

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Pusat pembinaan, pendidikan, dan pelatihan (Pusbindiklat) adalah sebuah lembaga yang berada di bawah naungan Lembaga Ilmu Pegetahuan Indonesia (LIPI). Pusbindiklat LIPI memiliki tiga fungsi utama yaitu sebagai pusat pembinaan jabatan fungsional peneliti, pusat pendidikan dan pelatihan bagi ASN yang duduk pada jabatan fungsional peneliti, serta pusat pendidikan dan pelatihan bagi civitas LIPI.

Sebagai sebuah lembaga yang bertugas menyelenggarakan pendidikan dan pelatihan (diklat) bagi ASN yang duduk pada jabatan fungsional peneliti, Pusbindiklat LIPI dituntut untuk senantiasa mengembangkan program pendidikan dan pelatihan yang berkualitas serta berdampak pada peningkatan kinerja peneliti. Program diklat bagi peneliti terbagi atas dua jenis. Diklat tersebut adalah diklat teknis dan diklat fungsional.

Diklat teknis merupakan diklat yang didesain untuk meningkatkan kompetensi teknis peneliti. Diklat ini ditujukan bagi para peneliti namun tidak bersifat wajib. Contoh diklat teknis adalah diklat penyusunan proposal penelitian, diklat metode penelitian, diklat pengumpulan dan pengolahan data, serta diklat penulisan karya tulis ilmiah.

Diklat fungsional merupakan diklat yang wajib diikuti oleh peneliti dan calon peneliti. Diklat ini dikenal dengan nama Pelatihan Pembekalan Jabatan Fungsional Peneliti (PPJFP). PPJFP terbagi atas dua tingkatan yaitu tingkat pertama dan tingkat lanjutan. PPJFP tingkat pertama merupakan syarat bagi calon peneliti yang akan duduk pada jabatan fungsional peneliti. Sedangkan PPJFP tingkat lanjutan merupakan syarat bagi peneliti yang akan naik ke jenjang peneliti madya.

PPJFP tingkat pertama dirancang secara khusus untuk mempersiapkan calon peneiti sebelum duduk pada jabatan fungsional peneliti. Pada program ini, peneliti dibekali dengan berbagai materi. Materi yang diberikan diantaranya

seputar etika dalam dunia penelitian, materi teknis penelitian, serta pembekalan jenjang karir peneliti. Selain mendapatkan berbagai materi seputar dunia penelitian, di akhir diklat ini para peserta juga dituntut untuk menghasilkan produk akhir berupa proposal penelitian. Selanjutnya proposal tersebut akan diujikan dengan cara diseminarkan di depan para widyaiswara yang juga berperan sebagai penguji.

Berdasarkan pengalaman peneliti, masih terdapat permasalahan pada proses penilaian proposal penelitian. Peneliti menilai bahwa instrumen penilaian yang digunakan oleh penguji masih lemah ditinjau dari sisi reliabilitasnya.

Penilaian proposal penelitian dilakukan dengan menggunakan beberapa indikator penilaian. Indikator penilaian yang digunakan saat ini meliputi pendahuluan, pendekatan/metode penelitian, analisis data dan sintesa, originalitas ide dan *novelty* substantif, serta penulisan daftar pustaka.

Permasalahan muncul pada saat penguji menilai proposal penelitian. Penilaian proposal penelitian dilakukan oleh enam orang penguji. Masing-masing penguji bertanggung jawab untuk menilai lima buah proposal penelitian.

Berdasarkan pengamatan peneliti, pada umumnya tiap penguji memiliki standar penilaian yang berbeda. Tidak jarang juga peneliti menemukan adanya selisih yang tinggi antar nilai rata-rata dari setiap penguji. Hal ini menjadi indikasi bahwa setiap penguji memiliki standar yang berbeda.

Perbedaan standar dalam proses penilaian seharusnya tidak terjadi. Proses evaluasi harus dilakukan dengan berpatokan pada kriteria penilaian yang sama. Standar yang dipakai seharusnya sesuai dengan tujuan pembelajaran yang telah dirumuskan sebelumnya. Perbedaan standar antar penguji tentu beresiko pada tingkat validitas data yang diperoleh.

Kriteria penilaian merupakan suatu yang penting untuk dirumuskan dalam proses evaluasi. Hal tersebut dibutuhkan terutama dalam menilai kualitas produk dan kinerja. Penilaian produk dan kinerja sangat umum dilakukan dalam konteks pelatihan di dunia kerja. Hal ini sangat wajar karena pada umumnya pelatihan di dunia kerja dilakukan dengan tujuan meningkatkan kinerja dan kualitas produk.

Merinci kriteria penilaian sebetulnya bukanlah sesuatu yang baru dalam sebuah proses penilaian. Perincian kriteria lazim dijumpai pada instrumen penilaian dengan format rubrik. Penilaian dengan menggunakan instrumen rubrik dikenal juga dengan istilah penilaian autentik (*authentic assessment*). Sebagaimana dijelaskan melalui namanya, penilaian ini dikembangkan dengan tujuan untuk mendapatkan nilai yang lebih otentik.

Permasalahan di atas ternyata tidak hanya terjadi di Pusbindiklat LIPI. Dalam artikel ilmiah yang ditulis oleh O'Donnell, Oakley, Haney, O'Neill, dan Taylor (2011), diceritakan bahwa permasalahan serupa juga terjadi pada latar pendidikan tinggi. Dalam tulisannya, mereka mendeskripsikan permasalahan dalam proses penilaian praktikum mahasiswa kedokteran gigi.

O'Donnell et al. (2011) menjelaskan bahwa pada umumnya pengajar pada program pendidikan kedokteran gigi terdiri dari praktisi dan ilmuwan. Para pengajar memang menguasai keilmuan secara substansi, namun pada umumnya mereka tidak cukup menguasai disiplin ilmu pendidikan. Sebagai akibatnya, mereka kesulitan ketika harus menilai kemampuan mahasiswanya. Permasalahan ini semakin bertambah berat karena mereka tidak memiliki acuan yang bisa dijadikan standar atau kriteria dalam melakukan penilaian. Akhirnya nilai yang diberikan oleh pengajar menjadi cenderung subjektif. Berdasarkan latar belakang serta kajian yang telah dilakukan sebelumnya, O'Donnell et al. (2011) merekomendasikan untuk mengembangkan instrumen penilaian dengan format rubrik.

Perumusan kriteria penilaian pada rubrik juga dapat kita lihat pada penelitian yang dilakukan oleh Furze, Gale, Black, Cochran, dan Jensen (2015). Latar belakang dari penelitian ini adalah kesulitan yang dialami oleh para pengajar dalam menilai kemampuan *clinical reasoning* mahasiswa jurusan *physical therapist* di *Creighton University, California*.

Furze et al. (2015) menuliskan bahwa ada beberapa instrumen yang sudah digunakan sebelumnya. Instrumen tersebut diantaranya adalah *Health Science Instrument Test* dan *California Critical Thinking Skill Test*. Walaupun instrumen

tersebut sudah mengukur sampai dengan level berfikir kritis (*high order thinking skill*), namun instrumen tersebut dinilai masih belum memadai.

Furze et al. (2015) mengatakan walaupun *critical thinking* adalah kemampuan yang penting, namun *clinical reasoning* merupakan konsep yang lebih kompleks. *Clinical reasoning* tidak hanya berfokus pada aktivitas mental saja namun juga berhubungan dengan kinerja dan perilaku yang berhubungan dengan pasien, praktisi kesehatan, dan lingkungan kerja. Hal tersebut kemudian menjadi alasan dikembangkannya instrumen penilaian dengan format rubrik.

Pengembangan instrumen rubrik dianggap sebagai solusi yang tepat bagi permasalahan di atas. Pengembangan instrumen tersebut sangat membantu pengajar untuk melakukan penilaian terhadap kinerja *clinical reasoning* mahasiswa. Dengan dikembangkannya instrumen rubrik, pengajar mendapatkan data yang lebih valid dan otentik.

Selain penelitian di atas masih terdapat contoh penelitian lain yaitu penelitian yang dilakukan oleh Rakedzon dan Tsabari (2017). Jika pada penelitian Furze et al. (2015) rubrik penilaian yang dikembangkan digunakan untuk menilai kinerja, maka pada penelitian selanjutnya kita akan melihat contoh pengembangan instrumen rubrik yang digunakan untuk menilai produk. Penelitian yang dilakukan oleh Rakedzon dan Tsabari (2017) bertujuan untuk menilai karya tulis mahasiswa dalam bahasa inggris.

Penelitian yang dilakukan oleh Rakedzon dan Tsabari (2017) bertujuan untuk mengembangkan sebuah rubrik penilaian. Rubrik tersebut akan digunakan untuk menilai produk berupa tulisan ilmiah dalam bahasa inggris. Ide penelitian ini berasal dari banyaknya program pelatihan yang melatih pesertanya untuk membuat karya tulis dalam bahasa inggris. Walaupun program serupa telah banyak ditemui namun pada umumnya mereka tidak memiliki instrumen penilaian yang tepat. Instrumen tes bahasa inggris yang ada seperti TOEFL, IELTS, dan TOEIC hanya mengukur pada aspek pengetahuan saja dan tidak dapat digunakan untuk menilai produk akhir pembelajaran. Oleh karena itu mereka berinisiatif untuk mengembangkan instrumen dengan format rubrik untuk menilai produk akhir berupa karya tulis ilmiah dalam bahasa inggris.

Selain beberapa contoh di atas, ternyata rubrik juga dapat digunakan untuk mendukung sebuah penelitian. Schweikhard, Hoberecht, Peterson, dan Peterson (2018) melakukan sebuah penelitian tentang pengaruh *library tutorial online* terhadap kemampuan literasi informasi mahasiswa. Penelitian tersebut bermaksud untuk melihat sejauh mana tutorial tersebut dapat membantu mahasiswa dalam melakukan penelusuran literasi. Pengumpulan data dilakukan dengan cara menilai karya tulis mahasiswa yang telah menggunakan tutorial tersebut. Penilaian tersebut dilakukan dengan menggunakan instrumen dengan format rubrik. Sampel pada penelitian adalah 180 orang mahasiswa pascasarjana di *University of Oklahoma*.

Berdasarkan contoh penelitian di atas, kiranya kita dapat mengambil solusi yang dapat diterapkan di Pusbindiklat LIPI. Pusbindiklat LIPI perlu juga untuk mengembangkan instrumen penilaian dengan format rubrik. Instrumen tersebut nantinya akan membantu penguji dalam menilai proposal penelitian yang disusun oleh peserta Program Pelatihan Jabatan Fungsional Peneliti (PPJFP). Dengan menggunakan rubrik penilaian diharapkan nilai yang diberikan kelak menjadi lebih valid dan otentik.

Permasalahan lainnya adalah dari sisi teknis yang selama ini peneliti rasakan secara langsung. Proses input nilai masih dilakukan dengan cara yang manual. Proses ini menjadi tidak efektif manakala hasil rekapitulasi nilai dilakukan dalam waktu yang sangat sempit. Biasanya penginputan nilai dilakukan setelah seminar berakhir.

Proses input nilai harus segera selesai karena data tersebut dibutuhkan dalam rangkaian pekerjaan yang lain. Data tersebut masih harus dibahas dalam rapat kelulusan. Setelah data tersebut selesai dirapatkan, selanjutnya data tersebut harus dicetak pada sertifikat kelulusan dan surat pengembalian yang masing-masing berjumlah 30 lembar. Apabila disimpulkan maka kegiatan rekapitulasi nilai, rapat kelulusan, pencetakan 30 lembar sertifikat, serta pencetakan 30 lembar surat pengembalian semuanya harus diselesaikan dalam waktu sekitar dua jam saja.

Alokasi waktu pekerjaan yang sempit berdampak pada penurunan kinerja (*performance*) sdm yang bersangkutan dan memiliki resiko kesalahan pekerjaan (*human error*). Oleh karena itu berdasarkan pengalaman yang peneliti rasakan, peneliti merasa perlu adanya inovasi untuk menyelesaikan masalah tersebut. Pekerjaan tersebut harus diintervensi dengan alat (*tools*) yang dapat membantu penyelesaian pekerjaan tersebut.

1.2. Pembatasan Penelitian

Agar penelitian ini lebih fokus dan terarah, peneliti merasa perlu melakukan pembatasan masalah. Penelitian ini tidak akan membahas seluruh permasalahan pada Program PPJFP, namun hanya dibatasi pada mata diklat penulisan proposal penelitian.

Setelah melakukan pembatasan masalah, maka peneliti perlu juga untuk melakukan identifikasi masalah. Identifikasi masalah pada mata diklat penulisan proposal penelitian adalah sebagai berikut:

1. Permasalahan kinerja Widyaiswara dalam menilai proposal penelitian peserta pelatihan.
2. Perbedaan pada standar penilaian yang digunakan oleh masing-masing Widyaiswara,. Hal ni dapat dilihat dari tingginya selisih nilai dari tiap Widyaiswara. Akibatnya proses penilaian menjadi sangat subjektif tergantung pada siapa yang menjadi melakukan penilaian.
3. Instrumen penilaian yang digunakan saat ini belum bisa menghasilkan nilai yang reliabel (konsisten).
4. Metode pengolahan nilai yang masih tradisional dan dilakukan dalam waktu yang sangat terbatas dapat memicu terjadinya *human error*.

Berdasarkan uraian di atas peneliti merasa perlu untuk mencari solusi dalam meningkatkan kinerja Widyaiswara dalam melakukan penilaian proposal penelitian. Peningkatan kinerja tersebut dilakukan dengan cara mengembangkan instrumen rubrik penilaian yang valid dan reliabel.

Produk yang dikembangkan juga harus dapat meringkas waktu pekerjaan. Dengan demikian pekerjaan mengolah hasil nilai dapat dilakukan dengan efektif dan efisien. Untuk itu, maka instrumen penilaian harus dikembangkan dalam bentuk elektronik. Dengan demikian proses penilai menjadi lebih cepat, tepat, dan akurat.

1.3. Rumusan Masalah

Berdasarkan identifikasi masalah yang telah disampaikan sebelumnya, maka rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana rancangan rubrik elektronik penilaian proposal penelitian yang akan digunakan pada pelatihan pembentukan jabatan fungsional peneliti?
2. Bagaimana kelayakan rubrik elektronik penilaian proposal penelitian yang akan digunakan pada pelatihan pembentukan jabatan fungsional peneliti?
3. Bagaimana efektivitas rubrik elektronik penilaian proposal penelitian yang akan digunakan pada pelatihan pembentukan jabatan fungsional peneliti?

1.4. Tujuan Penelitian

Pada dasarnya sebuah penelitian bertujuan untuk mengembangkan suatu disiplin ilmu tertentu. Hasil akhir dari sebuah penelitian harus dapat berkontribusi bagi perkembangan suatu disiplin ilmu. Oleh karenanya di dalam sebuah penelitian dikenal istilah tujuan teoritis. Tujuan teoritis dari penelitian ini adalah untuk memberikan kontribusi terhadap disiplin ilmu teknologi pendidikan khususnya pada kawasan penilaian.

Penelitian ini juga diharapkan dapat mendatangkan manfaat bagi komunitas ilmiah khususnya di bidang teknologi pendidikan. Peneliti berharap kelak hasil penelitian ini dapat dijadikan rujukan bagi penelitian-penelitian selanjutnya khususnya yang berhubungan dengan topik penilaian berbasis rubrik dan kemampuan dalam menyusun proposal penelitian.

Sebagaimana telah dijelaskan di atas bahwa *output* dari penelitian ini adalah sebuah rubrik penilaian berbasis elektronik untuk menilai proposal penelitian.

Peneliti berharap kelak produk tersebut dapat digunakan pada program pelatihan baik yang diadakan di LIPI maupun juga oleh lembaga lainnya.

Teknologi Pendidikan adalah sebuah disiplin ilmu yang bertujuan untuk menyelesaikan permasalahan dalam hal belajar dan pembelajaran. Oleh karenanya disiplin ilmu teknologi pendidikan masuk dalam rumpun ilmu terapan. Dengan demikian pada umumnya penelitian yang dilakukan akan memiliki tujuan praktis. Adapun secara praktis, tujuan umum dari penelitian ini adalah sebagai berikut.

1. Menghasilkan produk elektronik penilaian yang layak digunakan untuk menilai proposal penelitian pada pelatihan pembentukan jabatan fungsional peneliti.
2. Menganalisis kelayakan rubrik elektronik penilaian proposal penelitian yang akan digunakan pada pelatihan pembentukan jabatan fungsional peneliti.
3. Menganalisis efektivitas rubrik elektronik penilaian proposal penelitian yang akan digunakan pada pelatihan pembentukan jabatan fungsional peneliti.

1.5. State of The Art

State of the art dapat diartikan sebagai gambaran aktual dari sebuah topik penelitian. Pencarian *state of the art* dilakukan dalam rangka mencari *novelty* (nilai kebaruan) dari sebuah penelitian.

Novelty dapat diperoleh berdasarkan dua kriteria. Kriteria yang pertama adalah kebaruan berdasarkan produk. Kriteria yang kedua ialah kebaruan berdasarkan proses atau metode (Pusbindiklat LIPI, 2019). Pada bagian ini, peneliti akan memaparkan *state of the art* sekaligus *novelty* penelitian baik berdasarkan produk maupun proses.

1.5.1. Kebaharuan Penelitian Berdasarkan Produk

Untuk mencari *novelty* (nilai kebaruan penelitian) dari segi produk, peneliti telah melakukan penelusuran terhadap beberapa penelitian terdahulu. Penelusuran tersebut meliputi beberapa penelitian dengan tema penilai karya tulis. Melalui penelusuran ini, peneliti juga berharap dapat mencari format penilaian yang paling cocok untuk digunakan pada Program PPJFP LIPI.

Penelitian yang pertama adalah penelitian yang dilakukan oleh Parra G. dan Calero S. (2019). Penelitian tersebut bertujuan untuk menguji pengaruh *Automated Writing Evaluation (AWE)* terhadap kemampuan membuat tulisan berbahasa Inggris mahasiswa di Ekuador. *AWE* adalah sebuah perangkat lunak yang dikembangkan untuk mengoreksi tulisan berbahasa Inggris. Dalam penelitian ini mahasiswa menggunakan *AWE* dengan nama *Grammarly* dan *Grammark*. Dengan bantuan alat tersebut diharapkan mahasiswa dapat mengetahui kesalahan dari tulisan yang telah mereka buat untuk kemudian mereka pelajari.

Penelitian ini merupakan jenis penelitian eksperimen. Pengujian dilakukan dengan cara membandingkan hasil belajar antara mahasiswa yang belajar dengan *AWE* dan tidak.

Peneliti tertarik untuk mencari tahu perihal instrumen yang digunakan dalam penelitian tersebut. Terdapat dua buah instrumen yang digunakan dalam penelitian ini. Instrumen yang digunakan meliputi pengumpulan data berupa hasil belajar dan persepsi mahasiswa. Namun demikian peneliti hanya akan fokus pada instrumen yang digunakan untuk menilai hasil belajar setelah mahasiswa menulis dengan bantuan *AWE*.

Instrumen yang digunakan untuk mengumpulkan data berupa hasil belajar pada penelitian ini adalah *versant standardized placement test* yang dikembangkan oleh Pearson pada tahun 2016 (Parra G. dan Calero S. 2019).

Versant standardized placement test adalah sebuah instrumen yang dikembangkan untuk mengukur kemampuan berbahasa Inggris. Instrumen tersebut menguji beberapa keterampilan berbahasa diantaranya kemampuan membaca lantang, pengulangan, membangun kalimat, percakapan, menulis, penyelesaian kalimat, pendiktean, rekonstruksi bagian, dan terakhir mengujikan kemampuan dalam membuat kesimpulan dan pendapat. Dari semua bagian tersebut penelitian di atas hanya mengujikan beberapa bagian yang berhubungan dengan keterampilan menulis saja.

Peneliti berpendapat bahwa instrumen yang digunakan dalam penelitian di atas masih memiliki kelemahan. Evaluasi dilakukan untuk menilai ketercapaian tujuan pembelajaran. Oleh karena itu harus ada keterpaduan antara tujuan, evaluasi, dan pembelajaran itu sendiri. Instrumen yang digunakan pada penelitian di atas tidak dikembangkan berdasarkan tujuan dari pembelajaran yang diujicobakan.

Berdasarkan kritik di atas, peneliti merekomendasikan bahwa sebaiknya penelitian tersebut dilakukan dengan instrumen yang dikembangkan sesuai dengan tujuan pembelajaran pada penelitian tersebut. Dengan demikian data yang dikumpulkan akan semakin valid. Validitas instrumen akan berpengaruh juga pada tingkat validitas data yang dikumpulkan.

Penelitian selanjutnya adalah penelitian yang dilakukan oleh Halim, Hartati, dan Lestari (2019). Penelitian kualitatif ini bertujuan untuk melihat penggunaan portofolio dalam proses penilaian kemampuan siswa dalam membuat karya tulis berbahasa Inggris. Penelitian dilakukan di SMP Negeri 1 Pageran Cianjur.

Berdasarkan temuan peneliti diketahui bahwa guru melakukan seluruh prosedur penilaian portofolio sebagaimana yang disampaikan oleh Brown & Abeywickrama (Halim et al., 2019). Prosedur tersebut meliputi desain tujuan, bahan, waktu

pengumpulan produk, waktu melakukan review, dan tempat portofolio.

Terdapat beberapa permasalahan yang dialami oleh guru ketika melakukan penilaian dengan menggunakan portofolio. Permasalahan yang pertama adalah tidak cukupnya waktu untuk memberikan umpan balik bagi seluruh siswa. Penilaian berbasis portofolio menyita banyak waktu pekerjaan. Permasalahan yang kedua adalah guru mengalami kesulitan untuk menontrol dan membimbing seluruh siswa selama proses pembelajaran. Permasalahan yang ketiga adalah guru tidak menggunakan rubrik untuk menilai karya tulis siswa.

Penggunaan portofolio sebagai cara untuk melakukan penilaian ternyata memiliki kekurangan. Panjangnya proses penilaian membutuhkan waktu pengerjaan yang tidak singkat. Akibatnya waktu pengerjaan menjadi tidak efisien. Hal tersebut selanjutnya akan mengganggu kinerja guru.

Hambatan terhadap kinerja guru tentu bertentangan dengan prinsip dan tujuan teknologi pendidikan. Sejatinya peningkatan kinerja guru (*improving performance*) adalah salah manfaat yang dapat dirasakan apabila guru menerapkan praktek teknologi pendidikan (Januszewski & Molenda, 2008). Peneliti berasumsi bahwa permasalahan ini berkaitan dengan temuan pada poin ketiga yaitu guru tidak menggunakan rubrik untuk menilai karya tulis siswa.

Penelitian yang ketiga adalah penelitian yang dilakukan oleh Abdillah (2017). Penelitian ini berjudul *Implementasi Fuzzy Simple Additive Weighting (SAW) Sebagai Pendukung Keputusan pada Beasiswa Penelitian*. Melalui penelitian tersebut, Abdillah mencoba mengintegrasikan *Fuzzy Simple Additive Weighting (SAW)* pada proses seleksi penerima beasiswa Riset Inovatif Produktif (RISPRO) dari Lembaga Pengelola Dana Pendidikan (LPDP). Seleksi

dilakukan dengan cara menilai proposal penelitian calon penerima beasiswa.

Dari penelitian tersebut diketahui bahwa selama ini proses penilaian proposal penelitian dilakukan dengan menilai beberapa kriteria. Kriteria penilaian proposal penelitian meliputi kualitas riset, luaran riset, kemutakhiran riset, dan rekam jejak periset. Lebih lengkapnya akan ditampilkan pada tabel 1.

Tabel 1.1
Kriteria Penilaian Proposal Penelitian Calon Penerima Beasiswa
RISPRO LPDP

Aspek	Subaspek	Bobot (%)
Kualitas Riset	a. Roadmap riset terkait implementasi teknologi	20
	b. Kelayakan pelaksanaan (kredibilitas mitra, sarana penunjang, serta uraian tugas, anggaran dan jadwal)	20
Luaran Riset	a. Potensi hak kekayaan intelektual	15
	b. Nilai strategis dan komersial produk/teknologi	15
Kemutakhiran Riset	a. Kepustakaan yang dipakai/digunakan dan publikasi periset (terutama ketua periset)	5
	b. <i>State of the art</i>	5
	c. Metode pendekatan	10
Rekam Jejak Periset	a. Produktivitas riset dan publikasi	5
	b. Kerja sama riset dengan industri/UKM	5

Apabila dilihat dari bentuk instrumen yang digunakan, maka proses penilaian proposal penelitian dilakukan dengan menggunakan rubrik penilaian. Rubrik yang digunakan merupakan jenis rubrik holistik. Tidak seperti rubrik analitik, rubrik holistik tidak memiliki skala dan deskripsi kriteria. Skala dan deskripsi kriteria merupakan tingkatan kualitas pekerjaan yang dibutuhkan untuk menyelesaikan suatu tugas (Stevens & Levi, 2005).

Dalam pandangan peneliti, instrumen di atas memang cocok ketika digunakan sebatas sebagai alat untuk melakukan seleksi. Instrumen tersebut dapat digunakan sebagai alat untuk menilai produk. Berdasarkan nilai tersebut nantinya akan dipilih beberapa karya terbaik untuk kemudian diberikan penghargaan berupa

beasiswa. Namun demikian peneliti berpendapat bahwa instrumen tersebut tidak dapat digunakan dalam konteks menilai hasil belajar.

Sebagai telah dibahas sebelumnya bahwa evaluasi bertujuan untuk menilai ketercapaian hasil belajar. Oleh karenanya penilaian harus mengacu pada indikator ketercapaian pembelajaran. Instrumen yang digunakan pada proses seleksi di atas tidak memuat indikator pembelajaran dengan jelas. Akibatnya masing-masing penilai akan memiliki standar yang berbeda. Angka pada nilai yang diberikan akan beragam dan tidak memiliki bobot kualitas yang sama. Nilai 90 yang diberikan oleh seorang penilai belum tentu bermakna sama pada nilai 90 yang diberikan oleh penilai yang lain. Hal ini tentu bertentangan dengan prinsip penilai dengan acuan patokan (PAP) (Seels & Richey, 1994).

Di akhir bagian ini, peneliti akan mencari format penilaian yang paling cocok untuk menilai proposal penelitian pada program PPJFP LIPI. Pencarian tersebut didasarkan pada kritik dan komentar terhadap ketiga penelitian di atas.

Prinsip pengembangan yang pertama adalah bahwa instrumen yang akan dipakai nanti harus dikembangkan berdasarkan tujuan pembelajaran yang tertera pada kurikulum program PPJFP LIPI. Pemanfaatan instrumen yang sudah ada seperti yang dilakukan pada penelitian Parra G. and Calero S. (2019) tentu tidak dapat peneliti ikuti. Dengan menerapkan prinsip ini kelak akan melahirkan instrumen yang memiliki tingkat validitas yang tinggi.

Prinsip yang kedua adalah instrumen yang akan digunakan harus dapat dipakai untuk menilai produk hasil belajar. Hal ini peneliti simpulkan berdasarkan temuan pada penelitian Parra G. and Calero S. (2019). Instrumen *versant standardized placement test* yang digunakan pada penelitian tersebut dapat digunakan untuk mengukur kinerja namun tidak dapat digunakan untuk mengukur

kualitas produk hasil belajar. Dalam hal ini berupa sebuah karya tulis.

Prinsip yang ketiga berkaitan dengan kepraktisan dan efisiensi waktu. Penggunaan portofolio sebagaimana dalam penelitian Halim, Hartati, dan Lestari (2019) terbukti memiliki kekurangan dari segi lamanya waktu pekerjaan. Instrumen yang digunakan ternyata mengganggu kinerja guru. Hal ini tentu tidak kita harapkan.

Prinsip yang keempat berkaitan dengan konsistensi instrumen (reliabilitas). Hal ini didasarkan pada temuan yang peneliti dapatkan pada penelitian Abdillah (2017). Ketiadaan panduan penilaian berpotensi pada berbedanya standar penilaian yang digunakan oleh tiap penguji.

Prinsip yang terakhir adalah instrumen penilaian yang akan dikembangkan harus beracuan pada patokan (PAP). Dengan demikian instrumen tersebut dapat benar-benar digunakan untuk mengukur keberhasilan sebuah pembelajaran.

Kelima prinsip di atas masih dapat diperas lagi menjadi empat kata kunci. Prinsip pertama dan kedua berbicara tentang validitas. Prinsip yang ketiga berbicara kepraktisan. Prinsip yang keempat berkaitan dengan reliabilitas. Dan terakhir prinsip yang kelima berbicara terkait PAP. Dengan demikian kata kunci dalam pengembangan instrumen kelak adalah valid, praktis, reliabel, dan peracuan patokan (PAP).

Berdasarkan rumusan di atas maka peneliti berencana untuk mengembangkan sebuah rubrik analitik yang akan digunakan untuk menilai sebuah proposal penelitian. Pengembangan rubrik didasarkan pada prinsip valid, praktis, reliabel, dan beracuan pada patokan. Pembahasan mengenai keunggulan rubrik analitik akan dibahas lebih lanjut pada BAB II.

1.5.2. Kebaharuan Penelitian Berdasarkan Metode

Untuk menemukan nilai kebaruan (*novelty*) dari penelitian dari segi metode, peneliti telah melakukan penelusuran terhadap beberapa penelitian terdahulu. Berdasarkan penelusuran tersebut, peneliti menemukan bahwa model yang digunakan diantaranya adalah model 4D (*devine, design, develop, disseminate*) Thiagarjan, *Borg and Gall*, dan Pendekatan lain yang tidak dijelaskan namanya namun tetap memiliki prosedur yang jelas.

Beberapa penelitian pengembangan rubrik penilaian yang menggunakan model 4D dan *Borg and Gall* masing-masing akan ditampilkan dalam tabel di bawah ini.

Tabel 1. 2
Pengembangan Rubrik dengan Model 4D Thiagarjan dan Borg and Gall

No	Penulis	Judul Penelitian	Model Pengembangan	Jurnal / <i>Proceeding</i>
1	Fitriani & Yarmayani (2018)	Pengembangan Rubrik Berpikir Kreatif Siswa Menegah Atas Dala Menyelesaikan Masalah Matematika	4D Thiagarjan	Mosharafa Tahun 2018 Vol. 7
2	Aji, Hudha, Huda, Nandiyanto, & Abdullah (2018)	The Improvement of Learning Effectiveness in The Lesson Study by Using E-Rubric	4D Thiagarjan	Journal of Engineering Science and Technology Vol. 13, No. 5 Tahun 2018 1181 - 1189 © School of Engineering , Taylor's University

3	Hudha, Aji, & Huda (2018)	E-Rubric: Scientific Work Based on Android for Experimental Physic	4D Thiagarjan	IOP Conference Series: Materials Science and Engineering / Tahun 2018
4	Febriyanti, Harsiati, & Dermawan (2017)	Pengembangan Instrumen Asesmen Menulis Kreatif Cerita Fantasi untuk Siswa Kelas VII SMP	4D Thiagarjan	Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan Tahun 2017 Vol. 2
5	Hermawan et al., (2017)	Desain Rubrik Kemampuan Berkolaborasi Siswa SMP dalam Materi Pemantulan Cahaya	4D Thiagarjan	JPPPF- Jurnal Penelitian & Pengembangan Pendidikan Fisika Tahun 2017 Vol.3
6	Latifa, Rahman, Hamra, Jabu, & Nur (2015)	Developing a Practical Rating Rubric of Speaking Test for University Students of English in Parepare, Indonesia	Borg and Gall	English Language Teaching Tahun 2015 Vol. 8

Apabila ditinjau dari tahun terbitnya, penelitian di atas masih relevan untuk dijadikan rujukan. Penelitian di atas berasal dari rentang tahun 2015 sampai dengan 2018. Namun demikian walaupun penelitian di atas dilakukan dalam kurun waktu lima tahun

terakhir, tetapi model pengembangan yang digunakan adalah model yang sudah lama. Model *4D* dan *Borg and Gall* masing-masing diperkenalkan pada tahun 1974 dan 1989 (Arifin, 2011b). Dengan demikian pada tahun 2020 ini kedua model tersebut masing-masing telah berusia 46 dan 31 tahun. Disamping itu seluruh penelitian di atas juga ternyata dilakukan di dalam negeri. Dari sini kita dapat melihat kecenderungan penelitian pengembangan rubrik penilaian di dalam negeri yang cenderung menggunakan model *4D* dan *Borg and Gall*.

Pemilihan model pengembangan yang tepat dibutuhkan untuk menghasilkan rubrik penilaian yang valid. Untuk itu peneliti merasa perlu untuk melihat tahapan pengembangan baik pada model pengembangan *4D* maupun *Borg and Gall*.

Istilah *4D* merupakan singkatan dari tahapan dalam model itu sendiri. Tahapan pengembangan dengan model *4D* terdiri dari *define*, *design*, *develop*, dan *disseminate* (Arifin, 2011b). Tahapan *define* bisa disebut juga sebagai studi pendahuluan baik secara teoritik maupun empirik. Tahapan *design* dilakukan dengan merancang model dari program yang akan dikembangkan. Tahapan *develop* merupakan kegiatan mengkaji produk awal secara empirik terdiri dari uji coba, revisi, dan validasi. Selanjutnya tahapan terakhir yaitu *disseminate* merupakan upaya penyebarluasan produk akhir.

Model pengembangan *Borg and Gall* memiliki tahapan yang lebih panjang dibanding model *4D*. Tahapan pada model *Borg and Gall* terdiri dari 10 tahapan (Arifin, 2011b). Tahapan yang dimaksud terdiri dari:

- a. *Research and Information Collecting*,
- b. *Planning*,
- c. *Develop Preliminary Form of Product*,
- d. *Preliminary Field Testing*,
- e. *Main Product Revision*,

- f. *Main Filed Testing*,
- g. *Operational Product Revision*,
- h. *Operational Field Testings*,
- i. *Final Product Revision*,
- j. dan *Disseminatioan and Implementation*.

Sebagai mana telah dibahas sebelumnya, penelitian ini bertujuan untuk menghasilkan produk pendidikan berupa rubrik penilaian. Produk tersebut merupakan instrumen yang akan digunakan untuk menilai produk peserta diklat yaitu proposal penelitian. Dengan demikian, hasil akhir dari penelitian ini adalah instrumen penilaian.

Sebuah instrumen penilaian memiliki syarat yang mutlak untuk dimiliki. Sebuah instrumen harus memiliki tingkat validitas dan reliabilitas yang tinggi (Arifin, 2011a). Validitas berkaitan dengan ketepatan pemilihan instrumen dan reliabilitas berkaitan dengan tingkat konsistensi instrumen untuk menilai hasil belajar. Hal ini akhirnya mengantarkan kita pada pertanyaan, sejauh mana model pengembangan dapat menghasilkan produk berupa instrumen penilaian yang valid dan reliabel.

Apabila dilihat dari prosedur pengembangannya, baik model *4D* maupun *Borg and Gall*, sebetulnya dapat dipakai untuk menghasilkan instrumen penilaian yang valid. Tahapan *develop* pada model *4D* diakhiri dengan kegiatan validasi. Dalam konteks pengembangan instrumen penilaian, tahapan ini dapat diartikan sebagai proses uji validitas instrumen. Begitu pun dalam model *Borg and Gall* proses uji validitas instrumen mulai terakomodir dari tahapan ketiga yaitu *Develop Preliminary Form of Product*. Pada tahapan ini, peneliti harus melakukan uji validitas instrumen pada ahli mata pelajaran atau bidang studi. Namun demikian tidak satu pun dari kedua model tersebut yang merepresentasikan uji

reliabilitas instrumen. Oleh karenanya, dalam hal ini peneliti menemukan sebuah *gap teoritik*.

Uji reliabilitas sangat penting untuk dilakukan. Reliabilitas instrumen berhubungan dengan tingkat konsistensi instrumen dalam melakukan penilaian. Hal ini sangat penting terlebih dalam konteks penilaian berbasis rubrik. Instrumen penilaian dengan format rubrik akan digunakan oleh penguji yang berbeda-beda. Perbedaan penguji dikhawatirkan dapat berpengaruh pada objektivitas dan konsistensi penilaian. Oleh karenanya peneliti berpendapat bahwa pengembangan instrumen penilaian dengan format rubrik, harus disertai dengan uji reliabilitas. Berdasarkan uraian di atas, maka peneliti merasa perlu untuk melakukan uji reliabilitas pada produk penilaian yang kelak akan peneliti hasilkan.

Melihat hal tersebut peneliti merasa perlu melakukan penelusuran terhadap penelitian terdahulu. Penelusuran tersebut bertujuan untuk mencari model lain yang mungkin lebih relevan dan dapat digunakan untuk mengembangkan sebuah instrumen rubrik penilaian yang valid dan reliabel.

Berdasarkan penelusuran yang telah dilakukan, akhirnya peneliti menemukan beberapa penelitian pengembangan rubrik penilaian yang dilakukan dengan model yang berbeda. Penelitian tersebut diantaranya adalah penelitian yang dilakukan oleh Rakedzon & Tsabari (2017); Furze, et. al. (2015); dan Sumiyarrini, Rahayu, & Suhoyo (2017). Dua dari tiga penelitian ini telah peneliti singgung sebelumnya pada bagian latar belakang. Oleh karena itu, selanjutnya peneliti akan memfokuskan pembahasan pada model dan prosedur pengembangannya saja.

Dalam penelitiannya, Rakedzon & Tsabari (2017) menggunakan model pengembangan dari crusan. Model crusan terdiri dari lima tahapan. Lima tahapan dari model crusan adalah sebagai berikut.

- a. *Developing course goals* (Mengembangkan tujuan pembelajaran)
- b. *Choosing assessment tasks to fit these goals* (Memilih penugasan yang sesuai dengan tujuan pembelajaran)
- c. *Setting the standards for these tasks and goals* (Menentukan standar dari tiap penugasan)
- d. *Developing criteria to assess performance* (Mengembangkan kriteria untuk menilai kinerja)
- e. *Rating values for analytic scoring* (Melakukan pembobotan nilai)

Dilihat dari tahapan prosedur pengembangannya, model crusan sangat cocok untuk dipakai dalam pengembangan rubrik penilaian. Namun demikian, prosedur pengembangan crusan tidak membahas cara untuk menguji validitas dan reliabilitas instrumen.

Furze, et. al. (2015) memiliki cara yang berbeda dalam mengembangkan rubrik penilaiannya. Dalam tulisannya ia tidak mengatakan secara eksplisit perihal nama model maupun prosedur pengembangan yang dipakai. Dalam bagian metode pengembangan di tulisannya, Furze, et. al. (2015) hanya mengatakan bahwa rubrik yang ia kembangkan merupakan penjabaran dari kemampuan berpikir klinis. Namun demikian, pada bagian yang lain ia menjelaskan tahapan kegiatan penelitiannya. Informasi tersebut menjadi menarik karena menceritakan tahapan kegiatan pengembangan rubrik penilaiannya. Berikut adalah tahapan kegiatan penelitian yang dikutip dari Furze, et. al. (2015).

Tabel 1.3
Kegiatan Penelitian Klinis Furze, et. al. (2015)

<i>Date</i>	<i>Action</i>	<i>Comment</i>
<i>May 2010</i>	<i>Initial grading rubric developed.</i>	<i>Result of initial qualitative study exploring development of clinical reasoning skills in physical therapy students across time.</i>
<i>October 2010</i>	<i>Focus group. Grading rubric revised.</i>	<i>Investigators revised rubric based on feedback from a focus group of interprofessional educators (n = 5).</i>
<i>November 2010 – March 2011</i>	<i>Clinical Education Consortium. Grading rubric revised.</i>	<i>Revisions based on feedback from Clinical Education Consortium (n = 48) and further review by investigators.</i>
<i>April 2011 – July 2011</i>	<i>Faculty feedback.</i>	<i>Grading rubric used for CCPE. Revisions based on feedback from faculty.</i>
<i>November 2011</i>	<i>Grading rubric revised.</i>	
<i>January 2012</i>	<i>Received internal grant and initiated study.</i>	<i>Purpose of the grant was to further explore, revise, and establish content validity of the grading rubric.</i>
<i>March 2012</i>	<i>Survey sent to content experts. Grading rubric revised.</i>	<i>Revisions based on feedback from content experts (n = 6).</i>
<i>April 2013</i>	<i>Final grading rubric.</i>	

Sumber (Furze et al., 2015)

Sebagaimana telah disinggul di awal, penelitian yang dilakukan oleh Furze, et. al. (2015) tidak menjelaskan secara spesifik perihal cara dalam menyusun rubrik penilaiannya. Namun demikian, berdasarkan tabel di atas penelitian ini memiliki kekuatan terutama pada proses validitas instrumennya.

Berdasarkan uraian kegiatan pada tabel di atas, peneliti dapat berasumsi bahwa Furze, et. al. (2015) telah melakukan serangkaian aktivitas uji validitas. Uji validitas yang pertama dilakukan melalui kegiatan *focus group discussion* pada bulan Oktober 2010 dan dihadiri oleh lima orang praktisi ahli. Selanjutnya uji validitas yang kedua ditandai dengan adanya revisi pada rentang waktu November 2010 sampai dengan Maret 2011 dimana revisi tersebut didasarkan pada masukan dari 48 orang anggota *educational education consortium*. Uji validitas yang terakhir dapat diindikasikan dari adanya revisi pada bulan Maret 2012 di mana revisi tersebut didasarkan pada masukan dari enam orang ahli materi. Dengan melihat data di atas kita dapat melihat bahwa penelitian ini melakukan tiga kali uji validitas dengan melibatkan 59 orang pakar yang terkait.

Melihat fakta di atas, peneliti dapat berasumsi bahwa penelitian tersebut telah menghasilkan produk berupa rubrik penilaian dengan tingkat validitas yang tinggi. Namun demikian sebagaimana dapat dilihat bahwa ternyata penelitian ini juga belum menyinggung perihal reliabilitas soal.

Penelitian selanjutnya adalah penelitian yang dilakukan oleh Sumiyarrini, Rahayu, & Suhoyo (2017). Penelitian ini bertempat di Indonesia dan bertujuan untuk menghasilkan produk yang mereka dinamai *Instrumen Nursing Clinical Exercise (NCEx)*. *NCEx* adalah sebuah rubrik penilaian yang digunakan untuk menilai kinerja perawat. Seperti penelitian sebelumnya, penelitian ini tidak secara eksplisit mengatakan perihal nama metode atau prosedur

pengembangan yang digunakan. Peneliti juga tidak menemukan referensi yang dijadikan rujukan dalam prosedur pengembangan rubrik penilaian. Namun demikian, Sumiyarrini, et. al. (2017) dapat menjelaskan secara logis perihal prosedur yang mereka gunakan. Berikut adalah langkah-langkah pada penelitian tersebut.

- a. Menyusun draf area penilaian dengan mengacu pada standar kompetensi yang akan diujikan.
- b. Menyusun tingkatan keterampilan pada tiap area penilaian dengan bimbingan seorang ahli materi.
- c. Melakukan *focus grup discussion* bersama 7 orang yang terdiri dari praktisi dan ahli materi, kemudian dilanjutkan dengan melakukan revisi.
- d. Melakukan *uji content validity* dengan mengukur kesepakatan para ahli.
- e. Melakukan uji coba untuk melihat persepsi dari mahasiswa dan penilai.

Penelitian yang dilakukan oleh Sumiyarrini, et. al. (2017) sudah mulai menunjukkan upaya pengujian validitas dan reliabilitas instrumen. Uji validitas tergambar dari langkah penelitian poin c yaitu *focus grup discussion* bersama dengan tujuh orang ahli materi. Sementara itu, uji reliabilitas dapat terlihat dari langkah poin d yaitu melakukan *uji content validity* dengan mengukur kesepakatan para ahli. Walaupun tidak disebutkan bahwa langkah ini dilakukan untuk menguji reliabilitas instrumen, namun peneliti dapat menilai bahwa langkah ini juga dapat dinilai sebagai langkah untuk uji reliabilitas. Anggapan ini didasarkan pada pendapat dari Mckeown & Biss (2018) yang mengatakan bahwa reliabilitas instrumen dapat dicapai dengan melakukan *inter rater agreement (IRA)* sebagaimana telah dilakukan pada penelitian ini.

Dalam penelitiannya, Sumiyarrini, et. al. (2017) juga menjelaskan bahwa mereka mengumpulkan data berupa persepsi

dari mahasiswa dan dosen selaku penilai. Selanjutnya, data tersebut diolah secara statistik dengan menggunakan rumus cronbach alpha. Pada poin ini peneliti memberikan kritik terhadap penelitian Sumiyarrini, et. al. (2017). Seharusnya data yang diolah dengan rumus cronbach alpha bukanlah nilai persepsi. Seandainya pengolahan dilakukan terhadap nilai yang didapat dengan instrumen NCEx, maka hasil tersebut dapat dijadikan justifikasi sebagai hasil uji reliabilitas (Arifin, 2011a). Dengan demikian penelitian tersebut akan menjadi lebih utuh karena telah melakukan uji validitas dan uji reliabilitas instrumen.

Selain beberapa penelitian di atas, peneliti juga menemukan beberapa penelitian lain yang juga dapat dijadikan rujukan. Beberapa penelitian tersebut lebih menekankan pada uji reliabilitas rubrik penilaian. Dengan melakukan penelusuran ini, peneliti berharap dapat menemukan kekosongan untuk mengisi *gap teoritik* pada penelitian-penelitian sebelumnya. Hasil penelusuran akan ditambahkan pada tabel 1.4.

Penelitian yang dilakukan oleh Tam (2017) bertujuan untuk menguji tingkat reliabilitas dari ACAR. ACAR merupakan singkatan dari *Art Criticism Assessment Rubric*. Rubrik ini digunakan untuk menilai kemampuan siswa di Hong Kong dalam mengkritisi sebuah karya seni. Respon siswa terhadap karya seni disampaikan melalui sebuah essay. Selanjutnya, guru akan menilai kualitas essay dengan panduan dari ACAR.

Tabel 1.4
Penelitian Pengembangan Rubrik Penilaian dengan Uji Reliabilitas

No	Penulis	Judul Penelitian	Metode	Jurnal / <i>Proceeding</i>
1	Tam (2017)	<i>Evaluating Students ' Performance in</i>	<ul style="list-style-type: none"> • <i>Inter Rater Reliability</i> • <i>Intra Class Correlation</i> 	The International Journal of Art &

		<i>Responding to Art : The Development and Validation of an Art Criticism Assessment Rubric</i>		Design Education Tahun 2017
2	Aulet, Moore, Callas, Nicholas, & Hulme (2018)	<i>(En) trust me : Validating an assessment rubric for documenting clinical encounters during a surgery clerkship clinical skills exam</i>	<ul style="list-style-type: none"> • <i>Inter Rater Reliability</i> • <i>Intra Class Correlation</i> 	The American Journal of Surgery Tahun 2018
3	Sweeney, Rollins, Gantt, Swanson, & Ravitz (2019)	<i>Development and Reliability Testing of the Sweeney-Clark Simulation Evaluation Rubric Ó</i>	<ul style="list-style-type: none"> • <i>Inter Rater Reliability</i> • <i>Intra Class Correlation</i> 	Clinical Simulation in Nursing Tahun 2019

Penelitian tersebut melibatkan 99 orang siswa yang berasal dari delapan sekolah. Siswa yang terlibat berasal dari rentang usia 17-18 tahun. Selanjutnya siswa tersebut diminta untuk menilai dua buah karya seni. Penilaian tersebut selanjutnya dilaporkan dalam bentuk essay dengan pangjang rata-rata sebanyak 800 kata.

Setelah semua essay terkumpul, selanjutnya peneliti melakukan *Inter Rater Reliability (IRR)* dengan cara menyebarkan secara acak seluruh essay yang telah disamarkan identitasnya kepada

enam orang penilai. Keenam penilai dipilih dari *Hong Kong Curriculum Development Institute*. Mereka terdiri dari praktisi dan Widyaiswara yang telah berpengalaman menilai banyak essay serupa. IRR merupakan metode untuk melakukan uji reliabilitas instrumen melalui dua atau lebih *raters* (penilai). Pengujian dilakukan dengan membagi enam penilai ke dalam tiga pasang. Selanjutnya ketiga pasang nilai tersebut diujikan menggunakan *Intra Class Correlation (ICC)*. ICC dilakukan dengan tujuan untuk menghitung korelasi dari dua buah nilai yang berasal dari satu kelas sample yang sama. Pada akhir penelitian dapat disimpulkan bahwa ACAR layak untuk digunakan. Nilai untuk tiga pasangan pengujian masing-masing sebesar 0,91; 0,72; dan 0,69.

Penelitian selanjutnya adalah penelitian yang dilakukan oleh Aulet, et.al (2018). Penelitian tersebut dipublikasikan dalam sebuah artikel ilmiah dengan judul *(En) trust me : Validating an assessment rubric for documenting clinical encounters during a surgery clerkship clinical skills exam*. Penelitian yang melibatkan 57 orang mahasiswa dan dua orang penilai ini bertujuan untuk menguji reliabilitas rubrik dalam menilai kemampuan mahasiswa dalam membuat catatan medis. Nilai ujian yang diperoleh melalui rubrik tersebut selanjutnya diujikan dengan metode ICC. Reliabilitas rubrik penilaian mendapatkan predkat sangat baik dengan nilai ICC sebesar 0,86.

Penelitian yang terakhir adalah penelitian yang dilakukan oleh Sweeney, et. al. (2019). Penelitian ini berjudul *Development and Reliability Testing of the Sweeney-Clark Simulation Evaluation Rubric Ó*. Penelitian ini bertujuan untuk menghasilkan rubrik penilaian yang dapat dipakai untuk menilai kegiatan simulasi mahasiswa keperawatan. Validitas instrumen diujikan dengan metode *face validity (surface validity/validitas permukaan)* dengan melibatkan pihak fakultas, Widyaiswara dan administrator. Selain

uji validitas dilakukan juga uji reliabilitas instrumen. Seperti penelitian sebelumnya, penelitian ini juga melakukan pengujian reliabilitas instrumen. Reliabilitas instrumen diujikan melalui *inter rater reliability (IRR)* dan *intra class correlation (ICC)*.

Pada akhir bagian ini, peneliti menyimpulkan bahwasanya setiap penelitian yang telah dijabarkan di atas ternyata memiliki kelebihan dan kekurangannya masing-masing. Agar dapat menemukan *state of the art*, peneliti mengkategorikan penelitian-penelitian di atas menjadi tiga kategori. Kategori pertama adalah penelitian yang dilakukan dengan merujuk pada model/prosedur pengembangan tertentu. Kategori kedua adalah penelitian yang dilakukan dengan tanpa menyertakan rujukan model/prosedur pengembangan. Dan yang terakhir adalah kategori dengan penelitian yang memfokuskan pada pengujian reliabilitas. Lebih jelasnya, akan disajikan dalam tabel di bawah ini.

Tabel 1.5
Kategorisasi Berdasarkan Kekuatan dan Kelemahan Penelitian

	Kelebihan	Kekurangan
Kategori 1 <ul style="list-style-type: none"> • Fitriani & Yarmayani (2018) • Aji, Hudha, Huda, Nandiyanto, & Abdullah (2018) • Hudha, Aji, & Huda (2018) • Febriyanti, Harsiati, & Dermawan (2017) • Hermawan et al., (2017) 	<ul style="list-style-type: none"> • Penelitian dilakukan dengan model yang memiliki landasan literatur yang jelas. • Melakukan uji validitas instrumen. 	<ul style="list-style-type: none"> • Model yang digunakan terhitung lama dan tidak didesain secara khusus untuk mengembangkan rubrik [kecuali pada penelitian Rakedzon & Tsabari (2017)] • Tidak melakukan uji reliabilitas instrumen.

-
- Rakedzon & Tsabari (2017)
 - Latifa, Rahman, Hamra, Jabu, & Nur (2015)
-

Kategori 2

- Furze, et. al. (2015)
- Sumiyarrini, et. al. (2017)

- Prosedur pengembangan rubrik dijelaskan dengan logis.
- Melakukan uji validitas instrumen.

- Tidak menyertakan sumber rujukan model/prosedur pengembangan
 - Tidak melakukan uji reliabilitas instrumen.
-

Kategori 3

- Tam (2017)
- Aulet, Moore, Callas, Nicholas, & Hulme (2018)
- Sweeney, Rollins, Gantt, Swanson, & Ravitz (2019)

- Melakukan uji validitas instrumen.
- Melakukan uji reliabilitas instrumen.

- Tidak menjelaskan model /prosedur pengembangan rubrik.
-

Dari tabel di atas kita dapat melihat kekuatan dan kelemahan pada masing-masing penelitian. Apabila kita cermati, ternyata kelebihan dan kekurangan dari setiap penelitian dapat saling melengkapi satu sama lain.

Melalui penelitian ini, peneliti berharap dapat menjawab *gap teoritik* yang ada pada penelitian-penelitian sebelumnya. Peneliti berharap dapat menghadirkan *novelty* (kebaharuan) dari sisi metode penelitian. Model/prosedur pengembangan yang akan digunakan

tidak sekedar merujuk pada sumber rujukan yang jelas saja, namun juga sampai pada pengujian validitas dan reliabilitasnya.

1.6. Road Map Penelitian

Road map penelitian merupakan peta jalan dari sebuah proses penelitian. Berikut adalah gambaran *road map* penelitian yang meliputi *input*, proses, *output*, *outcome*, *impact* (dampak) dari penelitian ini.



Gambar 1.1
Road Map Penelitian

Tahapan input meliputi analisis masalah dan studi literatur baik dari segi teori maupun penelitian sebelumnya. Teori yang digunakan diantaranya berasal dari Mckeown dan Biss (2018) Stevens dan Levi (2005), Mislevy dan Knowles (2002), Zainul (2001), serta berbagai teori lainnya. Adapun penelitian sebelumnya yaitu Rakedzon dan Tsabari (2017), Tam (2017), Aulet et al. (2018), Sweeney et al. (2019), dan berbagai penelitian lainnya. Penelusuran penelitian terdahulu pada akhirnya dapat membantu peneliti dalam mencari *state of the art* dan *novelty* pada penelitian ini.

Tahapan selanjutnya adalah proses. Proses merupakan inti dari kegiatan penelitian. Penelitian bertujuan untuk menjawab rumusan masalah yang telah dirumuskan sebelumnya. Pada tahapan ini akan dilakukan pengembangan rubrik untuk menilai kualitas proposal penelitian.

Penelitian ini akan menghasilkan dua buah *output* penelitian. *Output* yang akan dihasilkan adalah produk berupa rubrik penilaian dan laporan penelitian berupa tesis dan publikasi artikel ilmiah.

Selain menghasilkan *output*, penelitian ini juga diharapkan akan melahirkan beberapa *outcome*. Melalui penelitian ini, peneliti berharap dapat memperbaiki proses penilaian proposal ilmiah pada beberapa program pelatihan di Pusbindiklat LIPI. Selain itu peneliti juga berharap dapat memberikan kontribusi pada perkembangan disiplin ilmu teknologi pendidikan, dan menjadi rujukan bagi peneliti lain yang akan melakukan penelitian sejenis. Dengan demikian penelitian ini memiliki kebermanfaatan dalam konteks *science for science*, *science for stakeholders*, dan *science for community*.

Pada akhirnya, peneliti berharap bahwa penelitian ini dapat memberikan *impact* (dampak) yang lebih jauh lagi. Melalui penelitian ini, peneliti berharap dapat memberikan kontribusi bagi dunia penelitian. Peneliti sadar betul bahwa untuk melahirkan dampak tersebut diperlukan kolaborasi dengan berbagai pihak. Oleh karenanya, monitoring dan evaluasi pada penelitian ini hanya dapat dilakukan sampai dengan tahapan *outcome* saja

