

Aplikasi Analisis Faktor Konfirmatori untuk Menentukan Reliabilitas Multidimensi

GAGUK MARGONO

Universitas Negeri Jakarta, Jl. Rawamangun Muka, Rawamangun, Jakarta 13220
Email: g_margono@yahoo.com

ABSTRACT

The aim of this paper is to draw on the application of Confirmatory Factor Analysis (CFA) in Structural Equation Modeling (SEM) to estimate of instrument for measuring attitudes toward statistics using semantic differential scale. The semantic differential scale is a type of survey question where respondents are asked to rate their opinion on a linear scale between 2 standpoints, theoretical typically with 7 levels. It has three dimensions such as evaluation, potency, and activity (EPA). This instrument has been tested to 133 students of Evaluation Research Program Education (Prodi PEP), Graduate Program in Jakarta State University (PPs UNJ). Reliability refers to accuracy and precision of a measurement instrument. Multidimensional measurement of internal consistency reliability is rarely used in the realm of research. Measurement and computation is described in this article uses an instrument measuring student attitudes toward statistics using semantic differential scale. It can be concluded that the instrument gauges student attitudes toward statistics using semantic differential scale using multidimensional reliability coefficient has a high accuracy when compared with a unidimensional reliability coefficient. Expected in advanced research used another formula multidimensional reliability, included if used SEM.

Keywords: CFA in SEM, multidimensional reliability coefficient, student attitudes toward statistics using semantic differential scale instrument.

1. PENDAHULUAN

Di bidang pendidikan, penilaian yang baik memerlukan pengukuran yang dapat diandalkan atau dipercaya. Menurut Naga (1992) untuk pengukuran pendidikan mencakup beberapa hal. Pertama, mengukur ciri terpendam yang tak kelihatan yang ada pada responden. Kedua, untuk mengukur ciri terpendam tersebut responden diberi stimulus berupa kuesioner atau alat ukur yang tepat. Ketiga, stimulus direspons oleh responden dengan harapan respons mencerminkan dengan benar ciri terpendam yang ingin diukur. Keempat, respons diskor dan dapat ditafsirkan secara memadai. Kemudian, perlu dipertanyakan sejauh manakah skor yang diperoleh dapat mencerminkan secara tepat ciri terpendam yang hendak diukur? Apakah instrumen yang dipakai sebagai stimulus itu mampu mengungkap secara benar ciri terpendam yang tak tampak itu? Kedua pertanyaan tersebut berkenaan dengan validitas. Sedang yang berkaitan dengan reliabilitas, apakah tanggapan yang diberikan oleh para peserta sudah dapat dipercaya untuk digunakan sebagai bahan penskoran bagi atribut psikologis itu?

Apapun yang digunakan untuk melakukan pengukuran disebut alat ukur (instrumen) yang seyogyanya terlebih dahulu divalidasi sebelum dipergunakan. Pada dasarnya ada dua macam instrumen, yaitu instrumen yang berbentuk tes untuk mengukur hasil belajar (kinerja maksimal) dan instrumen non tes untuk mengukur sikap (kinerja tipikal). Instrumen yang berupa tes jawabannya adalah salah atau benar, sedangkan instrumen nontes tidak ada salah atau benar tetapi bersifat positif atau negatif. Menurut Suryabrata (2000) untuk pengukuran non tes diperlukan respons jenis ekspresi sentimen, yaitu jenis respons yang tak dapat dinyatakan benar atau salah, seringkali dikatakan semua respons benar menurut alasannya masing-masing. Adapun tujuannya bukan untuk mengetahui apa yang mampu dilakukan melainkan apa yang akan cenderung dilakukan oleh seseorang. Di dalam penelitian ilmiah, instrumen yang baik diperoleh hanya melalui data dan diinterpretasikan dengan lebih baik bila diperoleh melalui proses pengukuran yang objektif, sah dan reliabel.

Menurut Wiersma (1986), reliabilitas ialah konsistensi suatu instrumen mengukur sesuatu yang hendak diukur. Reliabilitas menunjukkan sejauh mana hasil pengukuran dengan alat tersebut dapat dipercaya. Oleh karena itu reliabilitas merupakan indeks yang menunjukkan

sejauh mana suatu alat pengukur dapat dipercaya atau dapat diandalkan. Bila suatu instrumen dipakai berulang-ulang untuk mengukur gejala yang sama dan hasil yang diperoleh relatif stabil atau konsisten, maka instrumen tersebut terpercaya.

Secara garis besar ada tiga kategori besar dalam pengukuran reliabilitas: (1) tipe stabilitas (misalnya: tes ulang, bentuk paralel, dan bentuk alternatif), (2) tipe homogenitas atau internal konsistensi (misalnya: belah dua, Kuder-Richardson, *alpha* Cronbach, *theta* dan *omega*), dan (3) tipe ekuivalen (misalnya: butir-butir paralel pada bentuk alternatif dan reliabilitas antar penilai (*inter-rater reliability*)). Instrumen diberikan kepada sekelompok subjek satu kali lalu dengan cara tertentu dihitung estimasi reliabilitasnya. Pendekatan pengukuran satu kali ini menghasilkan informasi mengenai konsistensi internal instrumen. Konsistensi internal merupakan pernyataan-pernyataan tersebut mengukur aspek yang sama atau merefleksikan homogenitas butir-butir pernyataan.

Makin tinggi koefisien reliabilitas, makin dekat nilai skor amatan dengan skor yang sesungguhnya, sehingga nilai skor amatan dapat digunakan sebagai pengganti komponen skor yang sesungguhnya. Ukuran tinggi atau rendahnya koefisien reliabilitas tidak hanya ditentukan oleh nilai koefisien. Tafsiran tinggi rendahnya nilai koefisien diperoleh melalui perhitungan, ditentukan juga oleh standar pada cabang ilmu yang terlibat di dalam pengukuran itu. Makin tinggi koefisien reliabilitas suatu instrumen, maka kemungkinan kesalahan yang terjadi akan makin kecil kalau orang membuat keputusan berdasarkan skor yang diperoleh dalam instrumen tersebut.

Pada umumnya pengukuran karakteristik afektif memberikan koefisien reliabilitas yang lebih rendah daripada pengukuran ranah kognitif, karena karakteristik kognitif cenderung lebih stabil daripada karakteristik afektif. Menurut Gable (1986) koefisien reliabilitas instrumen ranah kognitif biasanya kira-kira 0,90 atau lebih, sedangkan koefisien reliabilitas instrumen ranah afektif kurang dari 0,70. Koefisien reliabilitas pada taraf 0,70 atau lebih biasanya dapat diterima sebagai reliabilitas yang baik (Litwin, 1995). Sedangkan menurut Naga (1992) koefisien reliabilitas yang memadai sebaiknya terletak di atas 0,75.

Pada setiap penelitian yang menggunakan pengukuran psikologis selalu menerapkan pengujian validitas dan reliabilitas. Namun dalam perjalanannya di bidang psikometri, para ahli belum ada kesepakatan tentang koefisien reliabilitas atau rumus yang mana untuk reliabilitas antar peneliti. Pertama, masih banyak peneliti yang dinilai cukup kompeten masih banyak yang kurang tepat dalam melaporkan reliabilitas hasil pengukuran mereka (Thompson, 1994).

Kedua, masalah yang muncul adalah penggunaan koefisien reliabilitas oleh para peneliti secara monoton tanpa mempertimbangan asumsi yang mendasari koefisien tersebut. Para peneliti tanpa sadar menggunakan koefisien alpha yang juga dengan tanpa sadar bahwa untuk koefisien ini memerlukan asumsi yang sulit dipenuhi. Jika asumsi tidak dipenuhi maka koefisien alpha yang dihasilkan adalah nilai di batas estimasi terendah. Banyak peneliti hanya terpaku pada penggunaan koefisien alpha dalam mengestimasi reliabilitas. Popularitas koefisien alpha Cronbach ini lahir karena faktor: 1) teknik komputasi relatif mudah, karena hanya memerlukan informasi berupa varians skor total, dan 2) distribusi sampling sudah diketahui sehingga penentuan interval kepercayaan pada populasi sangat dimungkinkan (Feld dan kawan-kawan, 1987).

Ketiga, permasalahan yang berhubungan dengan asumsi yang menjadi syarat dalam mengestimasi reliabilitas. Pada ranah empiris selain persyaratan adanya sifat paralel, persyaratan *tau-equivalent* merupakan tantangan yang cukup berat bagi peneliti dalam menyusun instrumen pengukuran. Hal ini didukung oleh Kamata dan kawan-kawan (2003) yang menemukan bahwa asumsi kesetaraan, daya diskriminasi antar komponen tes dan unidimensionalitas pengukuran merupakan hal relatif sulit dicapai. Jika asumsi *essentially tau-equivalent* tidak dapat dipenuhi maka koefisien alpha menghasilkan nilai reliabilitas yang sangat kecil, sehingga koefisien tersebut di bawah estimasi.

Keempat, wacana pengukuran adalah masalah unidimensionalitas pengukuran. Unidimensionalitas adalah aspek penting dalam mengestimasi reliabilitas. Hasil pengukuran psikologis yang bersifat unidimensi sangat sulit dicapai, terutama dalam konteks domain kepribadian yang kebanyakan memuat area varians-varians *traits* yang luas. Socan (2000) menulis bahwa analisis faktor yang dilakukan dari beberapa penelitian banyak kasus multidimensi dibanding dengan unidimensi.

Masalah asumsi bukan menjadi masalah utama dalam menyusun model konsistensi internal, namun masalah ini menjadi bahan kajian banyak peneliti dalam pengkajian reliabilitas. Seperti

penelitian Vehkahlati (2000) yang menyimpulkan bahwa asumsi yang tidak cukup realistis pada teori skor murni klasik adalah asumsi unidimensionalitas skor murni yang secara praktis sulit dibuktikan. Jadi kajian multidimensionalitas pengukuran muncul ke permukaan karena banyak kasus ditemui bahwa juga adanya korelasi antar butir di dalam dimensi tersebut kadang-kadang lebih tinggi dibanding dengan korelasi antar butir dalam tes.

Pada pengembangan instrumen pengukuran dalam bidang pendidikan banyak mengasumsikan penggunaan pengukuran yang bersifat unidimensi yang secara konseptual dirumuskan bahwa hanya ada satu jenis faktor kemampuan, kepribadian, sifat, maupun sikap yang diukur oleh satu instrumen pengukuran. Tetapi, banyak penelitian menunjukkan bahwa asumsi unidimensi tersebut sulit dipenuhi dengan ditemukannya beberapa faktor baru yang ikut diukur dalam satu instrumen. Dengan kata lain, instrumen yang bersifat psikologis yang sering dipakai peneliti cenderung bersifat multidimensi.

Beberapa alasan pentingnya pengukuran reliabilitas yang bersifat multidimensi seperti dikemukakan oleh Widhiarso (2010) dengan uraian sebagai berikut: pertama, karakteristik konstruk psikologis yang umumnya bersifat multidimensi. Kedua, adanya pelibatan aspek-aspek dalam penyusunan instrumen psikologis biasanya diawali dengan penurunan butir-butir dari beberapa aspek teoretis dan kecenderungannya bersifat multidimensi. Ketiga, jumlah butir di dalam instrumen. Jumlah butir yang terlalu banyak dapat menambah potensi penambahan varians *error* dalam butir sehingga memunculkan dimensi baru dari dimensi yang ditetapkan semula. Jumlah butir dan juga bentuk skala mempengaruhi sikap responden terhadap butir yang kemudian mempengaruhi tanggapan mereka terhadap instrumen. Keempat, teknik penulisan butir. Spector dan kawan-kawan (1997) menemukan bahwa teknik penulisan butir yang memiliki arah yang terbalik antara positif (*favorable*) dan negatif (*unfavorable*) dapat membentuk dimensi ukur baru padahal dalam pengambilan data banyak skala psikologi menggunakan teknik penulisan butir yang berbeda arah. Kelima, satuan pengukuran yang berbeda. Pengukuran bidang psikologis cenderung memiliki satuan ukur yang berbeda antara butir satu dengan butir lainnya memiliki kapabilitas yang berbeda sebagai indikator konstruk ukur. Kondisi ini akan menyebabkan hasil pengukuran cenderung bersifat multidimensi.

Dengan demikian dapat disimpulkan bahwa pengukuran psikologis, baik mengukur konstruk kognitif maupun nonkognitif sangat rentan terhadap kemajemukan atribut yang diukur (multidimensi). Selanjutnya dengan memahami kecenderungan pengukuran psikologis lebih pada model pengukuran multidimensi dibandingkan dengan model unidimensi, maka diharapkan proses pengukuran psikometris juga melibatkan teknik analisis yang menggunakan model multidimensi.

McDonald (1981) merumuskan sebuah koefisien reliabilitas yang kemudian diberi nama koefisien reliabilitas skor komposit McDonald yang juga dinamakan koefisien omega (ω). Koefisien reliabilitas ini berbasis pada analisis faktor konfirmatori yang merupakan bagiandari menu pemodelan *Structural Equation Modeling* (SEM). Reliabilitas skor komposit McDonald ini menjelaskan besarnya proporsi indikator dalam menjelaskan konstruk ukur. Adapun formula untuk mendapatkan koefisien reliabilitas konstruk adalah sebagai berikut:

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

Keterangan:

$$\lambda_i = \text{factor loading terstandarisasi indikator ke-}i$$

Menurut Latan (2012) SEM merupakan suatu teknik analisis *multivariate* generasi kedua yang menggabungkan antara analisis faktor dan analisis jalur sehingga memungkinkan peneliti untuk menguji dan mengestimasi secara simultan hubungan antara *multipleexogenous* dan *endogenous* variabel dengan banyak indikator. Hasil penelitian Joreskog pada tahun 1970an membawa teori statistika pada analisis struktural linear yang lebih dikenal dengan sebutan

model persamaan struktural atau SEM. Sumber penting yang digunakan dalam menganalisis adalah struktur kovarian sehingga terkadang pendekatan ini dinamakan dengan *covariant structure model* (CSM). Model yang disusun memuat variabel tak terukur yang dinamakan dengan konstruk laten yang dibangun oleh serangkaian variabel terukur yang dinamakan dengan konstruk terukur. *Error* pengukuran yang merefleksikan reliabilitas skor pengukuran dilihat sebagai konstruk unik dan menjadi bagian yang penting dalam analisis SEM, *error* pengukuran yang dilibatkan dalam analisis SEM inilah yang kemudian menjadi kelebihan SEM dibanding dengan teknik analisis lainnya (Capraro *et al.*, 2001). SEM dapat mengestimasi varians *error* skor hasil pengukuran secara aktual mengestimasi reliabilitas. Menurut Gefen *et al.*, (2001), SEM sebagai teknik statistik multivariat yang mengkombinasikan antara regresi berganda yang mengidentifikasi hubungan antara konstruk dan analisis faktor yang mengidentifikasi konsep tak terukur melalui beberapa indikator manifest yang keduanya dipakai secara simultan.

SEM memiliki beberapa kelebihan dibanding dengan teknik analisis lainnya. Dalam menguji hubungan antara variabel, SEM secara otomatis mereduksi efek *error* pengukuran. Capraro *et al.*, (2001) mengatakan bahwa pengaruh variabel independen terhadap variabel dependen dipengaruhi oleh efek atenuasi. Nilai efek ini tidak dapat melebihi batas koefisien reliabilitas skor tes yang digunakan. Pendekatan pertama adalah koreksi korelasi atenuasi yang disebabkan oleh *error* pengukuran dan pendekatan kedua adalah model persamaan struktural dalam konteks analisis faktor konfirmatori. Lee dan Song (2001) mengatakan bahwa SEM adalah salah satu pendekatan untuk menegaskan model pengukuran. Pada model pengukuran SEM menghubungkan antara konstruk laten dengan dengan konstruk empirik. Konstruk empirik dinyatakan oleh kombinasi konstruk laten. Disamping dapat mampu menangani *generalizability theory* dan *item response theory*, SEM mampu membandingkan model pengukuran dan memfasilitasi investigasi ketepatan model.

SEM memiliki sub model berupa analisis faktor. Analisis faktor berguna untuk mendeteksi dimensi instrumen pengukuran. Teknik ini diperkenalkan oleh Spearman yang bertujuan untuk mengeksplorasi faktor-faktor yang menjadi bagian dari inteligensi. SEM juga dapat mengidentifikasi reliabilitas konstruk yang terlihat melalui nilai butir *loading* yang dihasilkan. Berdasarkan perspektif SEM reliabilitas konstruk dapat dihitung melalui persamaan sebagai berikut:

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

Keterangan:

CR = reliabilitas konstruk

λ_i = *factor loading* terstandarisasi indikator ke- i

δ = *error* standar pengukuran

Reliabilitas konstruk ini memberikan hasil yang sama dengan reliabilitas skor komposit McDonald karena $\delta = 1 - \lambda^2$.

Koefisien reliabilitas multidimensi berikut merupakan koefisien reliabilitas konstruk yang dikembangkan oleh Hancock dan Mueller (2000), yang menunjukkan seberapa jauh indikator instrumen mampu merefleksikan konstruk yang hendak diukur. Koefisien ini merupakan modifikasi dari koefisien reliabilitas konstruk McDonald yang tidak mampu mengakomodasi bobot yang berbeda antardimensi. Hasil modifikasi dinamakan koefisien reliabilitas konstruk berbobot sebagai berikut:

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

Keterangan:

l_i = koefisien dimensi ke- i terstandar

Koefisien reliabilitas ini dapat diartikan sebagai korelasi kuadrat antara dimensi dengan skor komposit linier optimal, sehingga beberapa ahli menamakannya dengan reliabilitas maksimal (*maximal reliability*).

Pada penelitian Widhiarso dan Mardapi (2010), model multidimensi untuk koefisien reliabilitas memiliki ketepatan pengukuran yang tinggi bila dibandingkan dengan reliabilitas unidimensi. Oleh karena itu dalam penelitian ini, peneliti hanya difokuskan pada koefisien konsistensi internal seperti α untuk reliabilitas unidimensi dan ω , CR dan Ω_w .

Berdasarkan uraian di atas maka untuk organisasi pendidikan seperti Fakultas Teknik, Universitas Negeri Jakarta (UNJ) dimunculkan berbagai pertanyaan seperti: Bagaimanakah reliabilitas internal konsistensi multidimensi dari instrumen pengukur kepuasan mahasiswa sebagai pelanggan internal? Bagaimanakah komparasi antara reliabilitas multidimensi dan unidimensi? Manakah yang lebih akurat sebagai pengukur reliabilitas?

2. METODOLOGI PENELITIAN

Metode yang dipergunakan dalam penelitian ini adalah metode survei, di dalam pengembangan instrumen dengan menggunakan pendekatan respons. Penelitian dilaksanakan di Program Studi Penelitian dan Evaluasi Pendidikan (PEP), Program Pasca Sarjana (PPs), Universitas Negeri Jakarta, pada semester ganjil tahun 2012/2013. Populasi target adalah seluruh mahasiswa UNJ, sedangkan populasi terjangkau adalah seluruh mahasiswa PPs UNJ, namun yang dijadikan sampel dalam penelitian ini adalah mahasiswa Prodi PEP PPs UNJ yang mengambil matakuliah Statistika I dan Statistika II. Sampel diambil dengan cara pengambilan sampel acak sederhana (*simple random sampling*) dan dari 200 instrumen yang disebar diperoleh atau yang kembali sebanyak 133.

Agar mudah dipahami, maka konsep pengukur Statistika perlu dijabarkan terlebih dahulu ke dalam dimensi yakni (1) evaluasi (E) sebanyak 5 butir, (2) potensi (P) sebanyak 3 butir, dan (3) aktivitas (A) sebanyak 5 butir. Butir-butir tersebut diadaptasi dari buku Isaac dan Michael (1985). Variabel dalam penelitian ini sebagai kawasan yang dijadikan sasaran untuk diukur adalah sikap terhadap Statistika yaitu kecenderungan seseorang terhadap Statistika dengan segala potensi, evaluasi, dan aktifitasnya. Jenis respons dalam penelitian ini adalah kinerja tipikal (*typical performance*), oleh sebab itu respons yang diharapkan dapat diperoleh melalui instrumen tentang kebiasaan (*habit*) responden atau apa yang bisa orang kerjakan atau rasakan (*what a person usually does do or feel*) bila menghadapi situasi tertentu dalam kegiatan belajarnya. Juga biasanya disebut ekspresi sentimen (*expression of sentiment*) yakni jenis respons yang tidak dapat dinyatakan benar atau salah, atau sering kali dikatakan semua respons benar menurut alasannya masing-masing. Sesuai dengan karakteristik jenis respons, maka format alat ukur yang dipilih untuk menyajikan butir-butir instrumen adalah format pilihan terbatas. Untuk tiap-tiap butir memiliki 7 pilihan jawaban dengan rentang nilai 1 sampai dengan 7. Waktu untuk mengerjakan sekitar 5-10 menit. Kualitas atau status instrumen ini memiliki dua kutub kecenderungan, yaitu dari kutub positif sampai negatif sikap terhadap Statistika.

3. HASIL DAN PEMBAHASAN

Instrumen sikap terhadap statistika semula ini terdiri atas 15 butir pernyataan. Ketiga belas butir instrumen ini merupakan hasil penelitian dari peneliti sendiri yang semula 15 butir dan

gugur 2 butir. Instrumen yang terdiri dari 13 butir ini dapat dirinci sebagai berikut: 5 butir untuk dimensi evaluasi, 3 butir potensi, dan 5 butir aktivitas. Untuk reliabilitas konsistensi internal alpha Cronbach diperoleh langsung menggunakan program SPSS sebesar 0,710.

Untuk reliabilitas skor komposit McDonald, dengan menggunakan pemodelan analisis struktur

atau *structural equation modelling* (SEM) diperoleh: $\sum_{i=1}^i \lambda_i = 6,040$ dan $\sum_{i=1}^i 1 - \lambda_i^2 = 9,632$; jadi

$$\omega = \frac{(6,040)^2}{(6,040)^2 + (9,632)} = 0,791.$$

Reliabilitas konstruk diperoleh hasil yang sama sebagai berikut: $\sum_{i=1}^i \lambda_i = 6,040$ dan

$$\sum_{i=1}^i \delta = 9,620; \text{ jadi } CR = \frac{(6,040)^2}{(6,040)^2 + (9,620)} = 0,791.$$

Berikut untuk reliabilitas berbobot, dengan menggunakan pemodelan analisis struktur atau

SEM diperoleh: $\sum_{i=1}^p \frac{l_i^2}{(1-l_i^2)} = 5,148$, sehingga dapat dihitung sabagai berikut:

$$\Omega_w = \frac{5,148}{1 + 5,148} = 0,837$$

Dari hasil perhitungan di atas dapat dirangkum sebagai berikut:

Tabel 1. Rangkuman Hasil Perhitungan Koefisien Reliabilitas

Instrumen	Koefisien Reliabilitas Unidimensi	Koefisien Reliabilitas Multidimensi		
		CR	ω	Ω_w
Sikap	$\alpha = 0,710$	$CR = 0,791$	$\omega = 0,791$	$\Omega_w = 0,837$

Perhitungan untuk instrumen di atas diperoleh koefisien alpha Crobach jauh lebih kecil bila dibandingkan dengan reliabilitas konstruk, reliabilitas skor komposit McDonald, dan reliabilitas maksimal. Dengan selisih 0,081 dan 0,127. Namun selisih tersebut apakah mencerminkan keakuratan? Hal ini belum ada kesepakatan antar ahli psikometri. Namun kita di kalangan peneliti di Indonesia sebaiknya setelah mengetahui alat mana yang paling tepat sebaiknya kita mulai memakai/menggunakan alat tersebut secara benar dan memadai.

Memang sebagian besar peneliti di kalangan dosen maupun mahasiswa S2 maupun S3 belum mengetahui formula untuk menghitung koefisien reliabilitas berbasis SEM tersebut. Jadi kali ini saatnya untuk mengenalkan dan juga menggunakan formula tersebut. Dengan alasan sudah tahu rumusnya dan kebanyakan konstruk psikologis, kepribadian, pendidikan, dan sosial adalah multidimensi, sehingga seluruh peneliti baik mahasiswa maupun dosen berkembang dan makin berkembang untuk menggali lebih dalam lagi tentang koefisien reliabilitas yang lainnya.

Interpretasi koefisien reliabilitas merupakan evaluasi kecermatan skor tes, bukan sekedar keajegannya saja. Juga dalam menginterpretasikan tingginya koefisien reliabilitas, paling tidak ada dua hal yang perlu dipahami, yakni: (1) reliabilitas yang diestimasi dengan menggunakan suatu kelompok subjek dalam situasi tertentu akan menghasilkan koefisien yang tidak sama dengan estimasi tes tersebut pada kelompok subjek lain, dan (2) koefisien reliabilitas hanyalah mengindikasikan besarnya inkonsistensi skor hasil pengukuran, bukan menyatakan langsung sebab-sebab inkonsistensi itu.

Pengukuran bidang pendidikan merupakan sesuatu yang cukup rumit. Berbagai tulisan di dalam jurnal pengukuran pendidikan berkisar pada cara pengukuran yang diharapkan memberikan hasil yang valid, reliabel, dan akurat. Usaha para pakar tidaklah mudah karena

para pakar tersebut makin lama membawa pengukuran pendidikan itu jauh ke dalam kawasan matematika. Tanpa menguasai dengan baik matematika yang tinggi dan rumit, kita tidak dapat memahami berbagai jurnal pengukuran pendidikan.

Sejauh ini, kita sangat tertinggal di bidang pengukuran pendidikan. Sangat sedikit pakar ilmu pendidikan yang mampu memahami isi jurnal pengukuran pendidikan yang bertaburkan matematika tingkat tinggi. Oleh karena itu perlu diusahakan peningkatan para pakar ilmu pendidikan di bidang pengukuran pendidikan.

Usaha itu dapat dimulai dengan mengubah persepsi kita selama ini yang sejak lama, para pendidik dikalangan kita memiliki anggapan bahwa ilmu pendidikan tidak memerlukan matematika. Matematika adalah garapan MIPA dan Teknik dan bukan garapan ilmu pendidikan. Kini, berhadapan dengan pengukuran pendidikan para pendidik dikalangan kita perlu mengubah persepsi mereka terhadap matematika. Para pendidik perlu menyadari bahwa ada bagian ilmu pendidikan yang hampir tidak menggunakan matematika, tetapi ada juga bagian ilmu pendidikan yang sangat memerlukan matematika, seperti contoh di atas statistika multivariat yang memerlukan kemampuan matematika tinggi.

4. KESIMPULAN DAN SARAN

Dari hasil penelitian di atas dapat ditarik kesimpulan bahwa koefisien reliabilitas multidimensi lebih tepat (akurat) untuk mengestimasi reliabilitas dibandingkan dengan reliabilitas unidimensi. Saran yang dapat diajukan adalah sebagai berikut: pertama, estimasi instrumen ini perlu diuji lebih lanjut dengan menggunakan rumus lainnya yang peneliti kira masih banyak dan belum diterapkan di institusi UNJ. Kedua, oleh karena penelitian ini menggunakan skala tujuh maka bila perlu dilanjutkan menggunakan berbagai skala lain, misalnya skala Likert, skala dikotomi, skala Thurstone, dan sebagainya. Ketiga, instrumen ini perlu diuji dengan menggunakan sampel yang lebih besar dengan populasi dan *setting* yang lebih luas serta melibatkan beberapa propinsi sekaligus, juga dengan jenjang sekolah dan jenis universitas atau perguruan tinggi yang berbeda.

DAFTAR PUSTAKA

- [1]. Capraro, M. M., R. M. Capraro, dan R. K. Herson. "Measurement Error of Score on the Mathematics Anxiety Rating Scale Across Studies." *Educational and Psychological Measurement*, 61, 2001, 373-386.
- [2]. Feld, I. S., D. J. Woodruff, dan F. A. Salih. "Statistical Inference for Coefficient Alpha." *Applied Psychological Measurement*, 1987, II, 93 - 103.
- [3]. Gable, R. K. *Instrument Development in the Affective Domain*. Amsterdam: Kluwer Nijhoff Publishing, 1986.
- [4]. Geffen, D., D. W. Straub, dan M. D. Boudreau. "Structural Equation Modeling and Regression: Guidelines for Research Practice." *Communications of AIS*, Volume 4, 2001, Article 7.
- [5]. Hancock, G. R., dan R. O. Mueller. "Rethinking Construct Reliability within Latent Variable Systems." Di dalam *Structural Equation Modeling: Present and Future*, R. Cudek, S. H. C. duToit, dan D. F. Sorbom (Eds.), Chicago: Scientific Software International, 2000.
- [6]. Isaac, S. & Michael, W. B. *Handbook in Research and Evaluation: For Education and the Behavioral Sciences*. California: Edits Publishers, 1985.
- [7]. Kamata, A., A. Turhan, dan E. Darandari. "Estimating Reliability for Multidimensional Composite Scales Scores," *Makalah* disampaikan pada annual meeting of American Educational Research Association di Chicago, April 2003.
- [8]. Latan, Hengky. *Structural Equation Modelling Konsep dan Aplikasi Menggunakan Program Lisrel 8.80*. Bandung: Alfabeta, 2012.
- [9]. Lee, S. Y., dan X. Y. Song. "Hypothesis Testing and Model Comparison in Two-level Structural Equation Model." *Multivariate Behavioral Research*, Volume 36, Issue 4 January 2001, 639-655.
- [10]. Litwin, M. S. *How to Measure Survey Reliability and Validity*. London: Sage Publications, 1995.
- [11]. McDonald, R. P. "The Dimensionality of Test and Items." *British Journal of Mathematical and Statistical Psychology*, 1981, 34, 100 - 117.
- [12]. Naga, Dali S. *Teori Sekor*. Jakarta: Gunadarma Press, 1992.
- [13]. Socan, G. "Assessment of Reliability when Test Items are not Essentially t-Equivalent." Di dalam *Development in Survey Methodology*, Anuska Feligoj and Andrej Mrvar (Eds.). Ljubljana: FDV, 2000.

- [14]. Spector, P., P. Brannick, dan P. Chen. "When Two Factors Don't Reflect Two Constructs: How Item Characteristics Can Produce Artifictual Factors." *Journal of Management*, 1997, 23 (5), 659 – 668.
- [15]. Suryabrata, S. *Pengembangan Alat Ukur Psikologis*. Yogyakarta: Andi Offset, 2000.
- [16]. Thompson, B. "Guidelines for Author." *Educational and Psychological Measurement*, 1994, 54, 837 – 847.
- [17]. Vehkalahti, K. "Reliability of Measurement Scales Tarkkonnen's General Method Supersedes Cronbach's Alpha." *Academic Dissertation*. University of Helsinki, Finland, 2000.
- [18]. Widhiarso, Wahyu dan Djemari Mardapi. "Komparasi Ketepatan Estimasi Koefisien Reliabilitas Teori skor Murni Klasik." *Jurnal Penelitiandan Evaluasi Pendidikan*, 2010, 14 (1), 1 – 19.
- [19]. Widhiarso, Wahyu. "Koefisien Reliabilitas pada Pengukuran Kepribadian yang Bersifat Multidimensi." *Psikobuana*, 2009, Vol. 1 (1), 39 – 48.
- [20]. Wiersma, W. *Research Methods in Education: An Introduction*. Boston: Allyn and Bacon, Inc., 1986.