

PENGEMBANGAN INSTRUMEN PENILAIAN KINERJA SISWA PADA PRAKTIKUM FISIKA GETARAN DI SMP

Rhys Mauludin Almuflichan
SMP Negeri 1 Tana Lia, Kabupaten
Tana Tidung

Awaluddin Tjalla
Fakultas Ilmu Pendidikan
Universitas Negeri Jakarta

ABSTRACT

The purpose of this research is to develop an instrument to measure the performance of junior high school (SMP) students on the topic "Vibration" in Physics experiments. The technique used for choosing the sample was multistage random sampling. The development of statement of items in the instrument was based on the judgments of experts and panelists. The measurement scale of the instrument is Rating Scale. The result of experts' judgments produced 20 items for the instrument. The instrument was tested twice to 190 junior high school students in Babat District, Lamongan Regency, East Java Province. The data was analyzed by using Structured Equation Modelling (SEM). First try out (testing) resulted in the exclusion of 2 items. Based on appropriate model, second try out resulted in no exclusion of items, hence the instrument is found to be fit (model fit). As two items got dropped in the first try out, the final instrument consists of total 18 items. After having tested the validity and reliability of the instrument, it can be concluded that the instrument has an appropriate construct validity and reliability. Nevertheless further try outs are still needed to standardize the instrument.

Keywords

Instrument development, Structured Equation Modeling (SEM), performance assessment of vibration experiment

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sebuah instrumen yang mengukur kinerja siswa SMP pada kegiatan praktikum Fisika pokok bahasan getaran. Teknik pengambilan sampel pada penelitian ini adalah acak bertahap (*multistage random sampling*). Pengembangan butir pernyataan berdasarkan uji teoretik pada pakar dan panelis. Skala yang digunakan pada penelitian ini adalah *rating scale*. Berdasarkan uji pakar dan panelis, menghasilkan 20 butir pernyataan. Uji coba dilakukan dua tahap pada 190 siswa SMP di Kecamatan Babat, Kabupaten Lamongan, Provinsi Jawa Timur. Analisis data menggunakan model persamaan struktural (SEM). Hasil uji coba pertama, butir yang direduksi sebanyak 2 butir. Setelah sebelumnya model diperbaiki (*respesifikasi*), pada uji coba kedua berdasarkan uji kecocokan model, sudah memenuhi kriteria sebagai model *fit* dan tidak perlu diperbaiki lagi. Dengan butir yang direduksi sebanyak 2 butir pernyataan, sehingga instrumen akhir terdiri dari 18 butir. Setelah dilakukan uji validitas dan reliabilitas, maka dapat disimpulkan instrumen ini sudah valid dan reliabel. Uji coba selanjutnya masih diperlukan dengan jumlah sampel yang lebih banyak dan lebih beragam untuk menghasilkan instrumen yang baku.

Kata Kunci

Pengembangan instrumen, model persamaan struktural (SEM), penilaian kinerja praktikum getaran

Alamat Korespondensi

SMP Negeri 1 Tana Lia, Kabupaten
Tana Tidung, Kalimantan Utara
Indonesia
e-mail:
rhysm.s.pd@gmail.com

1. Pendahuluan

Pada pengembangan kurikulum KTSP (2006) untuk kompetensi masa depan adalah bahwa pelaksanaan penilaian mengacu pada Standar Penilaian Pendidikan. Salah satu prinsip penilaian yang tercantum dalam Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 tentang Standar Penilaian Pendidikan adalah menyeluruh dan berkesinambungan. Menyeluruh

berarti penilaian oleh guru mencakup semua aspek kompetensi (aspek kognitif, aspek psikomotor, dan aspek afektif) dengan menggunakan berbagai teknik penilaian yang sesuai. Berkesinambungan artinya penilaian dilakukan untuk memantau perkembangan kemampuan siswa.

Mata pelajaran IPA merupakan bagian dari pengembangan kurikulum KTSP (2006) dengan

pelaksanaan penilaian mengacu pada Standar Penilaian Pendidikan. Namun pada kenyataannya, penerapan kurikulum di sekolah kurang memperhatikan hal tersebut, sehingga kualitas pendidikan di Indonesia khususnya mata pelajaran IPA tergolong masih rendah. Hal ini dibuktikan dari hasil penelitian suatu badan internasional, *Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS)* tahun 2011, menunjukkan bahwa posisi prestasi IPA di Indonesia masih sangat rendah yang berada pada urutan ke 32 dari 32 Negara peserta.

Peneliti juga melakukan observasi di beberapa sekolah khususnya di SMP Negeri se-Kecamatan Babat. Hasil observasi menunjukkan bahwa cara atau teknik untuk menguji kemampuan berpikir kritis siswa, keterampilan (dalam IPA misalnya yang terkait dengan penelitian, percobaan atau praktikum) dan sikap jarang digunakan oleh guru. Selama ini, kemampuan-kemampuan tersebut dinilai melalui tes ataupun laporan tertulis dengan rambu-rambu penilaian yang kurang jelas dan belum konsisten. Hal ini didukung dengan pendapat Hamid bahwa sistem penilaian yang digunakan dalam pembelajaran IPA masih didominasi dengan penilaian *paper and pencil test*, sementara kinerja siswa maupun penilaian diri oleh siswa tidak pernah dilakukan oleh guru (Arifin, 2009).

Dampak dari penilaian yang menitikberatkan pada satu sisi kemampuan, misalnya pada pengetahuan saja, akan menyebabkan siswa beranggapan bahwa kemampuan lain yang tidak diuji tidak penting untuk dipelajari, dipahami, dan diterapkan. Hal tersebut juga mampu mengurangi motivasi dan keseriusan belajar siswa. Oleh karena itu, penilaian dari berbagai aspek kemampuan yang diajarkan termasuk kemampuan penelitian, percobaan dan praktikum, perlu diuji dan dinilai agar dipandang penting dan dipelajari secara serius oleh siswa.

Penilaian kinerja pada kegiatan praktikum merupakan bagian dari proses pembelajaran yang bertujuan agar siswa mendapatkan kesempatan untuk menguji dan melaksanakan dalam keadaan nyata apa yang diperoleh dalam teori. Kegiatan praktikum merupakan latihan aktivitas ilmiah baik berupa eksperimen, observasi, maupun demonstrasi yang menunjukkan adanya

ketertarikan antara teori dengan fenomena yang dilaksanakan baik di laboratorium maupun di luar laboratorium (Ardli, Mujdalipah, dan Ana, 2003). Kegiatan praktikum juga dapat memberikan pengalaman belajar secara nyata kepada peserta didik dengan mengembangkan keterampilan dasar bekerja di laboratorium seperti *scientist*, serta memberikan siswa kesempatan untuk berpartisipasi aktif sehingga memperoleh informasi dan kecakapan sains dengan cara observasi.

Penilaian kinerja merupakan ragam penilaian yang cukup luas yang menggambarkan seluruh kemampuan berpikir siswa semenjak awal kegiatan pembelajaran, kemampuan siswa bekerja selama proses pembelajaran, dan kemampuan siswa memahami di akhir pembelajaran. Berkaitan dengan definisi ini, penilaian kinerja senantiasa menggambarkan: (1) kebebasan siswa menentukan tugas yang akan dilakukan, (2) tugas yang menentukan siswa mengkolaborasi penggunaan proses belajar sebagai kunci dalam memahami materi inti pembelajaran, (3) tugas yang dirancang bukan hanya dapat dinilai guru melainkan dinilai orang lain (orang tua dan masyarakat), sistem penilaian yang eksplisit, dan proses pengukuran yang akurat sejalan dengan tugas terencana yang dibuat (Abidin, 2014).

Menurut Popham penilaian kinerja setidaknya memiliki tiga karakteristik umum, yaitu: (1) multi kriteria, kinerja siswa harus menggunakan yang memiliki lebih dari satu kriteria, (2) standar kualitas yang spesifik, masing-masing kriteria kinerja siswa dapat dinilai secara jelas dan eksplisit dalam memajukan evaluasi kualitas kinerja siswa, dan (3) adanya *judgment* penilaian, asesmen kinerja membutuhkan penilaian yang bersifat manusiawi untuk menilai bagaimana kinerja siswa dapat diterima secara nyata (*real*), bukan menilai dengan menggunakan angka pada komputer atau mesin (seperti pada tes baku) (Abidin, 2014).

Asesmen kinerja adalah suatu prosedur yang menggunakan berbagai bentuk tugas-tugas untuk memperoleh informasi tentang apa dan sejauh mana yang telah dipelajari siswa. Asesmen kinerja mensyaratkan siswa dalam menyelesaikan tugas-tugas kinerjanya menggunakan pengetahuan dan keterampilannya yang diwujudkan dalam bentuk perbuatan, tindakan, atau unjuk kerja. Tes unjuk

kerja meminta siswa mewujudkan tugas sebenarnya yang mewakili keseluruhan kinerja yang akan dinilai, seperti mempersiapkan alat, menggunakan alat/merangkai alat, menuliskan data, menganalisis data, menyimpulkan, menyusun laporan, dan sebagainya. Secara khusus penilaian kinerja menjelaskan kemampuan-kemampuan siswa, pemahaman konseptual, kemampuan untuk menerapkan pengetahuan dan keterampilan, kemampuan melaksanakan kinerja dan kemampuan melakukan suatu proses (Susila, 2012).

Penilaian kinerja adalah proses mengumpulkan data dengan cara pengamatan yang sistematis untuk membuat keputusan tentang individu. Penilaian kinerja terutama sangat sesuai dalam menilai keterampilan. Keterampilan peserta didik yang dapat dinilai meliputi keterampilan proses intelektual (seperti keterampilan observasi, berhipotesis, menerapkan konsep, merencanakan, serta melakukan penelitian, dan lain-lain). Penilaian kinerja sangat tepat bila digunakan dalam kegiatan praktikum. Bentuk penilaian kinerja, yaitu: kinerja klasikal, asesmen kinerja kelompok, dan asesmen kinerja personal (Ardli, Mujdalipah, dan Ana, 2003).

Tugas kinerja (*performance task*) adalah strategi penilaian dalam hal mana para siswa menciptakan, menghasilkan, melaksanakan, dan menghadirkan, dengan suatu cara yang melibatkan dunia nyata yang bermakna dan berkaitan dengan isu-isu atau masalah substantif, dalam upaya mempertunjukkan keterampilan atau kemahiran siswa (Basuki dan Hariyanto).

Penilaian kinerja dilakukan untuk menilai ketercapaian kompetensi yang menuntut siswa menunjukkan kinerjanya. Penilaian ini dilakukan dengan mengamati kegiatan siswa dalam melakukan sesuatu. Cara penilaian ini dianggap lebih autentik dari pada tes tertulis karena apa yang dinilai lebih mencerminkan kemampuan peserta didik yang sebenarnya. Untuk mata pelajaran IPA di SMP atau Fisika di SMA, penilaian semacam ini dapat dilakukan melalui kegiatan seperti pengujian atau penelitian, melakukan percobaan-percobaan, dan lain-lain. Dalam penilaian kinerja perlu dipertimbangkan hal-hal berikut: (1) identifikasi langkah-langkah kinerja yang diharapkan sesuai dengan tuntutan kompetensi, (2) kelengkapan dan ketepatan aspek

yang akan dinilai dalam kinerja tersebut, (3) upayakan kemampuan yang dinilai tidak terlalu banyak agar dapat diamati, dan (4) kemampuan yang dinilai diurutkan berdasarkan urutan yang diamati (Kuswanto, 2008).

IPA merupakan suatu mata pelajaran yang memberikan kesempatan kepada para siswa untuk berpikir kritis melalui kegiatan-kegiatan yang dilakukan sendiri oleh siswa. Kegiatan-kegiatan yang dimaksud dapat dilangsungkan di dalam laboratorium maupun di luar laboratorium seperti di kelas atau di alam terbuka, berkaitan dengan suatu bidang ilmu tertentu yang ditujukan untuk menunjang pembelajaran teori. Proses belajar mengajar yang demikian disebut sebagai kegiatan praktikum. Melalui kegiatan praktikum ini siswa diberi kesempatan untuk mengalami sendiri dalam mencari kebenaran, atau mencoba mencari suatu hukum atau dalil, dan menarik kesimpulan atas proses yang dialaminya itu.

Menurut Rustaman setidaknya ada 4 alasan yang menguatkan peran laboratorium dalam pembelajaran di sekolah, yaitu: (1) praktikum membangkitkan motivasi belajar IPA, (2) praktikum mengembangkan keterampilan dasar melakukan eksperimen, (3) praktikum menjadi wahana belajar pendekatan ilmiah, dan (4) praktikum menunjang materi pelajaran. Praktikum memberikan kesempatan bagi siswa untuk menemukan teori dan membuktikan teori (Parmin, dkk, 2008).

Laboratorium sains berperan penting dalam kegiatan pembelajaran, yakni dengan menumbuhkan dan mengembangkan aspek-aspek antara lain: (1) keterampilan dalam pengamatan, pengukuran, dan pengumpulan data, (2) kemampuan menyusun data dan menganalisis serta menafsirkan hasil pengamatan, (3) kemampuan menarik kesimpulan secara logis berdasarkan hasil eksperimen, mengembangkan model, dan menyusun teori, (4) kemampuan mengkomunikasikan secara jelas dan lengkap hasil-hasil percobaan, (5) keterampilan merancang percobaan, urutan kerja, dan pelaksanaannya, (6) keterampilan dalam memilih dan mempersiapkan peralatan dan bahan untuk percobaan, (7) keterampilan dalam menggunakan peralatan dan bahan, dan (8) kedisiplinan dalam mematuhi aturan dan tata tertib demi keselamatan kerja (Parmin, dkk).

Berdasarkan kajian teori yang disintesis menjadi konstruk, maka didapat dimensi dan indikator alat ukur penilaian kinerja siswa SMP pada kegiatan praktikum IPA (Fisika). Instrumen penilaian kinerja ini terdiri dari 3 dimensi dan 9 indikator. Dimensi pertama adalah kemampuan mempersiapkan kegiatan praktikum yang terdiri dari 3 indikator, yaitu: merencanakan kegiatan, mempersiapkan alat dan bahan, dan memperhatikan keselamatan kerja. Dimensi kedua adalah kemampuan melaksanakan kegiatan praktikum yang terdiri dari 4 indikator, yaitu: mengoperasikan/merangkai alat dan bahan, melakukan pengamatan, mencatat data hasil pengamatan, dan menganalisis data hasil pengamatan. Dimensi ketiga adalah kemampuan menyelesaikan atau megakhiri kegiatan praktikum yang terdiri dari 2 indikator, yaitu: kedisiplinan dalam kegiatan praktikum dan membereskan alat, bahan dan tempat setelah kegiatan.

Pada penelitian ini, analisis data menggunakan model persamaan struktural (SEM) dengan prosedur CFA. SEM merupakan kombinasi metodologi dua disiplin ilmu, yaitu: model analisis faktor konfirmatori yang diambil dari *psychometric* dan model persamaan *structural* (SEM) yang diambil dari *econometrics* (Latan, 2012)

Menurut Bollen dan Long terdapat 5 (lima) proses yang harus dilalui dalam analisis SEM, di mana setiap tahapan akan berpengaruh terhadap tahapan selanjutnya, yaitu: spesifikasi model, identifikasi model, estimasi model, evaluasi model, dan respesifikasi model (Latan, 2012).

Saat ini perkembangan piranti lunak perangkat komputer dapat mempermudah analisis SEM ini dengan memanfaatkan piranti lunak (*software*) LISREL (*Linear Structural Relationship*) yang merupakan *software* SEM yang pertama. LISREL diperkenalkan pertama kali oleh Karl Joreskog, kemudian beliau mengembangkan dan menyempurnakan piranti lunak ini bersama Dag Sorbom.

Program LISREL terus mengalami perkembangan dan penyempurnaan hingga saat ini LISREL 8.0. dengan bahasa pemrograman SIMPLIS, maka lebih dipermudah lagi karena dapat menghasilkan keluaran (*output*) dalam bentuk SIMPLIS, LISREL, dan kombinasi antara SIMPLIS dan LISREL. Keluaran program LISREL ini dihasilkan sesuai dengan perintah pada sintaknya.

Keluaran pada LISREL dengan bahasa SIMPLIS ada dua macam, yaitu: *printed output* yang berupa tulisan, tanpa tergantung apakah program ini berhasil atau tidak dan diagram lintasan (*part diagram*) jika program ini berhasil, dan tidak ada gambaran yang ditampilkan di layar jika program gagal.

Analisis terhadap program SIMPLIS ini yang kita periksa adalah: *offending estimate*, seperti *negative error variance* dan *standardized loading factor* yang lebih dari 0,1 serta nilai *standar error* yang sangat besar. Bila ada salah satu dari nilai-nilai *offending estimate* yang tidak sesuai kriteria, maka dilakukan respesifikasi model, jika tidak ada maka dilanjutkan dengan analisis validitas model pengukuran, yaitu: dengan memeriksa *t-value* dari *standardized loading factor* (λ) dari variabel yang teramati dalam model. Bila $< 1,96$, maka dilakukan respesifikasi model, jika sebaliknya, maka dilanjutkan dengan melihat *standardized loading factor* (λ) dari variabel yang teramati dalam model. Menurut Igbaria et al jika $< 0,3$, maka dilakukan respesifikasi model, jika nilai tersebut lebih dari 0,30, maka dilanjutkan dengan uji kecocokan model. Bila sebagian besar nilai pada ukuran-ukuran kecocokan model tidak memenuhi kriteria nilai *cut off*, maka dilakukan respesifikasi model. Terakhir, dilanjutkan dengan analisis reliabilitas model pengukuran, dengan menghitung nilai *construct reliability* (CR) dan *error variance* (VE). Model dikatakan reliabel jika nilai $CR \geq 0,70$ dan $VE \geq 0,50$. Jika di bawah kedua nilai tersebut, maka model harus diperbaiki karena tidak reliabel (Wijanto, 2008).

2. Metode Penelitian

Secara umum, penelitian ini bertujuan untuk mengembangkan instrumen penilaian kinerja siswa pada kegiatan praktikum Fisika di SMP yang valid dan reliabel. Tujuan umum ini dapat dirinci menjadi tujuan yang lebih operasional, sebagai berikut: (1) menjabarkan dimensi dan indikator yang mendasari penilaian kinerja pada praktikum Fisika di SMP; (2) mengkonstruksi instrumen penilaian kinerja pada praktikum Fisika di SMP yang valid; dan (3) mengkonstruksi instrumen penilaian kinerja pada praktikum Fisika di SMP yang reliabel.

Berdasarkan hasil telaah teori yang terkait dengan kinerja siswa SMP pada kegiatan praktikum, maka dapat dirumuskan kisi-kisi instrumen sebagai berikut.

Tabel 1. Kisi-kisi Instrumen

Dimensi	Indikator	Deskriptor
Kemampuan mempersiapkan kegiatan praktikum	Merencanakan kegiatan	Membaca dan memahami prosedur kerja Mengecek peralatan praktikum
	Mempersiapkan alat dan bahan	Memilih alat praktikum Memilih bahan praktikum
	Memperhatikan keselamatan kerja	Keselamatan alat dan bahan Keselamatan manusia
Kemampuan melaksanakan kegiatan praktikum	Mengoperasikan/ merangkai alat dan bahan	Mengukur massa beban Memberikan simpangan Mengukur waktu
	Melakukan Pengamatan	Mengamati sistem bandul saat bergetar harmonis Mengamati waktu
	Mencatat data hasil pengamatan	Mencatat massa beban Mencatat waktu
	Menganalisis hasil	Melakukan perhitungan Membuat kesimpulan Menyelesaikan permasalahan
Kemampuan menyelesaikan/mengakhiri kegiatan praktikum	Kedisiplinan dalam kegiatan praktikum	Memfokuskan perhatian pada kegiatan praktikum Memiliki minat/interes terhadap aktivitas praktikum
	Membereskan alat, bahan, dan tempat setelah kegiatan	Membersihkan alat dan bahan Membersihkan tempat praktikum

Uji validitas teoretis pada instrumen dilakukan oleh 3 orang pakar dan 30 panelis. Dalam uji kesesuaian yang dilakukan pakar dilakukan analisis kualitatif terhadap kesesuaian instrumen yang didasarkan pada tiga hal pokok yang dinilai, yaitu: (1) kesesuaian variabel dengan dimensi yang dikembangkan, (2) kesesuaian indikator dengan dimensi yang dikembangkan, dan (3) kesesuaian deskriptor dengan indikator kinerja yang dikembangkan. Sedangkan untuk uji panelis, penilaian tingkat validitas setiap butir ditentukan berdasarkan nilai median dengan menggunakan skala *semantic differential*, di mana para panelis menilai kesesuaian menggunakan skala 1 sampai 9 dan pada uji kesesuaian standar kinerja butir dinilai berdasarkan 4 kriteria yang didasarkan pada nilai mean, yaitu: jika nilai mean berada pada rentang 1 sampai 1,9 maka butir tersebut digugurkan, jika nilai mean berada pada rentang 2 sampai 3 maka butir tersebut diperbaiki konsep

dan bahasa, jika nilai mean 3,1 sampai 4 maka butir tersebut diperbaiki bahasa, dan jika nilai mean 4,1 sampai 5 maka butir tersebut baik atau sangat baik. Uji reliabilitas menggunakan rumus reliabilitas Hoyt (reliabilitas interrater).

Pada uji coba empiris dilakukan 2 kali, dengan masing-masing uji coba untuk tahap pertama adalah 100 responden. Karena jumlah butir pada uji coba pertama adalah 20 butir, maka $5 \times 20 = 100$ responden. Responden pada uji coba pertama adalah siswa SMP Negeri 1 Babat dan SMP Negeri 2 Babat. Sedangkan uji coba kedua berasal dari SMP Negeri 3 Babat dan SMP Negeri 4 Babat, Kecamatan Babat, Kabupaten Lamongan, Provinsi Jawa Timur.

Setelah diperoleh data hasil uji coba, selanjutnya dilakukan analisis daya beda atau daya diskriminasi butir sekaligus seleksi butir dengan menggunakan rumus korelasi butir total *product moment pearson*.

Untuk menentukan validitas konstruk uji coba empiris digunakan analisis dengan teknik analisis faktor melalui penggunaan metode SEM dengan perangkat lunak program LISREL. Terakhir, dilanjutkan dengan analisis reliabilitas model pengukuran, dengan menghitung nilai *construct reliability* (CR) dan *error variance* (VE).

3. Hasil Penelitian dan Pembahasan

Uji validitas teoritis pada instrumen dilakukan oleh 3 orang pakar dan 30 panelis yang dimaksudkan untuk menguji kesesuaian antara variabel, dimensi, indikator, dan deskriptor. Hasil telaah pakar yang telah direvisi ini didapat 20 butir dengan 3 dimensi dan 9 indikator serta 100 standar kinerja yang dikembangkan dari deskriptor yang selanjutnya akan dilakukan uji teoritis tahap kedua.

Selanjutnya, setelah *draft* instrumen direvisi kemudian dilakukan uji coba validitas teoritis kepada panelis sebanyak 30 orang. Pada kegiatan validasi teoritis, butir pernyataan dalam instrumen dinilai berdasarkan 5 kriteria, yaitu: skor 1 dengan

kriteria sangat buruk (konsep salah atau konsep tidak ada hubungan dengan deskriptor), skor 2 dengan kriteria buruk (konsep benar, rumusan bahasa salah, bukan kinerja inti deskriptor), skor 3 dengan kriteria cukup (konsep benar, rumusan bahas salah, kinerja inti deskriptor), skor 4 dengan kriteria baik (konsep benar, rumusan bahasa benar, bukan kinerja inti deskriptor) dan skor 5 dengan kriteria sangat baik (konsep benar, rumusan bahasa benar, kinerja inti deskriptor). Hasil uji coba panelis, bahwa 20 butir dengan 3 dimensi dan 9 indikator serta 100 standar kinerja menunjukkan semua butir valid dan dapat dipakai untuk uji coba empiris.

Pada uji coba empiris pertama dengan jumlah 20 butir, 3 dimensi dan 9 indikator serta 100 standar kinerja. Hasil uji kecocokan model (*goodness of fit*) konstruk instrumen penilaian kinerja siswa pada praktikum fisika getaran dengan menggunakan *Second Order Confirmatory Factor Analysis* (2ndCFA) sesudah adanya respesifikasi, maka dapat dibuat ringkasan seperti pada Tabel 2.

Tabel 2. Ringkasan Hasil Uji Kecocokan Model (*Goodness of Fit*) Konstruk Instrumen Penilaian Kinerja Siswa pada Praktikum Fisika Getaran dengan Menggunakan *Second Order CFA* Sebelum Respesifikasi Model pada Uji Coba I

Ukuran <i>Goodness of Fit</i>	Nilai	Tingkat Kecocokan yang dapat diterima	Keterangan
A. Absolute Fit Measures			
1. Chi Square	136,22 (p=0.00)	rendah (p ≥ 0,05)	Kurang Fit
2. GFI	0,77	≥ 0,90	Kurang Fit
3. RMSEA	0,22	< 0,08	Kurang Fit
4. RMR	0,44	< 0,05	Good Fit
B. Incremental Fit Measures			
1. AGFI	0,56	0,80 ≤ AGFI < 0,90	Kurang Fit
2. NFI	0,76	≥ 0,90	Kurang Fit
3. NNFI	0,69	≥ 0,90	Kurang Fit
4. CFI	0,79	≥ 0,90	Kurang Fit
5. IFI	0,80	≥ 0,90	Kurang Fit
6. RFI	0,65	≥ 0,90	Kurang Fit
C. Parsimonious Fit Measures			
1. AIC	178,22	< 90, 00 (<i>Saturated</i>) dan < 585,22 (<i>Independence Model</i>)	Kurang Fit
2. CAIC	253,93	< 252,23 (<i>Saturated</i>) dan < 617,70 (<i>Independence Model</i>)	Kurang Fit
3. ECVI	1,80	< 0,96 (<i>Saturated model</i>) dan < 5,91 (<i>Independence Model</i>)	Kurang Fit
4. PGFI	0,41	> 0,60	Kurang Fit

Pada Tabel 2, dapat dilihat bahwa hasil uji kecocokan model pada konstruk kinerja siswa SMP hanya 1 ukuran GOF (kecocokan model) yang memenuhi kriteria nilai *cut off* yang dipersyaratkan, yaitu: RMR. Sedangkan 13 ukuran GOF (kecocokan model) tidak memenuhi kriteria nilai *cut off* yang dipersyaratkan untuk model *fit* (tepat). Sehingga dapat disimpulkan bahwa kecocokan keseluruhan model pada uji

coba I untuk konstruk kinerja siswa SMP belum bisa dikatakan *fit*. Maka dalam mendapatkan model *fit*, perlu dilakukan perbaikan model (respesifikasi model) dengan memanfaatkan saran *modification indeks* pada hasil data LISREL, yang dapat menurunkan nilai chi-kuadrat. Hasil uji kecocokan model yang sudah diperbaiki dapat dilihat pada Tabel 3.

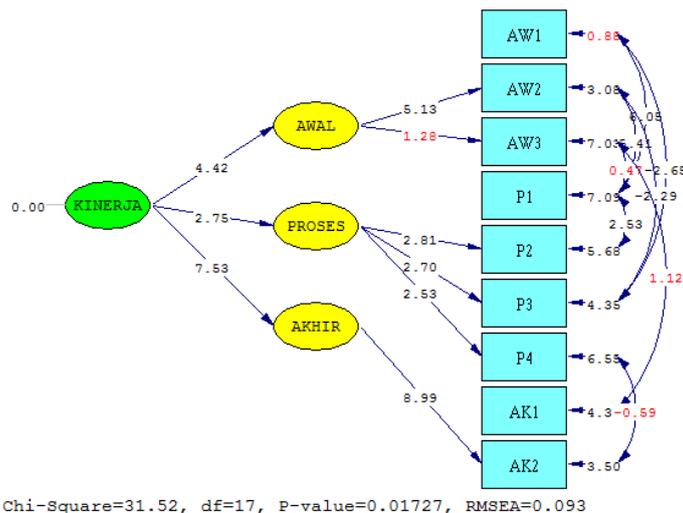
Tabel 3. Ringkasan Hasil Uji Kecocokan Model (*Goodness of Fit*) Konstruk Instrumen Penilaian Kinerja Siswa pada Praktikum Fisika Getaran dengan Menggunakan *Second Order CFA* Sesudah Respesifikasi pada Uji Coba I

Ukuran <i>Goodness of Fit</i>	Nilai	Tingkat Kecocokan yang dapat diterima	Keterangan
A. Absolute Fit Measures			
1. <i>Chi Square</i>	31,52 ($p=0.017$)	rendah ($p \geq 0,05$)	Kurang <i>Fit</i>
2. GFI	0,93	$\geq 0,90$	<i>Good Fit</i>
3. RMSEA	0,093	$< 0,08$	Kurang <i>Fit</i>
4. RMR	0,22	$< 0,05$	Kurang <i>Fit</i>
B. Incremental Fit Measures			
5. AGFI	0,83	$0,80 \leq AGFI < 0,90$	<i>Marginal Fit</i>
6. NFI	0,94	$\geq 0,90$	<i>Good Fit</i>
7. NNFI	0,93	$\geq 0,90$	<i>Good Fit</i>
8. CFI	0,97	$\geq 0,90$	<i>Good Fit</i>
9. IFI	0,97	$\geq 0,90$	<i>Good Fit</i>
10. RFI	0,87	$\geq 0,90$	Kurang <i>Fit</i>
C. Parsimonious Fit Measures			
1. AIC	87,52	$< 90,00$ (<i>Saturated</i>) dan $< 585,25$ (<i>Independence Model</i>)	<i>Good Fit</i>
2. CAIC	188,46	$< 252,23$ (<i>Saturated</i>) dan $< 617,70$ (<i>Independence Model</i>)	<i>Good Fit</i>
3. ECVI	0,88	$< 0,91$ (<i>Saturated model</i>) dan $< 5,91$ (<i>Independence Model</i>)	<i>Good Fit</i>
4. PGFI	0,35	$> 0,60$	Kurang <i>Fit</i>

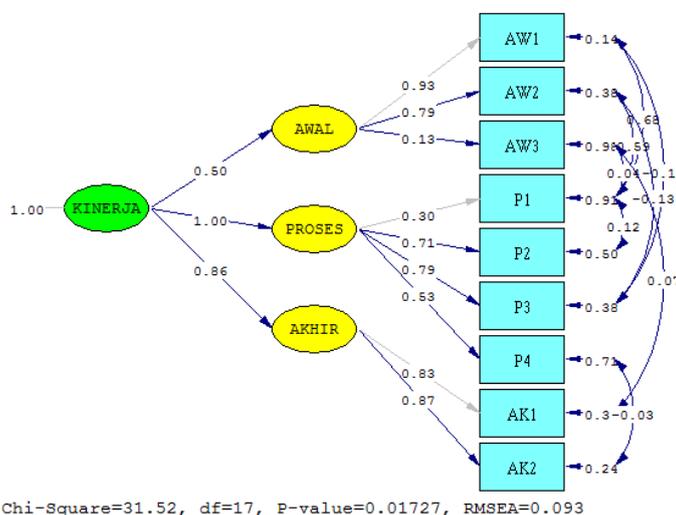
Pada Tabel 3, dapat dilihat bahwa terdapat 5 ukuran GOF yang menunjukkan hasil kecocokan yang kurang baik, yaitu: *chi-square*, RMSEA, RMR, RFI, PGFI, dan 1 ukuran GOF yang menunjukkan hasil kecocokan yang mendekati baik (*marginal fit*), yaitu: AGFI, dan 9 ukuran GOF yang menunjukkan kecocokan yang baik. Hal tersebut menandakan bahwa walaupun terdapat beberapa ukuran GOF yang menunjukkan kecocokan yang kurang baik, namun sebagian

besar ukuran GOF menunjukkan kecocokan yang baik sehingga dapat disimpulkan bahwa kecocokan keseluruhan model adalah baik (*model fit*).

Hasil estimasi *Second Order Confirmatory Factor Analysis* (2ndCFA) pada uji coba I dapat ditunjukkan dengan Part Diagram *Standardized Solution* dan Part Diagram *T-value* tampak pada Gambar 1 dan Gambar 2.



Gambar 1. Part Diagram Standardized Solution pada Uji Coba I



Gambar 2. Part Diagram T-Value pada Uji Coba I

Validitas korvergen dapat dilihat dari nilai *standardized loading factor* (terlihat pada Gambar 1) tiap-tiap indikator dari masing-masing dimensi dengan kriteria, yaitu: nilai *standardized loading factor* di atas 0,30 dapat diterima. Nilai

standardized loading factor dan nilai t_{hitung} atau *T-value* (terlihat pada Gambar 2) masing-masing indikator pada tiap-tiap dimensi dapat dilihat pada Tabel 4.

Tabel 4. Ringkasan Nilai *Standardized Loading Factor* (SLF) dan t_{hitung} (*T-value*) Masing-Masing Indikator Tiap-Tiap Variabel pada Model Hasil Uji Coba I

Nama Dimensi	Indikator	SLF (Standardized Loading Factor)	t_{hitung} (T-value)	R ²	Keterangan
AWAL	AW1	0,91	-	0,86	Valid
	AW2	0,79	5,13	0,62	Valid
	AW3	0,13	1,28	0,018	Tidak Valid
PROSES	P1	0,30	-	0,089	Valid
	P2	0,71	2,81	0,50	Valid
	P3	0,79	2,70	0,62	Valid
	P4	0,53	2,53	0,29	Valid
AKHIR	AK1	0,83	-	0,70	Valid
	AK2	0,87	8,99	0,76	Valid

Berdasarkan ringkasan data pada Tabel 4 tersebut, diketahui bahwa hasil analisis faktor dengan *Second Order Confirmatory Factor Analysis* (2ndCFA) yang dilakukan terhadap 9 indikator, 8 indikator yang menghasilkan muatan faktor atau *standardized loading factor* (λ) > 0,3 dan hanya 1 indikator yang tidak menghasilkan muatan faktor atau *standardized loading factor* (λ) < 0,3, yaitu: indikator AW3 yang juga memiliki nilai muatan faktor yang terendah dibanding indikator lainnya. Sehingga indikator AW3 (memperhatikan keselamatan kerja) memiliki hubungan yang tidak signifikan dengan konstruk kinerja siswa SMP. Hal ini mengingat nilai estimasi muatan faktor yang baik menunjukkan bahwa indikator penyusun dapat dengan baik menjelaskan variabel latennya. Maka indikator AW3 tidak diikutsertakan lagi

dalam uji empiris yang kedua. Selain itu, nilai muatan faktor di bawah 0,3 menunjukkan bahwa indikator tersebut tidak valid atau memiliki validitas yang rendah. Oleh karena itu, dapat disimpulkan bahwa untuk melakukan uji coba yang kedua hanya menggunakan 8 indikator dan 18 butir pernyataan dalam *draft* instrumen.

Setelah dilakukan uji validitas konstruk, maka langkah selanjutnya adalah melakukan uji reliabilitas konstruk atau *Construct Reliability* (CR) dan *Variance Extracted* (VE). Nilai CR dan VE sangat menentukan reliabel atau tidaknya suatu konstruk dan model pengukuran. Uji reliabilitas diperlukan juga untuk mengetahui akurasi, konsistensi, dan ketepatan instrumen dalam mengukur konstruk latennya. Hasil perhitungan nilai CR dan VE dapat dilihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Nilai *Construct Reliability* (CR) dan *Variance Extracted* (VE) Model Pengukuran Hasil Uji Coba I

Construct Reliability	Variance Extracted
0,9534	0,7274

Pada Tabel 5 tersebut, dapat dilihat bahwa nilai *Construct Reliability* (CR) sebesar 0,9534 dan nilai *Variance Extracted* (VE) sebesar 0,7247. Hal ini dapat dikatakan bahwa nilai CR dan CV telah memenuhi kriteria, dengan nilai CR > 0,7 dan VE > 0,5. Sehingga disimpulkan bahwa model pengukuran yang digunakan untuk mengukur kinerja siswa SMP dapat dipercaya dan mempunyai konsistensi yang baik. Maka model pengukuran yang digunakan untuk mengukur kinerja siswa SMP memiliki reliabilitas yang baik atau reliabel.

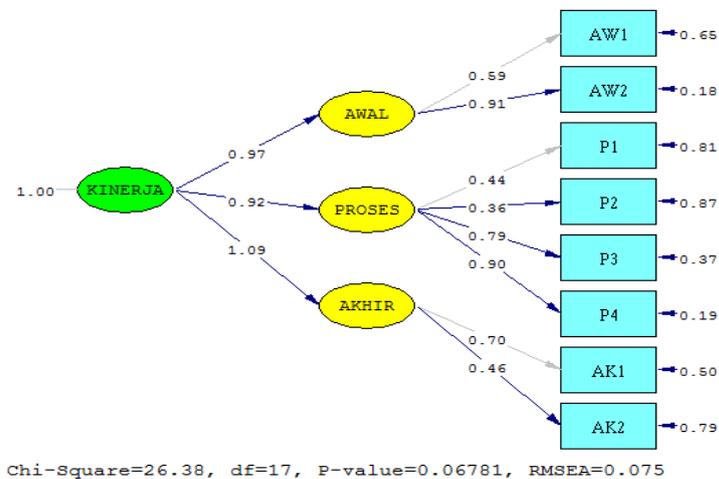
Berdasarkan hasil uji coba empiris pertama butir yang gugur ada 2 butir, sehingga menjadi 18 butir dengan 3 dimensi dan 8 indikator serta 90 standar kinerja. Selanjutnya, dilanjutkan uji coba empiris yang kedua, dengan hasil yang ditunjukkan pada Tabel 6.

Pada Tabel 6 tersebut, dapat dilihat bahwa terdapat 3 ukuran GOF yang menunjukkan hasil kecocokan kurang baik, yaitu: RMR dan PGFI. Kemudian, 1 ukuran GOF yang menunjukkan hasil kecocokan yang mendekati baik (*marginal fit*), yaitu: AGFI, dan 11 ukuran GOF yang menunjukkan kecocokan yang baik. Hal tersebut menandakan bahwa walaupun terdapat beberapa ukuran GOF yang menunjukkan kecocokan yang kurang baik, namun sebagian besar ukuran GOF menunjukkan kecocokan yang baik sehingga dapat disimpulkan bahwa kecocokan keseluruhan model adalah baik (*model fit*) dan tidak memerlukan untuk dilakukan perbaikan model (*respesifikasi model*).

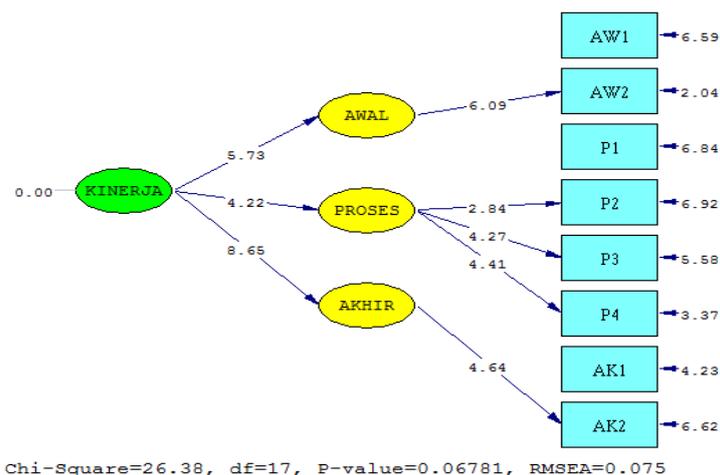
Hasil estimasi *second order CFA* ditampilkan dalam bentuk Part Diagram *Standardized Loading Factor* (SLF) dan *T-Value* yang dapat diperlihatkan pada Gambar 3 dan Gambar 4.

Tabel 6. Ringkasan Hasil Uji Kecocokan Model (*Goodness of Fit*) Konstruksi Instrumen Penilaian Kinerja Siswa Pada Praktikum Fisika Getaran dengan Menggunakan *Second Order CFA* pada Uji Coba 2

Ukuran <i>Goodness of Fit</i>	Nilai	Tingkat Kecocokan yang dapat diterima	Keterangan
Absolute Fit Measures			
<i>Chi Square</i>	26.38(p=0.068)	rendah ($p \geq 0,05$)	<i>Good Fit</i>
GFI	0,94	$\geq 0,90$	<i>Good Fit</i>
RMSEA	0,075	$< 0,08$	<i>Good Fit</i>
RMR	0,091	$< 0,05$	Kurang Fit
Incremental Fit Measures			
AGFI	0,87	$0,80 \leq AGFI < 0,90$	<i>Marginal Fit</i>
NFI	0,95	$\geq 0,90$	<i>Good Fit</i>
NNFI	0,97	$\geq 0,90$	<i>Good Fit</i>
CFI	0,98	$\geq 0,90$	<i>Good Fit</i>
IFI	0,98	$\geq 0,90$	<i>Good Fit</i>
RFI	0,92	$\geq 0,90$	<i>Good Fit</i>
Parsimonious Fit Measures			
AIC	64,36	$< 72,00$ (<i>Saturated</i>) dan $< 540,05$ (<i>Independence Model</i>)	<i>Good Fit</i>
CAIC	132,88	$< 201,79$ (<i>Saturated</i>) dan $< 568,89$ (<i>Independence Model</i>)	<i>Good Fit</i>
ECVI	0,65	$< 0,73$ (<i>Saturated model</i>) dan $< 6,45$ (<i>Independence Model</i>)	<i>Good Fit</i>
PGFI	0,44	$> 0,60$	Kurang Fit



Gambar 3. Part Diagram *Standardized Loading Factor (SLF)*



Gambar 4. Part Diagram T- value

Selanjutnya, untuk validitas konvergen dapat dilihat dari nilai *standardized loading factor* (terlihat pada Gambar 3) tiap-tiap indikator dari masing-masing dimensi dengan kriteria, yaitu: nilai *standardized loading factor* di atas 0,30 dapat

diterima. Nilai *standardized loading factor* dan nilai t-hitung atau *t-value* (terlihat pada Gambar 4) masing-masing indikator pada tiap-tiap dimensi dapat dilihat pada Tabel 7.

Tabel 7. Ringkasan Nilai *Standardized Loading Factor* (SLF) dan t_{hitung} (T-value) Masing-Masing Indikator Tiap-tiap Variabel pada Model Hasil Uji Coba 2

Nama Dimensi	Indikator	SLF (Standardized Loading Factor)	t_{hitung} (T-value)	R ²	Keterangan
AWAL	AW1	0,59	-	0,35	Valid
	AW2	0,91	6,09	0,82	Valid
PROSES	P1	0,44	-	0,19	Valid
	P2	0,36	2,84	0,13	Valid
	P3	0,79	4,27	0,63	Valid
	P4	0,90	4,41	0,81	Valid
AKHIR	AK1	0,70	-	0,50	Valid
	AK2	0,46	4,64	0,21	Valid

Berdasarkan ringkasan data pada Tabel 7 tersebut, diketahui bahwa hasil analisis faktor dengan *Second Order Confirmatory Factor Analysis* (2ndCFA) yang dilakukan terhadap 8 indikator, seluruhnya menghasilkan muatan faktor atau *standardized loading factor* (λ) > 0,3. Hal ini mengingat nilai estimasi muatan faktor yang baik menunjukkan bahwa indikator penyusun dapat dengan baik menjelaskan variabel latennya. Sehingga dapat disimpulkan bahwa seluruh indikator memiliki hubungan yang signifikan dengan konstruk kinerja siswa dan indikator tersebut dapat mengukur variabel latennya.

4. Kesimpulan

Berdasarkan hasil uji coba empiris yang telah dilakukan sebanyak dua kali, dapat disimpulkan bahwa instrumen penilaian kinerja terdiri dari 3 dimensi dengan 8 indikator dan 18 butir serta 90 standar kinerja sudah sesuai dengan teori yang membangun konstruk kinerja siswa SMP pada kegiatan praktikum.

Bersumber pada hasil uji kecocokan model dengan menggunakan model pengukuran *Second Order Confirmatory Analysis*, maka model akhir yang diperoleh sudah tepat atau cocok untuk mengukur kinerja siswa SMP pada kegiatan

praktikum. Dengan melihat nilai *goodness of fit* yang telah memenuhi kriteria nilai *cut off* yang dipersyaratkan. Indeks *goodness of fit* yang mengindikasikan model *fit* adalah nilai *chi-square* (χ^2) dan *p-value* (probabilitas), nilai GFI, nilai RMSEA, nilai NFI, nilai TLI/NNFI, nilai CFI, nilai IFI, nilai RFI, nilai AIC, nilai CAIC, dan nilai ECVI. Sehingga instrumen penilaian kinerja siswa SMP pada kegiatan praktikum dinyatakan sudah valid. Selain itu, instrumen penilaian kinerja siswa SMP pada kegiatan praktikum juga sudah reliabel, karena nilai *construct reliability* (CR) di atas nilai *cut off* ($>0,7$) dan *Varianced Extracted* (VE) di atas nilai *cut off* ($>0,5$). Jadi, instrumen penilaian kinerja siswa SMP pada kegiatan praktikum sudah valid dan reliabel, sehingga dapat digunakan sebagai alat ukur.

5. Daftar Pustaka

- Abidin, Yunus. (2014). *Desain Sistem Pembelajaran dalam Konteks Kurikulum*. Bandung: Refika Aditam.
- Ardli, Imam, Ade Gafar A., Siti Mujdalipah, dan Ana. (2003). *Perangkat Penilaian Kinerja Untuk Pembelajaran Teknik Pemeliharaan Ikan*: FPTK UPI Bandung.
- Arifin, Zainal. (2009). *Evaluasi Pembelajaran Prinsip Teknik Prosedur*. Bandung: Rosda Karya.
- Basuki, Ismet dan Hariyanto. (2014). *Asesmen Pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Kuswanto, Heru. (2008). *Penilaian Pembelajaran IPA*. http://staff.uny.ac.id/sites/default/files/tmp/PENILAIAN%20PEMBELAJARAN%20IPA_0.pdf (diakses 4/11-2014).
- Latan, Hengky. (2012). *Struktural Equation Modeling: Konsep dan Aplikasi Menggunakan Program Lisrel 8.80*. Bandung: Alfabeta.
- National Center for Education Statistic. Institute of Education Sciences, U.S Departemen of Education, *Highlights from Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS) 2011.* <http://nces.ed.gov/timss/results.asp>, (diakses 4/11-2014).
- Parmin dkk., Tim Instruktur Diklat Kepala Laboratorium IPA. (2008) *Menerapkan Gagasan, Teori dan Prinsip Kegiatan Laboratorium*. [http://staff.UNESA.ac.id/sites/default/files/tmp/Tim Instruktur Diklat Kepala Laboratorium IPA%20IPA_0.DOC](http://staff.UNESA.ac.id/sites/default/files/tmp/Tim%20Instruktur%20Diklat%20Kepala%20Laboratorium%20IPA_0.DOC) (diakses 4/11-2014).
- Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 tentang *Standar Penilaian Pendidikan*
- Susila, I. Ketut. *Pengembangan Instrumen Penilaian Unjuk Kerja (Performance Assement) Laboratorium Pada Mata Pelajaran Fisika Sesuai Kurikulum Tingkat Satuan Pendidikan SMA Kelas X di Kabupaten Gianyar*. <http://pasca.undiksha.ac.id/htmpublikasi/01-Susila.pdf>. (diakses pada 15 Desember 2012).
- Wijanto, Setyo Hari. (2008). *Structural Equation Modeling dengan Lisrel 8.8*. Yogyakarta: Graha Ilmu.