



Koefisien Reliabilitas Sebagai Karakteristik Sampel: Pengujian dengan instrument Potrait Values Questionnaire (PVQ) dari Schwartz

ABSTRAK Koefisien reliabilitas seringkali disalahpahami oleh peneliti sebagai indicator konsistensi alat ukur. Padahal, sebagaimana dinyatakan Vacha-Haase (1998) dan Thompson (1994), reliabilitas merupakan karakteristik sampel. Koefisien reliabilitas suatu alat ukur akan dipengaruhi oleh karakteristik sampel. Penelitian ini dilakukan untuk membuktikan pernyataan tersebut. Peneliti menggunakan Potrait Values Questionnaire (PVQ) Schwartz untuk melakukan pengambilan data pada 36 area pengambilan data (jumlah responden sebanyak 1.083). Koefisien reliabilitas PVQ pada masing-masing area dihitung kemudian diregresikan dengan proporsi laki-laki dan perempuan serta tingkat perkembangan. Hasil penelitian ini menunjukkan bahwa proporsi perempuan memiliki sumbangan efektif terhadap koefisien reliabilitas sebesar 34,8% ($\beta = -0,348$). Artinya, semakin besar proporsi responden perempuan maka koefisien reliabilitas PVQ semakin rendah. Hasil penelitian ini membuktikan bahwa karakteristik sampel yang digunakan mempengaruhi reliabilitas PVQ.

Kata kunci: Koefisien reliabilitas, Skala, Sampel, PVQ

Helly P. Soetjipto

Moh. Abdul Hakim

Dosen Fakultas Psikologi Universitas Gadjah Mada

Jl. Humaniora No. 1, Bulaksumur, Yogyakarta

Email: hakimpsi@yahoo.com

PSIKOISLAMIKA, Jurnal Psikologi Islam (JPI) Copyright © 2011 Lembaga Penelitian Pengembangan Psikologi dan Keislaman (LP3K). Vol 8 No. 2, Januari 2011 247-260



download the free trial online at nitropdf.com/professional

Muqadimah

Koefisien reabilitas digunakan oleh peneliti psikologi untuk menunjukkan seberapa konsisten parameter yang ia gunakan untuk mengukur konstruk psikologis tertentu pada sampel. Di dalam jurnal, skripsi, tesis, dan bahkan disertasi psikologi, banyak peneliti yang menyebutkan “reabilitas tes” atau menulis bahwa “skala yang digunakan reliabel”. Istilah reliabilitas alat tes tampaknya telah diterima secara luas di kalangan penelitian psikologi di Indonesia. Misalnya, dalam aturan penulisan, penyunting Jurnal *Phronesis* (2003) menyebutkan penulis diharuskan mencantumkan informasi “reliabilitas alat ukur”. Kemudian, dari hasil penelaahan artikel-artikel Jurnal *Anima* sepanjang tahun 2004, peneliti menemukan 6 artikel (Basoeki, dkk., 2004; Budiharjo, 2004; Djalali, 2004; Moesono, 2004; Simanjuntak, 2004; dan Amriang, dkk., 2004) menyebutkan koefisien reliabilitas Cronbach’s alpha sebagai koefisien reliabilitas instrumen.

Pernyataan-pernyataan di atas, menurut Tammi Vacha-Haase (1998), merupakan kesalah pahaman konsep skor reabilitas. Dalam aturan penulisan Jurnal *Educational and Psychological Measurement*, Thompson (1994) menegaskan bahwa reabilitas tidak merujuk pada alat tes atau skala melainkan pada skor-skor yang dihasilkan oleh alat ukur pada kelompok sampel yang diukur. Sementara Rowley (1976) mengemukakan bahwa instrumen tidak dapat dikatakan reliabel atau tidak reliabel. Ia menjelaskan sebuah instrumen dapat menghasilkan skor-skor reliabel dan skor-skor tidak reliabel pada beberapa kelompok sampel yang berbeda. Oleh karena itu, alat ukur yang sama bila digunakan pada kelompok sampel yang berbeda-beda, maka koefisien reabilitas yang dihasilkan pun akan berbeda-beda. Selanjutnya, peneliti sebaiknya menyebutkan koefisien reliabilitas sebagai reliabilitas skor atau reliabilitas pengukuran (Thompson, 1994 dan Vacha-Haase, 1998). Senada dengan pendapat ketiga ahli tersebut, Miller (2009) dalam menjelaskan validitas dan reliabilitas menulis:

“Reliability and validity are not global properties of an assessment. Instead, they are properties of specific uses and interpretations that are made from a set of test scores. A test could be valid for a particular use or interpretation and not for another. For example, a test might measure the curriculum covered in a school without providing valid estimates of student performance because of the

length of the tests or the nonequivalence of forms. The same is true for reliability. For example, a test might provide reliable scoring without being stable over time. In addition, reliability and validity are a matter of degree. Tests are not considered valid or invalid. Instead, they are valid to some degree. Similarly, a test is not considered reliable or unreliable, but is reliable to some degree."

Kekeliruan pemahaman skor reabilitas tidak dapat diabaikan begitu saja. Sebab, apabila peneliti tidak mencantumkan koefisien reliabilitas dari data penelitian yang sesungguhnya, maka tidak diketahui seberapa konsisten hasil pengukurannya itu. Sehingga, kesimpulan dari penelitian tidak sepenuhnya dapat dipercaya. Berdasarkan pengamatan peneliti terhadap hasil penelitian skripsi, tesis, dan laporan-laporan penelitian lain, peneliti mengidentifikasi tiga tipe kesalahan yang sering dilakukan oleh peneliti terkait dengan koefisien reabilitas.

Tipe kesalahan pertama, peneliti tidak menyebutkan koefisien reabilitas dalam laporan penelitiannya. Hal ini disebabkan dari hasil pengumpulan data, peneliti menemukan skor reabilitas yang tidak memuaskan dan cenderung fluktuatif. Oleh karena itu, peneliti berkesimpulan bahwa reabilitas skor tidak perlu dijadikan sebagai landasan analisis data. Fakta menunjukkan, tidak banyak peneliti yang mencantumkan koefisien reliabilitas dalam laporan penelitian mereka (Thompson, 2002), bahkan pada laporan penelitian yang sudah termuat di jurnal-jurnal ilmiah internasional Misalnya, Shields, dkk. (2007) mengungkapkan bahwa dari 470 laporan penelitian yang menggunakan *The Michigan Alcoholism Screening Test*, hanya 13,2% (62) diantaranya saja yang mencantumkan koefisien reliabilitas data penelitian. Selanjutnya, untuk instrumen *Teacher Efficacy Scale*, Henson, dkk (2001) menemukan hanya 40,38% laporan yang mencantumkan koefisien reliabilitas, sementara Beretvas, dkk. (2008) melaporkan sebanyak 76,9% penelitian yang menggunakan *Rotter's Locus of Control Scale* tidak mencantumkan koefisien reliabilitas data penelitiannya.

Tipe kesalahan kedua, peneliti menyebutkan koefisien reabilitas skala yang sama dari hasil penelitian lain. Kesalahan tipe kedua disebabkan secara langsung oleh kesalahpahaman peneliti yang menganggap bahwa koefisien reabilitas mengacu pada skala. Setelah peneliti melihat skala tertentu pada suatu penelitian menunjukkan

koefisien reliabilitas yang memuaskan, peneliti kemudian berkesimpulan bahwa skala yang sama cukup meyakinkan untuk digunakan dalam penelitiannya sendiri tanpa melalui pengujian reliabilitas dengan datanya sendiri.

Tipe kesalahan ketiga, peneliti hanya menyebutkan koefisien reabilitas dari data *try out* sebagai landasan analisis data penelitian yang sesungguhnya. Sebagai contoh, Yuliani M. dan Rostiana (2003) dalam laporan penelitiannya menulis:

".....dari hasil uji coba alat tes terhadap 35 responden didapat nilai α komitmen organisasi = 0,7778; α iklim organisasi = 0,8810; dan α motivasi berprestasi = 0,7794."

Walaupun dalam hal ini peneliti lebih berhati-hati terhadap tipe kesalahan kedua, namun peneliti tidak menyadari bahwa karakteristik kelompok sampel pada saat *try out* tidak sama dengan kelompok sampel pada penelitian sesungguhnya. Artinya, data *try out* dan data penelitian dihasilkan oleh dua penelitian yang berbeda. Dikecualikan dari kesalahan tipe ketiga ini bagi peneliti yang menggunakan baik skala maupun kelompok sampel yang sama persis antara *try out* dan penelitian sesungguhnya.

Kerangka Kerja Teoritik

Sekilas Tentang *Classical Test Theory*

Seperti yang telah umum diketahui, teori reliabilitas skor berakar pada *true score theory of measurement*. Dijelaskan oleh William M. K. Trochim (2006), skor-skor pada tiap pengukuran terdiri atas dua komponen, yaitu level responden sebenarnya pada pengukuran itu (*true ability*) dan kesalahan random (*random error*). Secara sederhana asumsi ini dirumuskan dengan

$$X = T + e_x$$

dimana X adalah pengukuran X, T adalah skor responden sebenarnya (*true score*), dan e_x merupakan kesalahan random pada pengukuran X. Skor T akan ditentukan oleh validitas pengukuran, sementara nilai e akan mempengaruhi reliabilitas pengukuran. Menurut *classical test theory* (CTT), *error* merupakan jumlah keseluruhan komponen-komponen random dan *true score* adalah jumlah keseluruhan efek yang konsisten.



nitroPDF[®]

professional

Pada kenyataannya, peneliti tidak mungkin mengetahui skor T dan e sebenarnya. Penting untuk mengetahui bahwa nilai e pada tiap-tiap pengukuran bersifat random. Bila dari pengukuran yang berulang-ulang dari satu orang sampel peneliti mendapatkan skor X yang bervariasi maka diasumsikan bahwa variasi ini disebabkan oleh variasi error ($\text{var}(e)$) sementara nilai T selalu tetap. Dengan demikian, peneliti dapat menghitung estimasi reliabilitas skor. Semakin nilai e mendekati nol maka reliabilitas skor semakin mendekati nilai 1. Dikarenakan e adalah random, maka semakin banyak pengukuran, nilai e akan semakin mendekati nol mengikuti asumsi kurve normal. Sehingga, semakin banyak item pada skala maka koefisien reliabilitas akan semakin baik (Collins, 2006; Embretson dan Reise, 2000). Pada keadaan ideal, nilai e sama dengan nol sehingga nilai X sama dengan nilai T . Hal ini akan ditunjukkan oleh koefisien reliabilitas yang bernilai 1.

Disamping kesalahan random, tidak menutup kemungkinan pengukuran dipengaruhi oleh kesalahan yang sistematis (*systematic error*). Kesalahan sistematis timbul sebab ada faktor tertentu yang mempengaruhi seluruh sampel. Misalnya, ketika peneliti melakukan pengukuran terhadap mahasiswa di kelas. Kemudian, ada keriuhan di luar kelas yang mempengaruhi konsentrasi seluruh mahasiswa dalam memberikan skor. Perbedaan mendasar antara *random error* dan *systematic error* adalah bahwa random error tidak mempengaruhi rata-rata skor, hanya berpengaruh terhadap variabilitas disekitar nilai rata-rata. Sementara *systematic error* berpengaruh terhadap nilai rata-rata skor sampel. Hal ini disebut sebagai *bias* (Trochim, 2006). Pengaruh *systematic error* dalam pengukuran dapat direduksi dengan prosedur penyusunan skala dan administrasi pengukuran yang dilaksanakan dengan baik.

Berdasarkan penjelasan di atas, maka para ahli mendefinisikan koefisien reliabilitas sebagai perbandingan antara true score variance ($\text{var}(T)$) dengan varians skor yang dihasilkan oleh pengukuran (*observed score variance*) (Miller, 2007; Trochim, 2006). Koefisien reliabilitas juga dapat diartikan sebagai rasio antara $\text{var}(T)$ dengan jumlah antara $\text{var}(T)$ dan *error variance* ($\text{var}(e)$). Oleh karena itu dapat disimpulkan bahwa koefisien reliabilitas adalah proporsi variasi pada skor-skor tes yang dapat diatribusikan pada pengukuran yang konsisten (Miller, 2007).

Created with



Reliabilitas dan Karakteristik Sampel

Sebagaimana telah dikemukakan pada awal tulisan, koefisien reliabilitas skor yang dihasilkan oleh instrumen psikologi dapat berubah-ubah sesuai dengan karakteristik sampel yang digunakan. Sebab, reliabilitas tidak hanya dipengaruhi oleh instrumen yang digunakan melainkan juga oleh sampel (Dawis, 1987). Hal inilah yang melandasi Eason (1991) menyebut koefisien reliabilitas sebagai karakteristik data. Senada dengan pendapat ini, Thompson (1994) menjelaskan, bila alat ukur yang sama diadministrasikan pada kelompok sampel yang lebih heterogen atau lebih homogen, akan menunjukkan skor reliabilitas yang berbeda-beda.

Shield dan Caruso (2003) mengidentifikasi enam jenis karakteristik yang mempengaruhi koefisien reliabilitas, yaitu rata-rata usia responden, jenis kelamin, tipe sampel, variabilitas skor, latar belakang etnis, dan koefisien reliabilitas otentik. Pengaruh karakteristik sampel terhadap skor reliabilitas telah dibuktikan oleh Tammi Vacha-Haase (1998). Dengan memperkenalkan metode *reliability generalization*, Vacha-Haase meneliti pengaruh karakteristik kelompok sampel pada koefisien reliabilitas skor *The Bem Sex Role Inventory* (BSRI) yang digunakan dalam 87 jurnal ilmiah. Hasilnya, koefisien reliabilitas skor BSRI dipengaruhi oleh beberapa karakteristik sampel, yaitu tipe reliabilitas, jender, dan jumlah sampel. Temuan serupa dikemukakan oleh banyak peneliti lain, seperti Henson, dkk (2001) untuk instrumen *Teacher Efficacy Scale* (TES), Beretvas, dkk. (2008) untuk instrumen *Rotter's Locus of Control Scales* dan *Nowicki-Strickland's Locus of Control Scales*, Shields, dkk (2007) untuk *The Michigan Alcoholism Screening Test*, dan Bachner dan O'Rourke (2007) untuk instrumen *The Zarit Burden Interview*.

Potrait Values Questionnaire (PVQ) Schwartz

PVQ merupakan instrument pengukuran nilai-nilai yang disusun oleh Schwartz berdasarkan teori konten dan struktur nilai yang ia kembangkan sendiri. Nilai adalah dorongan, tujuan yang bersifat trans situasional, yang berbeda-beda pada tingkat kepentingan, yang berfungsi sebagai pedoman-pedoman prinsip dalam hidup seseorang (Barnea dan Schwartz, 1998). Teori konten dan struktur nilai Schwartz terdiri atas 10 nilai motivasional yang disebut sebagai; *Power, Achievement, Hedonism, Stimulation, Self-Direction, Universalism, Benevolence, Tradition, Conformity, dan Security*. Delapan

nilai motivasional di antaranya (*Power, Achievement, Hedonism, Stimulation, Self-Direction, Benevolence Conformity, dan Security*) merepresentasikan tiga kebutuhan universal eksistensi manusia, yaitu kebutuhan sebagai organisme biologis, kebutuhan koordinasi dalam interaksi social, dan kebutuhan berkelompok untuk bertahan hidup dan mencapai kesejahteraan. Kemudian, berdasarkan studi empirisnya, Schwartz menambahkan dua nilai lainnya (*Universalism dan Tradition*) (Schwartz, 1992). Kesepuluh nilai ini telah diverifikasi di 47 negara. Penjelasan mengenai 10 nilai tersebut dapat dilihat pada Tabel 1.

Tabel 1:

Pengertian 10 Nilai Motivasional Schwartz (Barnea dan Schwartz, 1998)

POWER	Status sosial dan kehormatan, kontrol dan dominasi atas orang lain dan sumber daya (kekuatan sosial, otoritas)
ACHIEVEMENT	Sukses pribadi dengan menunjukkan kompetensi berdasarkan standar sosial (kesuksesan, kesejahteraan, ambisi)
HEDONISM	Kesenangan atau kenikmatan ragawi (kesenangan, menikmati hidup)
STIMULATION	Kegembiraan, kesenangan, dan tantangan dalam hidup (keberanian, hidup yang bervariasi, kehidupan yang penuh kegembiraan)
SELF-DIRECTION	Cara berpikir dan bertindak dalam memilih, menciptakan, dan menyelidiki secara independent (kreatifitas, kebebasan, keandirian, kemampuan, kebijaksanaan, dan dunia keindahan)
UNIVERSALISM	Pemahaman, apresiasi, toleransi, dan perlindungan kesejahteraan untuk semua orang dan alam (kejujuran, pikiran yang luas, pelestarian lingkungan, makna hidup)
BENEVOLENCE	Perlindungan dan peningkatan kesejahteraan orang-orang yang sering menjalin hubungan personal (bantuan, memaafkan, dan keadilan social)
TRADITION	Respek, komitmen, dan menerima adat dan ide-ide yang ditawarkan oleh agama dan budaya (kesalehan, penghormatan kepada tradisi, kerendahan hati, kehidupan spiritual)

CONFORMITY	Pengendalian perilaku, kecenderungan, dan keinginan yang dapat membahayakan dan mengganggu orang lain dan merusak norma-norma atau kepentingan social (disiplin diri, kesopanan, penghormatan kepada kedua orang tua dan orang-orang tua, kepatuhan)
SECURITY	Keamanan, harmoni, dan stabilitas masyarakat, hubungan antar warga, dan diri (keamanan keluarga, keamanan nasional, dan tata tertib sosial)

PVQ memuat 40 item yang terdiri atas 4 item untuk tipe nilai Conformity, 4 item Tradition, 4 item Benevolence, 6 item Universalism, 4 item Self-Direction, 3 item Stimulation, 3 item Hedonism, 4 item Achievement, 3 item Power, dan 5 item Security. Dalam pemaparannya mengenai teori Content and Structure of Values, Schwartz menjelaskan tentang aspek spiritualitas yang tidak ia rumuskan sebagai satu tipe nilai tersendiri. Pada umumnya, teoritikus nilai sering kali menjadikan spiritualitas sebagai tipe nilai independen (misalnya teori nilai Rokeach, 1968). Spiritualitas atau religiusitas selama ini memang terbukti menjadi factor psikologis yang penting bagi individu. Sebagai contoh, Orsolya Lelkes (2006) menemukan bahwa individu-individu yang religius cenderung lebih resisten terhadap depresi dan stress saat terjadi krisis ekonomi di Eropa Timur dibanding individu yang tidak religius. Mengenai aspek spiritualitas ini Schwartz mengatakan bahwa nilai spiritualism sudah tercakup di dalam tipe nilai universalism. Dengan teknik *multidimensional scaling*, Schwartz membuktikan bahwa nilai spiritualitas berada dalam satu dimensi dengan nilai-nilai kejujuran, pikiran yang luas, pelestarian lingkungan, dan pemaknaan hidup (nilai universalism) (Schwartz, 1992). Jadi, bagi Schwartz, spiritualitas atau religiusitas tidak mesti berkaitan dengan apakah seseorang memeluk agama tertentu, agnostic, ataupun atheis. Agama dalam hal ini ia kategorikan sebagai nilai tradisi sehingga spiritualitas lebih bersifat universal.

PVQ merupakan instrument pengukuran nilai yang relatif baru. Saat peneliti melakukan penelusuran tentang PVQ di database-database jurnal ilmiah internasional (EBSCO, PsyArticles, SpringerLink, ScienceDirect, dan Jstore) dengan kata kunci "Portrait Values Questionnaire" dan "Schwartz", peneliti hanya mendapatkan 9 artikel penelitian yang menggunakan PVQ. Karena terlalu



nitroPDF

professional

sedikitnya sampel yang dibutuhkan untuk dapat menggunakan metode reliability generalization (Vacha-Haase, 1998), maka dalam studi ini, peneliti hanya akan melakukan analisis korelasional dan regresi untuk mengetahui beberapa hubungan antara karakteristik sampel terhadap reliabilitas pengukuran dengan PVQ.

Berdasarkan uraian teori di atas, peneliti mengajukan hipotesis bahwa koefisien reliabilitas dipengaruhi oleh karakteristik sampel.

Metode

Data koefisien reliabilitas yang digunakan dalam penelitian ini diperoleh dari data survey kepemimpinan dan nilai yang dilakukan oleh Helly P. Soetjipto pada tahun 2006. Data survey terdiri atas 36 paket data yang masing-masing disebar dari berbagai daerah di Pulau Jawa, Sumatera, Kalimantan, dan Papua Barat. Pada saat penyebaran, tiap paket berisi 30-40 kuesioner. Saat pengumpulan data, jumlah kuesioner dari tiap paket menjadi 21 sampai dengan 40 bendel. Kemudian, peneliti menghitung koefisien reliabilitas masing-masing paket untuk digunakan sebagai data dalam penelitian ini.

Potrait Values Questionnaire (PVQ). Peneliti menggunakan PVQ versi bahasa Indonesia yang diadaptasi dari PVQ versi bahasa Inggris. Dalam prosedur penyajiannya, pertama, responden diminta untuk merespon 40 item pernyataan yang merepresentasikan 10 tipe nilai (Schwartz, 1992). Responden memberikan rating pada tiap item berkisar antara skor 6 untuk pernyataan yang "sangat serupa dengan diri saya" dan 1 untuk pernyataan yang "sama sekali tidak serupa dengan diri saya". Untuk mengetahui skor tiap nilai, item-item pada satu kategori nilai dihitung rata-ratanya.

Data demografis. Sebelum mengisi skala, responden diminta untuk mengisi data diri yang terdiri atas usia, jenis kelamin, dan tingkat pendidikan.

Untuk mengetahui pengaruh karakteristik sampel terhadap koefisien reliabilitas pengukuran, peneliti menggunakan teknik analisis regresi. Penghitungan data statistik dibantu dengan software SPSS versi 12.0.

Hasil

Variabel-variabel karakteristik sample yang dalam penelitian ini digunakan sebagai predictor koefisien reliabilitas pengukuran PVQ adalah jenis kelamin (laki-laki dan perempuan) dan tahap-tahap perkembangan (remaja, dewasa awal, dewasa tengah, dan dewasa akhir). Analisis regresi linier digunakan untuk menguji predictor-prediktor tersebut. Adapun hasilnya dapat dilihat pada table 2.

Tabel 2.

Hasil regresi variabel demografi terhadap koefisien reliabilitas pengukuran PVQ

N	Mean	SD	r	Beta	VIF
Laki2	15,97	4,025	0,253	-0.010	2,208
Perempuan	14,03	3,637	-0,348*	-0,348*	2,208
Remaja	3,93	2,495	0,046	-0,166	2,404
Dewasa awal	11,21	4,854	0,122	-0,213	4,639
Dewasa tengah	11,71	5,195	10,284	-0,428	2,744
Dewasa akhir	2,71	2,946	-0,062	-0,083	2,066

*signifikan, $p < 0,05$

Uji Prasyarat

Dua asumsi penting yang harus dipenuhi sebelum melakukan analisis regresi adalah uji linieritas dan uji kolinieritas. Hasil uji linieritas dengan menggunakan koefisien Pearson menunjukkan hanya variabel independen N perempuan yang linier dengan variabel koefisien reliabilitas PVQ ($r = -0,356$, $p < 0,01$). Sementara lima variabel independen lainnya, yaitu N laki-laki ($r = 0,253$, $p > 0,05$), N remaja ($r = 0,046$, $p > 0,05$), N dewasa awal ($r = 0,122$, $p > 0,05$), N dewasa tengah ($r = 10,28$, $p > 0,05$), dan N dewasa akhir ($r = -0,062$, $p > 0,05$) tidak menunjukkan hubungan dengan koefisien reliabilitas PVQ.

Selanjutnya peneliti melakukan uji kolinieritas untuk memeriksa apakah ada hubungan yang berlebihan antar variabel independen. Pengujian menggunakan koefisien VIF dengan asumsi tidak ada kolinieritas antar variabel independen bila $VIF < 4,00$ (Garson, 2009). Pada tabel 1 dapat dilihat hanya variabel N dewasa awal yang menunjukkan kolinieritas ($VIF = 4,639$), sementara lima variabel independen lainnya, yaitu N laki-laki ($VIF = 2,208$), N



nitro PDF

professional

perempuan ($VIF = 2,208$), N remaja ($VIF = 2,404$), N dewasa tengah ($VIF = 2,744$), dan N dewasa akhir ($VIF = 2,066$) tidak menunjukkan kolinieritas. Berdasarkan uji linieritas dan uji kolinieritas tersebut, dapat disimpulkan bahwa variabel N perempuan dapat dianalisis lebih lanjut dengan teknik analisis regresi.

Penelitian ini menggunakan teknik analisis regresi sederhana. Seberapa besar sumbangan efektif variabel independen sebagai prediktor terhadap variabel dependen diketahui dengan melihat skor beta. Sebagaimana yang telah disampaikan sebelumnya, pada akhirnya hanya variabel N perempuan yang dapat diuji regresi. Hasil regresi menunjukkan bahwa variabel N perempuan merupakan prediktor terhadap koefisien reliabilitas PVQ. Sumbangan efektif variabel independen N perempuan ($R = 0,348$) terhadap koefisien reliabilitas PVQ adalah sebesar 34,8 % sementara 65,2 % dipengaruhi oleh variabel-variabel lain yang tidak diketahui. Berdasarkan hasil analisis dapat disimpulkan bahwa hipotesis "terdapat pengaruh karakteristik sampel terhadap koefisien reliabilitas" diterima. Koefisien beta variabel N perempuan adalah sebesar $-0,348$. Hal ini mengindikasikan pengaruh negatif jumlah sampel perempuan terhadap koefisien reliabilitas PVQ. Semakin banyak jumlah responden perempuan akan semakin menurunkan konsistensi pengukuran dengan PVQ.

Diskusi

Penelitian ini membuktikan pengaruh karakteristik sampel pada koefisien reliabilitas PVQ. Hasil penelitian tersebut selaras dengan beberapa penelitian sebelumnya (Eason, 1991; Dawis, 1987; Thompson, 1994; Shield dan Caruso, 2003, dan Vacha-Haase, 1998). Shield dan Caruso mengidentifikasi enam karakteristik sampel yang dapat mempengaruhi reliabilitas pengukuran, yaitu rata-rata usia responden, jenis kelamin, tipe sampel, variabilitas skor, latar belakang etnis, dan koefisien reliabilitas otentik. Namun, penelitian ini menunjukkan hanya jumlah responden wanita yang mempengaruhi koefisien reliabilitas PVQ.

Pengaruh proporsi sampel wanita pada skor reliabilitas juga terungkap dari studi meta analitik Shields, Howell, Potter, dan Weiss (2007) pada *The Michigan Alcoholism Screening Test* (MAST) dan versi pendek skala ini, *Short Michigan Alcoholism Screening Test* (SMAST). Sama dengan hasil penelitian ini, Shields dkk menyimpulkan

bahwa semakin besar proporsi sampel laki-laki dibanding sampel perempuan, maka skor reliabilitas MAST dan SMAST semakin tinggi.

Secara umum, hasil penelitian ini memperkuat pendapat Eason (1991) dan Thompson (1994) yang menyatakan bahwa skor reliabilitas merupakan karakteristik data. Tingkat heterogenitas atau homogenitas latar belakang responden akan mempengaruhi tinggi rendah konsistensi pengukuran. Hal ini harus benar-benar dicermati oleh para peneliti yang melaporkan koefisien reliabilitas guna menghindari tiga tipe kesalahan pelaporan sebagaimana yang telah dipaparkan pada awal tulisan. Kesalahan yang paling umum dilakukan oleh peneliti adalah hanya melaporkan skor reliabilitas data try out. Semestinya, peneliti juga harus melaporkan skor reliabilitas data penelitian yang sesungguhnya mengingat heterogenitas karakteristik sampel berbeda dengan data try out.

Kesimpulan

1. Proporsi sampel wanita berpengaruh negatif terhadap koefisien reliabilitas pengukuran dengan *Potrait Values Questionnaire*. Semakin besar proporsi sampel wanita maka reliabilitas pengukuran PVQ semakin menurun.
2. Skor reliabilitas pengukuran merupakan karakteristik data penelitian.
3. Pada setiap pelaporan hasil penelitian, peneliti harus mencantumkan skor reliabilitas pengukuran dengan data sesungguhnya, bukan reliabilitas pengukuran data try out.

Daftar Pustaka

- Amriang, F., Probowati R, Y., dan Atmadji, Gunardi. 2004. *Anima: Indonesian Psychological Journal Vol. 18 No. 1 hlm 179-193*
- Bachner, Y. G. dan O'Rourke, N. 2007. Reliability generalization of responses by care providers to the Zarit Burden Interview. *Aging & Mental Health, November 2007; 11 (6): 678-685*
- Basoeki, L., Warsiki, E., Hanimah, F., Aryono, D., dan Soetjipto. 2004. Problems of Internally Displaced Children (IDC) in West Ketapang, Madura. *Anima: Indonesian Psychological Journal Vol. 18 No. 1 hlm 136-147* Created with

- Barnea, M.F. dan Schwartz, S. H. 1998. Values and Voting. *Political Psychology*, 19, 17-40
- Beretvas, S. Natasha, Suizzo, Marie-Anne, Durham, Jennifer A., dan Yarnell, Lisa M. 2008. A Reliability Generalization Study of Scores on Rotter's and Nowicky-Strickland's Locus of Control Scales. *Educational and Psychological Measurement*, Feb 2008; vol. 68: pp. 97-119
- Budiharjo, Andreas. 2004. Information System, Job Decision Latitude, Responsiveness, Planner Experience, and Planning Effectiveness. *Anima: Indonesian Psychological Journal* Vol. 19 No. 2 hlm 111-121
- Collins, Linda M. 2006. *Reliability for static and dynamic categorical latent variables*. In Linda M. Collins and line G. Soyer (Ed). *New Methods for The Analisis of Change*. American Psychological Association: Washington D. C
- Dawis, R. V. 1987. Scale construction. *Journal of Counseling Psychology*, 34(4), 481-489
- Djalali, M. As'ad. 2004. Tipe Kepribadian Kode Warna dan Kreativitas. *Anima: Indonesian Psychological Journal* Vol. 18 No. 1 hlm 24-30
- Editor. 2006. Tata Aturan untuk Penulis. *Jurnal Phronesis* Vol. 8 No. 2
- Embretson, Susan E. dan Reise, Steven P. 2000. *Item Response Theory for Psychologist*. Lawrence Erlbaum Associates, Publisher: London
- Moesono, Anggadewi. 2004. Variabel Pengasuhan Ibu yang Mampu Mencerdaskan Anak. *Anima: Indonesian Psychological Journal* Vol. 18 No. 1 hlm 12-23
- Simanjuntak, E., Sumargi, A. M., dan Apsari, Y. 2004. Metode Pengajaran Menggunakan Mailing List dan Motivasi Belajar Mahasiswa. *Anima: Indonesian Psychological Journal* Vol. 18 No. 1 hlm 167-178
- Yuliani M. dan Rostiana. 2006. Komitmen Organisasi Ditinjau Berdasarkan Iklim Organisasi dan Motivasi Berprestasi. *Jurnal Phronesis* Vol. 5 No. 9
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58, 6-20
- Thompson, B. 1994. Guideline for authors. *Educational and Psychological Measurement*, 54, 837-847

Created with

 **nitro**PDF[®] professional

- Thompson, B. 2002. What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, 31(3), 24-31
- Rowley, G. L. (1976). The reliability of observational measures. *American Educational Research Journal*, 13, 51-59.
- Miller, M. David. 2009. Reliability. *Encyclopedia of Educational Psychology*. 2007. SAGE Publications. 14 Jan. 2009.
- Shields, A. L., Howell, R. T., Potter, J. S., dan Weiss, R. D., 2007. The Michigan Alcoholism Screening Test and Its Shortened Form: A Meta-Analytic Inquiry Into Score Reliability. *Substance Use & Misuse*, 42:1783-1800
- Henson, R. K. Dan Thompson, B. 2002. Characterizing Measurement Error in Score Across Studies: Some Recommendations for Conducting "Reliability Generalization" Studies. *Measurement and Evaluation in Counseling and Development*. July 2008 Volume 35
- Trochim, W. dan Donnelly, J. P. 2007. *The Research Methods Knowledge Base*, 3rd edition. Atomic Dog Publishing: Ohio
- Schwartz, S. H. 1992. Universals in the content and structure of values: Theoretical advances and empirical test in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology*
- Shields, A.L., & Caruso, J.C. (2003). Reliability generalization of the alcohol use disorders identification test. *Educational and Psychological Measurement*, 63, 404-413
- Rokeach, M. 1968. *Beliefs, Attitudes and values ; a theory of organization and change*. Jossey-Bass: London
- Lelkes, Orsolya, 2006. Tasting freedom: Happiness, religion and economic transition. *Journal of Economic Behavior & Organization*, Elsevier, vol. 59(2), pages 173-194, February