

PENGEMBANGAN INSTRUMEN EVALUASI MODEL KIRKPATRICK LEVEL SATU (*REACTION*) TENTANG KEPUASAN LAYANAN PELATIHAN DI PPSDM GEOMINERBA

Harmoni Sofa, Zainal Arifin, dan Riche Cynthia Johan

Departemen Kurikulum dan Teknologi Pendidikan

Fakultas Ilmu Pendidikan

harmonisofa@student.upi.edu, zainalarifin.tp@gmail.com, riche@upi.edu

Abstract. *The application of program evaluation by experts is facilitated by the program evaluation model. As a training center and to control the quality of training carried out, training institutions carry out evaluation of training programs by referring to program evaluation models that have been developed by experts. This encourages the action to gain knowledge and see the implementation of evaluation theory, especially Kirkpatrick's evaluation model, then carried out research at PPSDM Geominerba by developing Kirkpatrick level one evaluation instrument model. Level one or reaction is a level that addresses user satisfaction, so how is the value of training needs met by the institution. The study was conducted using the 6-step Design Development Research (DDR) method. The DDR step starts from identifying problems, describing objectives, product design and development, product testing, trial evaluation, and communicating results is a series of programs that must be passed in developing instruments. The instrument developed was originally a hypothetical (temporary) instrument, which was the development of an analysis of institutions and users, then became a construct with a Likert rating scale as stated by Kirkpatrick. The instrument changes after testing. Instruments with 15 items developed, then re-developed into 20 items with 1 invalid item. The instruments developed are instruments that have been judged by experts and tested on users, the validity of the instruments using expert reviews and user reviews, item validity with CVR theory, construct validity of instruments with factor analysis, and reliability with Cronbach alpha. As a result, the instrument can be used but not yet perfect, the instrument can be used as enhanced research material, and tested until the value is near perfect.*

Keywords: *Program Evaluation, Kirkpatrick Evaluation Model, Design Development Research*

Abstrak. Penerapan evaluasi program oleh para ahli dimudahkan dengan adanya model evaluasi program. Sebagai pusat penyelenggara pelatihan dan untuk mengontrol kualitas pelatihan yang dilaksanakan, lembaga pelatihan melaksanakan evaluasi program pelatihan dengan mengacu kepada model evaluasi program yang telah dikembangkan ahli. Hal tersebut mendorong kepada tindakan untuk memperoleh pengetahuan dan melihat implementasi akan teori evaluasi khususnya model evaluasi Kirkpatrick, maka dilaksanakan penelitian di PPSDM Geominerba dengan mengembangkan instrumen evaluasi model Kirkpatrick level satu. Level satu atau reaksi adalah level yang membahas mengenai kepuasan pengguna, jadi bagaimana nilai akan kebutuhan-kebutuhan pelatihan yang dipenuhi oleh lembaga. Penelitian dilaksanakan menggunakan metode *Design Development Research (DDR)* 6 langkah. Langkah *DDR* dimulai dari identifikasi masalah, menggambarkan tujuan, perancangan dan pengembangan produk, ujicoba produk, evaluasi ujicoba, dan mengkomunikasikan hasil merupakan serangkaian program yang harus dilalui dalam mengembangkan instrumen. Instrumen yang dikembangkan mulanya merupakan instrumen hipotesis (sementara), yang merupakan pengembangan dari analisis terhadap lembaga dan pengguna, kemudian menjadi kisi-kisi dengan skala penilaian likert sebagaimana yang dikemukakan Kirkpatrick. Instrumen mengalami perubahan setelah dilakukan ujicoba. Instrumen dengan 15 item yang dikembangkan, kemudian kembali dikembangkan menjadi 20 item dengan 1 item tidak valid. Instrumen yang dikembangkan adalah instrumen yang telah di *judgement* oleh ahli dan diujicobakan kepada pengguna, validitas instrumen menggunakan *expert review* dan *user review*, validitas item dengan teori CVR, validitas konstruk instrumen dengan analisis faktor, dan reliabilitas dengan Cronbach alpha. Hasilnya, instrumen dapat digunakan namun belum sempurna, instrumen dapat dijadikan bahan penelitian yang disempurnakan, dan diujicobakan sampai nilainya mendekati sempurna.

Kata kunci: *Evaluasi Program Pelatihan, Model Evaluasi Kirkpatrick, Design Development Research*

A. PENDAHULUAN

Indonesia memiliki kondisi pendidikan dengan nilai indikator

development 67,2 (peringkat 53) berdasarkan penilaian *World Economic Forum (WEF)* dimana partisipasi

pendidikan dasar mencapai 92,9; indikator *capacity* didasarkan tingkat buta huruf dan kemampuan berhitung adalah 69,7 (peringkat 64); indikator *deployment* didasari nilai penyerapan SDM memiliki nilai 61,6 (peringkat 82); dalam ketersediaan SDM atau indikator *know-how* Indonesia memiliki peringkat 99, dengan golongan umur produktif 25-54 tahun. Peringkat Indonesia secara umum dalam hasil penilaian WEF mengenai pembangunan SDM dari 130 negara di dunia Indonesia berada pada peringkat 65, dengan tingkat penyerapan SDM dan tingkat kemampuan juga pengetahuan pekerja yang masih rendah. Peringkat umum tersebut juga belum mencerminkan keberhasilan, dimana Indonesia masih dibawah negara Asia Tenggara lainnya seperti Singapura, Malaysia, Thailand dan Filipina.

Pengembangan SDM yang baik bukan hanya pada level pendidikan dasar, melainkan level dewasa juga perlu diperhatikan. Pengembangan SDM level dewasa agar berkualitas di Indonesia dilakukan dengan mengadakan pendidikan dan latihan (diklat) atau pelatihan pada sektor pengembangan SDM. Lembaga penyedia layanan pelatihan tersebut kemudian dikenal juga dengan sebutan Pusat Pengembangan Sumber Daya Manusia (PPSDM). Banyaknya jumlah lembaga pengembangan SDM baik pemerintah atau swasta menjadikan lembaga penyedia layanan tersebut memiliki lembaga pelatihan tersendiri, salah satunya dalam pemerintahan masing-masing kementerian yang bekerja di bawah pemerintah memiliki pusat pengembangan SDM sendiri untuk mengontrol atau menjaga kualitas SDM yang dimiliki. Pusat Pengembangan Sumber Daya Manusia Geologi Mineral dan Batubara (PPSDM Geominerba) salah satu pusat pengembangan SDM dibawah Kementerian Energi dan Sumber Daya Mineral merupakan lembaga pelatihan

untuk meningkatkan kompetensi di bidang geologi, mineral dan batubara.

SDM yang berkualitas terwujud berbanding lurus dengan program pelatihan yang juga berkualitas, SDM yang berkualitas sebagai bentuk kontrol lembaga terhadap kemajuan lembaganya. Penyelenggaraan program menurut Sudjana (2007) akan memulai juga adanya evaluasi program, evaluasi program ini yang akan menghantarkan kepada manajemen kualitas program termasuk program pelatihan. Program pelatihan merupakan sebuah sistem yang membentuk siklus, mulai dari perencanaan, pengembangan, pelaksanaan hingga evaluasi. *Joint Committee on Standar for Educational Evaluation* pada tahun 1981 menyebutkan bahwa evaluasi program menilai aktivitas di bidang pendidikan dengan menyediakan data yang berkelanjutan. Evaluasi program dalam penerapannya oleh para ahli dimudahkan dengan adanya model evaluasi program.

Sebagai pusat penyelenggara pelatihan dan untuk mengontrol kualitas pelatihan yang dilaksanakan PPSDM melaksanakan evaluasi program pelatihan dengan mengacu kepada model evaluasi program yang telah dikembangkan ahli, beragam model evaluasi program mulai dari CIPP, yang biasa diterapkan adalah model evaluasi Kirkpatrick, Brinkerhoff, model *wheel* dan model evaluasi program lainnya. Berdasarkan observasi dan wawancara yang dilakukan pada bulan September-Oktober 2018 di PPSDM Geominerba model evaluasi program yang diterapkan adalah model evaluasi Kirkpatrick atau disebut juga model evaluasi empat level. Model Evaluasi Kirkpatrick banyak digunakan untuk evaluasi program dan dikatakan sederhana, sistematis juga efektif menurut penelitian yang telah dilakukan.

Penelitian yang dilakukan Fildzah Ikramina dan Aurik Gustomo pada tahun 2014 yang melakukan evaluasi Model Kirkpatrick pada PT. Bank Tabungan

Negara (Persero) Tbk. yang menyebutkan bahwa panduan yang dibuat Kirkpatrick merupakan cara terbaik yang dapat dilakukan untuk mengevaluasi, yang telah terbukti sewaktu-waktu menjadi metode evaluasi yang paling berguna yang digunakan oleh banyak perusahaan. Salah satu alasannya adalah karena ada kesenjangan antara desain teori dan desain yang ada di Bank Tabungan Negara pada Kirkpatrick level 3. Penelitian yang dilakukan Shahrooz Farjad pada tahun 2012 dengan judul *The Evaluation Effectiveness of training courses in University by Kirkpatrick Model (case study: Islamshahr university)* menyebutkan bahwa sebagai model evaluasi yang mudah Kirkpatrick juga dikenal dengan model yang mengklasifikasikan hasil evaluasi. Dalam penelitian ini digunakan untuk melihat efektifitas pada keempat levelnya pelatihan di Islamshahr University dan hasilnya pengaruh pada setiap level terlihat signifikan dan akan memberi dampak yang baik pada program.

Penelitian oleh Megan Paull, Craig Whitsed dan Antonia Girardi pada tahun 2016 mengenai penerapan Model Kirkpatrick untuk mengevaluasi kegiatan belajar pada penerapan kurikulum. Hasilnya Model Kirkpatrick adalah cara mengevaluasi penerapan interaksi untuk kerangka pembelajaran telah memberi hasil positif dalam waktu yang efisien bagi pendidik dan siswa, dengan kata lain model evaluasi ini dinilai efisien. Penelitian yang dilakukan oleh Mochammad Handal Widiansyah pada tahun 2016 yang membahas studi deskriptif pelatihan teknis kepala desa yang dilaksanakan oleh Pelatihan Provinsi Jawa Timur Tahun 2014 menggunakan model evaluasi program Kirkpatrick. Evaluasi dilakukan pada masing-masing tahapan dan menghasilkan hal-hal apa saja yang kurang dan perlu diperbaiki, berdasarkan penilaian setiap level tersebut menghasilkan masukan untuk perbaikan

program pelatihan, pelaksanaan pembelajaran dan hasil dari pelatihan.

Evaluasi dalam mengukur dan menghasilkan nilai memerlukan alat yang disebut instrumen evaluasi. Evaluasi Model Kirkpatrick yang terdiri atas empat level memiliki instrumen yang berbeda-beda untuk mengukur setiap levelnya. Berdasarkan observasi dan wawancara yang dilakukan salah satu instrumen yang digunakan untuk mengukur level pertama model Kirkpatrick atau *reaction* adalah dengan instrumen Survei Kepuasan Masyarakat (SKM). Penggunaan SKM di PPSDM Geominerba untuk mengukur level pertama model evaluasi Kirkpatrick ini didasari oleh Peraturan Pemerintah No. 39 Tahun 2006 tentang Tata Cara Pengendalian dan Evaluasi Pelaksanaan Rencana Pembangunan. Adapun acuan yang digunakan oleh PPSDM Geominerba, sebagai unit penyelenggara pelayanan publik dalam melakukan Survei Kepuasan Masyarakat adalah Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia Nomor 14 Tahun 2017 tentang Pedoman Penyusunan Survei Kepuasan Masyarakat Unit Penyelenggara Pelayanan Publik. Dalam Peraturan Menteri tersebut diatur mulai dari unsur SKM, penentuan responden, metode survei, pelaksanaan dan teknik survei, pengolahan data, penyusunan laporan, hingga rencana tindak lanjut.

Karakteristik instrumen evaluasi yang perlu dicerminkan dikemukakan Arifin (2017) harus valid, reliabel, relevan, representatif, praktis, deskriminatif, spesifik dan proporsional. Artinya instrumen harus menghasilkan informasi yang terukur, sesuai dengan apa yang akan diukur, mudah digunakan, mampu menunjukkan perbedaan sekecil apapun (deskriminatif), spesifik dan memiliki tingkatan yang benar. Untuk mendukung kepada pemencapaian karakteristik tersebut agar tercermin dalam instrumen perlu adanya penyusunan kisi-kisi agar

evaluasi yang dilakukan representatif dan relevan.

Observasi dan wawancara yang dilakukan terhadap penyusunan instrumen SKM, SKM dikembangkan PPSDM Geominerba secara langsung diturunkan dari aspek yang ditentukan regulasi kepada pertanyaan, tidak ada penyusunan kisi-kisi sebagaimana dikemukakan seharusnya. Evaluasi yang dihasilkan juga belum mampu menunjukkan perbedaan dimana diambil satu contoh terkait penyelenggaraan, menilai bagaimana penyelenggaraan pelatihan belum spesifik penyelenggaraan yang mana secara teknis (layanan) atau secara kebutuhan barang yang terpenuhi. Hal ini akan menyebabkan hasil evaluasi yang bias dalam penafsirannya, atau dengan kata lain aspek mana dalam penyelenggaraan yang sekiranya kurang masih dianggap bias. Berdasarkan permasalahan tersebut penulis tertarik untuk melakukan penelitian mengenai pengembangan instrumen evaluasi, beberapa pertanyaan penelitian yang beririsan dengan langkah dalam metode yang digunakan.

B. METODE PENELITIAN

Metode dan desain yang digunakan dalam penelitian ini bertujuan untuk pengembangan instrumen evaluasi Model Kirkpatrick Level Satu (*Reaction*) tentang kepuasan layanan pelatihan. Metode penelitian yang digunakan adalah metode penelitian desain dan pengembangan atau disebut juga *Design and Development Research* (DDR). Penelitian menggunakan desain DDR yang dikembangkan oleh Nunamaker dkk. pada tahun 1991 dan dilengkapi oleh Hevner dkk. pada tahun 2004 untuk menghasilkan produk. Penelitian DDR menurut Nunamaker dkk. (1991); dan Hevner dkk. (2004); (dalam Ellis & Levy, 2010, hlm. 111) memiliki enam langkah, yang terdiri atas; (1) *identifikasi masalah* merupakan langkah analisis awal yang berkaitan dengan pelayanan pelatihan bagaimana

pelatihan dikatakan puas dan tidak puas dari isis lembaga dan pengguna; (2) *menggambarkan tujuan* merupakan langkah representasi hasil sejak awal; (3) *perancangan dan pengembangan produk* dilaksanakan menggunakan tahapan Kirkpatrick dalam bukunya implementasi model empat level; (4) *ujicoba lapangan* kepada pengguna untuk melihat nilai validitas dan reliabilitas instrumen hasil pengembangan; (5) *evaluasi hasil ujicoba* untuk melihat kekurangan-kekurangan instrumen; dan (6) *mengkomunikasikan hasil* melihat perubahan sebelum dan sesudah pengembangan. Langkah-langkah tersebut dikatakan sebagai sebuah siklus yang tidak dapat terputus dan merupakan satu kesatuan utuh dalam pelaksanaan metode DDR.

Penelitian dilakukan di PPSDM Geominerba, dengan partisipan penelitian sebagai berikut.

Tabel 1. Partisipan Penelitian

Partisipan	Langkah DDR	Jumlah
Kepala Sub Bidang Evaluasi PPSDM Geominerba	<i>Identify The Problem Motivating The Research</i>	1
Ahli Evaluasi	<i>Design and Develop The Artifact</i>	2
Peserta Pelatihan	<i>Design and Develop The Artifact</i>	17
	<i>Subject the Artifact To Testing</i>	60
Jumlah		80

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah menggunakan beberapa jenis teknik, diantaranya wawancara, studi dokumentasi dan kuesioner dengan

instrument sebagai alat. Instrumen yang dikembangkan akan mengumpulkan data-data yang diperlukan untuk kemudian diolah menjadi informasi. Analisis data penelitian dilakukan dengan menggunakan software SPSS dan Microsoft Excel. Data yang didapatkan dari berbagai teknik pengumpulan data kemudian dianalisis. Analisis data yang didapatkan dilakukan dengan langkah yang dikemukakan Ali (2010, hlm. 329) meliputi: (1) penyusunan, (2) klasifikasi, (3) pemrosesan/pengolahan, dan (4) penafsiran dan penyimpulan.

C. HASIL DAN PEMBAHASAN

1. Identifikasi Masalah

Langkah pertama dalam penelitian DDR diawali dengan identifikasi masalah. Pelaksanaan langkah ini dilakukan dengan mengadakan studi pendahuluan untuk kepentingan pengembangan instrumen. Wawancara kepada kepala sub bidang evaluasi PPSDM Geominerba dan penyebaran kuesioner terbuka kepada peserta. Hasilnya, standar dalam kepuasan layanan pelatihan di PPSDM Geominerba adalah mengacu kepada Standar Pelayanan Publik PPSDM Geominerba, sedangkan untuk pengajar mengacu kepada Standar Pengajar dan Asesor PPSDM Geominerba. Standar-standar ini memiliki tujuan mewujudkan pelayanan prima di lingkungan PPSDM Geominerba. Hal-hal yang membuat peserta puas dengan pelatihan di PPSDM Geominerba adalah jenis diklat yang beragam, widyaiswara yang kompeten, sarana prasarana pelatih yang lengkap, biaya terjangkau, dan masih banyak hal lain yang termasuk pelayanan pelatihan. Hal-hal yang membuat peserta tidak puas dengan pelatihan di PPSDM Geominerba, materi diklat yang kurang up date dan sarana prasarana yang kurang memadai di hotel.

Tabel 2. Hasil Kuesioner Terbuka

Premis	Frekuensi
1	11

2	15
3	14
4	11
5	2

Kuesioner terbuka yang dibagikan kepada peserta pelatihan adalah mengenai layanan yang perlu dipertimbangkan dalam pelatihan, diklasifikasikan ke dalam kode yang merangkum seluruh masukan, yang terdiri atas:

P1 : Konten Pelatihan atau Materi

P2 : Kompetensi Pengajar atau Widyaiswara

P3 : Fasilitas atau Sarana Prasarana

P4 : Pengelolaan Kegiatan Pelatihan

P5 : Biaya Pelatihan

2. Menggambarkan Tujuan

Tujuan dari Design Development Research (DDR) diambil dari hasil identifikasi terhadap masalah, yakni pengembangan instrumen evaluasi model Kirkpatrick level satu mengenai reaksi. Penelitian ini bertujuan untuk mengembangkan kompetensi keilmuan dan juga pengetahuan mengenai evaluasi secara khusus sebagaimana telah dipelajari dibangku perkuliahan. Evaluasi diambil sebagai satu subjek kajian teknologi pendidikan yang berfungsi untuk mengembangkan program, program yang berkaitan dengan penelitian ini adalah program pelatihan. Penelitian juga bermaksud untuk menguji teori dari pengembangan instrumen mengikuti rules yang dibuat oleh pengembang teori yakni Kirkpatrick. Rangkaian penelitian terdiri atas; analisis keadaan instrumen yang digunakan, membuat rancangan pengembangan instrumen, dan menganalisis hasil pengembangan instrumen.

3. Perancangan dan Pengembangan Produk

Pengembangan instrumen dilakukan dengan kaidah pengembangan produk berdasarkan penjelasan langkah DDR dan Kirkpatrick. Membuat kerangka

konseptual, desain dan prototype dari produk yang hendak dikembangkan. Kisi-kisi kemudian dikembangkan menjadi instrumen evaluasi utuh yang terdiri dari judul sampai bagian penutup atau ruang rekomendasi dari peserta bagi lembaga untuk pengembangan program pelatihan. Instrumen yang dihasilkan bersifat sementara atau hipotesis karena belum melalui tryout atau ujicoba tahap 2 oleh expert. Penilaian expert tentang produk yang dikembangkan adalah awal dari pengembangan produk setelah peneliti dan sebelum dilakukan ujicoba lapangan kepada peserta. Penilaian dilakukan kepada tiga expert, yang terdiri atas; (1) Kepala Sub Bidang Evaluasi PPSDM Geominerba, dan (2) Dosen Psikologi FIP UPI yang memiliki konsentrasi dalam pengembangan instrumen serta penilaian layanan sebanyak dua orang. Hasil dari ujicoba terbatas instrumen kepada tiga expert adalah sebagai berikut.

Ujicoba Expert Tahap 1 ini menghasilkan beberapa poin penting yang menjadi acuan untuk memperbaiki instrumen selanjutnya, diantaranya; (1) beberapa item dinilai kurang sesuai dalam pemilihan kata atau redaksi, (2) ada juga item yang masih belum menunjukkan kepada sesuatu yang spesifik untuk diukur, (3) identitas dalam instrumen masih belum ada. Hasil tersebut jika direpresentasikan ke dalam pengambilan keputusan nilai CVR harus lebih besar sama dengan 0,5 ($CVR \geq 0,05$) maka kebanyakan item belum valid dan belum bisa untuk diujicobakan di lapangan. Pertimbangan selain itu pendapat Laswhe (1975) ketika setengah dari expert setuju dengan item tersebut, atau dalam kondisi ini dua orang setuju dan satu tidak maka bisa dipertimbangkan untuk diujicobakan di lapangan. Perbaikan dilaksanakan dengan ujicoba kedua.

Ujicoba Expert Tahap 2 memberikan 95 % hasil cocok terhadap 20 soal yang dikembangkan. Hasil ujicoba tahap 2 adalah instrumen yang kemudian akan

melalui ujicoba lapangan. Desain instrumen setelah ujicoba tahap 2 mengalami perubahan dari kisi-kisi sampai desain berdasarkan pertimbangan dan rekomendasi beberapa ahli. Perubahan desain ini didasari dari hasil penilaian expert terhadap produk yang dikembangkan, hal-hal yang berubah dari desain awal diantaranya: (1) bagian identitas pengisi kuesioner, (2) perubahan redaksi, dan (3) jumlah item soal bertambah untuk menambah spesifikasi penilaian. Pasca memperoleh penilaian terhadap butir item soal, penilaian yang juga diperoleh dari expert adalah penilaian validitas permukaan. Validitas permukaan yakni hasil tanggapan expert atas beberapa aspek yang dinilai dari instrumen reaction-sheet menggunakan skala likert.

Selain uji item soal juga ada *expert review*, atau penilaian ahli berdasarkan beberapa aspek menggunakan skala likert. Hasil lembar penilaian oleh para ahli, keterbacaan untuk jenis huruf dan ukuran dalam yang digunakan instrumen dengan representasi kelayakan 100%. Sangat layak untuk tata letak konten dengan representasi kelayakan 87%. Kisi-kisi instrumen dan indikator yang dikembangkan memiliki persentase kelayakan 87%, artinya sangat layak. Kelayakan komponen instrumen dengan persentase 93%, artinya sangat layak. Ketepatan bahasa yang digunakan pada setiap item dengan persentase nilai 93%, artinya sangat layak. Kesesuaian penggunaan skala dengan persentase nilai 93%, artinya sangat layak. Kemudahan mengisi dengan persentase 100%, artinya sangat layak. Representasi isi instrumen dengan persentase nilai 93%, artinya sangat layak. Sangat cukup waktu untuk mengisi dengan representasi kelayakan waktu 93%, artinya sangat layak.

4. Ujicoba Lapangan Produk yang Dikembangkan

Produk diujicobakan kepada peserta latihan non-sertifikasi di PPSDM

Geominerba. Rincian respondennya sebagai berikut.

Tabel 3. Persebaran Responden

No	Nama Pelatihan	Jumlah Peserta (orang)
1	Resolusi Konflik Pertambangan	12
2	Sistem Manajemen Keselamatan Pertambangan	14
3	Evaluasi Studi Kelayakan Pertambangan	13
4	Juru Ledak Kelas I	17
Jumlah Responden		60

Ujicoba dilakukan pada 60 orang peserta pelatihan non-sertifikasi, dengan pelaksanaan lapangan terdiri dari dua kegiatan; (1) penyebaran reaction-sheet untuk diisi peserta, dan (2) penyebaran kuesioner user review untuk menilai instrumen yang dikembangkan. Ujicoba lapangan diawali dengan pembukaan dan pengenalan peneliti kepada peserta pelatihan, kemudian penjelasan mengenai instrumen yang sudah dibagikan di meja peserta, selanjutnya pemaparan maksud dan tujuan pengumpulan data ini, dan terakhir meminta respon dari peserta untuk mengisi instrumen. Rangkaian penelitian tersebut dilakukan pada empat kelas pelatihan yang menjadi sampel.

Validitas konstruk dilakukan melalui pengujian analisis faktor terhadap instrumen yang dikembangkan. Validitas diuji dengan menggunakan rumus korelasi terhadap faktor-faktor dengan total dari setiap item, menggunakan Pearson Product Moment. Faktor dari instrumen yang dikembangkan terdiri atas lima faktor yang diturunkan dari dimensi, yakni; (1) penyelenggaraan pelatihan, (2) konten pelatihan, (3) instruktur yang kompeten, (4) fasilitas yang membantu pelatihan, dan (5) perubahan kompetensi

terhadap peserta. Hasil korelasi setiap faktor digambarkan sebagai berikut.

Tabel 4. Validitas Konstruk

Faktor (X)	Jumlah (Y) Total (Y1)	Hasil Korelasi
Penyelenggaraan pelatihan (X1)	X1Y1	0,75
Konten pelatihan (X2)	X2Y1	0,74
Instruktur yang kompeten (X3)	X3Y1	0,77
Fasilitas yang membantu pelatihan (X4)	X4Y1	0,78
Perubahan kompetensi terhadap peserta (X5)	X5Y1	0,72

Hasil dari validitas menggunakan analisis faktor adalah setiap faktor atau aspek dinyatakan valid, karena lebih dari 0,3 atau nilainya 0,3 keatas. Karena koefisien korelasi kelima faktor tersebut diatas 0,3 maka penyelenggaraan pelatihan, konten pelatihan, instruktur yang kompeten, fasilitas yang membantu pelatihan, perubahan kompetensi terhadap peserta merupakan konstruksi (construct) yang valid untuk layanan pelatihan.

Reliabilitas dari produk yang dikembangkan diukur dengan rumus Cronbach Alpha.

Tabel 5. Nilai Realibilitas Koef. Alpha

Cronbach's Alpha	N of Items
.918	20

Hasilnya nilai reliabilitas produk yang dikembangkan adalah 0.918 berarti berada pada rentang 0.80 s.d. 1.000, produk dikatakan memiliki reliabilitas sangat kuat.

Validitas Permukaan User Review

Hasil lembar penilaian oleh para peserta pelatihan, sebanyak 60 sampel. Pengolahan data instrumen menggunakan teknis analisis data dari rumus guttman di dapatkan hasil 97%. Lebih besar 3% dari

penilaian expert. Hasil 97% ketika di konversikan ke dalam kriteria berada pada rentang persentase 80 – 100, yang artinya Sangat Layak. Penilaian user berdasarkan beberapa aspek yang dinilai memiliki hasil akhir, sangat layak terhadap produk yang dikembangkan.

Tahap pengembangan instrumen yang sudah dilaksanakan yakni menyusun item pertanyaan, membuat item memiliki bentuk feedback menggunakan skala likert, penambahan ruang untuk rekomendasi dari peserta terkait pelayanan pelatihan yang didapatkan, tidak ada kolom nama, dan pengisian yang dianjurkan memperoleh 100% respon peserta pelatihan telah dilakukan dalam ujicoba. Ujicoba lapangan reaction-sheet memperoleh 100% respon. Respon dari keseluruhan jumlah responden yakni 60 orang, maka seluruh tahapan yang dikemukakan Kirkpatrick dalam bukunya telah terlaksanakan.

5. Evaluasi Hasil Ujicoba

Kepuasan layanan diukur melalui reaction-sheet memperoleh hasil per-item dari nilai mean (rata-rata) dan modus (nilai yang sering muncul) yakni sebagai berikut. Hasil dari reaction-sheet 100% jawaban peserta pelatihan memberikan tanggapan setuju terhadap pernyataan yang diajukan dan tingkat kepuasan dari pelatihan yang diadakan secara dominan adalah baik. Evaluasi terhadap hasil ujicoba produk dari dua pelaksanaan ujicoba melalui expert dan user. Hasilnya didapatkan dari catatan yang dituliskan oleh ahli dan user, terdiri atas: (1) perlu adanya spesifikasi terhadap item fasilitas diklat, fasilitas diklat bias ditambah dengan menanyakan kondisi ruangan, kondisi seminar kit, konsumsi, dan hal pribadi lain yang didapatkan, dan (2) perlu adanya garis bantu pembatas setiap item soal.

6. Mengkomunikasikan Hasil

Hasil dari keseluruhan tahapan penelitian desain dan pengembangan terhadap pengembangan instrumen

Kirkpatrick level satu (reaction) memberikan hasil akhir instrumen yang dikelompokan berdasarkan aspek instrumen antara instrumen yang diterapkan oleh PPSDM Geominerba dan instrumen yang dikembangkan sebagai produk penelitian, hasil dari ujicoba tersebut diperinci sebagai berikut.

Tabel 6. Hasil Akhir

Kondisi	Instrumen yang Diterapkan	Instrumen yang Dikembangkan
Skala yang Digunakan	Penggunaan skala dalam instrumen sebelumnya, atau yang digunakan dilapangan dikatakan kurang representatif. Alasannya, skala yang digunakan dalam satu instrumen variatif sesuai dengan fokus item pertanyaan. Sehingga, satu instrumen memiliki makna yang berbeda dalam tiap interval respon yang diisi.	Penggunaan skala memiliki satu makna yang sama dalam satu instrumen seperti teori dari Kirkpatrick. Skala yang digunakan adalah skala likert dengan keterangan; sangat tidak setuju, tidak setuju, netral, dan sangat setuju.
Tata bahasa	Ada dua aspek yang berbeda makna dalam satu item,	Sebelumnya, memiliki hal yang sama dengan

	<p>sehingga akan berakibat pada representasi hasil.</p>	<p>kondisi instrumen di lapangan. Setelah dinilai dan diujicobakan beberapa hal yang masih kurang pada instrumen multi diperbaiki. Sehingga instrumen lebih relevan dan representatif. Yang dibuktikan dengan hasil validitas konstruk oleh expert.</p>			<p>Kirkpatrick (2007) pada bukunya menyebutkan an untuk mendapatkan penilaian reaksi yang maksimal diakhir form isian ditambahkan ruang untuk komentar terhadap pelatihan untuk pengembangan program selanjutnya .</p>
<p>Kelengkapan instrumen</p>	<p>Kelengkapan dinilai lengkap sebagaimana teori sistematika instrumen yang terdiri atas; identitas, perintah pengisian, item soal, dan catatan untuk menambah komentar terhadap pelatihan.</p>	<p>Kelengkapan dinilai lengkap sebagaimana teori sistematika instrumen yang terdiri atas; identitas, perintah pengisian, item soal, dan catatan untuk menambah komentar terhadap pelatihan. Sebagaimana</p>		<p>Desain instrumen</p>	<p>Desain instrumen simple dan membutuhkan waktu yang sebentar untuk mengisi, namun hasil analisisnya tidak dapat merepresentasikan secara menyeluruh karena hasil per-item akan berbeda. Desain instrumen pasca ujicoba juga mengalami perubahan dari desain sebelum ujicoba, yakni penambahan garis bantu pembatas untuk setiap item soal. Desain instrumen akhir didapatkan setelah</p>

dilakukan
ujicoba
kepada 3
ahli dan 60
sampel
peserta
pelatihan.

Representasi pembasahan berdasarkan hasil yang diperoleh di lapangan adalah sebagai berikut.

A. Instrumen Evaluasi yang Digunakan oleh Pusat Pengembangan Sumber Daya Manusia Geologi Mineral dan Batubara (PPSDM Geominerba) untuk Mengukur Kepuasan Layanan Pendidikan dan Latihan

Instrumen yang digunakan disebut instrumen Survei Kepuasan Masyarakat (SKM). Instrumen yang ini memiliki struktur dengan kondisi skala instrumen yang berbeda-beda disesuaikan dengan fokus pertanyaan item. Jika kondisi skala seperti ini maka akan menghasilkan nilai yang tidak representatif untuk keseluruhan reaksi, dan keseluruhan kepuasan layanan. Hasilnya satu instrumen memiliki makna yang berbeda dalam tiap interval respon yang diisi. Ada dua aspek yang berbeda makna dalam satu item pertanyaan, sehingga akan berakibat pada representasi hasil. Desain instrumen sederhana dan membutuhkan waktu yang sebentar untuk mengisi, namun hasil analisisnya tidak dapat merepresentasikan secara menyeluruh karena hasil per-item akan berbeda.

Intrumen ini dapat dikatakan belum representatif sebagaimana ciri instrumen harus relevan dan representatif menurut Ihsan (2016). Instrumen ini juga ketika ditelaah ke langkah sebelum menjadi instrumen yakni pada kisi-kisi instrumen, instrumen belum memiliki kisi-kisi, padahal sebagaimana instrumen harus memiliki kisi-kisi sebagai blue print sebagai kontruk instrumen. Hal-hal ini

yang menjadi pertimbangan dan perbaikan dalam pengembangan produk penelitian, sehingga tujuan penelitian berdasarkan analisis yang dilakukan adalah pengembangan instrumen evaluasi Kirkpatrick level satu tentang reaksi.

Tujuan sebagai langkah DDR selanjutnya, dilakukan dengan menyusun strategi tentang pengembangan produk. Tujuan berdasarkan pendapat Ellis dan Levy (dalam Ellis dan Levy, 2010, hlm. 112) penelitian adalah pengembangan instrumen evaluasi model Kirkpatrick level satu (reaction) tentang kepuasan layanan. Produk dikembangkan untuk menguji teori dan melihat nilai dari produk yang dikembangkan yakni instrumen evaluasi model kirkpatrick level satu tentang kepuasan layanan di PPSDM Geominerba. Sasaran pengguna instrumen adalah peserta pelatihan yang merasakan langsung layanan dari PPSDM Geominerba.

B. Rancangan Pengembangan Instrumen Evaluasi Model Kirkpatrick Level Satu (Reaction) di Pusat Pengembangan Sumber Daya Manusia Geologi Mineral dan Batubara (PPSDM Geominerba)

Instrumen yang dihasilkan harus merepresentasikan layanan pelatihan sebagaimana Kirkpatrick (2007) bahwa dalam level satu hal yang diukur adalah reaksi atau customer satisfaction yakni kepuasan layanan terhadap pelatihan. Pengukuran kepuasan layanan pelatihan dirancang dengan menggunakan kisi-kisi dan desain awal sebagai hipotesis. Hasan (2003) Nunamaker et. Al (1992) (dalam Ellis dan Levy, 2010) menyebutkan bahwa dalam proses pengembangan produk DDR perlu untuk membuat kerangka konseptual, desain dan prototype dari produk yang hendak dikembangkan.

Kerangka konseptual, desain dan prototype dikerjakan dengan membuat kisi-kisi untuk instrumen yang akan dikembangkan sampai menjadi instrumen awal yang akan melalui tahapan tryout dan

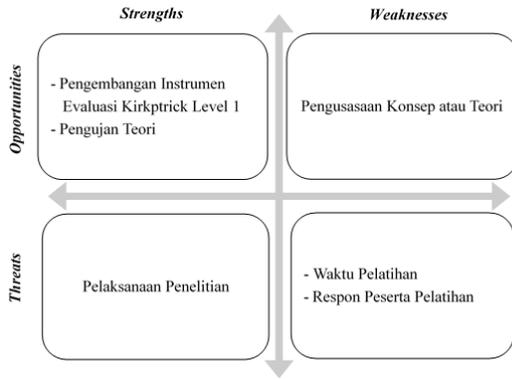
ujicoba lapangan kepada peserta pelatihan. Proses pengembangan produk mengacu kepada teori pengembangan instrumen menurut Kirkpatrick (2007, hlm. 25) memaparkan tahapan dalam pengembangan instrumen reaksi, yakni terdiri atas; (1) membuat daftar item atau soal yang membutuhkan feedback atau jawaban, (2) desain instrumen reaksi memiliki hasil yang bisa dihitung, menggunakan 5 poin skala (skala likert), (3) akhiri instrumen dengan bagian rekomendasi yang diisi oleh peserta untuk pengembangan program, (4) jangan membiarkan peserta mencantumkan nama, dan (5) berusaha untuk mendapatkan seluruh respon dari peserta pelatihan.

Instrumen yang dihasilkan dinamakan Reaction-Sheet sebagaimana Kirkpatrick menyebutnya. Proses pengembangan yang pertama adalah membuat daftar item reaksi, yang didapatkan dari identifikasi masalah melalui kuesioner terbuka dan wawancara secara langsung kepada Kepala Sub Bidang Evaluasi PPSDM Geominerba. Item yang dihasilkan terdiri atas hal-hal yang perlu dipertimbangkan untuk mengetahui reaksi peserta, yakni; (1) konten pelatihan atau materi, (2) kompetensi pengajar atau widyaiswara, (3) fasilitas atau sarana prasarana, (4) pengelolaan kegiatan pelatihan, dan (5) biaya pelatihan. Item-item ini yang kemudian menjadi aspek dalam dimensi penilaian kepuasan pelayanan pelatihan yang diuraikan dalam bentuk kisi-kisi awal. Tahap pengembangan instrumen yang sudah dilaksanakan yakni menyusun item pertanyaan, membuat item memiliki bentuk feedback menggunakan skala likert, penambahan ruang untuk rekomendasi dari peserta terkait pelayanan pelatihan yang didapatkan, tidak ada kolom nama, dan selanjutnya pengisian yang dianjurkan memperoleh 100% respon peserta pelatihan akan dilakukan dalam ujicoba.

C. Hasil Penerapan Instrumen Evaluasi Model Kirkpatrick Level Satu (Reaction) yang Dikembangkan Di Pusat Pengembangan Sumber Daya Manusia Geologi Mineral Dan Batubara (PPSDM Geominerba)

Penerapan atau ujicoba lapangan instrumen kepada peserta pelatihan dilaksanakan setelah pelatihan akan berakhir, tepatnya satu hari sebelum penutupan pelatihan. Sebagaimana Widoyoko (2017) menyebutkan bahwa waktu yang paling tepat dalam melaksanakan evaluasi Kirkpatrick level satu adalah sesaat setelah pelatihan berakhir atau sesaat sebelum berakhir. Instrumen dibagikan kepada peserta di kelas pelatihan masing-masing, dengan memaparkan maksud dan tujuan dari penyebaran instrumen. Waktu yang diberikan untuk mengisi tidak bertepatan dengan berakhirnya pemaparan hal-hal terkait instrumen, dikarenakan peserta masih memiliki kegiatan yang terbilang padat dari pagi sampai sore di satu hari sebelum penutupan sehingga waktu pengisian yang diberikan adalah dari pagi sejak dibagikan sampai sore harinya. Hal-hal yang dipaparkan adalah mengenai instrumen yang akan diisi, dari mulai identitas sampai kolom komentar, termasuk penggunaan skala dan pertanyaan soal.

Hasilnya, dari sejumlah instrumen yang dibagikan, masih terdapat beberapa peserta yang tidak mengisi. Hal tersebut menunjukkan bahwa respon peserta berada diluar kendali peneliti dan tidak dapat dipaksakan. Sebagaimana analisis terhadap sistem evaluasi yang dilaksanakan menggunakan SWOT selanjutnya. Untuk memperoleh gambaran yang komprehensif penulis akan menjelaskan dengan analisis SWOT yang dilaksanakan terhadap perancangan pengembangan instrumen evaluasi Kirkpatrick level satu mengenai kepuasan layanan pelatihan.



Strategi tentang pengembangan produk dari analisis SWOT yang dilakukan, tujuan berdasarkan pendapat Ellis dan Levy (dalam Ellis dan Levy, 2010, hlm. 112) penelitian adalah pengembangan instrumen evaluasi model Kirkpatrick level satu (reaction) tentang kepuasan layanan. Produk dikembangkan untuk menguji teori dan melihat nilai dari produk yang dikembangkan yakni instrumen evaluasi model kirkpatrick level satu tentang kepuasan layanan di PPSDM Geominerba. Sasaran pengguna instrumen adalah peserta pelatihan yang merasakan langsung layanan dari PPSDM Geominerba.

Analisis memberikan gambaran kelemahan dan ancaman dari penelitian, yakni; (1) waktu penelitian dan (2) respon peserta pelatihan. Waktu penelitian dikatakan sebagai kelemahan dan ancaman karena tidak dapat di setting oleh peneliti, termasuk didalamnya jumlah peserta dengan insidental sampel namun purposive. Respon peserta pelatihan menjadi kelemahan dan ancaman karena tidak dapat dikontrol dan mengikuti kehendak peneliti, jadi bisa saja pengembangan instrumen ini memiliki nilai yang sangat kecil.

Penelitian yang telah dilaksanakan diakhir kegiatan pelatihan memberikan hasil yang sesuai dengan analisis SWOT tentang perancangan pengembangan

instrumen. Penelitian dapat terhambat oleh waktu pelatihan dan respon peserta, dari sejumlah respon yang dihasilkan menghasilkan banyak respon yang tidak dapat digunakan karena responden hanya mengisi sebagian instrumen atau bahkan tidak mengisi. Hasilnya hanya diperoleh instrumen yang telah lengkap diisi.

Instrumen yang dihasilkan dari penelitian ini harus memenuhi prinsip instrumen yang baik. Instrumen harus valid dan reliabel, maka perlu diketahui ini **Gambar 1.** Analisis SWOT ini re. Pengembangan Produk yang dihasilkan. Hasil didapatkan berdasarkan ujicoba lapangan kepada responden berasal dari empat pelatihan non-sertifikasi di PPSDM Geominerba. Hasil yang didapatkan dari pengisian reaction-sheet oleh peserta terdiri atas; validitas konstruk dan reliabilitas produk, dan hasil lain yang digunakan sebagai nilai dari instrumen yang dikembangkan adalah validitas butir item dengan CVR dan validitas permukaan dari expert dan user. Selanjutnya, akan dibahas mengenai hasil instrumen dari penerapan kepada expert dan user di lapangan.

Tryout dari penilaian expert secara umum memperoleh hasil penilaian direpresentasikan per-item sebagian besar menilai sangat layak. Item dikatakan layak berdasarkan pendapat Laswhe (dalam Hendryadi, 2017) ketika nilainya lebih dari, 0,5 dan atau setengah dari expert setuju dengan item tersebut, atau dalam kondisi ini dua orang setuju dan satu tidak maka bisa dipertimbangkan untuk diujicobakan di lapangan dan nilai CVR yang dihasilkan lebih dari yang ditentukan. Rancangan pengembangan yang dihasilkan adalah instrumen dengan item pertanyaan yang valid. Hasil dari uji expert menggunakan CVR dari Laswhe adalah dari 15 item yang diajukan diawal masing dominan item yang kurang valid. Selanjutnya, dilakukan kembali penilaian expert mengenai item, hasilnya dari 15 item yang dikembangkan kemudian

menjadi 20. Hasil CVR dari 20 item satu item memperoleh hasil tidak layak dan diambil keputusan untuk dihapus.

Hasil lain dari penerapan instrumen adalah validitas permukaan oleh expert. Item-item penilaian adalah validitas permukaan yang terdiri dari keterbacaan huruf, kondisi ruang dalam instrumen, konstruk penurunan indikator menjadi item pertanyaan, fisibilitas dalam pengisian kuesioner dan komponen-komponen instrumen. Expert review ini menjadi hasil validitas permukaan untuk produk yang dikembangkan, hasilnya sangat layak. Rinciannya, sangat layak untuk tata letak konten. Kisi-kisi instrumen dan indikator yang dikembangkan memiliki nilai sangat layak. Kelayakan komponen instrumen dengan nilai sangat layak. Ketepatan bahasa yang digunakan pada setiap item nilai sangat layak. Kesesuaian penggunaan skala dengan nilai sangat layak. Kemudahan mengisi dengan nilai sangat layak. Representasi isi instrumen dengan nilai sangat layak. Sangat cukup waktu untuk mengisi dengan representasi kelayakan waktu sangat layak.

Validitas konstruk dilakukan dengan analisis faktor. Analisis faktor mengkorelasikan antara faktor-faktor yang diturunkan dari dimensi dengan skor total. Menurut Arifin (2014) faktor yang digunakan merupakan dimensi atau indikator dari variabel yang diukur. Hasilnya keseluruhan faktor yang ada dikatakan "valid". Selain validitas, hasil dari reaction-sheet adalah mengukur reliabilitas atau keajegan instrumen menggunakan cronbach-alpha, hasilnya instrumen "reliabel". Hasil keseluruhan menilai instrumen yang dikembangkan telah valid dan reliabel.

Instrumen juga dinilai secara permukaan oleh user atau pengguna. Penilaian user berdasarkan beberapa aspek yang dinilai memiliki hasil akhir, sangat layak terhadap produk yang

dikembangkan dan fisible untuk digunakan. Rinciannya, sangat layak untuk tata letak konten. Ketepatan bahasa yang digunakan pada setiap item nilainya layak. Kesesuaian penggunaan skala dengan nilai layak. Kemudahan mengisi dengan nilai sangat layak. Representasi isi instrumen dengan nilai layak. Sangat cukup waktu untuk mengisi dengan nilai sangat layak.

Hasil lain dari instrumen adalah tentang kepuasan layanan dari peserta adalah dominan peserta mengatakan setuju bahwa pelayanan yang diberikan baik. Hasil tersebut sejalan dengan pendapat Widoyoko (2017) selain untuk mengevaluasi proses pelatihan juga untuk melihat dampak langsung motivasi dan semangat peserta pelatihan. Sejumlah pertanyaan yang diajukan dalam reaction-sheet bahwa peserta setuju pelatihan memberikan perubahan kepada kompetensi peserta, dan tanggapan positif bahwa pelatihan bermanfaat untuk peserta.

Silberman (2006, hlm. 318-319) menyebutkan bahwa kualitas pelatihan dapat diukur melalui tingkat kepuasan pesertanya. Hasil dari pengembangan instrumen evaluasi Kirkpatrick level satu tentang kepuasan layanan pelatihan di PPSDM Geominerba selain menghasilkan instrumen yang valid dan reliabel juga memberikan hasil pengukuran mengenai kepuasan layanan yang diberikan. Hasil menunjukkan kepuasan yang baik dari peserta pelatihan terhadap layanan. Secara umum, hasil dari pengembangan instrumen adalah instrumen yang dikembangkan telah memiliki dasar yakni; (1) relevan dan representatif, (2) valid dan reliabel, serta (3) pengukuran kepuasan layanan pelatihan yang fisible.

Selain itu, dari hasil penelitian memperoleh juga kekurangan dari instrumen yang dikembangkan adalah uji expert yang minimum, hasil CVR yang kurang variatif dan dapat memunculkan keragu-raguan. Waktu ujicoba dilakukan

satu kali kepada user, sehingga hasilnya hanya merepresentasikan satu kali uji. Kekurangan lainnya masih ada item yang bisa dijadikan lebih spesifik untuk penilaiannya. Hasil dari kekurangan yang didapatkan bisa menjadi rekomendasi terhadap penelitian selanjutnya dan instrumen yang dikembangkan bisa digunakan untuk mengukur kepuasan layanan pelatihan di PPSDM Geominerba. Hasil penelitian ini sejalan dengan penelitian dari Shahrooz Farjad (2012) menyebutkan bahwa sebagai model evaluasi yang mudah Kirkpatrick juga dikenal dengan model yang mengklasifikasikan hasil evaluasi. Hasilnya pengaruh pada setiap level terlihat signifikan dan akan memberi dampak yang baik pada program. Selanjutnya, penelitian oleh John M. LaVelle dan Stewart I. Donaldson (2010), mengevaluasi program pelatihan yang ada di Amerika Serikat. Hasilnya sangat penting untuk memahami sifat program pelatihan agar mengetahui alternatif evaluasi yang sesuai dan akan diterapkan, dan penggunaan model evaluasi Kirkpatrick mengarahkan kepada terlihatnya konsistensi pelatihan pada hasil evaluasinya.

Kedua penelitian tersebut memperkuat bahwa hasil dari penelitian evaluasi Kirkpatrick dapat memperoleh hasil yang signifikan bilamana dilakukan dua kali uji. Untuk memperoleh hasil yang signifikan selain dilakukan dua kali uji lapangan, dilakukan juga uji lapangan kepada beberapa pelatihan terdapat hasil yang signifikan serupa dan memandang pelatihan sebagai hal yang baik. Hasil tersebut dilihat dari reaksi terhadap setiap aspek yang dinilai dalam instrumen reaction-sheet.

D. KESIMPULAN

Analisis dari pelaksanaan lapangan mengenai Pengembangan Instrumen Evaluasi Model Kirkpatrick Level Satu

(*Reaction*) menghasilkan kristalisasi dari hasil dan pembahasan sebagai berikut.

1. Hasil analisis terhadap instrumen yang digunakan di PPSDM Geominerba adalah instrumen belum representatif dalam mengukur kepuasan layanan pelatihan, sedangkan pada prinsipnya instrumen evaluasi harus representatif dan relevan untuk bisa mengukur hal-hal yang dievaluasi. Instrumen yang digunakan memiliki skala pengukuran yang berbeda-beda sehingga untuk diambil representasi hasil terhadap dimensi yang diukur tidak akan menghasilkan nilai yang representatif sama untuk semua item. Jadi, terdapat kesenjangan antara desain instrumen dengan desain yang seharusnya digunakan secara teoritis.
2. Merancang dan mengembangkan merupakan kegiatan inti penelitian, yang dilaksanakan berdasarkan prosedur Hevner desain DDR 6 langkah, pada kegiatan merancang dan mengembangkan beririsan dengan langkah *design and development of product* dan *testing of product*. Pengembangan instrumen menggunakan langkah yang dikembangkan oleh Kirkpatrick, yang terdiri dari penyusunan item, penyusunan kerangka konseptual, pembuatan prototype sampai mendapatkan keseluruhan respon peserta. Pengembangan instrumen diawali dengan instrumen sementara (hipotesis) yang dikembangkan oleh peneliti, untuk kemudian diujicobakan kepada *expert* dan *user*. Hasil dari pengembangan produk menjadi semakin kuat dengan adanya penilaian *expert* dan *user* secara komprehensif. Pengembangan produk melalui tahapan yang cukup panjang, dari pengembangan menghasilkan produk hipotesis oleh peneliti, *tryout* oleh *expert* yang hasilnya menambah jumlah item dan mempertimbangkan redaksi dalam item soal, dan ujicoba

lapangan oleh *user* untuk melihat validitas konstruk dan reliabilitas terhadap instrumen yang dikembangkan.

- Hasil dari penerapan produk yang telah dikembangkan berdasarkan karakteristiknya dari mulanya memiliki 15 item soal setelah melalui *tryout* oleh *expert* item soal bertambah menjadi 19 soal dengan pertimbangan instrumen harus *observable* dan *measurable*. Kualitas instrumen diuji dengan berbagai uji validitas, yang terdiri atas; validitas permukaan *user* dan *expert review*, validitas item soal dengan menggabungkan teori CVR dari Laswhe, dan validitas konstruk dengan analisis faktor. Kualitas instrumen juga dinilai reliabilitas hasilnya dengan *cornbach alpha*. Instrumen memiliki validitas yang dominan **valid** dan reliabilitas yang **reliabel** dengan kriteria sangat kuat. Hasilnya, instrumen yang dikembangkan layak digunakan untuk kepuasan layanan pelatihan di PPSDM Geominerba, ditambah instrumen sudah relevan dan representatif, valid dan reliabel, serta pengukuran kepuasan layanan pelatihan yang *fisible* untuk mengukur reaksi.

Penerapan evaluasi model Kirkpatrick pada program pelatihan memberikan hasil representatif dan mudah untuk dilaksanakan sebagaimana banyak penelitian yang telah dilaksanakan.

E. DAFTAR PUSTAKA

- Ali, M. (2010). *Metodologi dan aplikasi riset pendidikan*. Bandung: Pustaka Cendekia Utama.
- Arifin, Z. (2014). *Penelitian Pendidikan: metode dan paradigma baru*. Bandung: PT Remaja Rosdakarya.
- Arifin, Z. (2017). *Evaluasi Pembelajaran: Prinsip-Teknik-Prosedur, Cetakan ke-10*. Bandung: PT. Remaja Rosdakarya.
- Ellis, E.T dan Levy, Y. (2010). *A Guide for Novice Research Design and Development Reseach Methods*. Dalam *Proceedings Informing Science & IT Education Conference (InSITE)*. Florida: Nova Southeastern University.
- Farjad, S. (2012). The Evaluation Effectiveness of training courses in University by Kirkpatrick Model (case study: Islamshahr university). *Social and Behavioral Science*, 46, hlm. 2837-2841.
- Ihsan, H. (2016). Validitas Isi Alat Ukur Penelitian Konsep dan Panduan Penilainnya. *Pedagogia*, hlm. 266-273.
- Ikramina, F. dan Gustomo, A. (2014). Analysis Of Training Evaluation Process Using Kirkpatrick's Training Evaluation Model At PT. Bank Tabungan Negara (Persero) Tbk. *Journal of Business and Management*, 3 (1), hl. 102-111.
- Joint Committee on Standar for Educational Evaluation. (1981). *Standars for Evaluations of Educational Program, Project, and Mterials*. New York: McGraw-Hill.
- Kirkpatrick, D. L. dan Kirkpatrick, J. D. (2007). *Implementing the Four Levels: A Practical Guide for Effective Evaluation of Training Programs*. San Fransisco: Barrett-Koehler Pubilsher, Inc.
- LaVelle, M. dan Donaldson, S. I. (2010). University-Based Evaluation Training Programs in the United States 1980–2008: An Empirical Examination. *American Journal of Evaluation*, 31 (1) hlm. 9-23.
- Paull, M., Whitsed, Cr. dan Girardi, A. (2016). Applying The Kirkpatrick Model: Evaluating An Interaction For Learning Framework Curriculum Intervention. *Issues in Educational Research*, 26 (3), hlm. 490-507.
- Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia Nomor 14 Tahun 2017
- Peraturan Pemerintah No. 39 Tahun 2006

- Richey, R. C. & Klein, J. D. (2009). *Design and Development Methods, Strategies, and Issues*. New York: Routledge.
- Silberman, M. L. (2006). *Active Training: A Handbook of Technique, Designs, Case Examples, and Tips*. San Fransisco: Pfeiffer an Inprint of Wiley.
- Sudjana, D. (2007). *Sistem & Manajemen Pelatihan: Teori & Aplikasi*. Bandung: Falah Production.
- Sukmadinata, N. S. (2016). *Metode Penelitian Pendidikan. Cetakan ke-11*. Bandung: PT. Remaja Rosdakarya.
- Universitas Pendidikan Indonesia. (2018). *Pedoman Penulisan Karya Ilmiah*. Bandung: UPI.
- Widiansyah, M. H. (2016). Studi Deskriptif tentang Evaluasi Tahapan Hasil Pendidikan dan Pelatihan Teknis Kepala Desa oleh Badan Diklat Provinsi Jawa Timur Tahun 2014. *Kebijakan dan Manajemen Publik*, 4 (2), hlm. 126-134.
- Widoyoko, S. E. P. (2017). *Evaluasi Program Pelatihan*. Yogyakarta: Pustaka Pelajar.
- World Economic Forum. (2017). *The Human Global Capital Report 2017*. [Online]. Diakses dari: <https://www.weforum.org>