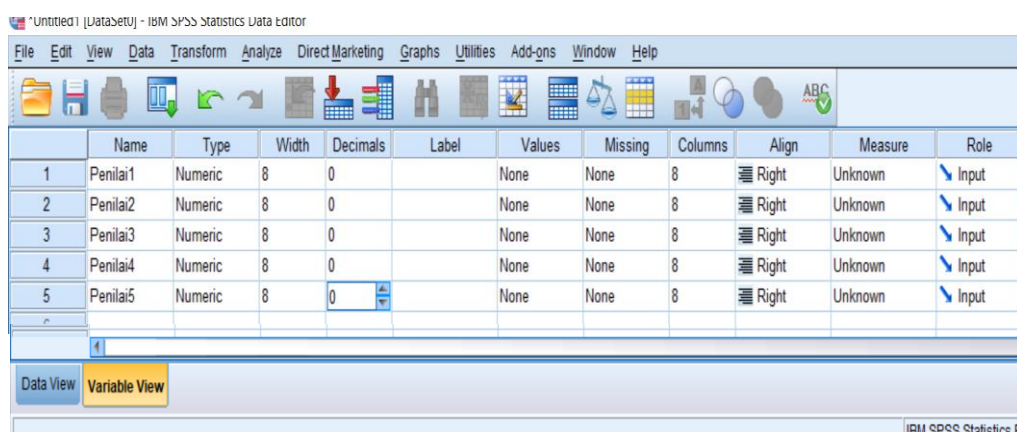
$$ICC = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2} \quad \dots\dots 3.3$$

dimana varians ( $\sigma^2$ ) adalah ukuran variasi, subskrip s adalah subjek, o adalah observer (penilai) dan e adalah *random error*. Persamaan 3.3 digunakan untuk menghitung indeks konsistensi penilaian antar penilai (*interrater reliability*). Konsistensi penilaian antar penilai menunjukkan sejauh mana penilai

mengkategorikan objek yang dinilai dengan cara yang sama. Indeks konsistensi ini juga sering disebut reliabilitas relatif (*relative reliability*).

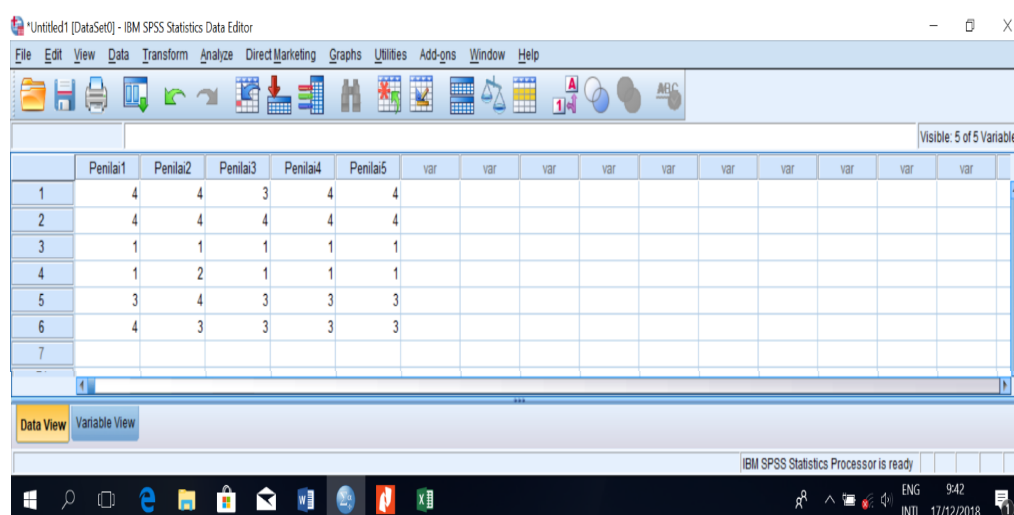
Pada penelitian ini menghitung besar indeks *ICC* dan indeks konkordansi *W Kendall*, keduanya menggunakan aplikasi *SPSS* versi 20. Estimasi *Intraclass Correlation Coefficients* dengan aplikasi *SPSS* versi 20, dilakukan dengan langkah sebagai berikut.

1. Memasukkan nama penilai1, penilai2, .....dan seterusnya pada *sheet variable view* seperti Gambar 3.8.



Gambar 3.8 Memasukkan Nama Penilai1, Penilai2, ... dan seterusnya pada *Sheet Variable View*

2. Memasukkan data pada *sheet data view* seperti Gambar 3.9



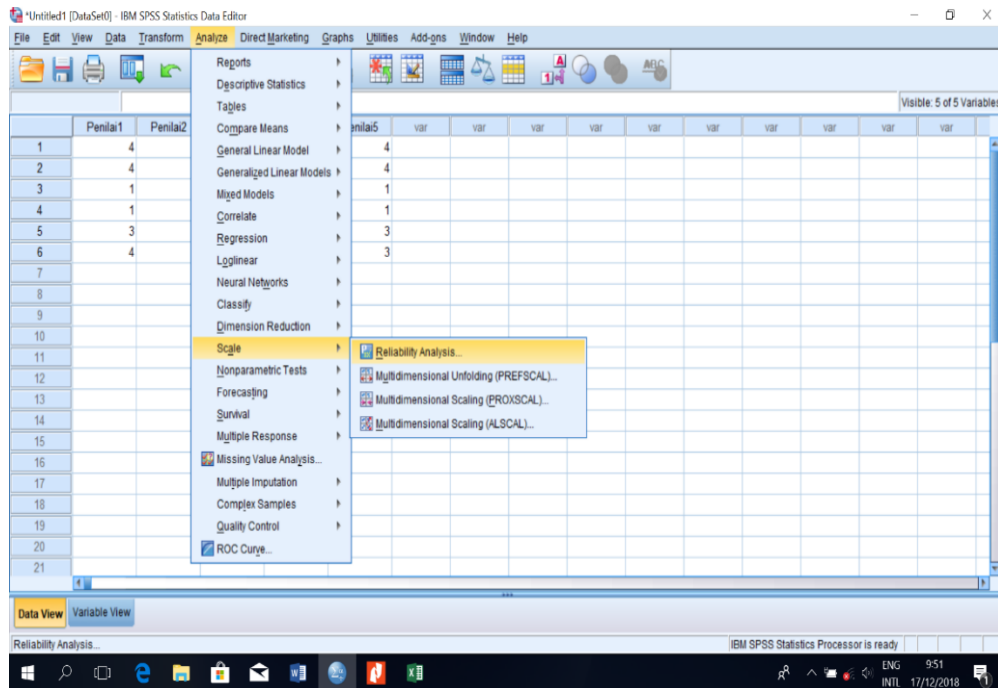
Gambar 3.9 Memasukkan Data pada *Sheet Data View*

3. Memilih menu *analyze-scale-reliability analysis*, seperti Gambar 3.10.

Erwin, 2019

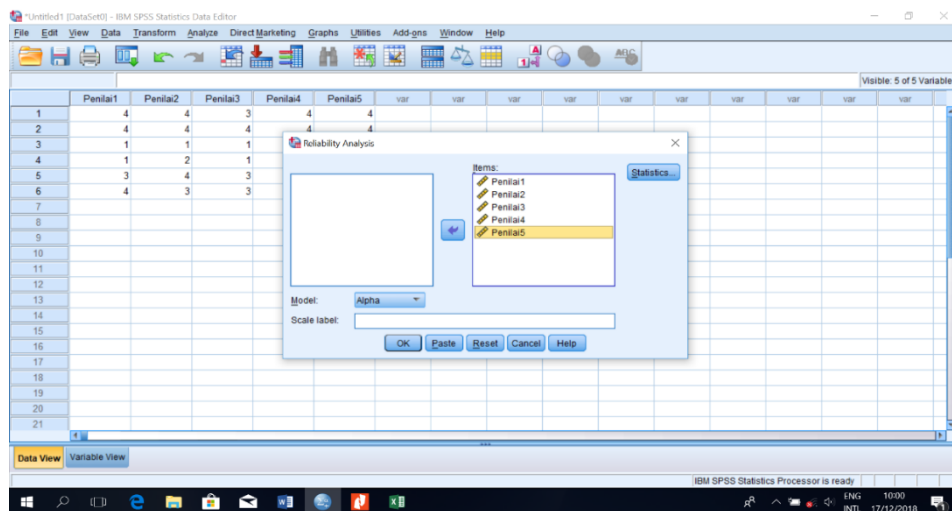
**PENGEMBANGAN INSTRUMEN ASESMEN PEDAGOGICAL CONTENT KNOWLEDGE CALON GURU FISIKA DALAM KONTEKS PEMBELAJARAN KETERAMPILAN KOMUNIKASI DAN PENYELESAIAN MASALAH SAINTIFIK**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu



Gambar 3.10 Memilih Menu *Analyze-Scale-Reliability Analysis*

- Memasukkan variabel penilai1, penilai2, .... dan seterusnya ke dalam kotak item, seperti Gambar 3.11



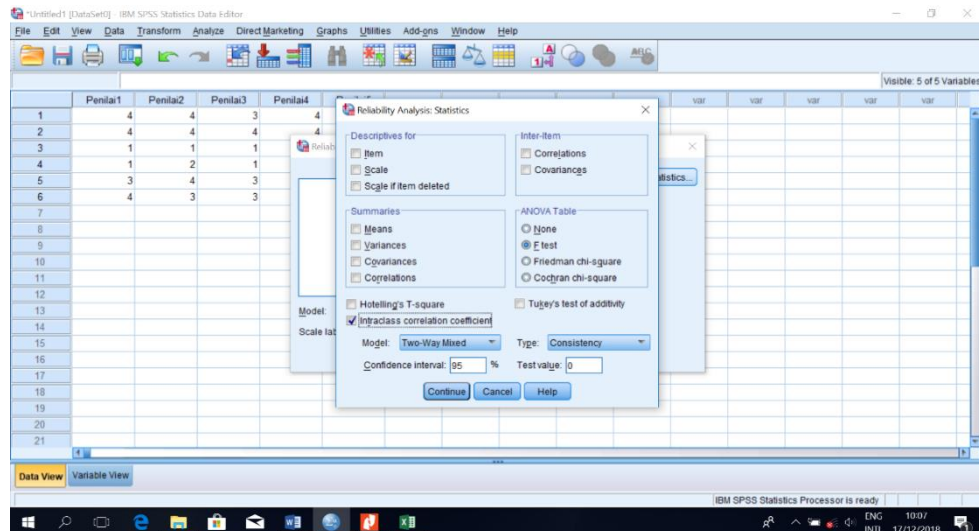
Gambar 3.11 Memasukkan Variabel Penilai1, Penilai2, .... dan seterusnya ke dalam Kotak *Item*

- Memilih *statistics*, dan mencentang *F-test* dan *intraclass correlation coefficients*. Memilih jenis analisis *two way mixed* dan *consistency* dengan *confidence interval 95%*, seperti Gambar 3.12

Erwin, 2019

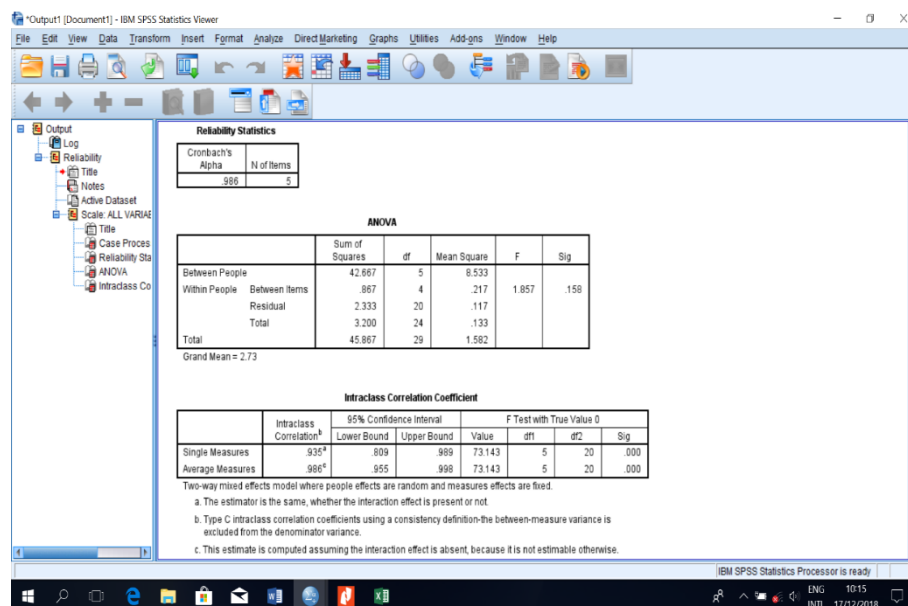
**PENGEMBANGAN INSTRUMEN ASESMEN PEDAGOGICAL CONTENT KNOWLEDGE CALON GURU FISIKA DALAM KONTEKS PEMBELAJARAN KETERAMPILAN KOMUNIKASI DAN PENYELESAIAN MASALAH SAINTIFIK**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu



Gambar 3.12 Memilih Model Statistik

6. Memilih *continue*, lalu mengklik ok. Setelah itu muncul *output* olahan data statistik seperti Gambar 3.13.



Gambar 3.13 Output Olahan Data Statistik ICC

7. Langkah selanjutnya adalah menginterpretasi *output* olahan data dari aplikasi SPSS.

Interpretasi indeks *ICC* dilakukan berdasarkan kriteria yang dikemukakan Koo dan Li (2016), apabila indeks *ICC* lebih kecil dari 0,5 menunjukkan reliabilitas yang jelek, indeks *ICC* berada antara 0,5 sampai dengan 0,75 menunjukkan reliabilitas yang sedang, indeks *ICC* berada pada nilai antara 0,75 sampai 0,9

Erwin, 2019

**PENGEMBANGAN INSTRUMEN ASESMEN PEDAGOGICAL CONTENT KNOWLEDGE CALON GURU FISIKA DALAM KONTEKS PEMBELAJARAN KETERAMPILAN KOMUNIKASI DAN PENYELESAIAN MASALAH SAINTIFIK**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

menunjukkan reliabilitas yang baik, dan indeks *ICC* dengan nilai lebih besar dari 0,9 menunjukkan reliabilitas yang istimewa.

### **3.6.3.2 Koefisien Konkordansi W Kendall (*Kendall's Coefficient of Concordance W*)**

Menilai kesepakatan antar penilai untuk data berbentuk peringkat (ordinal) dengan jumlah penilai lebih dari dua dapat menggunakan ukuran konkordansi W Kendall (Tran, Dolgun & Demirhan, 2018). Uji W Kendall merupakan uji non parametrik yang digunakan untuk menguji keselarasan terhadap penilaian yang diberikan oleh sekelompok penilai terhadap atribut-atribut yang dianggap penting. Nilai koefisien konkordansi W Kendall berada antara nol sampai dengan satu, dimana nol menunjukkan tidak adanya kesepakatan, sedangkan satu menunjukkan adanya kesepakatan sempurna antar penilai. Nilai W negatif tidak mungkin terjadi, karena ketidaksepakatan sempurna tidak dapat diperoleh dengan dua atau lebih *rater* (penilai) (Siegel & Castellen, 1988).

Koefisien konkordansi W Kendall telah banyak digunakan untuk penelitian dalam berbagai disiplin ilmu, seperti ilmu sosial dan ilmu yang berhubungan dengan perilaku manusia (Bajpai, Bajpai & Chaturvedi, 2015), ilmu administrasi farmasi (Gisev, Bell & Chen, 2013; Bentley *et al.*, 2017), Kesehatan ( Stausberg, Halim & Farber, 2011; Tognetto *et al.*, 2019) pertanian (Minero *et al.*, 2016), Biologi (Legendre, 2005; Gearhart *et al.*, 2013) dan ekonomi (Grothe dan Schmid, 2011), yang melibatkan sejumlah pakar sebagai penilai untuk meranking suatu item. Penelitian yang dilakukan George (2004) dalam bidang farmasi menggunakan koefisien konkordansi W Kendall, sebuah studi yang melibatkan lima orang pakar untuk meranking kerumitan cara pengobatan terhadap enam jenis pengobatan, yang dibandingkan dengan perkembangan terbaru indeks kerumitan cara pengobatan.

Penelitian lain dilakukan dengan melibatkan lima orang pakar multidisiplin, menilai hasil pemeriksaan kesehatan secara komprehensif terhadap anggota komunitas kesehatan mental, menggunakan skala Likert lima poin. Setiap panelis menilai temuan pemeriksaan secara independent, rekomendasi temuan, kemungkinan implementasi rekomendasi, harapan dampak klinis secara



keseluruhan. Kesepakatan antar panelis dilakukan dengan indeks W untuk setiap skala (Gisev *et al.*, 2010). Indeks W dapat juga digunakan untuk menentukan apakah kesepakatan telah tercapai dan juga kekuatan relatif dari kesepakatan.

Penelitian Bentley *et al.* (2017) yang berjudul kerangka kerja asesmen nilai validitas dan reliabilitas untuk obat kanker baru, bertujuan memahami sejauh mana kerangka kerja yang dikembangkan beberapa organisasi untuk mengases secara sistematis nilai obat baru, dapat memfasilitasi keputusan perawatan berbasis nilai dalam onkologi. Organisasi yang mengembangkan kerangka kerja untuk mengases secara sistematis nilai obat baru yaitu, *the American Society of Clinical Oncology (ASCO)*, *the European Society for Medical Oncology (ESMO)*, *the Institute for Clinical and Economic Review (ICER)*, dan *the National Comprehensive Cancer Network (NCCN)*. Delapan panelis melakukan penilaian nilai dari lima obat kanker paru-paru lanjut menggunakan kerangka ASCO, ESMO, dan ICER. Panelis menerima instruksi dan menerbitkan data klinis yang diperlukan untuk menyelesaikan penilaian. Skor kerangka kerja NCCN yang dipublikasikan dipisahkan. Koefisien W Kendall digunakan untuk mengukur validitas konvergen keempat kerangka kerja, kemudian koefisien korelasi intraclass digunakan untuk mengukur reliabilitas antar penilai antara kerangka ASCO, ESMO, dan ICER. Hasil pengujian terhadap kerangka kerja ASCO, ESMO, ICER, dan NCCN menunjukkan validitas konvergen, meskipun ada perbedaan dalam pendekatan konseptual yang digunakan. *Inter-rater reliability* kerangka kerja ASCO berada pada kategori tinggi, meskipun berpotensi dengan biaya beban pengguna. *inter-rater reliability* kerangka kerja ICER masuk dalam kategori buruk, mungkin karena kegagalannya untuk membedakan nilai diferensial di antara sampel obat yang diuji.

Penelitian yang dilakukan Stausberg, Halim dan Farber (2011) salah satu tujuannya untuk menjawab pertanyaan bagaimana konkordansi perbedaan indikator pada tingkat rumah sakit. Data rutin disediakan untuk tiga perusahaan yang menawarkan indikator keselamatan pasien, indikator yang ditetapkan dari HELIOS Hospital Group, dan pengukuran berdasarkan kemunculan Penyakit. Sepuluh rumah sakit di North Rhine-Westphalia, dengan total pasien rawat inap 151.960 pada tahun 2006, mengajukan diri untuk berpartisipasi dalam penelitian ini. Perusahaan memberikan laporan kualitas standar untuk sepuluh rumah sakit. Selain

indikator yang berbeda, nilai referensi yang berbeda dimungkinkan membangun beberapa kelompok pembanding. Konkordansi dan kekokohan dianalisis menggunakan koefisien korelasi non-parametrik dan W Kendall. Hasil penelitian menunjukkan kurangnya kesesuaian dalam mengestimasi kinerja rumah sakit ketika indikator kualitas yang berbeda digunakan.

Penelitian lain yang menggunakan W Kendall untuk menguji kesepatan antar penilai dilakukan oleh Tognetto *et al.* (2019), penelitian dengan tujuan untuk menilai gambaran makula sentral yang ditangkap dengan biometer optik berdasarkan pemindaian sumber mata-lengkap oCt (ss-oCt) yang dilakukan sepanjang mata sebagai strategi skrining untuk mengidentifikasi penyakit makula pada pasien yang dijadwalkan untuk operasi Katarak. 1.114 mata dari 749 pasien berturut-turut menjalani pemeriksaan biometrik dengan teknologi IOLMaster 700 SS-OCT (*Carl Zeiss*) dan *spectral-domain* OCT (SD-OCT) (*Spectralis* OCT, *Heidelberg*) analisis perangkat dilakukan pada hari yang sama. Tujuh penguji menilai pemindaian secara individual dalam mode *full-masked*. Analisis dilakukan menggunakan statistik non-parametrik dengan koefisien konkordansi W Kendall untuk menghitung tingkat kesepatan antara penguji yang berbeda.

Penelitian Minero *et al.* (2016) bertujuan untuk mengembangkan skala penilaian tetap pada penilaian kualitatif perilaku (QBA) keledai, untuk mengevaluasi reliabilitas antar-pengamat ketika diterapkan di lahan, dan untuk menilai apakah hasil QBA berkorelasi dengan ukuran kesejahteraan lainnya. Sebanyak 16 daftar deskriptor tetap dirancang berdasarkan konsultasi dalam kelompok fokus. Daftar tetap kemudian digunakan oleh empat pengamat terlatih untuk mencetak sembilan video 2 menit kelompok keledai yang dimiliki oleh enam peternakan dan peternakan mendapatkan 11 fasilitas keledai yang mewakili jenis fasilitas keledai paling umum di Eropa Barat. Di masing-masing peternakan, seorang penilai berpengalaman mengumpulkan pengukuran kesejahteraan yang berbeda untuk semua keledai dewasa. Skor QBA dan ukuran kesejahteraan dianalisis menggunakan principal component analysis (PCA, matriks korelasi, tanpa rotasi). W Kendall dan ANOVA digunakan untuk menilai reliabilitas antar-pengamat.

Grothe dan Schmid (2011) menggunakan W Kendall dalam bidang financial untuk mempelajari hubungan volatilitas pengembalian aset. Asosiasi pengembalian aset merupakan indikator yang berguna untuk diversifikasi portofolio aset. Volatilitas pengembalian aset merupakan aspek utama dalam penilaian derivatif yang sesuai, dan kemudian asosiasi volatilitas merupakan indikator yang berguna untuk pasar derivatif. Legendre (2005) juga menggunakan W Kendall dalam ekologi untuk mencari asosiasi spesies, yaitu kelompok spesies yang ditemukan bersama. Asosiasi spesies merupakan masalah penting dari ekologi komunitas karena asosiasi spesies dapat digunakan untuk memprediksi karakteristik lingkungan.

Penggunaan koefisien konkordansi W Kendall untuk mengases kesepakatan antara para pengamat terhadap gambar dengan resolusi sangat tinggi dalam penelitian Gearhart *et al.* (2013), dilatarbelakangi penggunaan metode monitoring vegetasi darat membutuhkan biaya mahal, memakan waktu dan terbatas dalam ukuran sampel. Gambar udara menarik bagi para manajer karena berkurangnya waktu dan biaya dan peningkatan ukuran sampel. Salah satu tantangan gambar udara adalah mendeteksi perbedaan di antara pengamat terhadap gambar yang sama. Enam pengamat menganalisis serangkaian sampel gambar udara vegetasi darat berjarak 1 mm untuk komposisi spesies graminoid dan karakteristik tutupan lahan yang penting. Koefisien konkordansi W Kendall digunakan untuk mengukur kesepakatan di antara para pengamat. Kelompok yang terdiri dari enam pengamat itu sepakat ketika dinilai sebagai kelompok. Ketika masing-masing pengamat dinilai secara independen terhadap lima lainnya, ditemukan kurangnya kesepakatan untuk spesies graminoid yang sulit diidentifikasi dalam gambar udara. Menghitung indeks W Kendall digunakan persamaan 3.4.

$$W = \frac{12S}{p^2(n^3-n)-pT} \quad \dots 3.4$$

Keterangan:

S = jumlah kuadrat dari jumlah baris peringkat  $R_i$

n = jumlah objek

p = jumlah penilai (*rater*)

T = faktor koreksi untuk peringkat terikat

Jumlah kuadrat dari jumlah baris peringkat  $R_i$ , dihitung menggunakan persamaan 3.5. selanjutnya faktor koreksi untuk peringkat terikat dihitung menggunakan persamaan 3.6.

$$S = \sum_{i=1}^n R_i^2 = SSR \quad \dots 3.5$$

$$T = \sum_{k=1}^m (t_k^3 - t_k) \quad \dots 3.6$$

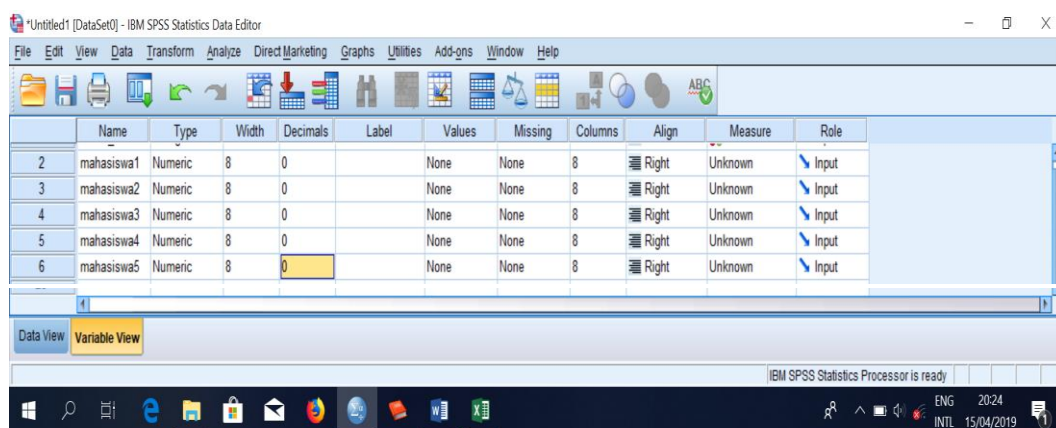
Keterangan:

$m$  = jumlah kelompok

$t_k$  = jumlah pangkat terikat dalam setiap ( $k$ ) kelompok  $m$

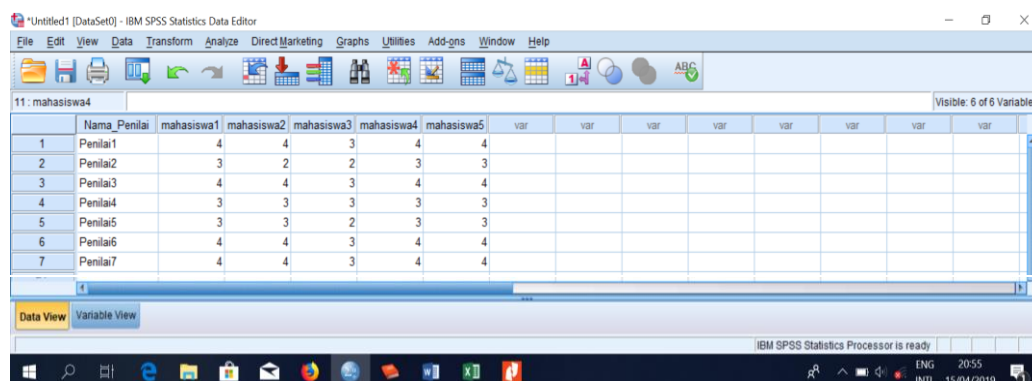
Pada penelitian ini, uji konsistensi penilaian antar penilai menggunakan indeks *Kendall's Coefficient Concordance W*, dengan memanfaatkan aplikasi SPSS versi 20, dilakukan dengan langkah-langkah berikut.

1. Memasukkan nama mahasiswa1, mahasiswa2, ... dan seterusnya pada *sheet variable view* seperti Gambar 3.14.



Gambar 3.14 Memasukkan Nama Mahasiswa1, Mahasiswa2, ... dan seterusnya pada *Sheet Variable View*

2. Memasukkan data pada *sheet data view* seperti Gambar 3.15



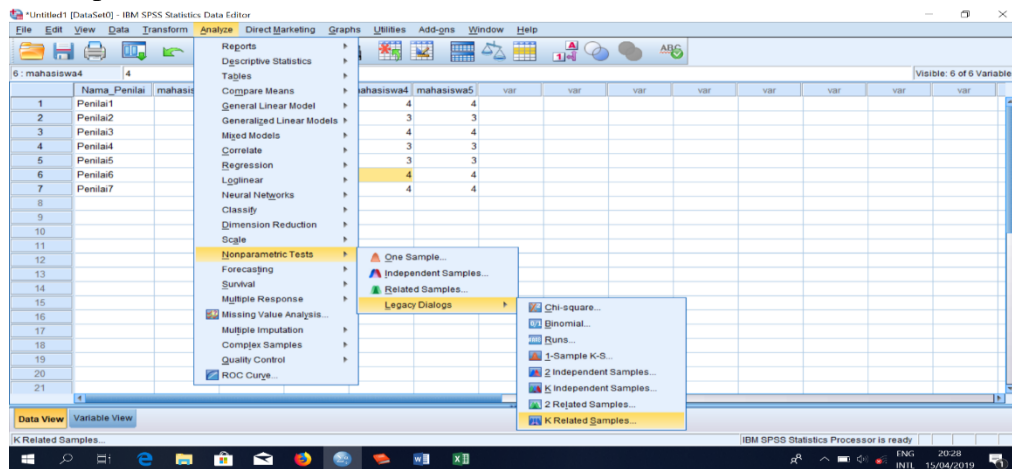
Gamba3.15 Memasukkan Data pada *Sheet Data View*

Erwin, 2019

**PENGEMBANGAN INSTRUMEN ASESMEN PEDAGOGICAL CONTENT KNOWLEDGE CALON GURU FISIKA DALAM KONTEKS PEMBELAJARAN KETERAMPILAN KOMUNIKASI DAN PENYELESAIAN MASALAH SAINTIFIK**

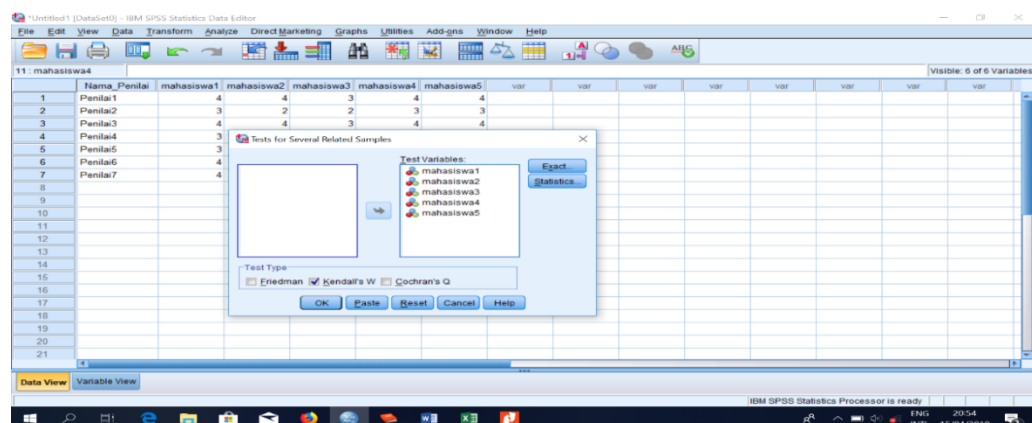
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

3. Memilih menu *analyze-nonparametric test-legacy dialogs-k related samples*, seperti Gambar 3.16.



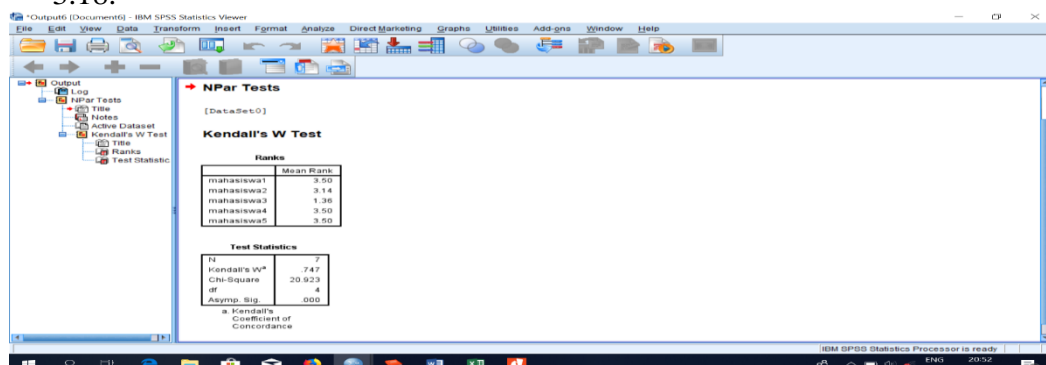
Gambar 3.16 Memilih Menu *Analyze-Nonparametric Test*

4. Memasukkan variabel mahasiswa1, mahasiswa 2, .... dan seterusnya ke dalam kotak *test variables* dan centang *Kendall's W*, seperti Gambar 3.17.



Gambar 3.17 Memasukkan Variabel Mahasiswa1, Mahasiswa2,... dan seterusnya ke dalam Kotak *Test Variables* dan mencentang *Kendall's W*

5. Mengklik ok. Setelah itu muncul *output* olahan data statistik seperti Gambar 3.18.



Gambar 3.18 *Output* Olahan Data Statistik *Kendall's W*

Erwin, 2019

**PENGEMBANGAN INSTRUMEN ASESMEN PEDAGOGICAL CONTENT KNOWLEDGE CALON GURU FISIKA DALAM KONTEKS PEMBELAJARAN KETERAMPILAN KOMUNIKASI DAN PENYELESAIAN MASALAH SAINTIFIK**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

6. Langkah selanjutnya adalah menginterpretasi *output* olahan data dari aplikasi SPSS.

Interpretasi kategori korelasi Spearman *rank* telah diperluas untuk menginterpretasi indeks W (Levitan *et al.*, 2008). Indeks W juga diinterpretasikan dengan menganalogikannya dengan koefisien korelasi yang lain (Gisev *et al.*, 2010). Namun demikian interpretasi kategori indeks W secara khusus dikemukakan dalam tulisan yang berjudul *Managing Delphi Surveys Using Nonparametric Statistical Techniques* (Schmidt, 1997).

Nilai-nilai Kendall W dapat bervariasi dari 0 (tidak ada kesepakatan sama sekali) hingga 1 (kesepakatan sempurna), dengan nilai lebih tinggi dari 0,6 yang menunjukkan kesepakatan substansial (Minero *et al.*, 2016). Indeks W Kendall diinterpretasi dengan cara mencocokkan nilai indeks W hitung dengan tabel kriteria kategori indeks W. nilai indeks W lebih kecil dari 0,11, diinterpretasikan kesepakatan sangat lemah; nilai indeks W antara 0,11 sampai dengan 0,30, diinterpretasikan kesepakatan yang lemah; nilai indeks W antara 0,31 sampai dengan 0,50, kesepakatan yang sedang; nilai indeks W antara 0,51 sampai dengan 0,70, diinterpretasikan adanya kesepakatan yang kuat; dan nilai indeks W lebih besar dari 0,71, diinterpretasikan adanya kesepakatan yang sangat kuat (Schmidt, 1997; Orsbon, Kaiser & Ross, 2014).