

INSTRUMEN PENILAIAN KINERJA PRAKTIKUM ILMU PENGETAHUAN ALAM FISIKA SMP MENGGUNAKAN PENDEKATAN SAINTIFIK

Abdul Rahman H

SMP Negeri 3 Ciruas, Kab. Serang-Banten
arhoke65@gmail.com

Abstrak

Tujuan penelitian ini adalah untuk mendapatkan instrumen penilaian kinerja praktikum Ilmu Pengetahuan Alam Fisika Smp kelas VII pada materi pemisahan campuran dengan menggunakan pendekatan saintifik yang valid dan reliabel. Penelitian ini terdiri atas 3 tahapan utama yaitu: a) pengembangan butir instrumen ditelaah pakar, yang berjumlah 5 orang dan dinilai oleh 25 panelis, b) uji coba tahap I dan II, dan c) tahap pelaporan hasil untuk pembakuan instrumen. Penelitian ini dilaksanakan di 9 SMP Negeri di Kabupaten Serang, Provinsi Banten yang melibatkan total 297 siswa-siswi. Adapun analisis data menggunakan structural equation modeling (SEM). Hasil penelitian instrumen penilaian kinerja dengan total 89 butir pada 6 kegiatan praktikum yang masing-masing kegiatan dibangun oleh 5 dimensi, yaitu; mengamati, menanya, mengumpulkan informasi/melakukan praktikum, mengasosiasikan/mengolah informasi, dan mengkomunikasikan. Seluruh butir termasuk dalam kategori validitas konstruk yang baik, dan memiliki reliabilitas konstruk berkategori tinggi. Implikasi penelitian ini adalah bahwa Instrumen ini sangat mudah digunakan oleh guru yang mau menilai kinerja siswa dalam praktikum pemisahan campuran karena telah berisi petunjuk penilaian, pedoman penskoran dan disertai rubrik (kriteria penilaian).

Kata Kunci: Penilaian Kinerja, Pendekatan Saintifik, Praktikum Pemisahan Campuran

Abstract

The purpose of this study was to obtain a practicum performance assessment instrument for grade VII Junior High School in the of Physic's, discussing separation of mixtures by using a scientific approach that are valid and reliable. This study consists of three main stages, namely: a) development of instrument items examined by experts amounting to 5 people and judged by 25 panelists, b) test phase I and II, and c) reporting stage on results to standardized instruments. This study was conducted in 9 Junior High School in the district of Serang, Banten involving a total of 297 students. Analysis of data using structural equation modeling (SEM). This study achieved a performance assessment instrument for a total of 89 items on 6 practicum with each activity built by five dimensions, namely: observe, ask, collect information, conducting practicum, associating/processing information, and communicating. All items are within the category of good construct validity, and has high reliability construct. Implication of this research is that the instrument is very easy to be used by teachers to assess student performance in separation of the mixture practicum having assessment guidelines, scoring guidelines, and rubrics (criteria for assessment).

Keywords: performance assessments, scientific approach, separation of mixtures practicum.

PENDAHULUAN

Pada dasarnya kegiatan pendidikan dilaksanakan untuk menghasilkan perubahan tingkah laku dari orang – orang yang mengikuti pendidikan. Perubahan tingkah laku yang dimaksud adalah dapat berupa bertambahnya pengetahuan, keahlian, keterampilan dan perubahan sikap dan perilaku. Sejalan dengan Undang - Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional berfungsi mengembangkan kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa,

bertujuan untuk berkembangnya potensi peserta didik agar menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha esa, berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri, dan menjadi warga negara yang demokratis serta bertanggungjawab.

Pendidikan mencakup tiga aspek yaitu kategori psikomotorik, afektif dan kognitif. Aspek psikomotorik meliputi pengontrolan otot – otot sehingga orang dapat melakukan gerakan yang tepat sehingga sasarannya adalah agar orang tersebut memiliki keterampilan fisik tertentu. Kategori afektif meliputi perasaan,

nilai, sikap, sedangkan aspek kognitif adalah proses intelektual seperti mengingat, memahami, menganalisis sehingga pendidikan membuat orang mempunyai pengetahuan dan keterampilan berpikir. Rivai dan Murni. (2009:5).

Demikian halnya pelaksanaan pendidikan Ilmu Pengetahuan Alam seharusnya memiliki 3 aspek di atas. Proses Ilmu Pengetahuan Alam berkaitan dengan keterampilan proses dan prosedur kegiatan ilmiah. Contoh kegiatan yang berkaitan dengan hal tersebut adalah penelitian, percobaan, kegiatan laboratorium atau praktikum. Kegiatan tersebut merupakan bentuk pembelajaran aktif dan bermakna, yang dalam padanya tercakup berbagai kegiatan aktif sesuai pengalaman misalnya untuk pengembangan kemampuan mengamati, pengukuran, interpretasi, dan menerapkan konsep melalui percobaan: merumuskan masalah, menguji hipotesa, mengumpulkan data, bereksperimen dan prediksi. Dalam konteks ini Ilmu Pengetahuan Alam bukan sekedar cara bekerja, melihat dan cara berpikir, melainkan “ilmu pengetahuan alam tidak sekedar pengetahuan tapi jauh lebih dari itu”. Artinya Ilmu Pengetahuan Alam sebagai proses juga dapat meliputi kecenderungan sikap/tindakan, keingintahuan, kebiasaan berpikir, dan seperangkat prosedur. Sementara nilai-nilai Ilmu Pengetahuan Alam berhubungan dengan tanggung jawab moral, nilai-nilai sosial, manfaat Ilmu Pengetahuan Alam dan kehidupan manusia, serta sikap dan tindakan, Siahaan dkk., (2011:11). Seperti tertuang dalam Peraturan Menteri Pendidikan dan Kebudayaan Nomor 65 Tahun 2013 tentang Perbaikan Proses Pembelajaran yang sesuai dengan harapan kurikulum 2013. Guru yang tugas utamanya mendidik, mengajar, membimbing, mengarahkan bergeser menjadi guru sebagai fasilitator, navigator, motivator, mengutamakan pendidikan karakter dan berfokus pembelajaran dengan pendekatan ilmiah (*scientific*), yaitu pendekatan pembelajaran dengan urutan kegiatan: a) mengamati, b) menanya, c) mengumpulkan informasi/melakukan eksperimen d) mengasosiasikan, dan e) mengkomunikasikan.

Berkaitan dengan instrumen penilaian praktikum Ilmu Pengetahuan Alam sampai saat ini belum ada penilaian instrumen praktikum yang baku. Belum ada instrumen penilaian yang telah berstandar, umumnya setiap sekolah

dengan sekolah yang lain berbeda penilaiannya dan biasanya diserahkan kepada guru Ilmu Pengetahuan Alam masing – masing, walaupun ada baru hanya hasil diskusi dari guru – guru musyawarah guru ilmu pengetahuan alam (MGMP IPA) yang belum teruji dengan kebenarannya dan kadang hanya di buat dengan seadanya.

Pengembangan dan kalibrasi instrumen perlu dilakukan berdasarkan prosedur yang diajukan ahli dan telah dicobakan beberapa kali agar dapat memenuhi syarat validitas dan reliabilitas. Pengujian instrumen dimulai dengan proses uji coba instrumen terhadap sejumlah responden. Hasil yang diperoleh kemudian dianalisis untuk menguji instrumen tersebut. Ukuran umum digunakan adalah validitas dan reliabilitas. Candiasa dkk., (2001:139) Syarat validitas dan reliabilitas perlu dilakukan untuk memenuhi syarat metodologi penelitian.

Djaali dan Muljono (2008:60-62) menguraikan, langkah-langkah penyusunan dan pengembangan instrumen sebagai berikut: (1) berdasarkan sintesis dari teori yang di kaji dibuatlah konstruk variabel, (2) berdasarkan konstruk, maka dikembangkan dimensi dan indikator variabel, (3) membuat kisi instrumen dalam bentuk tabel spesifikasi, (4) menetapkan besaran atau parameter yang bergerak dalam suatu rentangan kontinum dari suatu kutub ke kutub lain yang berlawanan, (5) menuliskan butir-butir instrumen, (6) butir-butir yang telah ditulis merupakan konsep instrumen yang harus melalui proses validasi, (7) tahap validasi pertama yang ditempuh adalah validasi teoritik, yaitu melalui pemeriksaan pakar atau melalui panel, (8) revisi atau perbaikan, (9) penggantian instrumen, (10) uji coba instrumen, (11) pengujian validitas dilakukan dengan menggunakan kriteria, kriteria internal maupun kriteria eksternal, (12) berdasarkan kriteria tersebut diperoleh kesimpulan mengenai valid atau tidaknya sebuah butir, (13) butir-butir yang tidak valid dikeluarkan atau diperbaiki untuk diujicoba ulang, sedangkan butir-butir yang valid dirakit kembali menjadi sebuah perangkat instrumen untuk melihat kembali validitas kontennya berdasarkan kisi-kisi, (14) selanjutnya dihitung koefisien reliabilitas. (15) perakitan butir – butir instrumen yang valid untuk dijadikan instrumen final (2008:60-62).

Menurut Kirkendal, Gruber dan Johnson agar dapat diperoleh data yang valid, instrumen

atau alat untuk mengevaluasinya harus valid (1980:69). Validitas menunjuk pada penyesuaian alat pengukur dengan tujuan yang hendak diukur. Tes dikatakan memiliki validitas apabila tes itu betul – betul mengukur apa yang hendak diukur. Lebih lanjut Mcmillan (2008:19) menyatakan validitas dapat didefinisikan seperti keseluruhan dari evaluasi. Berdasarkan Candiasa dkk., terdapat 4 jenis validitas, yaitu validitas isi, validitas konstruk, validitas ramalan atau prediksi dan validitas kesamaan (2001:140-141).

Reliabilitas merujuk pada konsistensi dari suatu pengukuran. Reliabilitas dapat pula ditafsirkan sebagai seberapa tingginya korelasi antara skor tampak pada dua tes yang paralel. Azwar (2012:28). Suparno (2010:35) menyatakan bahwa, reliabilitas menunjuk pada level konsistensi internal dari alat ukur sepanjang waktu. Reliabilitas merujuk pada hasil yang didapat melalui sebuah instrumen tes, bukan merujuk kepada instrumennya sendiri, reliabilitas merupakan syarat perlu, tetapi belum cukup untuk syarat validitas. Cara atau metode mencari besarnya reliabilitas seperti yang dinyatakan Kirkendall, Gruber, dan Johnson (1980:74), yaitu: (a) metode bentuk paralel (*equivalen-forms method*), (b) metode tes ulang (*test – retest method*), (c) metode belah dua (*split half method*), sedangkan menurut Gronlund(2010:89) ditambah dua cara lagi yaitu Kuder-Richarson *method* dan metode tes ulang dengan keseimbangan (*test-retest with equivalent form*).

Secara umum tujuan penelitian ini adalah untuk mengembangkan dan memvalidasi instrumen penilaian kinerja praktikum Ilmu Pengetahuan Alam Fisika SMP. Tujuan ini dapat dirinci menjadi tujuan operasional, yaitu: menemukan indikator-indikator instrumen penilaian kinerja praktikum Ilmu Pengetahuan Alam Fisika SMP, menghasilkan instrumen penilaian kinerja praktikum Ilmu Pengetahuan Alam Fisika SMP yang valid dan reliabel, mengembangkan instrumen penilaian kinerja praktikum Ilmu Pengetahuan Alam Fisika SMP yang secara teoritis valid dan reliabel dan secara praktis mudah digunakan oleh guru.

METODE

Penelitian ini dilakukan pada sekolah-sekolah yang sudah menerapkan kurikulum 2013 di Kabupaten Serang-Banten, yaitu: SMP N.1 Ciruas, SMP N.1 Cikande, SMP N.1

Kragilan, SMP N.1 Cinangka, SMP N. 1 Kramat Watu, SMP N. 1 Anyer, SMP N.1 Bojonegara, SMP N.1 Pabuaran dan SMP N.1 Baros. Tahap pengembangan instrumen penilaian kinerja praktikum IPA Fisika SMP dilakukan awal 2015. Tahap pengujian rasional dilakukan di Jakarta dan di kota Serang pada bulan September - Oktober 2015 yang melibatkan 30 orang panelis yang terdiri atas orang yang berkompeten dibidangnya. Tahap pengujian empiris pertama (ujicoba tahap I) dilakukan di kab. Serang pada November tahun 2015 yang melibatkan 132 orang siswa-siswi SMP. Selanjutnya tahap pengujian empiris kedua (ujicoba tahap II) dilakukan di kab. Serang provinsi pada bulan Februari–Maret tahun 2016 yang melibatkan 164 orang.

Teknik pengambilan sampel dalam penelitian ini adalah acak bertahap (multistage random sampling). Pengacakan dilakukan terhadap sekolah-sekolah SMP yang sudah menerapkan kurikulum 2013 di Kab. Serang Provinsi Banten, pengacakan terhadap kelas dan yang terakhir adalah pengacakan terhadap siswa-siswi SMP kelas VII.

Adapun prosedur analisis data pertama, dalam penelitian ini telaah pakar difokuskan pada dua aspek yaitu mengetahui ketepatan dimensi dari konstruk, seberapa jauh indikator merupakan jbaran yang tepat dari dimensi dan seberapa jauh butir-butir instrumen yang dibuat secara tepat dapat mengukur indikator sebagaimana yang dimaksud di atas dan kejelasan bahasa. Kedua, pengujian terhadap penilai (pakar dan panelis) dengan menggunakan koefisien ICC. Hartzoulakis (www.enl.auth.gr). Ketiga, Analisis faktor dilakukan dalam dua tahap. Pertama, menggunakan langkah awal pada analisis faktor eksploratori dengan aplikasi analisis faktor dari SPSS dan kedua, menggunakan analisis faktor konfirmatori dengan aplikasi SEM. Pada tahap ini prosedur yang dilakukan: (1) menguji kelayakan sampel dan (2) menyajikan matriks korelasi. Syarat dilakukan analisis ditunjukkan dengan: (1) sampel mencukupi apabila butir yang direspons minimal 5 kali jumlah butir, (2) kecukupan sampel atau data menggunakan rumus Kaiser-Meyer-Olkin (KMO). (3) data berdistribusi normal yang ditunjukkan melalui uji normalitas Barlett's (Bartlett test of spherity), dengan Chi-Square hasil perhitungan lebih kecil dari Chi-Square tabel pada taraf signifikan yang

ditetapkan. Apabila ketiga hal tersebut terpenuhi analisis faktor dapat dilanjutkan.

Keempat, butir yang mempunyai AIC dengan measure of sampling adequacy (MSA) $\geq 0,50$ merupakan butir yang layak dianalisis lebih lanjut. Santoso (2012:45).Kelima, hasil analisis faktor dari uji SEM untuk mengetahui dan memastikan apakah butir-butir atau indikator yang dibuat benar-benar menjelaskan dimensi yang telah dikonstruksi dari variabel. Adapun jenis pengujian yang dilakukan adalah pengujian Confirmatory Faktor Analysis (CFA). Pada tahapan kedua ini, dilakukan tiga pengujian, yaitu: (1) kecocokan data dengan model, (2) validitas dan reliabilitas model, dan (3) uji signifikansi koefisien-koefisien dari model struktural. Keenam, setelah kecocokan model dan data terpenuhi, maka dilakukan uji kecocokan model pengukuran, dan kecocokan model struktural (structural model fit) dengan melakukan evaluasi terhadap setiap konstruk terhadap validitas dan reliabilitas. Menurut Fornier dan Larcker seperti dikutip Wijanto, nilai reliabilitas dengan menggunakan composite reliability (CR) dan variance extracted (VE). Menurut Hair seperti dikutip Yamin (2014:30), suatu konstruk mempunyai nilai yang baik apabila nilai CR $\geq 0,70$ dan nilai VE $\geq 0,50$.

HASIL PENELITIAN

Hasil Penelaahan Pakar

Penilaian instrumen yang telah dikembangkan peneliti dimintakan kepada 5 pakar dan 25 panelis. Adapun hasil penelaahan pakar terhadap instrumen yang dikembangkan ada beberapa hal yang dapat peneliti sampaikan: 1) Secara umum substansi instrumen yang dikembangkan sudah mengukur keterampilan siswa dalam praktikum, 2) Instrumen yang dikembangkan sudah memadai dilihat dari kegiatan siswa mulai dari persiapan praktikum, pelaksanaan dan tindak lanjut, 3) Penyempurnaan bahasa/kalimat butir-butir instrumen yang dicoret para pakar, 4) Adanya penambahan dan pengurangan butir-butir instrumen.

Hasil Penilaian Panelis

Penilaian butir-butir instrumen dilaksanakan bulan Agustus - November 2015. Skala yang digunakan bergerak dari 1 sampai dengan skala 5. Skala 1 menunjukkan butir-butir instrumen sangat tidak tepat dan skala 5 menunjukkan

butir-butir instrumen sangat tepat. Adapun nilai ICC kegiatan 1 sampai kegiatan 6 berturut-turut, yaitu 0,708; 0,651; 0,509; 0,557; 0,497; 0,703. Terlihat bahwa nilai ICC paling rendah adalah praktikum kegiatan 5 sebesar 0,497 dan nilai ICC paling tinggi adalah terjadi pada kegiatan 1 sebesar 0,708.

Hasil Uji Coba Tahap I

Hasil uji kelayakan sampel dan butir kegiatan 1 sampai kegiatan 6 pada uji coba Tahap I dipaparkan seperti pada Tabel 1 berikut ini.

Tabel 1 Nilai *KMO* dan *Bartlett's Test* Untuk Hasil Uji Coba I

No	Nomor Kegiatan	Nilai KMO	Bartlett's Test	Ket.
1	Kegiatan 1	0,574	0,002	Signifikan
2	Kegiatan 2	0,557	0,000	Signifikan
3	Kegiatan 3	0,572	0,001	Signifikan
4	Kegiatan 4	0,547	0,001	Signifikan
5	Kegiatan 5	0,551	0,001	Signifikan
6	Kegiatan 6	0,553	0,003	Signifikan

Berdasarkan Tabel 1 nilai KMO berada di atas 0,5 (kategori cukup menurut Kaiser). Sementara nilai signifikansi atau peluang (p) mulai dari kegiatan 1 sampai kegiatan 6 $< 0,05$. Dengan demikian sampel layak untuk dianalisis menggunakan analisis faktor dan faktor-faktornya dapat dianalisis lebih lanjut. Tabel 2 adalah rangkuman nilai MSA yang kurang dari 0,5 sehingga butir dibuang dan tidak diikuti lagi pada ujicoba tahap II.

Setelah butir yang tidak layak dikeluarkan maka dilanjutkan dengan analisis faktor konfirmatori. Jumlah butir setiap kegiatan yang sebelumnya 18 butir menjadi 15 butir, kecuali kegiatan 3 jumlah butirnya menjadi 14 butir karena ada 4 butir yang dibuang.

Tabel 3 memperlihatkan rangkuman parameter hasil estimasi respesifikasi data uji coba tahap I pada kegiatan 1 seperti di bawah ini.

Tabel 3 Rangkuman Parameter GoF Hasil Estimasi Respesifikasi Data Uji Coba I pada Kegiatan 1

Tabel 2 Rangkuman Butir Nilai MSA < 0,5 Hasil Uji Coba I

No	Keg	Butir	Nilai MSA	Ket
1	Keg. 1	11	0,444	Dibuang
		16	0,495	Dibuang
		18	0,471	Dibuang
2	Keg. 2	7	0,413	Dibuang
		11	0,483	Dibuang
		18	0,435	Dibuang
3	Keg. 3	6	0,438	Dibuang
		11	0,463	Dibuang
		16	0,494	Dibuang
		18	0,497	Dibuang
4	Keg. 4	12	0,461	Dibuang
		15	0,455	Dibuang
		18	0,489	Dibuang
		18	0,478	Dibuang
5	Keg. 5	16	0,48	Dibuang
		18	0,424	Dibuang
		18	0,477	Dibuang
6	Keg. 6	8	0,434	Dibuang
		16	0,434	Dibuang
		18	0,494	Dibuang

Berdasarkan Tabel 3 di atas 12 dari 14 parameter *GoF* yang memenuhi kriteria. Kemudian parameter nomor 2 dan parameter nomor 13 mendekati cocok. Hal ini menunjukkan bahwa model yang direncanakan merupakan model yang baik untuk mempresentasikan hasil penelitian pada kegiatan 1 ini. Khusus untuk parameter nomor 1 merupakan uji untuk mengetahui apakah model yang dirancang tersebut cocok dengan model hasil estimasi (prediksi). Uji kelayakan model dengan hipotesis:

$$H_0 : \Sigma = \Sigma(\theta)$$

$$H_1 : \Sigma \neq \Sigma(\theta)$$

Keterangan:

Σ adalah matriks kovarians sampel data

$\Sigma(\theta)$ adalah matriks kovarian hasil prediksi model

Berdasarkan data tersebut diperoleh *p-value* lebih besar dari 0,05. Hal ini menunjukkan bahwa H_0 diterima, yang berarti tidak ada perbedaan antara matriks kovarians sampel data dan matriks kovarian hasil prediksi. Dengan perkataan lain model yang direncanakan sama dengan model yang diestimasi.

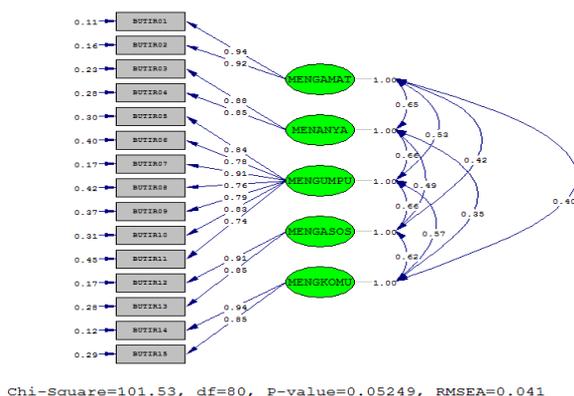
Hasil Uji Coba Tahap II

Nilai Bartlett's Test mulai kegiatan 1 sampai kegiatan 6 adalah 0,000, sedangkan nilai KMO berturut-turut 0,897; 0,907; 0,884; 0,898; 0,896; dan 0,897. Berdasarkan data tersebut nilai *KMO*

No	Ukuran Kecocokan	Kriteria	Hasil Estimasi Respesifikasi	Kategori
1	p-Chi-Square	$P > 0,05$	$P = 0,67$	Cocok
2	RMSEA	$0,05 < RMSEA < 0,08$	0,043	Mende-kati cocok
3	Saturated ECVI		1,83	
	ECVI	$< \text{Saturated ECVI}$	1,22	Cocok
4	Saturated AIC		240,00	
	AIC	$< \text{Saturated AIC}$	179,68	Cocok
5	Saturated CAIC		705,94	
	CAIC	$< \text{Saturated CAIC}$	335,0	Cocok
6	NFI	$\geq 0,90$	0,98	Cocok
7	NNFI	$\geq 0,90$	1,00	Cocok
8	CFI	$\geq 0,90$	1,00	Cocok
9	IFI	$\geq 0,90$	1,00	Cocok
10	RFI	$\geq 0,90$	0,98	Cocok
11	RMSR	$< 0,08$	0,017	Cocok
12	GFI	$\geq 0,90$	0,91	Cocok
13	AGFI	$\geq 0,90$	0,86	Mende-kati cocok
14	PGFI	$> 0,60$	0,61	Cocok

semua di atas 0,5 (kategori cukup menurut Kaiser). Sementara nilai signifikansi atau peluang (*p*) mulai dari kegiatan 1-6 adalah $< 0,05$. Dengan demikian sampel layak untuk dianalisis menggunakan analisis faktor dan faktor-faktornya dapat dianalisis lebih lanjut.

Gambar 1 menunjukkan diagram jalur untuk butir setiap dimensi, nilai muatan faktor, dan kesalahan pengukuran setiap butir serta nilai korelasi antara variabel laten.



Gambar 1 Diagram Jalur Hasil Kegiatan 1 Uji Coba II

Berdasarkan Gambar 1 tergambar bahwa, semua butir memiliki nilai $t > 1,96$ dan muatan faktor (*faktor loading*) $> 0,5$.

HASIL DAN PEMBAHASAN

Kualitas instrumen ditentukan oleh dua hal, pertama validitas, yang kedua adalah reliabilitas. Instrumen yang baik adalah instrumen yang mampu menghasilkan pengukuran sama dengan harga sebenarnya dan akan menunjukkan hasil yang sama atau hampir sama bila dilakukan pengukuran berulang-ulang. Cara untuk menguji validitas instrumen pendidikan dapat dibedakan menjadi dua menurut Soeprijanto yaitu, panel ahli (*expert judgement*) dan uji coba lapangan (2010:139). Tabel 1 terdapat rangkuman nilai *ICC* kegiatan 1 sampai dengan kegiatan 6. Nilai *ICC* terendah adalah kegiatan 5 (memisahkan campuran dengan metode dekantasi) yaitu 0,497 dan nilai *ICC* tertinggi adalah kegiatan 1 (memisahkan garam dapur dari pengotor dengan cara penyaringan) yaitu 0,708. Menurut Fleiss dalam Widhiarso (<http://widhiarso.staff.ugm.ac.id>) mengkategorikan tingkat reliabilitas antar rater menjadi 4 kategori, yaitu; jika $< 0,4$ kategori buruk (*bad*); $0,4 - 0,60$ kategori cukup (*fair*); $0,6 - 0,75$ kategori memuaskan (*good*); dan $> 0,75$ kategori istimewa (*excellent*). Sedangkan menurut Landis dan Koch interpretasi nilai *ICC* ada 6 kategori, yaitu: $r < 0,00$ (*poor agreement*); $0,00 \leq r \leq 0,20$ (*slight*); $0,21 \leq r \leq 0,40$ (*fair*); $0,41 \leq r \leq 0,59$ (*moderate*); $0,60 \leq r \leq 0,79$ (*substantial*); $0,80 \leq r \leq 1,00$ (*almost perfect*) (http://www_enl_auth_gr_gala.pdf).

Berdasarkan referensi tersebut, sehingga pada penelitian ini kegiatan 3, kegiatan 4, dan

kegiatan 5 nilai *ICC* berkategori cukup sedangkan kegiatan 1, kegiatan 2 dan kegiatan 6 nilai *ICC* berkategori memuaskan (baik).

Berdasarkan Tabel 2 dan 6 tersebut tergambar bahwa semua kegiatan praktikum baik pada uji coba tahap I maupun uji coba tahap II nilai *KMO MSA* $> 0,5$. Pada Uji Coba tahap I nilai *KMO MSA* terendah sebesar 0,547 (kegiatan 4=pemurnian alkohol dengan cara destilasi) dan tertinggi sebesar 0,574 (kegiatan 1) serta nilai *Bartlett's test* yang ditunjukkan oleh nilai *Chi-Square* terendah sebesar 206,112 (kegiatan 6) dan tertinggi sebesar 235,374 (kegiatan 2) pada masing-masing tingkat signifikansi $p < 0,05$. Demikian pada uji coba tahap II nilai *KMO* terendah sebesar 0,884 (kegiatan 3) dan tertinggi sebesar 0,907 (kegiatan 2) serta nilai *Bartlett's test* yang ditunjukkan oleh *Chi-Square* terendah sebesar 1765,670 (kegiatan 3) dan tertinggi sebesar 1820,639 (kegiatan 1) serta masing-masing tingkat signifikansi sebesar $p=0,000$. Hal ini menunjukkan bahwa indikator-indikator sebagai variabel laten yang telah dirancang layak dianalisis dengan analisis faktor menurut Wijanto (2008:33-34). Variabel laten yang dimaksud dalam penelitian ini adalah 5 dimensi yaitu Dimensi A (mengamati), Dimensi B (menanya), Dimensi C (mengumpulkan informasi), Dimensi D (mengasosiasikan/mengolah informasi) dan Dimensi E (mengkomunikasikan) pada masing-masing kegiatan 1 (memisahkan garam dapur dari pengotor dengan cara penyaringan), kegiatan 2 (memisahkan kapur barus dari pengotor dengan metode sublimasi), kegiatan 3 (kromatografi kertas), kegiatan 4 (pemurnian alkohol), kegiatan 5 (metode dekantasi) dan kegiatan 6 (kristalisasi).

Menurut Santoso (2006:20) jika nilai *MSA* $< 0,5$; maka butir harus dikeluarkan atau dieliminasi sebelum dilakukan analisis faktor. Dengan kata lain *MSA* $< 0,5$ variabel tidak bisa diprediksi dan tidak bisa dianalisis lebih lanjut, atau dikeluarkan dari variabel lainnya. Sedangkan menurut Djaali (2008:81) jika nilai *KMO-MSA* dibawah 0,5 maka penggunaan analisis tidak tepat untuk dilanjutkan analisis faktor. Berdasarkan referensi tersebut, ke-19 butir dibuang dan tidak diikuti sertakan dalam ujicoba tahap II. Sedangkan pada hasil uji coba tahap II menunjukkan nilai *MSA* seluruh butir $> 0,5$. Hal ini berarti bahwa hasil analisis ujicoba

II menunjukkan keseluruhan butir yang dikembangkan layak untuk dianalisis dengan analisis faktor.

Setelah butir yang tidak layak dibuang maka dilanjutkan dengan analisis faktor tahap 2 dengan *SEM*. Ada sejumlah alasan kenapa menggunakan *SEM*, 2 diantaranya; *pertama SEM* dapat meneliti variabel atau konstruk yang tidak teramati atau tidak dapat diukur secara langsung (*unobserved variables*), *kedua SEM* menguji kesalahan pengukuran (*measurement error*) untuk variabel atau konstruk yang teramati (*observed variables*). Latan (2012:7) Tahapan pertama melakukan uji kecocokan model keseluruhan. Hasil uji tersebut di dapat parameter hasil estimasi yang berkenaan dengan uji kecocokan model keseluruhan yang menunjukkan secara keseluruhan semua parameter memenuhi kriteria *goodness of fit*, cocok pada uji coba tahap I maupun uji coba tahap II.

Evaluasi model pengukuran atau *CFA* dilakukan terhadap masing-masing kegiatan, dimensi dan butir-butir instrumen yang ada di dalam model. Berkenaan dengan nilai muatan faktor, hasil uji coba tahap I menunjukkan bahwa seluruh butir benar-benar mengukur indikator-indikatornya atau butir-butir yang ada mengukur apa yang seharusnya diukur. Nilai rata-rata muatan faktor perkegiatan adalah sebesar 0,743 (kegiatan 3) - 0,793 (kegiatan 1). Demikian juga dengan hasil uji coba tahap II menunjukkan bahwa seluruh butir benar-benar mengukur apa yang seharusnya diukur. Nilai rata-rata muatan faktor perkegiatan mulai dari 0,809 (kegiatan 2) - 0,864 (kegiatan 3). Hal ini ditunjukkan oleh semua butir memiliki muatan faktor (*SLF*) > 0,5 dengan nilai t statistik di atas 1,96. Dengan perkataan lain semua butir baik hasil analisis uji coba tahap I maupun analisis hasil uji coba tahap II memenuhi kriteria sebagai butir yang memiliki validitas konstruk yang baik. Validitas konstruk adalah validitas yang mempermasalahkan seberapa jauh item-item tes mampu mengukur apa yang benar-benar hendak diukur sesuai dengan konsep khusus atau definisi konseptual yang telah ditetapkan. Djaali (2008:51).

Kemudian jika dilihat dari kesalahan pengukurannya (*error*) pada uji coba tahap I, nilai kesalahan pengukuran paling kecil sebesar 0,22 terdapat pada kegiatan 1 dan kegiatan 5, serta kesalahan pengukuran (*error*) paling besar

adalah sebesar 0,67 pada kegiatan 3, sedangkan pada uji coba tahap II nilai kesalahan pengukuran paling kecil sebesar 0,07 pada kegiatan 3 dan kesalahan terbesar 0,71 pada kegiatan 2. Kesalahan pengukuran ini dapat disebabkan oleh berbagai hal, misalnya faktor kejenuhan penilai, faktor psikologis penilai, dan perubahan kriteria penilaian saat penilai menilai menurut Sapriati (2005:271) serta bisa juga disebabkan oleh faktor dari siswa pada saat praktikum.

Seberapa besar setiap butir dapat dijelaskan oleh indikator ditunjukkan oleh koefisien determinasi, atau koefisien determinasi digunakan untuk mengetahui seberapa besar varian dimensi menjelaskan varian indikator atau seberapa besar varian indikator menjelaskan butir-butir instrumen. Nilai rata-rata koefisien determinasi perkegiatan pada uji coba tahap II antara 0,46 - 0,49. Hal ini berarti 46% - 49% butir-butir yang ada pada dimensi mampu dijelaskan oleh indikator yang bersesuaian.

Koefisien RK sebagai perangkat instrumen, dimensi baik hasil uji coba tahap I maupun uji coba tahap II, pada umumnya termasuk dalam kategori baik. Adapun syarat yang diajukan Hair et.al (1998: 56) sebuah konstruk mempunyai reliabilitas yang baik jika $RK \geq 0,70$ dan $VE \geq 0,5$. *VE* yang kurang dari 0,50 menunjukkan bahwa lebih banyak *error* yang terdapat pada item-item indikator daripada varians yang dapat dijelaskan oleh konstruk laten. Usman (2014:193).

PENUTUP

Simpulan

Pertama, prosedur pengembangan instrumen penilaian kinerja praktikum Ilmu Pengetahuan Alam terdiri atas: (1) merumuskan definisi konseptual dan operasional, (2) mengembangkan spesifikasi dan penulisan pernyataan dan pengamatan, (3) penelaahan pernyataan dan pengamatan, (4) ujicoba, (5) analisa, (6) revisi instrumen, dan (7) perakitan instrumen untuk menjadi instrumen baku.

Kedua, Instrumen Penilaian Kinerja Praktikum IPA Fisika SMP Menggunakan Pendekatan Saintifik yang terdiri atas enam kegiatan, dibangun oleh lima dimensi yang dimulai dari mengamati (A), menanya (B), mengumpulkan informasi (C),

mengasosiasikan/mengolah informasi (D), dan mengkomunikasikan (E).

Ketiga, hasil uji coba instrumen tahap I ada 19 butir nilai MSA < 0,5, Kesembilan belas butir tersebut dibuang dan tidak diikutsertakan pada uji coba tahap II. Keempat, hasil uji coba instrumen tahap II menunjukkan bahwa secara empiris seluruh butir pada setiap kegiatan memenuhi validitas konstruk yang baik. Sehingga seluruh butir pada kegiatan 1, 2, 4, 5 dan 6 masing-masing 15 butir instrumen dan pada kegiatan 3 ada 14 butir instrumen penilaian adalah valid. Koefisien Reliabilitas Konstruk (RK) baik hasil uji coba tahap I maupun uji coba tahap II pada umumnya berkategori baik.

Soeprijanto, *Pengukuran Kinerja Guru Praktik Kejuruan: Konsep dan Teknik Pengembangan Instrumen*. Jakarta: Tursina, 2010.

Santoso, Singgih. *Aplikasi SPSS Pada Statistika Multivariat*. Jakarta: Elex Media Kom, 2012

Uno, Hamzah B., dan Satrio Koni. *Assesment Pembelajaran*. Jakarta: Bumi Aksara, 2012.

Wijanto, Setyo Hari. *Struktural Equation Modeling Dengan Lisrel 8.8: Konsep & Totorial*. Yogyakarta: Graha Ilmu, 2008.

Yamin, Sofyan. *Rahasia Olah Data Lisrel*. Jakarta: Mitra Wacana Media, 2014.

DAFTAR PUSTAKA

Azwar, Saifudin. *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar, 2012.

Djaali dan Pudji Muljono. *Pengukuran dalam Bidang Pendidikan*. Jakarta: Grasindo, 2008.

Dachlan, Usman. *Panduan Lengkap Structure Equation Modeling*. Semarang: Lentera Ilmu, 2014.

Gronlund, Norman E. *Measurement and Evaluation in Teaching*. New York: Macmillan Publishing Company, 1985.

Kirkendall, Don R, Joseph J. Gruber, dan Robert E. Johnson. *Measurament and Evaluation for Physical Educator*. Boston: Wm. C. Brown Company Publishers, 1980.

Latan, Hengky. *Aplikasi Analisis Data Statistik Untuk Ilmu Sosial Sains Dengan Stata*. Bandung: Alfabeta, 2012

Mcmillan, James H. *Aessment Essentials for Standar - Based Education*. California: Corwin Press, 2008.

Murni, Sylviana dan Vaithzal Rivai. *Education Management: Analisis Teori dan Praktik*. Jakarta: RajaGrafindo Persada, 2009

Purwanto. *Instrumen Penelitian Sosial dan Pendidikan*. Yogyakarta: Pustaka Pelajar, 2010.

Sapriati, Amalia. "Pengembangan Instrumen Penilaian Praktikum IPA-Studi Pengembangan Instrumen Penilaian Praktikum IPA-Biologi Pada Siswa SMP di Kota Jogyakarta." *Disertasi*, UNJ, 2005