

VALIDITAS KONSTRUK INSTRUMEN EVALUASI *OUTCOME* LEMBAGA PENDIDIKAN GURU VOKASIONAL

Nurbening Yuniarti ^{1*}, Soenarto ¹

¹Fakultas Teknik, Universitas Negeri Yogyakarta

¹Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

* Corresponding Author. Email: nurbening@uny.ac.id

Abstrak

Penelitian ini bertujuan untuk mendapatkan bukti validitas konstruk dari instrumen yang digunakan untuk kegiatan evaluasi *outcome* dari lembaga pendidikan guru vokasional. Instrumen ini terdiri atas 3 jenis instrumen. Teknik analisis data yang digunakan adalah *confirmatory factor analysis (CFA)*. Kriteria yang digunakan untuk melihat kecocokan model adalah: *p-value*, *normed Chi-square (X^2/df)*, *RMSEA (Root Mean Square Error of Approximation)*, *RMR (Root Mean-square Residual)*, *GFI (Goodness-of-Fit Index)*, *NFI (Normed Fit Index)*, *Non-Normed Fit Index (NNFI)*, dan *CFI (Comparative Fit Index)*. Bukti validitas instrumen dapat dilihat pada besarnya *faktor loading* dan *t-value*. Jika *loading factor* lebih besar dari 0,3 dengan *t-value* lebih besar dari 1,96 maka butir pernyataan dapat dikategorikan valid. Berdasarkan hasil analisis diperoleh hasil: (1) pada instrumen I terdapat 26 butir yang dinyatakan valid; (2) pada instrumen II terdapat 23 butir yang dinyatakan valid; dan (3) pada instrumen III terdapat 18 butir yang dinyatakan valid.

Kata kunci: *validitas konstruk, evaluasi, lembaga pendidikan guru vokasional*

CONSTRUCT VALIDITY OF OUTCOME EVALUATION INSTRUMENT IN VOCATIONAL TEACHER EDUCATION INSTITUTION

Abstract

This research aims to get the construct validity evidence of the outcome evaluation instrument of vocational teacher education institution. The instrument consists of 3 kinds of instruments. The analysis technique that is used to test the construct validity is confirmatory factor analysis (CFA). The criteria used to determine the goodness of fit were: *p-value*, *normed Chi-square (X^2/df)*, *RMSEA (Root Mean Square Error of Approximation)*, *RMR (Root Mean-square Residual)*, *GFI (Goodness-of-Fit Index)*, *NFI (Normed Fit Index)*, *Non-Normed Fit Index (NNFI)*, and *CFI (Comparative Fit Index)*. The evidence of the construct validity was based on the standardized loading factor and *t-value*. If the loading factor > 0.3 with *t-value* > 1.96 , the item is valid. Based on the analysis, the research results are: (1) 26 item of the instrument I are categorized valid; (2) 23 items of instrument II are categorized valid; and (3) 18 items of instrument III are categorized valid.

Keywords: *construct validity, evaluation, vocational teacher education institution*

Permalink/DOI: <http://dx.doi.org/10.21831/pep.v20i2.8448>

Pendahuluan

Kongres UNESCO ke 3 dalam bidang TVET menyepakati bahwa sektor pendidikan vokasional akan menjadi pendorong utama pertumbuhan ekonomi dunia. Konsekuensi logisnya adalah perlunya peningkatan kualitas pendidikan vokasional. Berbicara mengenai peningkatan pendidikan vokasional, maka peran Lembaga Pendidikan Tenaga Kependidikan (LPTK) vokasional sebagai lembaga penghasil guru vokasional menjadi sangat dominan. Hal ini disebabkan karena guru vokasional banyak berasal dari LPTK vokasional. Lembaga pendidikan guru vokasional mempunyai tanggung jawab menghasilkan tenaga pendidik yang profesional untuk mengembangkan pendidikan vokasi di Indonesia. Hasil-hasil penelitian menunjukkan bahwa salah satu faktor penting yang menentukan keberhasilan pendidikan adalah pendidik. Secara tegas Direktorat Tenaga Kependidikan (2008, p. 1) menjelaskan bahwa begitu pentingnya peran guru dalam mentransformasikan *input-input* pendidikan, sampai-sampai banyak pakar menyatakan bahwa di sekolah tidak akan ada perubahan atau peningkatan kualitas tanpa adanya perubahan dan peningkatan kualitas guru. Pendapat yang sama juga disampaikan oleh Wijayati, Suyata, & Sumarno (2013) yang menyatakan bahwa mutu pembelajaran dapat dicapai apabila guru yang menjadi ujung tombak pembelajaran memiliki mutu yang baik. Berdasarkan uraian tersebut maka guru memiliki peran yang sangat dominan dalam menentukan keberhasilan pendidikan.

Keberhasilan sistem pendidikan di LPTK ditentukan oleh beberapa faktor, antara lain: kualitas calon mahasiswa, tenaga pendidik, sumber daya, kurikulum, sarana prasarana, proses belajar mengajar, sistem evaluasi dan sebagainya. Dengan memperhatikan faktor-faktor tersebut dapat ditentukan upaya-upaya peningkatan kualitas LPTK.

Berdasarkan pendekatan sistemik dan sistematis dalam sistem pendidikan guru maka isi program LPTK didasarkan pada tugas yang akan ditunaikan oleh guru. Isi program LPTK mengacu pada: (a) misi

pendidikan nasional dan pembangunan, (b) tujuan kelembagaan menurut jenis dan jenjang sekolah sasaran tempat lulusan bekerja, (c) sasaran penguasaan bidang ajaran yang akan dibina lulusan (Dedi Supriadi, 2003, p. 826). Program pendidikan yang dilaksanakan di LPTK terdiri dari: (a) kelompok Mata Kuliah Dasar Umum yang berlaku untuk semua perguruan tinggi (b) kelompok Mata Kuliah Dasar Kependidikan (MKDK) dengan sasaran pembentukan dasar keahlian; (c) kelompok Mata Kuliah Bidang Studi (MKBS); dan (d) kelompok Mata Kuliah Proses Belajar Mengajar (MKPBM). Penjelasan tersebut memberikan keyakinan bahwa pendidikan di LPTK memberikan bekal yang cukup untuk mengantarkan seseorang menjadi guru yang berkualitas. Selain pengetahuan dan keterampilan, LPTK juga membekali dengan ilmu dasar kependidikan.

Pengakuan *stakeholder* terhadap keberhasilan misi yang diemban oleh LPTK vokasional menjadi sesuatu yang sangat penting dan dapat dijadikan indikator keberhasilan. Kinerja lulusan LPTK vokasional di dunia kerja merupakan unsur utama yang akan dilihat oleh *stakeholder*. Oleh karena itu, perlu dilakukan evaluasi *outcome* LPTK vokasional. Menurut Suranto, Muhyadi, & Mardapi (2014, p. 100), "evaluasi merupakan salah satu rangkaian kegiatan untuk meningkatkan kualitas, kinerja, maupun produktivitas suatu lembaga dalam melaksanakan kegiatan." Berdasarkan definisi tersebut maka kegiatan evaluasi dapat meningkatkan produktivitas dan kualitas lembaga. Melalui kegiatan evaluasi akan diperoleh informasi tentang sejauh mana suatu kegiatan dapat dilaksanakan, selanjutnya informasi tersebut dapat digunakan untuk menentukan langkah perbaikan.

Terkait dengan *outcome*, Rossi, Lipsey, & Freeman (2004, p. 204), menyatakan bahwa "*an outcome is the state of the target population or the social conditions that a program is expected to have changed.*" Lain halnya dengan Lunenburg & Ornstein (2000, p. 101) yang menyampaikan bahwa:

Outcomes, classified as first or second level, are the end results of certain work behaviors.

First-level outcomes refer to some aspect of performance and are the direct result of expending some effort on the job. Second-level outcomes are viewed as consequences to which first-level outcomes are expected to lead. That is, the end result of performance (first-level outcomes) is some type of reward for work goal accomplishment.

Berdasarkan kedua definisi tersebut, dapat dirumuskan bahwa *outcome* merupakan perilaku dari target populasi atau kondisi sosial yang terjadi sebagai akibat dari sebuah program yang dapat dilihat dari kinerja maupun aspek lain sebagai konsekuensi dari kinerja. Berdasarkan konsep tersebut maka *outcome* difokuskan pada perubahan yang dirasakan oleh sebuah populasi atau sebuah kondisi sosial yang dapat dilihat dari beberapa indikator.

Langkah awal dalam mengembangkan pengukuran *outcome* sebuah program adalah melakukan identifikasi secara spesifik apakah *outcome* itu relevan untuk diukur. Untuk itu, evaluator harus mempertimbangkan perspektif *stakeholder* tentang *outcome* yang diharapkan.

Evaluasi terhadap *outcome* sangat penting untuk dilakukan, hal ini disebabkan karena dengan melakukan evaluasi *outcome* dapat diketahui bagaimana tingkat ketercapaian tujuan dari sebuah program serta dapat mengungkap kebermanfaatannya bagi orang atau lingkungan yang menjadi target program tersebut. Beberapa pendapat tentang evaluasi *outcome* dapat disimak dari pernyataan berikut ini:

Outcome evaluations provide information on how well your programme is accomplishing its goals. Outcome evaluations measure how clients and their circumstances change, and whether the treatment experience has been a factor in causing this change. In other words, outcome evaluations aim to assess treatment effectiveness. (WHO, 2000).

Pendapat lain disampaikan oleh Myers & Barner (2005) "*Evaluation that asks questions about what has changed as a result of the programme and its activities. Outcomes can be either short-term or long-term and identifying such*

outcomes will be an integral part of demonstrating the value of a service, activity or programme." Dengan adanya evaluasi *outcome* maka dapat dihasilkan informasi bagaimana tujuan program dapat tercapai. Evaluasi *outcome* dapat dilakukan jangka pendek (*short-term*), maupun jangka panjang (*long-term*).

Kegiatan evaluasi *outcome* perlu memperhatikan prinsip-prinsip penilaian *outcome*. Menurut Samuel Ball (Fernandes, 1984, p. 19) prinsip yang harus diperhatikan dalam melakukan penilaian terhadap *outcome* adalah sebagai berikut: (1) tujuan program harus dilihat tingkat pencapaiannya; (2) memastikan bahwa *outcome behavior* seperti *afektif, motivasi attitude, psikomotorik* tidak diabaikan; (3) menggunakan "*medical model*" bukan "*engineering model*"; (4) evaluator harus mempertimbangkan fakta dalam interaksi antara program dan kelompok siswa yang berbeda; dan (5) evaluator harus mempertimbangkan tujuan jangka panjang. Jika prinsip-prinsip tersebut digunakan sebagai dasar dalam melakukan evaluasi *outcome* maka kegiatan evaluasi tersebut akan memberikan hasil yang akurat.

Berbicara masalah pendidikan, maka evaluasi *outcome* ini juga menjadi bagian dari upaya peningkatan kualitas. Salah satu indikator keberhasilan sebuah program pendidikan dapat dilihat dari *outcome* yang dihasilkan, sedangkan untuk mengetahui *outcome* tersebut dapat dilakukan dengan evaluasi. Evaluasi terhadap *outcome* yang paling banyak digunakan dalam pendidikan tinggi sejauh ini cenderung berfokus *outcome* jangka pendek (*short-term outcomes*) yang dapat diukur saat mahasiswa masih di perguruan tinggi. Namun, ada juga lembaga pendidikan tinggi yang fokus pada perubahan jangka panjang (*longer term changes*).

Outcome dari sebuah program akan menjawab pertanyaan "*What happened as a result of the program?*" Selanjutnya, Sihvonen (1999, p. 12) mengaitkan antara *outcome* dengan peningkatan kualitas pendidikan seperti berikut ini:

In practice the evaluation of educational quality is largely concerned with educational outcomes. Here, the emphasis lies on the evalua-

tion of the management and the appropriateness of the processes, that is, on their efficiency, whereas in the case of outcomes the focus is on the whole, and above all, on the effects produced by the activities concerned.

Pendapat tersebut menekankan bahwa *outcome* dari lembaga pendidikan dapat digunakan sebagai indikator dari kualitas pendidikan yang diselenggarakan dan mengomunikasikan dampak dari penggunaan sumber daya.

Stufflebeam (Isaac, 1981, p. 2) menyatakan bahwa *“the purpose evaluation is to improve, not to prove.”* Pendapat tersebut memberikan kejelasan bahwa tujuan kegiatan evaluasi adalah untuk peningkatan program dan pembuatan beberapa kebijakan terkait peningkatan program. Selain untuk peningkatan program, evaluasi *outcome* juga dapat memberikan informasi sejauh mana sebuah program memberikan dampak terhadap perubahan individu atau sebuah populasi. Menurut Trochim, (2006) *“An outcome evaluation is a type of evaluation that investigates whether changes occur for participants in a program and if these changes are associated with a program or an activity.”* Berdasarkan pendapat tersebut, maka dapat dinyatakan bahwa evaluasi *outcome* sangat bermanfaat untuk mengetahui efektivitas sebuah program. Jika program yang dimaksud adalah program pendidikan, maka dengan melakukan evaluasi *outcome* maka dapat dilihat bagaimana efektivitas program pendidikan yang diselenggarakan. Hal ini dapat dilihat dari perubahan perilaku peserta didik setelah menyelesaikan program pendidikan tersebut.

Astin (1993, p. 43) menawarkan sebuah skema konseptual untuk mengembangkan pengukuran *outcome* bagi sebuah institusi. Skema konseptual tersebut meliputi tiga dimensi: jenis *outcome*, jenis data, dan waktu. Menurut jenisnya *outcome* dalam bidang pendidikan dibedakan menjadi 2 yaitu: kognitif dan afektif. Jenis data berhubungan dengan jenis-jenis informasi yang dikumpulkan untuk menilai hasil kognitif atau afektif. Ditinjau dari dimensi waktu, *outcome* dibedakan menjadi *short-term* dan *long-term*. Pendapat

tersebut harus diperhatikan oleh evaluator untuk melakukan kegiatan evaluasi *outcome*.

Menurut Rossi, Lipsey, & Freeman (2004), ada beberapa alasan mengapa perlu melakukan evaluasi terhadap *outcome* yaitu: alokasi sumber daya yang efektif, perbaikan dan peningkatan program, pertanggungjawaban penggunaan sumber daya, mengembangkan bukti dasar yang efektif, meningkatkan pelayanan, serta memberikan bukti terkait apa yang telah dilakukan. Hal ini sejalan dengan pendapat Myers & Barner (2005) yang menyatakan bahwa evaluasi *outcome* berguna untuk: (1) *effective decision making, such as allocation of resources*, (2) *reshaping and programme improvement*, (3) *accountability for resources used*, (5) *developing an effective evidence base*, (5) *delivering better services*, and (6) *building an evidence base of what works*. Kedua pendapat tersebut memberikan gambaran yang jelas bahwa evaluasi *outcome* akan memiliki banyak informasi dan manfaat bagi sebuah program. Dengan demikian, maka evaluasi *outcome* dapat memberikan kontribusi terhadap peningkatan kualitas sebuah program.

Terkait dengan teknis pelaksanaan evaluasi *outcome*, Allen & Tinkew (2008, p. 3) menyatakan bahwa:

Once a program decides to pursue an outcome evaluation, certain steps should be undertaken. The steps involved in planning an outcome evaluation are not always linear. It may be necessary to return to previous steps to reevaluate decisions that were made based on the availability of resources or on the feasibility of the evaluation activities. Overall, programs should be responsive to the changing needs of the evaluation design and flexible enough to create a better design when necessary.

Pendapat tersebut memberikan keleluasaan bagi evaluator untuk melakukan tugasnya, namun harus tetap responsif terhadap kebutuhan evaluasi.

Evaluasi *outcome* yang dimaksud dalam penelitian ini ditujukan untuk mengetahui perubahan tingkah laku dari lulusan lembaga pendidikan guru vokasional yang meliputi: penghargaan yang diterima, motivasi kerja, pengembangan karier, kompetensi dalam menjalankan profesinya, kemampuan

dalam mengelola administrasi sekolah, kontribusi terhadap pengembangan sekolah, serta kreativitas dan inovasi dalam bekerja, penguasaan bidang keahlian, pemanfaatan media pembelajaran, penggunaan strategi pembelajaran, serta evaluasi dan penilaian. Aspek-aspek tersebut merupakan representasi kinerja lulusan dari pendidikan guru vokasional di dunia kerja.

Berdasarkan pendapat tersebut, evaluasi *outcome* bagi dunia pendidikan khususnya lembaga pendidikan guru vokasional menjadi hal yang sangat penting. Oleh karena itu, perlu dikembangkan instrumen yang tepat untuk mengukur *outcome* lembaga pendidikan guru vokasional. Kegiatan evaluasi *outcome* yang dilakukan akan memberikan hasil yang valid jika instrumen yang digunakan memenuhi kaidah pengukuran. Salah satu kaidah yang harus dipenuhi dalam pengukuran adalah validitas. Rossi, Lipsey, & Freeman (2004, pp. 218-220) menyampaikan bahwa: *key properties of measurement procedures in outcome evaluation consist of: (1) reliability, (2) validity, dan (3) sensitivity*. Evaluasi validitas merupakan salah satu langkah yang harus ditempuh dalam pengembangan sebuah instrumen evaluasi *outcome*.

Menurut Azwar (2014), kegiatan pengukuran dikatakan memiliki validitas yang tinggi jika menghasilkan data yang secara akurat memberikan gambaran mengenai variabel yang diukur sesuai dengan tujuan pengukuran. Pernyataan tersebut mengandung pengertian bahwa validitas tidak bisa diberlakukan pada semua pengukuran, sehingga perlu diikuti dengan penjelasan yang terkait dengan tujuan pengukuran. Penjelasan yang dimaksud adalah instrumen dikatakan valid untuk mengukur apa.

Validasi kontrak membuktikan apakah hasil pengukuran yang diperoleh melalui *item-item* tes berkorelasi tinggi dengan kontrak teoretik yang mendasari penyusunan tes tersebut. (Azwar, 2014, p. 116). Lebih lanjut, Cronbach & Meehl (1995) menyatakan bahwa untuk membuktikan validitas kontrak dapat dilakukan melalui: (1) mengartikulasikan serangkaian konsep teoretik dan interrelasinya, (2) mengembangkan cara

untuk mengukur kontrak hipotetik yang diteorikan, dan (3) menguji secara empirik hubungan hipotetik di antara kontrak tersebut dan manifestasinya yang nampak. Dengan demikian, validitas kontrak adalah proses pengumpulan bukti untuk mendukung penggunaan dari sebuah tes sebagai alat ukur kontrak. Terkait dengan hal tersebut, studi ini bertujuan memperoleh bukti validitas kontrak dari instrumen evaluasi *outcome* bagi lembaga pendidikan guru vokasional.

Metode Penelitian

Pelaksanaan uji empiris dilakukan pada bulan Maret sampai dengan Mei 2015 di 62 SMK negeri dan swasta yang berada di provinsi Daerah Istimewa Yogyakarta (DIY) dan Jawa Tengah.

Responden yang digunakan untuk uji empiris ini terdiri atas 3 kelompok responden yakni guru (lulusan pendidikan guru vokasional), kepala sekolah dari SMK tempat lulusan bekerja, dan siswa SMK yang mengikuti pembelajaran yang dilaksanakan oleh guru. Jumlah keseluruhan responden adalah 94 guru, 87 kepala sekolah, dan 100 siswa SMK.

Teknik analisis data yang digunakan adalah *second order confirmatory factor analysis (2nd CFA)* dengan bantuan *software* Lisrel 8.80. Sebelum melakukan evaluasi validitas kontrak, dilakukan pengkajian kecocokan keseluruhan model (*overall model fit*) dan dilanjutkan dengan pengkajian kecocokan model pengukuran (*measurement model fit*). Dalam mengevaluasi kecocokan keseluruhan model tersebut (*goodness of fit*) digunakan beberapa kriteria. Kriteria yang digunakan adalah: *p-value*, *normed chi-square (X²/df)*, *RMSEA (Root Mean Square Error of Approximation)*, *RMR (Root Mean-square Residual)*, *GFI (Goodness-of-Fit Index)*, *NFI (Normed Fit Index)*, *Non-Normed Fit Index (NNFI)*, dan *CFI (Comparative Fit Index)*. Kriteria yang digunakan untuk kecocokan model pengukuran atau evaluasi validitas kontrak adalah *loading factor (lf)* dan besarnya *t-value*. Sebuah *item* dikategorikan valid apabila memiliki nilai *loading factor* lebih besar dari 0,3 dengan nilai *t-value* lebih besar dari 1,96.

Kriteria validitas yang digunakan mengacu pada pendapat Hair, Black, Babin, Anderson, & Tatham (2010, p. 119) yang menyatakan bahwa “*factor loadings ± 0,3 to 0.4 are minimally acceptable*”.

Hasil Penelitian dan Pembahasan

Pada penelitian ini dilakukan pengkajian evaluasi kecocokan keseluruhan model (*overall model fit*) dan evaluasi kecocokan model pengukuran (*measurement model fit*). Evaluasi kecocokan keseluruhan model (*overall model fit*) dilakukan untuk mengetahui kecocokan model berdasarkan kriteria yang telah ditentukan, sedangkan evaluasi kecocokan model pengukuran (*measurement model fit*) dilakukan untuk mengetahui validitas konstruk.

Untuk mengetahui kecocokan keseluruhan model dan kecocokan model pengukuran, digunakan *second order confirmatory factor analysis*. Wijayanto, (2008, p. 190) menjelaskan bahwa *second order confirmatory factor analysis (2ndCFA)* merupakan sebuah model pengukuran yang terdiri atas dua tingkat. Tingkat pertama adalah CFA yang menunjukkan hubungan antara variabel-variabel teramati sebagai indikator-indikator dari variabel laten terkait. Tingkat kedua menunjukkan hubungan antara variabel-variabel laten pada tingkat pertama sebagai indikator-indikator dari sebuah variabel laten tingkat kedua.

Kriteria yang digunakan adalah: *p-value*, *normed chi-square* (X^2/df), *RMSEA* (*Root Mean Square Error of Approximation*), *RMR* (*Root Mean-square Residual*), *GFI* (*Goodness-of-Fit Index*), *NFI* (*Normed Fit Index*), *Non-Normed Fit Index* (*NNFI*), dan *CFI* (*Comparative Fit Index*). Penjelasan untuk setiap kriteria seperti berikut ini. Kriteria pertama yaitu *p-value*, kriteria ini sangat mudah digunakan untuk melihat kecocokan model karena dapat langsung dilihat pada *path*. Model dapat dikatakan baik jika *p-value* lebih besar dari 0,05. *Normed chi-square* adalah perbandingan antara *chi-square* dengan *degree of freedom*. *RMSEA* merupakan indikator model fit yang paling informatif. *RMR* mewakili rerata residual yang diperoleh dari

mencocokkan matrik varian-kovarian data sampel. *GFI* merupakan ukuran mengenai ketepatan model dalam menghasilkan matriks kovarian. *NFI* memiliki tendensi untuk merendahkan fit pada sampel yang kecil. Nilai *NNFI* digunakan untuk mengatasi permasalahan yang timbul akibat kompleksitas model.

Evaluasi Kecocokan Keseluruhan Model (*Overall Model Fit*)

Tahap pertama yang dilakukan ditujukan untuk mengetahui secara umum derajat kecocokan atau *Goodness-of-Fit (GOF)* antara data yang diperoleh dengan model pengukuran. Jika evaluasi kecocokan keseluruhan model pengukuran sudah dilakukan maka dapat dilanjutkan dengan evaluasi kecocokan model pengukuran. Hasil evaluasi kecocokan keseluruhan model ditetapkan berdasarkan 8 kriteria yakni: *p-value*, *normed χ^2* , *RMSEA*, *RMR*, *GFI*, *NFI*, *NNFI*, dan *CFI*.

Berikut ini disajikan hasil uji kecocokan keseluruhan model (*overall model fit*) untuk tiga paket instrumen yaitu: instrumen I, instrumen II, dan instrumen III.

Instrumen I

Instrumen I ditujukan kepada lulusan pendidikan guru vokasional. Instrumen ini digunakan untuk mengungkap *outcome* pendidikan guru vokasional berdasarkan pendapat lulusan pendidikan guru vokasional khususnya yang mengampu mata pelajaran produktif. Tabel 1 berikut menyajikan hasil evaluasi kecocokan keseluruhan model (*overall model fit*) untuk instrumen I.

Tabel 1. Kecocokan Model Instrumen I.

No	Ukuran GOF	Hasil Estimasi	Tingkat Kecocokan
1	p value	0,15016	Good fit
2	Normed χ^2	1,085	Good fit
3	RMSEA	0,030	Close fit
4	RMR	0,058	Close fit
5	GFI	0,79	Marginal fit
6	NFI	0,95	Marginal fit
7	NNFI	0,99	Marginal fit
8	CFI	0,99	Good fit

Tabel 1 menunjukkan bahwa nilai *p-value* sebesar 0,15016 ($p \geq 0,05$), sehingga tingkat kecocokan model dapat dikategorikan baik (*good fit*). Kriteria berikutnya adalah *normed Chi-square* (*Normed χ^2*), yang merupakan rasio antara *Chi-square* dan *degree of freedom*. Nilai yang disarankan adalah antara batas bawah sebesar 1,00 dan batas atas sebesar 2,00, dengan nilai sebesar 1,085 maka ukuran *Normed χ^2* dikategorikan baik (*good fit*). Ukuran kesesuaian model selanjutnya adalah nilai RMSEA sebesar 0,030 (RMSEA < 0,05) menandakan tingkat kecocokan model adalah *close fit* (Browne & Cudeck, 1993). Kriteria berikutnya adalah nilai RMR, dengan $N \leq 250$ maka model akan mempunyai kecocokan yang baik (*good fit*) jika mempunyai nilai *Standardized RMR* $\leq 0,09$, dan nilai CFI > 0,92 (Hair, Black, Babin, Anderson, & Tatham, 2006). Selain itu, kriteria GFI berkisar antara 0 (*poor fit*) hingga 1 (*perfect fit*). Untuk model yang dikembangkan, perolehan nilai GFI sebesar 0,79 ($0,80 \leq \text{GFI} \leq 0,90$) sehingga tergolong *marginal fit*. Kriteria ukuran GFI ini berlaku juga untuk NFI dan NNFI. Berdasarkan evaluasi model, secara keseluruhan dapat disimpulkan bahwa kecocokan seluruh model adalah baik (*good fit*).

Instrumen II

Instrumen II ditujukan kepada kepala sekolah selaku pengguna lulusan LPTK kejuruan. Instrumen II ini digunakan untuk mengungkap *outcome* LPTK vokasional dari penilaian kepala sekolah. Kepala sekolah merupakan pimpinan langsung dari guru sehingga dapat memberikan penilaian yang akurat terkait dengan kinerja guru lulusan LPTK vokasional. Tabel 2 berikut menyajikan hasil evaluasi kecocokan keseluruhan model (*overall model fit*) untuk instrumen II.

Tabel 2 menunjukkan bahwa nilai *p-value* sebesar 0,6842 ($p \geq 0,05$), sehingga tingkat kecocokan model dapat dikategorikan baik (*good fit*). Kriteria berikutnya adalah *normed Chi-square* (*Normed χ^2*), yang merupakan rasio antara *Chi-square* dan *degree of freedom*. Nilai yang disarankan adalah antara batas bawah sebesar 1,00 dan batas atas sebesar

2,00, dengan nilai sebesar 1,145 maka ukuran *Normed χ^2* dikategorikan baik (*good fit*). Ukuran kesesuaian model selanjutnya adalah nilai RMSEA sebesar 0,041 (RMSEA < 0,05) menandakan tingkat kecocokan model adalah *close fit* (Browne & Cudeck, 1993). Kriteria berikutnya adalah RMR, dengan $N \leq 250$ maka model akan mempunyai kecocokan yang baik (*good fit*) jika mempunyai nilai *Standardized RMR* $\leq 0,09$ dan nilai *comparative fit index* CFI > 0,92 (Hair, Black, Babin, Anderson, & Tatham, 2006). Selain itu, kriteria *Goodness-of-Fit Index* (GFI) berkisar antara 0 (*poor fit*) hingga 1 (*perfect fit*). Untuk model yang dikembangkan, perolehan nilai GFI sebesar 0,80 ($0,80 \leq \text{GFI} \leq 0,90$) sehingga tergolong *good fit*. Kriteria ukuran GFI ini berlaku juga untuk *Normed Fit Index* (NFI) dan *Non-Normed Fit Index* (NNFI). Secara keseluruhan dapat disimpulkan bahwa kecocokan seluruh model adalah baik (*good fit*).

Tabel 2. Kecocokan Model untuk Instrumen II

No	Ukuran GOF	Hasil Estimasi	Tingkat Kecocokan
1	p value	0,6842	Good fit
2	Normed χ^2	1,145	Good fit
3	RMSEA	0,041	Close fit
4	RMR	0,019	Close fit
5	GFI	0,80	Good fit
6	NFI	0,91	Marginal fit
7	NNFI	0,97	Marginal fit
8	CFI	0,98	Good fit

Instrumen III

Instrumen III ditujukan kepada siswa SMK yang mengikuti pembelajaran yang dilaksanakan oleh guru lulusan LPTK kejuruan. Instrumen ini digunakan untuk mengungkap *outcome* pendidikan guru vokasional dari penilaian siswa SMK. Hal ini dengan pertimbangan bahwa, siswa SMK merupakan pengguna lulusan yang secara langsung dapat memberikan penilaian terhadap kinerja guru lulusan pendidikan guru vokasional. Tabel 3 berikut menyajikan hasil evaluasi kecocokan keseluruhan model (*overall model fit*) untuk instrumen III.

Tabel 3. Kecocokan Model untuk Instrumen III

No	Ukuran GOF	Hasil Estimasi	Tingkat Kecocokan
1	p value	0,16564	Good fit
2	Normed χ^2	1,120	Good fit
3	RMSEA	0,035	Close fit
4	RMR	0,023	Close fit
5	GFI	0,86	Good fit
6	NFI	0,86	Marginal fit
7	NNFI	0,97	Marginal fit
8	CFI	0,98	Good fit

Tabel 3 menunjukkan bahwa nilai *p-value* sebesar 0,6842 ($p \geq 0,05$), sehingga tingkat kecocokan model dapat dikategorikan baik (*good fit*). Kriteria berikutnya adalah *normed Chi-square* (*Normed χ^2*), yang merupakan rasio antara *Chi-square* dan *degree of freedom*. Nilai yang disarankan adalah antara batas bawah sebesar 1,00 dan batas atas sebesar 2,00, dengan nilai sebesar 1,120 maka ukuran *Normed χ^2* dikategorikan baik (*good fit*). Ukuran kesesuaian model selanjutnya adalah nilai RMSEA sebesar 0,035 ($RMS- EA < 0,05$) menandakan tingkat kecocokan model adalah *close fit* (Browne & Cudeck, 1993). Selanjutnya adalah nilai RMR, dengan $N \leq 250$ maka model akan mempunyai kecocokan yang baik (*good fit*) jika mempunyai nilai *Standardized RMR* $\leq 0,09$, dan nilai $CFI > 0,92$ (Hair, Black, Babin, Anderson, & Tatham, 2006). Selain itu, kriteria GFI berkisar antara 0 (*poor fit*) hingga 1 (*perfect fit*). Untuk model yang dikembangkan, perolehan nilai GFI sebesar 0,86 ($0,80 \leq GFI \leq 0,90$) sehingga tergolong *good fit*. Kriteria ukuran GFI ini berlaku juga untuk *NFI* dan *NNFI*. Berdasarkan delapan kriteria tersebut maka secara keseluruhan dapat disimpulkan bahwa kecocokan seluruh model adalah baik (*good fit*).

Berdasarkan hasil uji keseluruhan model pengukuran pada ketiga jenis instrumen maka dapat disimpulkan bahwa ketiga jenis instrumen memiliki tingkat kecocokan model pada delapan kriteria ukuran GOF. Dengan demikian, ketiga model pengukuran dalam kategori baik (*good fit*) oleh karena itu

dapat dilanjutkan dengan evaluasi kecocokan model pengukuran untuk mengetahui validitas konstruk dari ketiga instrumen tersebut.

Evaluasi Kecocokan Model Pengukuran (*Measurement Model Fit*)

Setelah kecocokan model keseluruhan dinyatakan baik maka langkah selanjutnya adalah melakukan evaluasi terhadap kecocokan model pengukuran. Dalam penelitian ini evaluasi kecocokan model pengukuran yang dilakukan adalah dengan melakukan evaluasi terhadap validitas dari model pengukuran. Validitas berkaitan dengan apakah sebuah variabel dapat mengukur apa yang seharusnya diukur. Hal ini sesuai dengan pendapat Hair, Black, Babin, Anderson, & Tatham (2006, p. 771) yang menyatakan bahwa “*construct validity extent to which a set of measured variables actually represent the theoretical latent construct they are designed to measure*”. Dengan demikian, validitas sebuah instrumen merupakan hal yang sangat penting dalam kegiatan pengukuran, karena dengan pengukuran yang baik maka sebuah penelitian akan memberikan hasil yang representatif. Lebih lanjut, Mardapi (2005, p. 19) menjelaskan bahwa validitas sebuah tes tidaklah berlaku umum untuk semua tujuan ukur. Pendapat tersebut menegaskan bahwa validitas harus disesuaikan dengan tujuan ukur.

Berdasarkan cara estimasinya, validitas terbagi menjadi 3 tipe yaitu: (1) validitas isi, (2) validitas konstruk, dan (3) validitas berdasarkan kriteria. Validitas isi dilakukan melalui pengujian terhadap isi dengan analisis rasional melalui *expert judgment*. Validitas konstruk dimaksudkan untuk mengetahui sejauh mana instrumen dapat mengungkap konstruk. Validitas kriteria diestimasi berdasarkan kriteria yang telah ditentukan. Jenis validitas yang dapat dievaluasi dalam penelitian ini adalah validitas konstruk. Validitas ini dapat dilihat sebagai istilah menyeluruh untuk menilai keabsahan prosedur pengukuran yang digunakan untuk mengukur konstruk yang diberikan.

Berikut ini disajikan hasil uji kecocokan model pengukuran (*measurement model fit*)

untuk tiga paket instrumen yang digunakan untuk mengukur *outcome* pendidikan guru vokasional. Uji kecocokan model pengukuran ini dilakukan dengan analisis faktor yakni *second order confirmatory factor analysis* atau disebut dengan 2ndCFA. Gambar lintasan yang dihasilkan pada 2ndCFA lebih banyak memberikan informasi dan terlihat lebih kompleks jika dibandingkan dengan *confirmatory factor analysis*. Gambar lintasan dalam 2ndCFA dapat menunjukkan: (1) hubungan antarvariabel teramati dengan variabel laten, dan (2) hubungan antarvariabel laten. Cara yang digunakan untuk menentukan apakah setiap item memenuhi syarat validitas konstrak atau tidak dilakukan dengan cara melihat output dari 2ndCFA. Kriteria validitas konstrak yang digunakan terdiri dari muatan faktor (*loading factor/SLF*) dengan nilai *t-val* yang dimiliki. Jika nilai *loading factor* lebih besar dari 0,3 dan nilai *t-value* lebih besar dari 1,96 maka *item* dikategorikan sebagai *item* yang valid. Kriteria ini mengacu pada pendapat (Hair, Black, Babin, Anderson, & Tatham, 2006, p. 129) yang menyatakan bahwa “*factor loadings ± 0.3 to 0.4 are minimally acceptable.*” Nilai *loading factor* menunjukkan besarnya muatan faktor pada setiap *item*, sedangkan *t-value* sebesar 1,96 merujuk pada taraf signifikansi sebesar 95%.

Evaluasi kecocokan model pengukuran dilakukan pada ketiga instrumen yaitu: instrumen I, instrumen II, dan instrumen III.

Instrumen I

Analisis faktor dilakukan pada instrumen I yang terdiri atas 26 *item* untuk menentukan 3 variabel laten yaitu penghargaan yang diterima (W_App), motivasi kerja (W_Mot), dan pengembangan karier (CarDev). Ketiga variabel laten tersebut dapat digunakan untuk mengungkap *outcome* lembaga pendidikan vokasional dengan responden lulusan pendidikan guru vokasional yang menekuni profesi sebagai guru di SMK.

Berikut ini diberikan penjelasan untuk setiap variabel laten yang dimaksud dalam penelitian. Penghargaan yang diterima (W_App) adalah penghargaan yang diterima

oleh lulusan pendidikan guru vokasional selama menekuni profesi guru SMK. Penghargaan ini berupa penghargaan yang bersifat finansial dan nonfinansial. Motivasi kerja (W_Mot) adalah dorongan dalam diri seseorang yang dipengaruhi oleh faktor internal dan eksternal untuk melaksanakan tugas sebagai guru SMK. Motivasi kerja merupakan salah satu aspek yang digunakan untuk mengungkap *outcome*, karena motivasi kerja akan mempengaruhi keberhasilan dalam bekerja dan kepuasan pengguna layanan. Sementara itu, pengembangan karier (CarDev) adalah proses peningkatan kemampuan kerja yang dilakukan dalam rangka mencapai sebuah tujuan karier dalam pendidikan kejuruan. Pengembangan karier akan terkait dengan peningkatan kesejahteraan lulusan dalam menjalani profesinya sebagai guru SMK.

Tabel 4 berikut menyajikan hasil evaluasi validitas konstrak untuk instrumen I.

Tabel 4. Validitas Konstrak Instrumen I

Faktor	item	Second Order CFA		Ket
		SLF	t-Val	
W_App	WA_1	0,66	***	Valid
	WA_2	0,67	5,91	Valid
	WA_3	0,58	5,24	Valid
W_Mot	WM_2	0,77	***	Valid
	WM_4	0,63	6,35	Valid
	WM_5	0,76	7,90	Valid
	WM_7	0,65	6,54	Valid
	WM_8	0,61	6,10	Valid
	WM_9	0,72	7,42	Valid
	WM_10	0,66	6,68	Valid
	WM_11	0,59	5,89	Valid
	WM_13	0,72	7,46	Valid
	WM_14	0,73	7,53	Valid
CarDev	WM_15	0,69	7,09	Valid
	CD_1	0,73	***	Valid
	CD_2	0,66	6,45	Valid
	CD_3	0,72	6,99	Valid
	CD_4	0,72	7,05	Valid
	CD_5	0,65	6,35	Valid
	CD_6	0,69	6,74	Valid
	CD_7	0,73	7,09	Valid
	CD_8	0,71	6,97	Valid
	CD_9	0,69	6,73	Valid
	CD_11	0,69	6,71	Valid
	CD_12	0,72	7,00	Valid
	CD_13	0,73	7,15	Valid

Tabel 4 menunjukkan ringkasan evaluasi validitas konstrak yang dilakukan dengan *software* Lisrel 8.8. Hasil evaluasi validitas konstrak dengan menggunakan 2ndCFA menunjukkan bahwa: 3 butir pernyataan pada aspek penghargaan yang diterima (WApp), 11 butir pernyataan pada motivasi kerja (WMot), dan 12 butir pernyataan pada pengembangan karier (CarDev) memiliki *loading factor*>0,3 dan *t-value*>1,96 sehingga semua butir pernyataan pada setiap aspek pada instrumen I sudah memenuhi kaidah validitas konstrak. Dengan demikian, 26 butir pernyataan pada instrumen I tersebut dinyatakan valid untuk mengukur *outcome* lembaga pendidikan vokasional berdasarkan pendapat lulusan.

Instrumen II

Analisis faktor dilakukan pada instrumen II yang terdiri atas 23 *item* untuk menentukan 4 variabel laten yaitu: kompetensi mengajar (TComp), kemampuan dalam menangani administrasi Sekolah (SAdm), kontribusi terhadap pengembangan lembaga (ContSD), serta kreativitas dan inovasi (CreInn). Keempat variabel laten tersebut dapat digunakan untuk mengungkap *outcome* lembaga pendidikan vokasional dengan responden kepala sekolah selaku pengguna lulusan.

Berikut ini diberikan penjelasan untuk setiap variabel laten yang dimaksud dalam penelitian. Kompetensi guru (TComp) adalah kompetensi yang dimiliki oleh lulusan pendidikan guru vokasional khususnya dalam menjalankan tugasnya dengan mengacu pada empat kompetensi guru yaitu: kompetensi pedagogik, kepribadian, sosial, dan profesional. Keempat kompetensi ini dapat mencerminkan sosok guru yang memenuhi standar kualifikasi profesi. Kemampuan dalam menangani administrasi sekolah (SAdm) adalah kemampuan yang dimiliki lulusan LPTK untuk menyelesaikan administrasi sekolah. Dalam menjalankan tugasnya, guru tidak terlepas dari penyelesaian administrasi baik yang menyangkut tugasnya secara pribadi maupun tugas lembaga. Kontribusi terhadap pengembangan lembaga (ContSD) adalah kontribusi yang diberikan lulusan

lembaga pendidikan guru vokasional untuk mengembangkan SMK tempat lulusan bekerja. Peran serta guru untuk memajukan sekolah sangat diharapkan. Hal ini bisa tercermin dari bagaimana guru dapat menyumbangkan pemikiran, tenaga, dan waktu yang dimiliki untuk kemajuan lembaga. Sementara itu kreativitas dan inovasi (CreInn) adalah kemampuan untuk mengembangkan ide-ide baru dan cara-cara baru dalam pemecahan masalah serta menemukan peluang demi kemajuan SMK. Melalui kreativitas dan inovasi, guru dapat menyelesaikan permasalahan yang dihadapi dalam menjalankan tugasnya dengan cara-cara yang efektif dan efisien.

Tabel 5 berikut menyajikan hasil evaluasi validitas konstrak untuk instrumen II.

Tabel 5. Validitas Konstrak Instrumen II

Faktor	item	Second Order CFA		Ket
		SLF	t-Val	
TComp	TC_1	0,66	***	Valid
	TC_2	0,80	9,53	Valid
	TC_3	0,83	6,41	Valid
	TC_4	0,82	6,38	Valid
	TC_5	0,61	5,04	Valid
	TC_6	0,68	5,48	Valid
	TC_8	0,73	5,84	Valid
	TC_9	0,72	5,80	Valid
	SAdm	SA_1	0,63	***
SA_2		0,94	6,99	Valid
SA_3		0,90	6,81	Valid
SA_4		0,58	15,49	Valid
SA_5		0,91	6,84	Valid
ContSD	CiS_1	0,92	***	Valid
	CiS_2	0,92	12,93	Valid
	CiS_3	0,67	7,41	Valid
	CiS_4	0,66	7,28	Valid
	CiS_5	0,52	5,27	Valid
	CiS_6	0,63	6,73	Valid
CreInn	CI_1	0,80	***	Valid
	CI_2	0,65	5,44	Valid
	CI_3	0,71	5,81	Valid
	CI_4	0,53	4,47	Valid

Tabel 5 menunjukkan ringkasan evaluasi validitas konstrak yang dilakukan de-

ngan *software* Lisrel 8.8. Hasil evaluasi validitas konstrak dengan menggunakan 2ndCFA menunjukkan bahwa: 8 butir pernyataan pada aspek kompetensi guru (TComp), 5 butir pernyataan pada aspek kemampuan dalam administrasi sekolah (SAdm), 6 butir pernyataan pada aspek kontribusi terhadap pengembangan lembaga (ContSD), serta 4 butir pernyataan pada aspek kreativitas dan inovasi (CreInn) memiliki *loading factor*>0,3 dan *t-value*>1,96 sehingga 23 butir pada instrumen II sudah memenuhi kaidah validitas konstrak. Dengan demikian, 23 butir pernyataan pada instrumen II tersebut dinyatakan valid untuk mengukur *outcome* lembaga pendidikan vokasional dari berdasarkan pendapat kepala sekolah sebagai pengguna lulusan.

Instrumen III

Analisis faktor dilakukan pada instrumen III yang terdiri atas 18 *item* untuk menentukan 4 variabel laten yaitu: penguasaan materi ajar (Mat), pemanfaatan media pembelajaran (Med), penguasaan strategi pembelajaran (Stra), serta evaluasi dan penilaian (EvAss).

Berikut ini diberikan penjelasan untuk setiap variabel laten yang dimaksud dalam penelitian. Penguasaan materi ajar (Mat) adalah kemampuan lulusan lembaga pendidikan vokasional dalam menguasai setiap materi yang disampaikan kepada siswa SMK. Seorang guru yang memiliki penguasaan materi baik akan berakibat pada tumbuhnya kepercayaan siswa terhadap guru tersebut sehingga siswa akan lebih termotivasi dalam belajar. Pemanfaatan media pembelajaran (Med) adalah kemampuan lulusan lembaga pendidikan vokasional dalam mengembangkan dan memanfaatkan media dalam pembelajaran di SMK. Media dapat digunakan guru untuk membantu siswa dalam peningkatan pemahaman materi yang disampaikan. Penguasaan strategi pembelajaran (Stra) adalah kemampuan lulusan lembaga pendidikan vokasional dalam menerapkan strategi pembelajaran yang tepat di SMK. Dalam mengajar, guru harus memiliki kemampuan dalam menerapkan strategi pembelajaran

yang disesuaikan dengan materi pembelajaran, karakteristik peserta didik, dan tujuan pembelajaran. Sementara itu, evaluasi dan penilaian (EvAss) adalah kemampuan lulusan lembaga pendidikan vokasional dalam melakukan evaluasi dan penilaian terhadap hasil belajar yang diperoleh siswa SMK. Evaluasi dan penilaian guru mencerminkan hasil pembelajaran yang telah dilaksanakan. Hasil evaluasi dan penilaian yang dilakukan guru dapat dijadikan sebagai alat ukur kepuasan siswa terhadap kinerja seorang guru.

Tabel 6 berikut menyajikan hasil evaluasi validitas konstrak untuk instrumen III.

Tabel 6. Validitas Konstrak Instrumen III

Faktor	item	Second Order CFA		Ket
		SLF	t-val	
Master	Mat_1	0,55	***	Valid
	Mat_2	0,58	4,11	Valid
	Mat_3	0,60	4,23	Valid
	Mat_4	0,81	4,66	Valid
Med	Med_1	0,72	***	Valid
	Med_2	0,81	7,64	Valid
	Med_3	0,81	7,64	Valid
	Med_4	0,86	8,01	Valid
Stra	Stra_1	0,76	***	Valid
	Stra_2	0,35	3,13	Valid
	Stra_3	0,83	6,61	Valid
	Stra_5	0,50	4,50	Valid
	Stra_6	0,55	4,90	Valid
	EvAss	EvAs_1	0,60	***
	EvAs_2	0,55	3,78	Valid
	EvAs_4	0,57	3,84	Valid
	EvAs_5	0,46	3,36	Valid
	EvAs_6	0,51	3,58	Valid

Tabel 6 menunjukkan ringkasan evaluasi validitas konstrak yang dilakukan dengan *software* Lisrel 8.8. Hasil evaluasi validitas konstrak dengan menggunakan 2ndCFA menunjukkan bahwa: 4 butir pernyataan pada aspek penguasaan materi ajar (Mat), 4 butir pernyataan pada aspek pemanfaatan media (Med), 5 butir pernyataan pada aspek penguasaan strategi pembelajaran (Stra), serta 5 butir pernyataan pada aspek evaluasi dan penilaian (EvAss) memiliki *loading factor*>0,3 dan *t-value*>1,96 sehingga 18 butir

pada instrumen III sudah memenuhi kaidah validitas konstrak. Dengan demikian, 18 butir pernyataan pada instrumen III tersebut dinyatakan valid untuk mengukur *outcome* lembaga pendidikan vokasional berdasarkan penilaian siswa SMK.

Berdasarkan analisis tersebut dihasilkan 3 paket instrumen yang memenuhi validitas konstrak. Dengan demikian, ketiga paket instrumen tersebut sudah dinyatakan memiliki validitas konstrak dan dapat digunakan untuk mengukur *outcome* lembaga pendidikan vokasional melalui tiga sumber data yaitu: lulusan lembaga pendidikan vokasional, kepala sekolah, dan siswa SMK. Aspek yang diungkap pada ketiga jenis instrumen juga disesuaikan dengan tujuan pengukuran dengan harapan dapat memberikan data yang akurat.

Simpulan dan Saran

Simpulan

Berdasarkan hasil analisis dapat dirumuskan kesimpulan sebagai berikut: (1) pada instrumen I sebanyak 26 item memiliki nilai *loading factor* yang lebih besar dari 0,3 dengan nilai *t-value* lebih besar dari 1,96 sehingga 26 butir tersebut dinyatakan valid; (2) pada instrumen II sebanyak 23 item memiliki nilai *loading factor* yang lebih besar dari 0,3 dengan nilai *t-value* lebih besar dari 1,96 sehingga 23 butir tersebut dinyatakan valid; dan (3) pada instrumen III sebanyak 18 item memiliki nilai *loading factor* yang lebih besar dari 0,3 dengan nilai *t-value* lebih besar dari 1,96 sehingga 16 butir tersebut dinyatakan valid. Dengan demikian, maka dapat disimpulkan bahwa instrumen yang telah disusun dan diuji empiris tersebut dinyatakan valid untuk digunakan untuk mengukur *outcome* lembaga pendidikan vokasional.

Saran

Instrumen ini dapat digunakan sebagai alat ukur yang digunakan untuk mengungkap *outcome* lembaga pendidikan guru vokasional. Untuk pengembangan penelitian lebih lanjut perlu dilakukan uji empirik

dengan jumlah sampel yang lebih besar agar diperoleh validitas konstrak yang lebih baik.

Daftar Pustaka

- Allen, T., & Tinkew, J.B. (2008). *Outcome evaluations: A guide for out-of-school time practitioners: Part 4 in a series on practical evaluation methods*. Retrieved March 27, 2015, from http://www.childtrends.org/wp-content/uploads/2008/01/Child_Trends-2008_01_07_OutcomeEvaluation.pdf
- Astin, A.W.(1993). *Assessment for excellence. The philosophy and practice of assessment and evaluation in higher education*. New York: The Oryx Press.
- Azwar, S. (2014). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models*. Newbury Park, CA: Sage.136–162
- Cronbach, L. J. & Meel, P. E. (1995): Construct validity in psychological test, *Psychological Bulletin*, 52, 281-302
- Direktorat Tenaga Kependidikan. (2008). *Penilaian kinerja guru*. Jakarta: Departemen Pendidikan Nasional.
- Fernandes, H. J. X. (1984). *Evaluation of educational programs*. Jakarta: National Education Evaluation and Curriculum Development.
- Hair, Jr. J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis*. (6th ed.). New Jersey: Person Prentice Hall.
- Issac, S., & Michael, W. B. (1981). *Handbook in research and evaluation*. (2nd ed.). California: Edits Publisher.
- Lunenburg, F. C. & Ornstein, A. C. (2000). *Educational administration: concept and*

- practices*. (3th ed.). Belmont: Wadsworth/Thomson Learning.
- Mardapi, D. (2005). *Pengembangan instrumen penelitian pendidikan*. Yogyakarta: Pascasarjana UNY.
- Myers, P. & Barner, J. (2005). *Measuring Outcomes: guidance on outcome evaluation for Sure Start Local Programmes*. NESS. Institute for the Study of Children, Families and Social Issues, Birkbeck, University of London
- Rossi, P. H., Lipsey, M.W., & Freeman, H.E. (2004). *Evaluation: A systematic Approach*. (7th ed.). London: SAGE Publications.
- Sihvonen, R. J. (1999). *A framework for Evaluating Educational Outcome in Finland*. Helsinki: National Board of Education.
- Supriadi, D. (2003). *Guru di Indonesia: pendidikan, pelatihan dan perjuangan sejak zaman colonial hingga era reformasi*. Jakarta: Departemen Pendidikan Nasional, Direktorat Jenderal Pendidikan Dasar dan Menengah, Direktorat Tenaga Kependidikan.
- Suranto, S., Muhyadi, M., & Mardapi, D. (2014). Pengembangan instrumen evaluasi uji kompetensi keahlian (UKK) administrasi perkantoran di SMK. *Jurnal Penelitian dan Evaluasi Pendidikan*, 18(1), 98-114. Retrieved from <http://journal.uny.ac.id/index.php/jpep/article/view/2127>
- Trochim, W. M. K. (2006). *Introduction to evaluation*. Retrieved March 27, 2015, from <http://www.socialresearchmethods.net/kb/interval.php>
- Wijayanto, S. H. (2008). *Structural Equation modeling dengan lisrel 8.8: konsep dan tutorial*. Yogyakarta: Graha Ilmu.
- Wijayati, P., Suyata, S., & Sumarno, S. (2013). Model evaluasi pembelajaran berbasis kaizen di sekolah menengah atas. *Jurnal Penelitian dan Evaluasi Pendidikan*, 17(2), 318-332. Retrieved from <http://journal.uny.ac.id/index.php/jpep/article/view/1703>
- World Health Organization. (2000). *Outcome evaluations: evaluation of psychoactive substance use disorder treatment*.