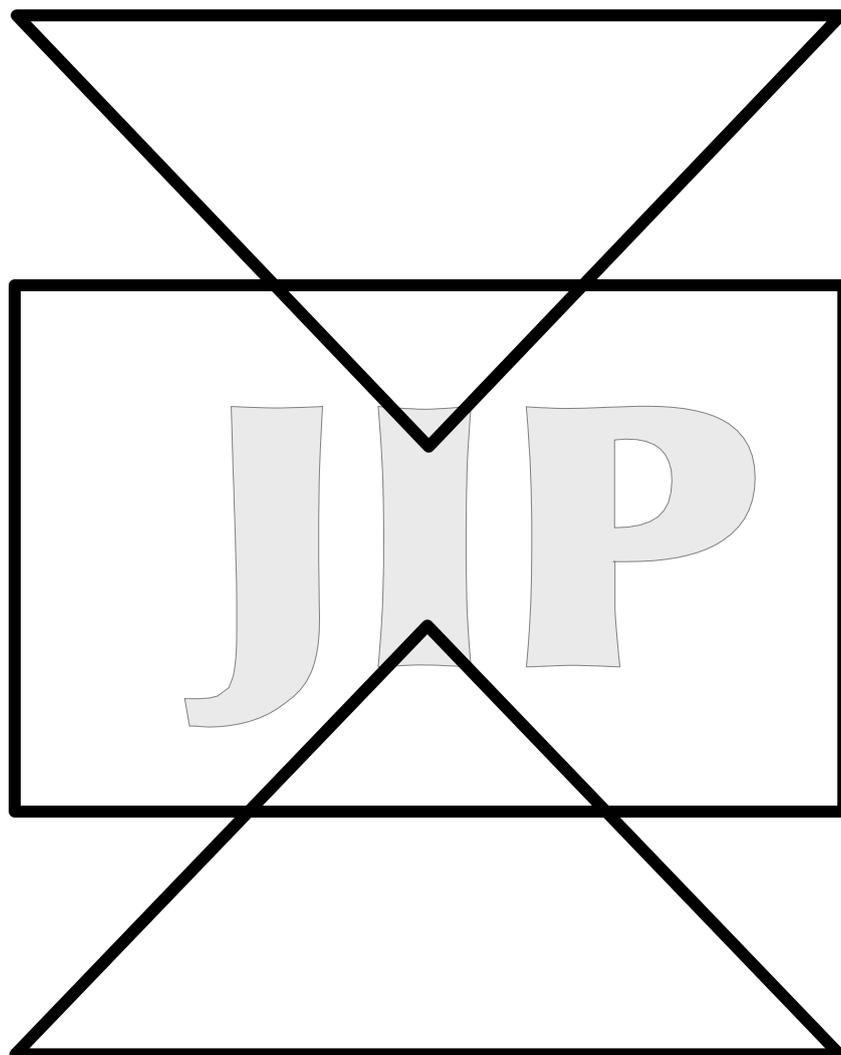


JURNAL ILMU PENDIDIKAN



Diterbitkan oleh:
Lembaga Pendidikan Tenaga Kependidikan (LPTK)
dan
Ikatan Sarjana Pendidikan Indonesia (ISPI)

JIP	Jilid 23	Nomor 2	Halaman 91-182	Malang Desember 2017	ISSN 0215-9643
------------	----------	---------	-------------------	-------------------------	-------------------

JIP

JURNAL ILMU PENDIDIKAN

ISSN 0215-9643

Jilid 23, Nomor 2, Desember 2017, hlm. 91-176

Terbit dua kali setahun pada bulan Juni dan Desember. Berisi tulisan yang diangkat dari hasil penelitian di bidang pendidikan. Artikel telaah (*review article*) dimuat atas undangan. ISSN 0215-9643.

Ketua Penyunting

Ali Saukah

Wakil Ketua Penyunting

Imam Agus Basuki

Penyunting Pelaksana

Margono

Amat Mukhadis

Ery Tri Djatmika

Ahmad Samawi

Yazid Basthomi

Utami Widiati

Subandi

Effendy

Pelaksana Tata Usaha

Suwadi

Rahadi

Abd. Wahed

Prihatini Retnaningsih

Arif Wicaksono

Pembantu Pelaksana Tata Usaha

Lina Sumarlin

Harris Ferdiansyah

Alamat Penyunting dan Tata Usaha: Subag Program, LP2M, Universitas Negeri Malang, Gedung H-4 Lantai II
Jln. Semarang 5 Malang 65145 Telepon (0341) 551-312 pesawat 484, sambungan langsung dan Fax. (0341) 552-115.
Homepage: <http://www.um.ac.id>. *E-mail:* jip@um.ac.id/jurnalilmupendidikan@gmail.com.

JURNAL ILMU PENDIDIKAN diterbitkan sejak 21 Februari 1994 oleh Lembaga Pendidikan Tenaga Kependidikan (LPTK) dan Ikatan Sarjana Pendidikan Indonesia (ISPI).

Penyunting menerima sumbangan tulisan yang belum pernah diterbitkan dalam media lain. Naskah diketik di atas kertas HVS kuarto spasi ganda sepanjang lebih kurang 20 halaman, dengan format seperti tercantum pada halaman belakang ("Petunjuk bagi Calon Penulis JIP"). Naskah yang masuk dievaluasi dan disunting untuk keseragaman format, istilah, dan tata cara lainnya.

Dicetak di Percetakan UM Press. Isi di luar tanggung jawab Percetakan

JIP
JURNAL ILMU PENDIDIKAN
ISSN 0215-9643
Jilid 23, Nomor 2, Desember 2017, hlm. 91-182

DAFTAR ISI

Pengaruh Gaya Pengasuhan Orang Tua, Konsep Diri, dan Regulasi Diri terhadap Motivasi Berprestasi Siswa. Bambang Suryadi & Eha Soriha (<i>UIN Syarif Hidayatullah Jakarta</i>) Yuli Rahmawati (<i>Universitas Negeri Jakarta</i>)	91-98
Pertimbangan Pemilihan Konsentrasi pada Program Studi Manajemen Jenjang Sarjana. Bayu Ilham Pradana (<i>Universitas Brawijaya</i>)	99-107
Identifikasi Model Manajemen Pengetahuan Universitas Swasta di Jakarta Bayu Koen Anggoro (<i>Universitas Negeri Malang</i>) Musa Hubeis & Illah Sailah (<i>Institut Pertanian Bogor</i>)	108-116
Efektifitas Model Pembelajaran <i>Argument Driven Inquiry</i> (ADI) dalam Meningkatkan Pemahaman Konsep Siswa pada Materi Bentuk dan Kepolaran Molekul Billy Ambrocious Kalay (<i>SMA Negeri 3 Ambon</i>), Subandi & Endang Budiasih (<i>Universitas Negeri Malang</i>)	117-125
Validitas dan Reliabilitas Konstruk Skor Tes Kemampuan Calon Mahasiswa Heri Retnawati (<i>Universitas Negeri Yogyakarta</i>)	126-135
Peran Input Kelembagaan dan Input Kemahasiswaan terhadap Capaian Akademik Mahasiswa I Gusti Ayu Manuati Dewi, Eka Ardani Sisdyani, Ni Made Adi Erawati, Ni Nyoman Ayu Diantini, Surya Dewi Rustariyuni (<i>Universitas Udayana</i>)	136-151
Analisis Kebutuhan Model Pelatihan untuk Meningkatkan Kompetensi Guru SD Daerah Terpencil Piter Joko Nugroho (<i>Universitas Palangkaraya</i>)	152 - 163
Model Pembelajaran Pembentukan Perilaku Moral pada Anak Kelompok Bermain Wirdatul 'Aini (<i>Universitasnegeri Padang</i>)	164 - 170
Model Pembelajaran Berbasis Proyek dalam Menanamkan Karakter Kewirausahaan, Keterampilan Proses Sains dan Keterampilan Berpikir Tingkat Tinggi Mahasiswa (Wiwi Wikanta & Yuni Gayatri (<i>Universitas Muhammadiyah Surabaya</i>))	171 - 175
Pengaruh Strategi Membaca Dialogis terhadap Kemampuan Membaca Pelajar Pemula Bahasa Inggris Sebagai Bahasa Asing Nunung Suryati, Furaidah & Ali Saukah (<i>Universitas Negeri Malang</i>)	176-182
Indeks Subjek	
Indeks Pengarang	
Indeks Mitra Bebestari	

VALIDITAS DAN RELIABILITAS KONSTRUK SKOR TES KEMAMPUAN CALON MAHASISWA

Heri Retnawati

Universitas Negeri Yogyakarta, Jl. Colombo No.1, Yogyakarta
e-mail: retnawati_heriuny@yahoo.co.id

Abstract: On the Construct Validity and Reliability of the Candidate Students' Proficiency Scores. This study aims to prove the construct validity and estimate the construct reliability of the scores gathered from 893 prospective students (applicants) to the same entrance test administered in 2013, 2014, and 2015. The construct validity was evident as shown by confirmatory factor analysis and cross-validation. This was followed with the estimation of its construct reliability. The similarity of reliabilities was based on analysis using H* test. The study shows that by utilizing cross-validation, confirmatory factor analysis using data from 2013, 2014, and 2015 produced fit models. The invariance test of factor loading shows that the same value form year to year, but based on the cross-validation, it is only scores from 2013 and 2015 which show the same results. The subtest reliabilities, estimated using construct reliability, ω -coefficient, and Ω -coefficient, resulted in mediocre level, but the maximum reliability formula resulted in coefficient of high category..

Keywords: construct validity, cross-validation, construct reliability

Abstrak: Validitas dan Reliabilitas Konstruksi Skor Tes Kemampuan Calon Mahasiswa. Studi ini bertujuan untuk membuktikan validitas konstruk tes pengukur kemampuan awal mahasiswa dan mengestimasi reliabilitas skornya antar waktu. Penelitian ini menggunakan data dokumentasi selama 3 tahun (2013-2015) dari 893 calon mahasiswa yang merespons tes yang sama. Validitas konstruk dibuktikan dengan analisis faktor konfirmatori kemudian dilakukan kros validasi, kemudian diestimasi reliabilitas konstruksinya. Kesamaan reliabilitas dianalisis dengan uji H*. Hasil studi menunjukkan bahwa dengan menggunakan krosvalidasi, analisis faktor konfirmatori dengan menggunakan data tahun 2013, 2014, dan 2015 menghasilkan model yang fit. Hasil uji invariansi muatan faktor menunjukkan bahwa muatan-muatan faktor bernilai sama antar tahun, namun berdasarkan hasil krosvalidasi model hasilnya sama pada tahun 2013 dan 2015. Reliabilitas subtes yang diestimasi dengan reliabilitas konstruk, reliabilitas ω , menghasilkan koefisien pada level sedang, namun pada reliabilitas maksimal menghasilkan koefisien pada kategori tinggi.

Kata kunci: validitas konstruk, krosvalidasi, reliabilitas konstruk

Dalam pendidikan tinggi, calon mahasiswa ditentukan melalui seleksi. Jika terjadi kasus banyaknya calon mahasiswa yang berminat lebih besar dari daya tampungnya, maka lembaga tersebut perlu melaksanakan seleksi masuk. Selain karena alasan daya tampung, seleksi dilakukan untuk memilih calon mahasiswa yang sesuai dengan karakteristik yang sesuai dengan bidang ilmu yang dipelajari, dan juga calon yang diinginkan oleh lembaga (Ekwoaba, Ikeje, Ufoma, 2015). Selain memilih calon mahasiswa yang berhak untuk melanjutkan pendidikan pada jenjang tersebut, seleksi digunakan untuk me-

meroleh informasi mengenai kemampuan awal calon mahasiswa baru. Informasi ini diperoleh melalui asesmen terhadap calon mahasiswa baru, yang dapat dimanfaatkan untuk pemilihan jurusan atau program studi, pengelompokan calon, maupun kepentingan lainnya. Agar dapat diperoleh informasi untuk menggambarkan kemampuan calon yang sebenarnya, untuk melaksanakan asesmen ini diperlukan instrumen yang valid dan reliabel (Wright, 2008).

Ada beberapa kemampuan dasar calon mahasiswa yang biasa diperlukan oleh lembaga pend-

dikan tinggi. Kemampuan dasar tersebut berupa potensi akademis (Dodge & Derwin, 2008; Asrijanti, 2014) dan kemampuan bahasa Inggris (Pali, 2009; Martirosyan, Hwang, Wanjohi, 2015). Tes potensi akademis terdiri dari 3 subtes, yaitu subtes yang mengukur kemampuan verbal, subtes yang mengukur kemampuan penalaran, dan subtes yang mengukur kemampuan kuantitatif. Tes bahasa Inggris meliputi struktur kalimat dan membaca. Informasi mengenai kemampuan dasar tersebut diperoleh melalui pelaksanaan tes tertulis atau tes berbantuan komputer bagi calon mahasiswa baru. Dengan menggunakan respons calon mahasiswa terhadap tes-tes tersebut, dapat diketahui karakteristik instrumen, misalnya validitas dan reliabilitas.

Validitas instrumen terkait dengan interpretasi skor yang diperoleh dengan menggunakan instrumen tersebut. Gregory (2007) menyatakan bahwa validitas mencerminkan suatu penilaian tentang seberapa memadai sebuah tes mengukur atribut yang dirancang untuk diukur. Kemudahan penilaian terhadap instrumen tersebut terkait dengan empat prinsip yaitu interpretasi, kegunaan, nilai, dan konsekuensi hasil penilaian (Nitko & Brookhart, 2007). Validitas itu dapat dikelompokkan menjadi tiga tipe, yaitu (1) isi, (2) validitas konstruk, dan (3) validitas kriteria (*criterion-related*) (Gregory, 2007). Validitas ini dapat diketahui dengan menggunakan berbagai sumber baik isi instrumen, proses respons, struktur internal dalam instrumen, hubungan variabel yang diukur dalam satu instrumen dengan variabel lain yang bersesuaian dengan instrumen tersebut, dan juga pelaksanaan tes (AERA, APA, and NCME, 1999). Keberadaan validitas dari suatu instrumen ini dapat diketahui melalui analisis isi tes dan analisis empiris dari skor tes data respons peserta terhadap butir instrumen tersebut (Lissitz & Samuelsen, 2007; Kumaidi, 2014).

Validitas isi instrumen terkait dengan keterwakilan aspek dari keseluruhan substansi yang akan diukur dan sejauh mana butir dalam instrumen mencerminkan karakteristik yang akan diukur (Gregory, 2007). Ada dua jenis validitas kriteria, yaitu validitas prediktif dan validitas konkuren. Pemanfaatan skor instrumen untuk memprediksi masa mendatang merupakan validitas prediktif, sedangkan pemanfaatan skor instrumen untuk memprediksi skor instrumen lain dengan tenggang waktu yang hampir bersamaan merupakan validitas konkuren (Gregory, 2007).

Validitas konstruk terkait dengan klaim instrumen untuk mengukur atribut psikologi yang kompleks. Konstruk tersebut dibuktikan setelah mencocokkannya dengan data empiris. Apabila hasilnya

sesuai dengan teori yang digunakan, maka instrumen itu dianggap memiliki validitas konstruk yang baik. Validitas konstruk digunakan untuk membuktikan konstruk teoretis tertentu yang akan diukur dalam suatu instrumen (Wright, 2008). Sejalan dengan hal tersebut, Cronbach & Meehl menyatakan bahwa investigasi validitas konstruk termasuk prosedur untuk mengembangkan dan melakukan konfirmasi terhadap teori tertentu (Haig, 2012). Validitas konstruk ini dapat dibuktikan dengan menggunakan analisis faktor, baik analisis faktor eksploratori (*exploratory factor analysis, EFA*) maupun konfirmatori (*confirmatory factor analysis, CFA*). Analisis ini bertujuan untuk mengurangi faktor (Izquierdo, Olea, Abad, 2014), menguji sifat dan hubungan antar konstruk laten (Jackson, Gillaspay, Stephenson, 2009), dan juga merupakan cara untuk mengidentifikasi apakah sekumpulan butir reliabel mengukur suatu variabel laten (ten Holt, van Duijn, Boomsma, 2010).

Analisis faktor digunakan untuk mereduksi data, dengan menemukan hubungan antar variabel yang saling bebas (Hardle & Hlavka, 2007), yang kemudian terkumpul dalam variabel yang jumlahnya lebih sedikit untuk mengetahui struktur dimensi laten (Garson, 2006), yang disebut dengan faktor. Faktor ini merupakan variabel yang baru, yang disebut juga dengan variabel laten, variabel konstruk dan memiliki sifat tidak dapat diketahui langsung (*unobservable*). Analisis faktor dapat dilakukan dengan dua cara, yakni analisis faktor eksploratori (*exploratory factor analysis*) dan analisis faktor konfirmatori (*confirmatory factor analysis*).

Alasan utama untuk melakukan analisis faktor baik eksploratori maupun konfirmatori adalah mereduksi banyaknya variabel dari suatu instrumen. Pada analisis faktor konfirmatori, ada dua jenis variabel, yakni variabel teramati (*observable variabel*) dan variabel laten (Schreiber, Nora, Stage, Barlow, King, 2006). Misalkan variabel awalnya adalah x_1, \dots, x_q yang merupakan variabel teramati, yang selanjutnya akan ditemukan himpunan faktor laten ξ_1, \dots, ξ_n (dengan $q > n$). Variabel *observable* tergantung pada kombinasi linear faktor laten ξ_i yang dinyatakan dengan

$$X_i = \lambda_{i1} \xi_1 + \lambda_{i2} \xi_2 + \dots + \lambda_{in} \xi_n + \delta_i$$

dan δ_i (kesalahan pengukuran) merupakan bagian unik dari x_i yang diasumsikan tidak berkorelasi dengan $\xi_1, \xi_2, \dots, \xi_n$. Untuk $i \neq j$, maka $\delta_i \neq \delta_j$. Bagian unik terdiri dari faktor khusus s_i dan suatu kesalahan pengukuran acak e_i . Pada analisis faktor eksploratori, faktor-faktor ditemukan berdasarkan data empiris. Namun pada analisis faktor konfirma-

tori, faktor ini ditentukan berdasarkan teori atau informasi *a priori* (Garson, 2006) kemudian dikonfirmasi menggunakan data empiris. Faktor-faktor ini biasanya lebih sedikit jumlahnya dibandingkan dengan variabel terobservasi (Joreskog dan Sorbom, 1993; Retnawati, 2016). Analisis faktor eksploratori digunakan ketika peneliti belum mengetahui banyaknya faktor, model yang akan dipilih, dan juga struktur hubungan antar faktor, sedangkan CFA ini digunakan untuk menguji model teoretis dengan struktur yang teramati dalam suatu sampel, digunakan jika peneliti telah yakin dengan banyaknya faktor, hubungan antar faktor, dan hubungan antar faktor dengan variabel yang terukur (Ullman, 2006). Hubungan antar variabel dalam CFA direpresentasikan dengan grafis. Ada bentuk-bentuk khusus yang digunakan untuk menyimbolkan variabel-variabel dalam CFA. Bentuk oval digunakan untuk menggambarkan variabel laten, dan pesegi panjang digunakan untuk menyatakan variabel yang diamati, dan hubungan antar variabel dinyatakan dengan panah, baik panah satu arah yang menyatakan hubungan variabel hubungan langsung sebab akibat, maupun panah ganda untuk menyatakan hubungan bolak-balik (Milfont & Fisher, 2010).

Selain validitas, suatu instrumen pengukuran dikatakan baik jika memiliki sifat reliabel. Reliabilitas merupakan derajat keajegan (*consistency*) di antara dua skor hasil pengukuran pada objek yang sama (Reynold, Livingstone, & Wilson, 2010). Untuk mengetahui derajat keajegan ini, Gregory (2007) menyatakan bahwa tes dikatakan reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor yang sebenarnya, dan nilai korelasi ini dinamai dengan koefisien reliabilitas (*reliability coefficient*, ρ). Suatu alat ukur dikatakan memiliki koefisien reliabilitas tinggi manakala digunakan untuk mengukur hal yang sama pada waktu berbeda hasilnya konsisten.

Reliabilitas (ρ) suatu tes pada umumnya diekspresikan secara numerik dalam bentuk koefisien yang besarnya $-1,00 \leq \rho \leq +1,00$ (Nitko & Brookhart, 2007). Koefisien tinggi menunjukkan reliabilitas tinggi. Sebaliknya, jika koefisien suatu skor tes rendah maka reliabilitas tes rendah. Nilai ini terkait pula dengan kesalahan pengukuran. Reliabilitas tinggi menunjukkan kesalahan yang kecil dalam memperoleh hasil pengukuran. Semakin besar reliabilitas suatu instrumen, akan semakin kecil kesalahan pengukuran, demikian pula sebaliknya, semakin kecil reliabilitas skor, akan semakin besar kesalahan pengukurannya (Gregory, 2007). Estimasi reliabilitas tes yang dapat dilakukan dengan beberapa cara,

konsistensi eksternal, konsistensi internal (Gregory, 2007; Nitko & Brookhart, 2007; Wright, 2008), reliabilitas komposit (Gregory, 2007), reliabilitas konstruk (Kamata, Turhan, Darandari, 2003), reliabilitas interrater (Nitko & Brookhart, 2007), dan estimasi reliabilitas dengan teori generalisabilitas (*Generalizability theory*) (Retnawati, 2016). Estimasi reliabilitas konstruk disajikan berikut ini.

Reliabilitas konstruk ini dapat diestimasi setelah pengembang instrumen membuktikan validitas konstruk dengan analisis faktor konfirmatori sampai memperoleh model yang cocok (model yang fit) (Retnawati, 2016). Dengan analisis faktor ini, pengembang instrumen dapat memperoleh muatan faktor (*factor loading*) tiap indikator yang menyusun instrumen (λ) dan indeks kesalahan unik dari tiap indikator (δ). Estimasi reliabilitas dapat dilakukan dengan tiga cara, estimasi reliabilitas dengan reliabilitas konstruk (*construct reliability*, *CR*) (Geldhof, Preacher, Zyphur, 2014), reliabilitas ω (Kamata, Turhan, Darandari, 2003), dan reliabilitas maksimal (Ω) (Penev & Raykov, 2006). Ketiga formula tersebut mengukur hal yang sama, ketiganya menggunakan muatan faktor dalam proses estimasi, namun formulanya dinyatakan dengan cara yang berbeda.

Jika instrumen yang digunakan untuk melakukan asesmen merupakan instrumen yang sama dari tahun ke tahun, informasi mengenai validitas dan reliabilitasnya perlu diketahui. Pada validitas konstruk, membuktikan validitas antar tahun atau lintas kelompok dapat dilakukan dengan krosvalidasi (*crossvalidation*) (Karimi, Karanika-Murray, & Meyer, tth). Krosvalidasi digunakan untuk menguji apakah ada kesamaan konstruk antartahun pada model teori hasil analisis faktor konfirmatori. Krosvalidasi dilakukan dengan memanfaatkan muatan faktor dari analisis faktor konfirmatori, reliabilitas konstruk instrumen antar tahun juga dapat diestimasi, kemudian dibandingkan menggunakan statistik untuk menguji kesamaannya.

Milfont & Fisher (2010) menyatakan dua jenis krosvalidasi, yaitu model invariansi pengukuran (*measurement invariance model*) dan model invariansi struktural (*structural invariance model*). Perbedaan antara keduanya yakni model invariansi pengukuran mengetahui invarians dari konstruk, muatan faktor, intersep butir dan varians kesalahan, biasanya digunakan untuk perbandingan lintas grup, sedangkan model invariansi struktural mengakses invariansi dari varians, kovarians, dan rerata dari variabel laten. Lebih lanjut Milfont & Fisher (2010) menyatakan bahwa menguji model invariansi pengukuran ini meliputi menguji invariansi konfigurasi, invari-

ansi varians kesalahan, invariansi metrik, invariansi varians faktor, invariansi scalar, invariansi kovarians faktor, invariansi rerata faktor. Menguji kesamaan konfigurasi merupakan langkah pertama. Langkah ini digunakan untuk membuktikan tidak adanya perbedaan lintas grup, diindikasikan dengan penggunaan grup yang berbeda namun mengukur hal yang sama. Metrik invarians digunakan untuk membuktikan kesamaan konstruk lintas grup. Invariansi scalar atau disebut juga inariansi intersep membandingkan rerata variabel laten, yang menunjukkan kesamaan konstruk variabel laten lintas grup. Invariansi varians kesalahan menguji kesamaan kesalahan pengukuran lintas grup. Invariansi varians faktor ditunjukkan dengan rentang dari skor pada variabel laten tidak berbeda lintas grup. Invariansi kovarians faktor menunjukkan stabilitas hubungan antar variabel laten lintas grup. Invariansi rerata faktor menunjukkan grup tidak berbeda dengan menggunakan konstruk yang sama, yang dilakukan dengan membatasi kesamaan rerata lintas grup.

Analisis faktor konfirmatori dapat digunakan untuk menguji model teoritis dengan struktur yang teramati dalam suatu sampel. Indikator suatu model dikatakan baik jika memenuhi kecocokan model. Kecocokan model (*goodness of fit indices*) digunakan untuk menyatakan derajat konsistensi model teoretis dengan data empiris (Retnawati, 2016). Hu & Bentler menyatakan bahwa salah satu diantaranya dengan uji perbandingan likelihood (atau uji *chi-square*), dan indikator lainnya adalah RMSEA (*root mean square error of approximation*) dekat dengan 0,06 (Milfont & Fisher (2010).

Beberapa penelitian mengenai krosvalidasi telah dilakukan peneliti sebelumnya. Theodorakis (2014) menggunakan krosvalidasi untuk membuktikan validitas konstruk dari *other costumers perceptions scale* (OCP) pada konteks *fitness centers*. Studi ini memberikan hasil bahwa OCP merupakan instrumen yang valid lintas pengguna/tahun, sehingga dapat direkomendasikan penggunaan instrumen dari tahun ke tahun. Dengan krosvalidasi pula, Karimi, Kranikan-Murray, Murray (tth) membuktikan adanya capaian skor *work organization assessment questionair* lintas jenis kelamin, yang memberikan rekomendasi instrumen valid digunakan pada responden laki-laki dan perempuan. Studi ini menggunakan data respons peserta terhadap tes potensi akademik calon mahasiswa sekolah pascasarjana. Bagi pengelola sekolah pascasarjana, tes ini sangat berarti, karena memberikan informasi karena skor perolehannya dapat digunakan untuk mencirikan kemampuan calon (Howel, Sorenson, Jones, 2014). Dengan diketahuinya karakteristik kemampuan ca-

lon, pengelola dapat memberikan perlakuan yang tepat untuk calon sehingga dapat menyelesaikan studi dengan baik.

Terkait dengan pelaksanaan seleksi, jika instrumen yang digunakan memenuhi sifat valid dan reliabel, keberhasilan mahasiswa dapat diprediksikan. Keberhasilan mahasiswa pascasarjana dalam menyelesaikan studi dipengaruhi berbagai factor. Terkait dengan keberhasilan mahasiswa pascasarjana, Shavelson & Hung mengatakan bahwa "*framework learning outcome meliputi intelligence (general, fluid, crystallized), general reasoning (verbal, quantitative, spatial), broad abilities (reasoning, critical thinking, problem solving, decision making, communicating in broad domain), and knowledge, understanding and reasoning in major field sand professional*" (Shavelson, 2007). Hal tersebut diperkuat Wang (2013) yang membuktikan bahwa GRE (*graduate record examination*) merupakan prediktor untuk memprediksi indeks prestasi (*graduate grade point average, GGPA*) mahasiswa. Senada dengan pendapat-pendapat tersebut, Kuncel, Hezlett, Ones (2001) dan Howell, Sorenson, dan Jones (2014) menyatakan bahwa GRE dan *undergraduate point average* merupakan predictor yang kuat memprediksi GGPA.

Sejalan dengan pelaksanaan seleksi yang menggunakan instrumen yang sama dari tahun ke tahun, diperlukan informasi mengenai karakteristik instrumen yang digunakan tersebut. Karakteristik instrumen dapat diketahui diantaranya dengan menginvestigasi validitas dan reliabilitas konstruksya dari tahun ke tahun. Terkait dengan latar belakang tersebut, studi ini bertujuan untuk (1) menginvestigasi validitas konstruk tes kemampuan awal mahasiswa dengan menguji kecocokan model, mengestimasi faktor loading, melakukan krosvalidasi antartahun, menguji kesamaan muatan faktor antar tahun, dan (2) mengestimasi reliabilitas konstruk skor kemampuan awal mahasiswa dengan menggunakan metode reliabilitas konstruk, koefisien omega, dan reliabilitas maksimal.

METODE

Studi ini merupakan studi deskriptif eksploratif dengan menggunakan pendekatan kuantitatif. Data penelitian berupa data dokumentasi dari skor calon mahasiswa terhadap tes yang mengukur kemampuan awal yang mengambil master di sekolah pascasarjana di Yogyakarta. Data dikumpulkan dengan metode dokumentasi data selama 3 tahun. Selama tahun 2013 didokumentasikan 191 calon mahasiswa, tahun 2014 sebanyak 589 calon mahasiswa,

dan tahun 2015 sebanyak 113 calon mahasiswa, yang semuanya mengerjakan tes yang sama.

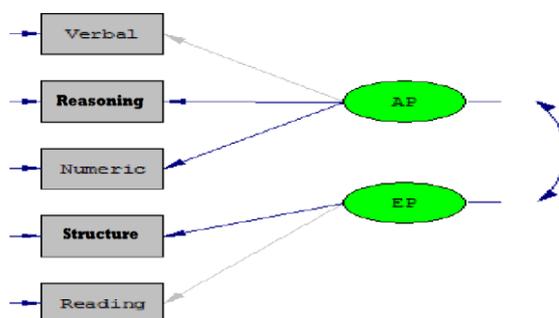
Tes kemampuan awal untuk seleksi di perguruan tinggi biasanya disebut dengan tes potensi akademik atau tes bakat skolastik (Asrijanti, 2014). Tes tersebut tersusun atas beberapa subtes. Subtes tersebut yakni potensi akademik (*Academic Proficiency*, AP) dan kemampuan bahasa Inggris (*English Proficiency*, EP). Tes potensi akademik dan kemampuan bahasa Inggris ini merupakan kemampuan yang dianggap menjadi prediktor keberhasilan calon dalam menyelesaikan tugas akademik di perguruan tinggi dengan memanfaatkan literatur dalam bahasa Inggris (Pali, 2009). Tes kemampuan akademik tersusun atas tes kemampuan verbal (*Verbal*) sebanyak 25 butir, tes kemampuan penalaran (*Reasoning*) sebanyak 30 butir dan tes kemampuan numerik (*Numeric*) sebanyak 30 butir. Subtes Kemampuan bahasa Inggris tersusun atas tes *structure* sebanyak 30 butir dan tes *reading* sebanyak 45 butir. Tes berbentuk pilihan ganda, dengan penskoran dikotomi. Peserta tes yang menjawab butir dengan betul diskor 1 dan jika salah, diberikan skor 0. Skor butir dari tiap subtes tes dijumlahkan, kemudian dientrikan untuk dianalisis.

Analisis dilakukan untuk membuktikan validitas konstruk dan mengestimasi reliabilitasnya. Analisis validitas konstruk dilakukan dengan menggunakan analisis faktor konfirmatori, karena komponen tes potensi akademik telah jelas dan penyusunan instrument disadarkan pada konstruk yang dibangun berdasarkan teori tertentu. Analisis faktor konfirmatori dilakukan dengan bantuan Lisrel 8.8 (Joreskog & Sorbom, 1993), dilakukan secara terpisah untuk tiap tahun. Analisis ini menghasilkan kecocokan model, muatan faktor dan kesalahan (*error*) dari tiap variabel terobservasi. Hasil ini digunakan untuk membuktikan validitas konstruk dan mengestimasi reliabilitas. Untuk mengetahui stabilitas

muatan faktor dan juga membuktikan kesamaan model, dilakukan krosvalidasi antartahun. Hasil analisis disajikan dalam tabel-tabel dan grafik-grafik, untuk diinterpretasikan. Hubungan antar variabel amatan dan variabel laten untuk analisis faktor konfirmatori disajikan dalam Gambar 1. Kecocokan model dilihat dari nilai p-value yang melampaui 0,05. Muatan faktor dibandingkan antar tahun. Krosvalidasi dilakukan tiga kali, 2013-2014, 2013-2015, dan 2014-2015, untuk menguji kesamaan model maupun kesamaan muatan faktor.

Kesamaan model maupun kesamaan muatan faktor pada dasarnya menguji hipotesis H_0 yang menyatakan terjadinya kesamaan model konstruk pada CFA antar tahun. Analisis ini dilakukan dengan bantuan Lisrel (Mels, 2006). Hipotesis alternatif untuk H_0 tersebut adalah H_1 untuk menyatakan terjadinya ketidaksamaan model konstruk pada CFA antar tahun. Selanjutnya dianalisis untuk model awal, dan model alternatif. Perbedaan nilai khi-kuadrat, derajat kebebasan dan nilai signifikansi digunakan untuk membuat keputusan menerima atau menolak H_0 , dengan kriteria H_0 ditolak jika nilai signifikansi kurang dari 0,05. Demikian pula ketika menguji invariansi muatan faktor.

Dengan menggunakan muatan faktor, reliabilitas konstruk kemudian diestimasi (Geldhof, Preacher, Zyphur, 2014; Kamata, Turhan, Darandari, 200; Penev & Raykov, 2006). Reliabilitas yang menggunakan muatan faktor yang dihasilkan dari analisis faktor konfirmatori yakni (*construct reliability*, CR), reliabilitas ω , dan reliabilitas maksimal (Ω). Pada studi ini ketiga cara estimasi validitas konstruk digunakan, kemudian dibandingkan. Estimasi CR menggunakan muatan faktor (*factor loading*) tiap indikator yang menyusun instrumen (λ) dan indeks kesalahan unik dari tiap indikator (δ). Formulasinya sebagai berikut (Geldhof, Preacher, Zyphur, 2014).



Gambar 1. Konstruk Teori dari Tes Kemampuan Awal Calon Mahasiswa yang Terdiri dari Subtes Potensi Akademik (AP) dan Kemampuan Bahasa Inggris (EP)

$$CR = \frac{\left(\sum_{i=1}^i \lambda_i\right)^2}{\left(\sum_{i=1}^i \lambda_i\right)^2 + \left(\sum_{i=1}^i \delta_i\right)^2} \quad (1)$$

Estimasi dengan reliabilitas ω dilakukan hanya dengan menggunakan muatan faktor (λ) saja. Formula untuk estimasi reliabilitas ω sebagai berikut (Kamata, Turhan, Darandari, 2003).

$$\omega = \frac{\left(\sum_{i=1}^i \lambda_i\right)^2}{\left(\sum_{i=1}^i \lambda_i\right)^2 + \left(\sum_{i=1}^i 1 - \lambda_i^2\right)} \quad (2)$$

Pada estimasi reliabilitas maksimal, muatan faktor disimbolkan dengan l , dengan formula sebagai berikut (Penev & Raykov, 2006).

$$\Omega_w = \frac{\sum_{i=1}^p \frac{l_i^2}{(1 - l_i^2)}}{1 + \sum_{i=1}^p \frac{l_i^2}{(1 - l_i^2)}} \quad (3)$$

Setelah reliabilitas diestimasi, selanjutnya dilakukan analisis uji kesamaan reliabilitas antar tahun. Uji ini digunakan untuk mengetahui apakah koefisien reliabilitas suatu tes lebih tinggi dibandingkan dengan koefisien reliabilitas tes yang lain, dapat digunakan uji kesamaan dua koefisien reliabilitas. Kesamaan dua koefisien reliabilitas dapat diketahui dengan uji-H* yang dikembangkan oleh

Feldt (Feldt & Brennan, 1989). Nilai H* untuk menguji kesamaan dua koefisien reliabilitas $\hat{r}_{x_1x_1}$ dan $\hat{r}_{x_2x_2}$ memenuhi persamaan

$$H^* = \frac{(1 - \hat{r}_{x_1x_1})}{(1 - \hat{r}_{x_2x_2})} \quad (4)$$

dengan H* berdistribusi F dengan derajat kebebasan $N_x - 1$ dan $N_y - 1$.

Hasil membuktikan validitas dan reliabilitas konstruk antar tahun kemudian disajikan dan diinterpretasikan dalam hasil studi.

HASIL DAN PEMBAHASAN

Dengan menggunakan model pada Gambar 1, skor-skor tiap subtes kemampuan awal mahasiswa tiap tahun dianalisis dengan analisis faktor konfirmatori. Analisis ini menghasilkan nilai *Chi-square* dan derajat kebebasan (*degree of freedom, df*) dan *p-value* untuk menguji kecocokan konstruk didukung oleh data empiris, dan juga muatan faktor dan *error* dari tiap subsub kemampuan. Hasil analisis disajikan pada Tabel 1.

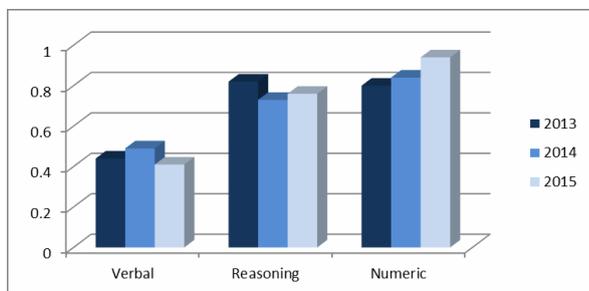
Mencermati Tabel 1 tersebut, diperoleh bahwa konstruk pada Gambar 1 merupakan model yang cocok (*fit*), dengan indikasi nilai *p-value* melampaui taraf signifikansi sebesar 0,05. Hal ini menunjukkan bahwa data yang didokumentasikan selama 3 tahun tersebut mendukung konstruk teori. Ini berarti

Tabel 1. Hasil Analisis Faktor Konfirmatori

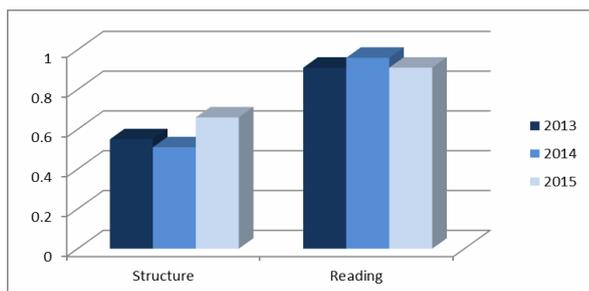
Tahun	Variabel Teramati (Observable Variable)	Variabel Laten (Latent Variable)	Muatan Faktor (Factor Loading)	Delta (δ)	Chi-Square	df	p-value
2013	Verbal		0,44	0,8	4,66	4	0,324
	Reasoning	Potensi akademik	0,82	0,33			
	Numeric		0,8	0,36			
	Structure	Kemampuan	0,55	0,7			
	Reading	Bahasa Inggris	0,91	0,18			
2014	Verbal		0,49	0,76	7,82	4	0,098
	Reasoning	Potensi akademik	0,73	0,47			
	Numeric		0,84	0,29			
	Structure	Kemampuan	0,51	0,74			
	Reading	Bahasa Inggris	0,96	0,07			
2015	Verbal		0,41	0,83	1	4	0,910
	Reasoning	Potensi akademik	0,76	0,42			
	Numeric		0,94	0,12			
	Structure	Kemampuan	0,66	0,57			
	Reading	Bahasa Inggris	0,91	0,18			

bahwa tes kemampuan awal mahasiswa, yang terdiri subtes AP dapat diukur dengan sub-sub tes kemampuan verbal, kemampuan reasoning, dan kemampuan numeric. Demikian pula kemampuan bahasa Inggris, terbukti secara empiris dapat diukur menggunakan subsubtes structure dan reading. Hal yang perlu menjadi perhatian adalah nilai *p-value*, pada hasil CFA tahun 2014. Nilai *p-value* pada tahun 2014 sebesar 0,098 yang jauh lebih rendah jika dibandingkan dengan hasil tahun 2013 dan tahun 2015.

Muatan faktor pada Tabel 1 disajikan ulang dengan bar diagram untuk melihat polanya pada tiap subtes. Pada potensi akademik disajikan pada Gambar 2 dan pada kemampuan bahasa Inggris disajikan pada Gambar 3. Mencermati kedua gambar tersebut, nampak bahwa ada kekonsistenan hasil estimasi muatan faktor pada tiap subsubtes, baik pada potensi akademik maupun kemampuan bahasa Inggris. Hal ini menunjukkan bahwa berdasarkan hasil estimasi, diperoleh kesamaan hasil muatan faktor tiap subsubtes tahun 2013, 2014, dan tahun 2015.



Gambar 2. Muatan Faktor Subtes Potensi Akademik



Gambar 3. Muatan Faktor Subtes Kemampuan Bahasa Inggris

Konsistensi muatan faktor kemudian diuji. Uji dilakukan dengan menggunakan selisih *Chi-Square*. Hasil selengkapnya disajikan pada Tabel 2. Hasil analisis menunjukkan bahwa *Significance* pada hi-

potesis mengenai *difference* pada tahun 2013-2014, 2013-2015, dan 2014-2015 menghasilkan nilai lebih dari 0,05. Hal ini berarti bahwa koefisien muatan faktor yang sama antar tahun besarnya sama.

Tabel 2. Hasil Analisis Uji Invariansi dari Muatan Faktor

Invariansi Muatan Faktor Tahun	Hipotesis	Chi-Square	df	Significance
2013-2014	H ₀	34,03	16	0,005
	H ₁	25,08	12	0,054
	<i>Difference</i>	8,95	4	0,062
2013-2015	H ₀	13,99	16	0,600
	H ₁	11,24	12	0,508
	<i>Difference</i>	2,75	4	0,600
2014-2015	H ₀	22,33	16	0,007
	H ₁	26,92	12	0,008
	<i>Difference</i>	4,59	4	0,332

Hasil yang agak berbeda nampak pada hasil krosvalidasi model antar tahun. Pada krosvalidasi menggunakan data 2013 dan 2015, dihasilkan signifikansi sebesar 0,819 yang melampaui 0,05. Namun signifikansi ini menjadi kurang dari 0,05 ketika menggunakan data 2014, yaitu pada analisis krosvalidasi 2013-2014 yakni sebesar 0,002 dan 2014-2015 sebesar 0,012. Hasil selengkapnya disajikan pada Tabel 3. Hal tersebut menunjukkan bahwa model validasi instrumen 2013 dan 2014 merupakan model yang berbeda, demikian pula model validasi instrumen 2014 dan 2015. Namun hal yang berbeda terjadi pada krosvalidasi menggunakan data tahun 2013 dan 2015, yang menghasilkan model yang sama.

Tabel 3. Hasil Analisis Krosvalidasi

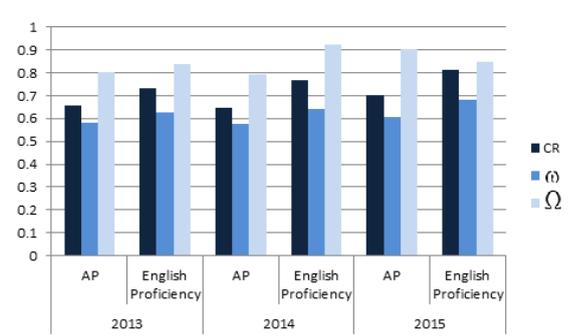
Krosvalidasi Tahun	Hipotesis	Chi-Square	Df	Significance
2013-2014	H ₀	26,94	14	0,020
	H ₁	45,95	19	0,001
	<i>Difference</i>	19,01	5	0,002
2013-2015	H ₀	12,09	14	0,600
	H ₁	14,3	19	0,766
	<i>Difference</i>	2,21	5	0,819
2014-2015	H ₀	27,67	14	0,016
	H ₁	42,25	19	0,002
	<i>Difference</i>	14,58	5	0,012

Dengan menggunakan formula (1), (2), dan (3), serta menginputkan nilai muatan faktor dan

kesalahan (*error*) dari hasil analisis faktor konfirmatori, dapat diestimasi reliabilitas konstruk. Hasil analisis menunjukkan bahwa menggunakan rumus CR dan ω , menghasilkan reliabilitas yang sedang untuk semua subtes. Estimasi reliabilitas dengan menggunakan rumus reliabilitas maksimal (Ω) menghasilkan koefisien yang paling tinggi dibandingkan estimasi dengan rumus CR dan ω . Ada kecenderungan, hasil yang relatif stabil pada estimasi antar tahun, baik pada subtes AP maupun EP. Hasil selengkapnya disajikan pada Tabel 4 dan Gambar 4.

Tabel 4. Hasil Estimasi Reliabilitas dengan Rumus Reliabilitas Konstruk (CR), Reliabilitas Omega (ω), dan Reliabilitas Maksimal (Ω)

Tahun	Subtes	CR	ω	Ω
2013	Potensi Akademik	0,657	0,580	0,803
	Kemampuan Bahasa Inggris	0,734	0,627	0,840
2014	Potensi Akademik	0,647	0,575	0,794
	Kemampuan Bahasa Inggris	0,767	0,642	0,924
2015	Potensi Akademik	0,703	0,606	0,902
	Kemampuan Bahasa Inggris	0,814	0,681	0,848



Gambar 4. Grafik Hasil Estimasi Reliabilitas dengan Rumus Reliabilitas Konstruk, Reliabilitas Omega, dan Reliabilitas Maksimal

Selanjutnya, kesamaan reliabilitas diuji dengan menggunakan uji-F dari Feld & Brennan persamaan (4). Nilai H^* dengan membandingkan reliabilitas hasil estimasi tahun 2014 dan tahun 2015 dengan hasil estimasi tahun 2013 maupun hasil estimasi pada data tahun 2014 dengan data tahun 2013, baik pada perangkat AP maupun *English Proficiency*. Hasil selengkapnya disajikan pada Ta-

bel 5 untuk CR, Tabel 6 untuk ω dan Tabel 7 untuk Ω . Hasil pada tabel ini kemudian dibandingkan dengan Nilai F tabel dengan taraf signifikansi 5% dan derajat kebebasan 74 untuk pembilang dan 74 untuk penyebut (banyaknya butir soal pada tiap perangkat 75). Nilai F (0,05; 74,74) sebesar 1,469.

Tabel 5. Nilai F untuk Perbandingan Reliabilitas CR

Tahun	Subtes	Tahun		
		2013	2014	2015
2013	Potensi Akademik	1	1,026	0,863
	Kemampuan Bahasa Inggris	1	0,874	0,697
2014	Potensi Akademik		1	1,026
	Kemampuan Bahasa Inggris		1	1,113
2015	Potensi Akademik			1
	Kemampuan Bahasa Inggris			1

Tabel 6. Nilai F untuk Perbandingan Reliabilitas ω

Year	Subtes	Tahun		
		2013	2014	2015
2013	Potensi Akademik	1	1,011	0,937
	Kemampuan Bahasa Inggris	1	0,958	0,855
2014	Potensi Akademik		1	1,011
	Kemampuan Bahasa Inggris		1	1,055
2015	Potensi Akademik			1
	Kemampuan Bahasa Inggris			1

Tabel 7. Nilai F untuk Perbandingan Reliabilitas Ω

Year	Subtest	Tahun		
		2013	2014	2015
2013	Potensi Akademik	1	1,045	0,499
	Kemampuan Bahasa Inggris	1	0,477	0,949
2014	Potensi Akademik		1	1,045
	Kemampuan Bahasa Inggris		1	0,615
2015	Potensi Akademik			1
	Kemampuan Bahasa Inggris			1

Berdasarkan hasil estimasi H^* pada Tabel 5, Tabel 6, dan Tabel 7 tersebut, diperoleh bahwa hasil nilai H^* kurang dari 1,469. Hal ini menunjukkan

bahwa dengan menggunakan 3 cara estimasi reliabilitas konstruk yang berbeda, yaitu dengan formula CR, ω , maupun Ω diperoleh hasil perbandingan kesamaan reliabilitas konstruk.

Berdasarkan hasil analisis, diperoleh bahwa muatan faktor antartahun pada perangkat tes seleksi masuk sekolah pascasarjana baik pada subtes potensi akademik maupun pada subtes bahasa Inggris. Hasil ini menunjukkan konsistensi, instrumen mengukur hal yang sama dan telah terbukti secara statistik. Hal ini memperkuat bukti bahwa instrumen yang digunakan untuk mengukur potensi mahasiswa sebelum kuliah ini merupakan instrumen yang mengukur konstruk yang sama, didukung oleh data empiris.

Pada hasil analisis krosvalidasi, diperoleh bahwa model untuk membuktikan validitas konstruk menggunakan data tahun 2013 berbeda dengan model menggunakan data tahun 2014. Demikian pula pada krosvalidasi menggunakan data tahun 2014, dihasilkan model yang berbeda dengan menggunakan data tahun 2015. Namun yang unik, krosvalidasi menggunakan data tahun 2013 dan 2015 memberikan hasil, model konstruk kedua tahun ini merupakan model yang sama. Mencermati lebih jauh hasil ini, terlebih dibandingkan dengan muatan faktor yang hasilnya relatif sama dalam 3 tahun, cara mengetahui kecocokan model perlu dijadikan perhatian.

Untuk mengetahui kecocokan model pada krosvalidasi studi ini menggunakan uji *Chi-square* (Hu & Bentler, 1999) dan perbandingan *Chi-square* kedua model (Garson, 2003). Mengingat khi-kadrat merupakan uji statistik yang sensitif terhadap ukuran sampel (Mislevy & Bock, 1999), maka dapat dipahami banyaknya peserta tes yang berbeda jauh menyebabkan adanya perbedaan signifikansi model. Pada studi ini, tahun 2013 menggunakan atau dengan kata lain, peserta tes tahun 2014 relatif banyak jika dibandingkan dengan 2013 dan 2015.

DAFTAR RUJUKAN

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 1999. *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Asrijanti, A. 2014. Validitas Prediktif Bakat Skolastik dan Prestasi Belajar Sebagai Kriteria Seleksi Masuk Perguruan Tinggi. *Jurnal Pendidikan dan Kebudayaan*, 20 (4), 515-534.
- Dodge, L. & Derwin, E.B. 2008. Overcoming Barriers of Tradition Through an Effective New Graduate admission Policy. *The Journal of Continuing Higher Education*, 56(2), 2-11.
- Ekwoaba, J.O., Ikeije, U.U., Ufoma, N. 2015. The Impact of Recruitment and Selection Criteria on Organizational Performance. *Global Journal of Human Resource Management*, 3(2), 22-33.
- Garson, D. 2006. *Factor Analysis, Path Analysis & SEM*. Diambil tanggal 24 September 2006 dari <http://www2.chass.ncsu.edu/garson/pa765/index.htm>.
- Geldhof, G.J., Preacher, K., Zyphur, M.J. 2014. Reliability Estimation in a Multilevel Confirmatory Factor Analysis Framework. *Psychological Methods*, 19(1), 72-91.
- Gregory, R.J. 2007. *Psychological Testing: History, Principles, and Applications*. Boston, MA: Pearson.

Mengenai hasil perbandingan reliabilitas, diperoleh bahwa reliabilitas konstruk dengan formula Ω memberikan hasil yang lebih besar dibandingkan estimasi dengan formula ω dan formula CR. Hal ini disebabkan karena pembagi pada formula Ω melibatkan $(1-li)$ dan ini yang menyebabkan formula Ω berbeda jauh dengan formula CR dan formula ω . Berdasarkan formula CR (1) dan formula ω (2) diperoleh $(1-\lambda^2)$ lebih besar dibandingkan dengan δ pada studi ini sehingga menghasilkan hasil estimasi reliabilitas konstruk ω yang lebih kecil dibandingkan estimasi reliabilitas dengan formula CR.

SIMPULAN

Dengan menggunakan krosvalidasi, analisis faktor konfirmatori dengan menggunakan data tahun 2013, 2014, dan 2015 menghasilkan model yang fit. Hasil uji invariansi muatan faktor menunjukkan bahwa muatan-muatan faktor bernilai sama antar tahun, namun berdasarkan hasil krosvalidasi model hasilnya sama pada tahun 2013 dan 2015. Reliabilitas subtes yang diestimasi dengan reliabilitas konstruk, reliabilitas ω , menghasilkan koefisien pada level mediocare, dan namun pada reliabilitas maksimal menghasilkan koefisien pada kategori tinggi.

Selanjutnya, penelitian mengenai sifat krosvalidation perlu dilakukan dengan memanfaatkan banyaknya peserta tes yang berbeda-beda. Banyaknya peserta tes yang perlu diselidiki misalnya peserta tes yang cukup banyak dengan peserta tes yang sedikit, apakah berpengaruh pada kecocokan model dan berdampak pada hasil estimasi muatan faktor. Demikian pula penelitian mengenai karakter dan perbandingan estimasi reliabilitas konstruk dengan formula Ω , formula CR, dan formula ω . Studi ini dapat dilakukan dengan studi simulasi, untuk memperkuat pembahasan mengenai sifat-sifat estimasi reliabilitas konstruk.

- Haig, B.D. 2012. From Construct Validity to Theory Validation, Measurement. *Interdisciplinary Research and Perspectives*, 10:1-2, 59-62, DOI: 10.10-80/15366367.2012.681975
- Hardle, W.K. & Hlavka, Z. 2007. *Multivariate Statistics: Exercises and Solutions*. New York, NY: Springer.
- Howell, L. L., Sorenson, C. D., Jones, M.R. 2014. "Are Undergraduate GPA and General GRE Percentiles Valid Predictors of Student Performance in an Engineering Graduate Program?" *All Faculty Publications*. Paper 1343. Tersedia di <http://scholarsarchive.byu.edu/facpub/1343>.
- Izquierdo, I., Olea, I. J., Abad, F.J. 2014. Exploratory Factor Analysis in Validation Studies: Uses and Recommendations. *Psicothema*, 26(3):395-400. Doi: 10.7334/psicothema2013.349.
- Jackson, D.L. Gillaspy, A. J., Stephenson, R.P. 2009. Reporting Practices in Confirmatory Factor Analysis: An Overview and Some Recommendations. *Psychological Methods*, 14(1), 6-23.
- Joreskog, K. & Sorbom, D. 1993. *Lisrel 88: Structural Equation Modeling with the SIMPLIS Command Language*. Hillsdale, NJ: Scientific Software International.
- Kamata, A., Turhan, A., Darandari, E. 2003. *Estimating Reliability for Multidimensional Composite Scale Scores*. Paper. Presented at The Annual Meeting of American Educational Research Association, Chicago, April 2003.
- Karimi, L., Karanika-Murray, M., & Meyer, D. tth. Cross-Validation of the Work Organization Assessment Questionnaire Across Genders: a Study in Australian healthcare organization. *Journal of Occupational & Environmental Medicine* (accepted).
- Kumaidi. 2014. Validitas dan Pemvalidasian Instrumen Penilaian Karakter. *Makalah* disampaikan dalam Seminar Nasional Pengembangan Instrumen Penilaian Pendidikan Karakter yang Valid, diselenggarakan Fakultas Psikologi, Universitas Muhammadiyah Surakarta, 24 Mei 2014.
- Kuncel, N.R., Hezlett, S.A., Ones, D.S. 2001. Comprehensive Meta-Analysis of the Predictive Validity of the Graduate Record examinations: Implications for Graduate Student Selection and Performance. *Psychological Bulletin*. 2001. 127(1): 162-181.
- Lissitz, W. & Samuelsen, K. 2007. Further clarification regarding validity and education. *Educational Researcher*, 36(8): 482-484.
- Martirosyan, N. M., Hwang, E., Wanjohi, R. 2015. Impact of English Proficiency on Academic Performance of International Students. *Journal of International Students*, 5(1), 60-71. <http://jistudents.org/>
- Mels, G. 2006. *LISREL for Windows: Getting Started Guide*. Lincolnwood, IL: Scientific Software International, Inc.
- Milfont, T.L., Fischer, R., 2010. Testing Measurement Invariance Across Groups: Applications in Cross-Cultural Research. *International Journal of Psychological Research*, 3(1), 111-121.
- Nitko, A.J. 2007. Educational Assessment of Students. Upper Saddle River, NJ: Pearson.
- Pali, M. 2009. Kontribusi Prediktor-Prediktor Utama Karakteristik Psikologis dan Iklim Sekolah Terhadap Keberhasilan Belajar Siswa. *Jurnal Pendidikan dan Pembelajaran*, 16(2), 131-139.
- Penev, S. & Raykov, T. 2006. On the Relationship between Maximal Reliability and Maximal Validity of Linear Composites. *Multivariate Behavioral Research*, 41(2), 105-126.
- Retnawati, H. 2014. *Teori Respons Butir dan Penerapannya (Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana)*. Yogyakarta: Parama.
- Retnawati, H. 2016. *Analisis Kuantitatif Instrumen Penelitian*. Yogyakarta: Parama.
- Reynold, C. R., Livingstone, R. B. & Wilson, V. 2010. *Measurement and Assessment in Education*. New York, NY: Pearson.
- Schreiber, J.B., Nora, A., Stage, F.K., Barlow, E.A., King, J. 2006. Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, 99(6): 323-337.
- Shavelson, R.J. 2007. A Brief History of Student Learning Assessment: *How We Got Where We are and a Proposal for Where to Go Next*. Washington: Association of American Colleges and Universities.
- Ten Holt, J.C., van Duijn, M.A.J., Boomsma, A., 2010. Scale Construction and Evaluation in Practice: A Review of Factor Analysis Versus Item Response Theory Applications. *Psychological Test and Assessment Modeling*, 52(3): 272-297.
- Theodorakis, N. D. 2014. A Cross-Validation Study of the Other Customers Perceptions Scale in the Context of Sport and Fitness Centres. *RICYDE. Revista Internacional de Ciencias del Deporte*, 35(10), 63-74. <http://dx.doi.org/10.5232/ricyde2014.03505>
- Ullman, J.B., 2006. Structural Equation Modeling: Reviewing the Basics and Moving Forward. *Journal of Personality Assessment*, 87(1), 35-50.
- Walpole, R.E., Mers, R.H., Myers, S.L. et al. 2002. *Probability and Statistics for Engineers and Scientists*. Upper Saddle River, NJ: Prentice Hall.
- Wang, W. 2013. "Testing the Validity of GRE Scores on Predicting Graduate Performance for Engineering Students" *Open Access Theses and Dissertations from the College of Education and Human Sciences*. Paper 192. <http://digitalcommons.unl.edu/cehsdiss/192>.
- Wright, R.J. 2008. Educational Assessment: Tests and Measurement In The Age of Accountability. Thousand Oaks, CA: Sage.