

UJI VALIDITAS KONSTRUK PADA ALAT UKUR INTEGRITAS DENGAN METODE *CONFIRMATORY FACTOR ANALYSIS* (CFA)

Masitah¹

Universitas Indonesia

Abstrak:

Integritas merupakan kekuatan karakter yang mempengaruhi kesehatan mental, kesejahteraan psikologis dan keefektifan hubungan interpersonal. Integritas sangat dibutuhkan dalam dunia pekerjaan terutama dalam hal promosi. Namun, alat ukur integritas lebih banyak dikembangkan di luar negeri sehingga kurang sesuai dengan konteks masyarakat Indonesia. Selain itu, umumnya alat ukur integritas dikembangkan menggunakan pendekatan klasik yang memiliki beberapa kelemahan. Penelitian ini menguji validitas konstruk dari teori integritas Carl Rogers. Peneliti mengembangkan alat ukur integritas menggunakan pendekatan polytomous Item Response Theory dengan melibatkan 1210 pekerja Indonesia. Hasil uji coba menunjukkan alat ukur terbukti valid dan memenuhi asumsi unidimensionalitas. Berdasarkan pengujian 26 item, terdapat satu item yang tidak fit, sehingga item tersebut dikeluarkan. Hasil pengujian kembali terhadap 25 item menunjukkan bahwa model fit dan seluruh item fit mengukur integritas. Analisis menggunakan differential item functioning menunjukkan 1 item memiliki bias respon jenis kelamin. Dengan demikian, item yang dipertahankan dalam alat ukur integritas ini berjumlah 24 item

Kata kunci : Integritas, Rating Scale Model, Polytomous Item Response Theory

Pendahuluan

Integritas merupakan kekuatan karakter yang berlaku di segala aspek kehidupan seperti pendidikan, p e n e l i t i a n d a n p e k e r j a a n . Sebagaimana dikemukakan Peterson dan Seligman (2004), integritas bersifat universal dan dibutuhkan di berbagai peran. Hal ini didukung oleh Schlenker, Miller dan Johnson (2009) yang menyatakan bahwa integritas telah dikembangkan dalam setiap lapisan masyarakat karena berdampak penting bagi hubungan sosial. Penjelasan Deci dan Ryan (2000) mengenai kontribusi integritas

terhadap kesehatan mental, kesejahteraan psikologis serta keefektifan hubungan interpersonal juga memperkuat pernyataan tersebut. Khalil (2004) menambahkan bahwa integritas berkaitan dengan apa yang disetujui individu secara implisit atau eksplisit.

Integritas sendiri didefinisikan Rogers (1961) sebagai kondisi yang terjadi ketika individu mampu menerima serta bertanggung jawab terhadap perasaan, niat, komitmen dan perilaku, termasuk mampu mengakui kondisi itu kepada orang lain bila

¹ Penulis adalah alumni program Magister Pemintatan Psikometri - Fakultas Psikologi UI tahun 2012.

Korespondensi tentang artikel ini dapat menghubungi : redaksi_jp3i@yahoo.co.id

diperlukan. Carter (1996) memperkuat definisi tersebut dengan menyatakan bahwa individu yang memiliki integritas bersedia menanggung konsekuensi dari keyakinannya, meskipun hal itu sulit dilakukan, konsekuensinya tidak menyenangkan, bahkan tidak mendapat kerugian jika tidak mempertahankan integritasnya. Menurut Khalil (2004), integritas ditegaskan ketika individu memiliki pilihan untuk melanggar komitmen dan janji yang ia buat sendiri karena pelanggaran tersebut mendatangkan rasa malu terhadap dirinya sendiri.

Beberapa tahun terakhir banyak yang tertarik meneliti integritas sebagai kekuatan karakter. Deci dan Ryan (2000) menyatakan bahwa manusia memiliki kebutuhan psikologis terhadap integritas. Selain itu, Miller dan Schlenker (2011) menjelaskan bahwa integritas dalam diri individu berkaitan dengan pandangan hidup yang lebih positif, orientasi yang lebih positif terhadap orang lain, spiritualitas yang lebih tinggi serta minimnya tindakan irasional.

Dunia kerja adalah salah satu yang paling menuntut pentingnya integritas. Alat ukur integritas diperlukan dalam membina sumber daya manusia di dalam perusahaan karena sumber daya manusia dituntut dapat bekerja secara produktif untuk mendukung tujuan-tujuan yang ingin dicapai perusahaan. Namun kenyataan menunjukkan bahwa sumber daya manusia di perusahaan, yang disebut pekerja, sering kali justru berperilaku

sebaliknya, yaitu perilaku

kontraproduktif seperti korupsi dan mencuri waktu untuk keperluan pribadi. Perilaku kontraproduktif menunjukkan rendahnya kualitas pekerja pada suatu perusahaan. Hal ini diketahui berdasarkan *The Global Competitiveness Report 2011-2012* dari *World Economic Forum* yang merilis bahwa peringkat daya saing Indonesia menurun dari 44 menjadi 46 dari 142 negara yang di survei. Peringkat ini menunjukkan bahwa profesionalisme, kehadiran dan kuantitas pekerjaan pekerja Indonesia masih tergolong rendah. Untuk itu diperlukan unsur penting yang dapat meningkatkan efektivitas sikap serta kualitas dalam bekerja. Istilah yang paling menggambarkan unsur tersebut secara tepat adalah integritas (Impelman, 2006). Hal ini sesuai dengan pernyataan Marchus dan Schuler (2004) bahwa pekerja yang memiliki level integritas tinggi akan menghasilkan produktivitas kerja yang baik.

Untuk mendapatkan informasi mengenai integritas yang dimiliki pekerja, perusahaan perlu melakukan pengukuran. Waktu yang tepat mengukur integritas, menurut Mumford (2000) ditentukan berdasarkan tujuan yang diharapkan dari pengukuran. Mengukur integritas ketika rekrutmen memungkinkan perusahaan menyeleksi calon karyawan dengan tingkat integritas yang baik. Hal ini penting karena

karyawan yang memiliki integritas rendah dapat merusak keseluruhan kinerja perusahaan, *image* perusahaan, kepercayaan kolega dan pemegang saham, menurunkan performa finansial perusahaan serta mempengaruhi kondisi ekonomi masyarakat secara keseluruhan (Mount, Iles dan Johnson, 2006). Selain itu, integritas juga dapat diukur pada waktu promosi menjadi pimpinan. Sebagai figur yang mengemban peran dan tanggung jawab lebih besar dalam perusahaan, tuntutan terhadap pemimpin yang memiliki integritas sangat besar (Erhard, Jensen dan Zaffron, 2011). Pimpinan yang memiliki integritas tinggi merupakan solusi mencegah perilaku kontraproduktif karyawan (Wanek, Sackett dan Ones, 2003). Jika karyawan menilai pimpinan menunjukkan perilaku yang memiliki integritas, karyawan akan mengupayakan perilaku yang sama (Ones dan Viswesvaran, 2001), sehingga interaksi lingkungan kerja menjadi lebih efektif (Palanski dan Yammarino, 2007).

Sebagaimana umumnya variabel psikologis, integritas bersifat laten dan hanya dapat diamati melalui sampel perilaku. Merancang *item-item* yang mampu menggal integritas merupakan tantangan besar yang dihadapi dalam perancangan alat ukur integritas. Kebutuhan akan alat ukur integritas yang valid dan reliabel, khususnya dalam dunia kerja, sudah

sedemikian besar dan akan semakin besar (Barry, Sackett dan Wiemann, 2007; Ones, Viswesvaran dan Schmidt, 2003; Sackett, Burris dan Callahan, 1989).

Gambaran Umum Alat Ukur Integritas

Selama beberapa dekade terakhir, penggunaan tes-tes integritas dalam konteks pekerjaan semakin berkembang di negara lain (Erhard, Jensen dan Zaffron, 2011; Egberink dan Veldkamp, 2007; Palanski dan Yammarino, 2007; Impelman, 2006; Barrett, 2001; Ones dan Viswesvaran, 2001; US Congressional Office of Technology Assessment, 1990; Martelli, 1988; Harris, 1987) serta dianggap sebagai salah satu alat ukur yang paling valid dan tidak memberi dampak yang merugikan (Barry, Sackett dan Wiemann, 2007; Ones, Viswesvaran dan Schmidt, 1995; Wanek, Sackett dan Ones, 2003; Hunter, in preparation). Hal ini didukung oleh Ones, Viswesvaran dan Schmidt (1995) yang menyatakan bahwa tes integritas dapat memprediksi performa kerja serta perilaku kontraproduktif dalam pekerjaan. Penelitian Ones (1993) juga menemukan bahwa tes integritas berkorelasi dengan usia dan *gender*. Individu yang berusia lebih muda cenderung berperilaku kontraproduktif karena kecerobohan dan keinginan coba-coba/eksperimen.

Berkaitan dengan *gender*, rata-rata wanita dilaporkan memiliki skor integritas yang lebih tinggi daripada pria.

Integritas merupakan karakteristik positif yang memiliki sifat universal, pekerja di Amerika dan Indonesia sama-sama dituntut memiliki integritas dalam pekerjaannya. Namun kurang tepat jika alat ukur integritas yang dikembangkan pada latar belakang Amerika digunakan untuk mengukur integritas di Indonesia. Terlebih jika hasil pengukuran tersebut akan menjadi dasar keputusan-keputusan penting. Indonesia bahkan memiliki perbedaan mendasar dengan bangsa Asia lainnya. *Self* bagi bangsa Asia umumnya tidak ditekankan sebagai kesatuan yang independen karena menganggap *self* ditentukan lingkungan dan konteks sosial (Markus, Kitayama dan Heiman, 1996), hal ini sesuai dengan penelitian yang dilakukan Suh (2001) terhadap bangsa Korea. Hasil penelitian Iwao (1997) juga menambahkan bahwa partisipan yang berasal dari Jepang tidak mengungkapkan ketidaksetujuan yang mereka rasakan segamblang partisipan yang berasal dari Amerika. Temuan ini mengindikasikan bahwa mempresentasikan diri secara jujur tidak terlalu penting dalam kebudayaan Asia. Tidak berarti bangsa Asia memiliki tingkat integritas yang rendah atau mengkhianati nilai-nilai dan komitmen yang mereka anut, itu hanya indikasi perbedaan latar

belakang budaya.

Perbedaan tersebut memunculkan tuntutan yang semakin besar terhadap alat ukur integritas yang dirancang berdasarkan kondisi Indonesia. Alasan tersebut juga menarik minat peneliti untuk merancang alat ukur integritas yang sesuai dengan konteks perusahaan di Indonesia. Alat ukur integritas yang dirancang dalam penelitian ini ditujukan bagi kepentingan promosi pimpinan, khususnya level manager pada perusahaan. Alat ukur ini tepat diberikan kepada partisipan yang telah memiliki pengalaman sebagai pekerja karena *item-item* alat ukur dirancang berdasarkan situasi yang dialami pekerja sehari-hari.

Di Indonesia masih sedikit yang mengembangkan penelitian mengenai alat ukur integritas, salah satunya Permatasari (2011). Tetapi alat ukur tersebut memiliki kelemahan dari sisi teori dan psikometris. Kelemahan di sisi teori terjadi karena Permatasari (2011) merancang alat ukur integritas menggunakan *moral identity theory* dari Blasi (2004) yang tidak membahas integritas secara spesifik. Blasi (2004) adalah tokoh filsafat yang bahasan utamanya moral, bukan kondisi psikologis. Teori yang dikemukakan Rogers (1961), seorang psikolog humanistik, tentu lebih mampu menggambarkan integritas dari sisi psikologis. Maka alat ukur ini dirancang berdasarkan konsep integritas yang diutarakan Rogers

(1961).

Kelemahan lain alat ukur Permatasari (2011) dapat dijelaskan melalui sisi psikometris. Meski dirancang berdasarkan kondisi *Indonesian*, pengembangan nya menggunakan pendekatan klasik sehingga menghasilkan alat ukur yang bersifat *sample bound*. Hambleton, Swaminathan dan Rogers (1991) menjelaskan bahwa alat ukur yang tergolong *sample bound* seolah-olah memiliki tingkat kesulitan yang tinggi ketika diberikan pada kelompok subjek yang memiliki kemampuan rendah dan seolah-olah memiliki tingkat kesulitan yang rendah ketika diberikan pada kelompok subjek yang memiliki kemampuan tinggi. Pengembangan alat ukur menggunakan pendekatan klasik juga menyebabkan tidak diketahuinya parameter *item*. Analisis kualitas *item* berubah tergantung siapa yang mengerjakan alat ukur tersebut. Alat ukur yang kualitas *item*nya tidak diketahui secara jelas tentu tidak tepat dijadikan dasar keputusan-keputusan penting.

Untuk mengatasi kelemahan psikometris tersebut, penelitian ini menggunakan pendekatan yang lebih modern, akurat dan bebas *sample bound* yaitu *Item Response Theory* (IRT). Kelebihan penggunaan IRT adalah diperoleh karakteristik item yang tidak tergantung pada kemampuan individu yang menemukannya (*sampling invariant*).

Item-item alat ukur dianalisis menggunakan pendekatan *Rasch Model* agar bisa dipakai berulang-ulang, ini juga menjadi alasan pentingnya memiliki *item-item* yang *sampling invariant*.

Alat ukur integritas yang dirancang dalam penelitian ini menggunakan pendekatan *Item Response Theory* (IRT). Hambleton, Swaminathan dan Rogers (1991) menyatakan bahwa pendekatan IRT dapat diterapkan apabila data memenuhi asumsi *unidimensionality* dan *local independence*. Kedua asumsi tersebut akan memastikan bahwa hanya ada satu faktor dominan yang mempengaruhi skor partisipan. Pengujian asumsi diperlukan untuk membuktikan bahwa alat ukur yang dirancang memang hanya mengukur integritas.

Berkaitan dengan penerapan IRT, alat ukur integritas yang dikembangkan dalam penelitian ini menggunakan model pengukuran *polytomous IRT*. Model ini ditentukan berdasarkan kategori respon serta jumlah parameter item yang terlibat. Mengingat alat ukur dirancang dalam bentuk skala likert yang memiliki empat kategori respon, model politomi yang paling tepat adalah *Rating Scale Model* (Embretson dan Reise, 2000). Selain itu, penerapan IRT memberi informasi akurat mengenai kualitas *item-item* yang dirancang. Keputusan yang diambil berdasarkan alat ukur integritas yang dirancang

menggunakan pendekatan IRT menjadi lebih reliabel. Kelebihan IRT berikutnya adalah mampu mendeteksi *Different Item Functioning* (DIF) yang dapat menjelaskan apakah item yang sama akan memiliki kecenderungan berbeda ketika ditempuh oleh kelompok partisipan yang berbeda, namun memiliki tingkat integritas yang sama. Peneliti menguji DIF berdasarkan jenis kelamin karena Ones (1993) menemukan bahwa gender berpengaruh terhadap ada atau tidaknya pengukuran integritas.

Mengingat masih sedikit alat ukur integritas yang dirancang berdasarkan konteks pekerjaan di Indonesia, maka penelitian ini berupaya merancang alat ukur integritas untuk kepetingan promosi level manager di perusahaan. Meskipun alat ukur dirancang menggunakan teori Rogers (1961) yang mengamati integritas dari sisi psikologis, belum pernah ada yang melakukan analisis faktor untuk membuktikan bahwa teori integritas tersebut memiliki konstruk yang valid. Untuk itu, alat ukur integritas dalam penelitian ini dikembangkan menggunakan *metode confirmatory factor analysis* untuk membuktikan bahwa teori integritas yang dikemukakan Rogers (1961) memenuhi asumsi unidimensionalitas. Unidimensionalitas sangat penting diteliti untuk membuktikan bahwa alat ukur yang dirancang dalam penelitian ini memang hanya mengukur integritas.

Selain itu, dilakukan pengujian untuk membuktikan apakah alat ukur integritas yang dirancang dalam penelitian ini terdiri dari *item-item* yang *fit* karena *item-item* yang *fit* akan memberikan hasil ukur yang valid. Setelah asumsi unidimensionalitas terpenuhi dan alat ukur terbukti memiliki *item-item* yang *fit*, pengujian terakhir dilakukan untuk mendeteksi kemungkinan adanya *item* yang memiliki kecenderungan bias respon. Ini merupakan kelebihan penggunaan pendekatan IRT, tetapi semuanya tidak pernah diuji pada pendekatan klasik.

Item-item alat ukur integritas yang dikonstruksi dalam penelitian ini memiliki 4 pilihan respon mulai dari “Sangat Sesuai” sampai “Sangat Tidak Sesuai”. Alat ukur dirancang berdasarkan teori integritas yang dikemukakan oleh Rogers (1961). Faktor-faktor pembentuk integritas menurut psikolog humanistik Rogers (1961) terdiri dari kejujuran; keteguhan; memiliki *self-control* yang kuat; dan memiliki *self-esteem* yang tinggi. Faktor merupakan kesatuan utuh (koherensi) yang tidak bisa dipecah-pecah karena saling terkait satu sama lain. Individu dikatakan memiliki integritas apabila memiliki seluruh faktor.

Awalnya alat ukur memiliki 40 *item* yang ditetapkan berdasarkan hasil wawancara dengan beberapa expert. Sebelumnya telah dilakukan studi pendahuluan dan analisis statistik untuk menjadi acuan memperbaiki

item-item yang kurang baik. Kemudian peneliti menguji coba alat ukur integritas kepada 100 partisipan yang hasilnya diolah kembali untuk mendapatkan hasil uji psikometrinya. Alat ukur integritas yang telah melalui proses uji coba selanjutnya melewati proses elisitasi untuk merevisi *item-item* yang kurang sesuai sebelum digunakan untuk pengambilan data.

Metode

Partisipan Penelitian

Untuk menguji validitas alat ukur integritas dalam konteks pekerjaan di Indonesia, peneliti memilih pekerja sebagai partisipan penelitian. Alat ukur diberikan kepada individu yang menduduki level manager, supervisor dan staf dengan masa kerja minimal satu tahun di perusahaan tempat bekerja saat ini. Persyaratan masa kerja minimal satu tahun digunakan agar pekerja memiliki pengetahuan dan pengalaman bekerja di perusahaan. Partisipan juga disyaratkan memiliki tingkat pendidikan minimal strata satu (S1) dalam rentang usia 24-56 tahun. Patokan itu merujuk pada usia pensiun yang ditetapkan Undang-Undang Ketenagakerjaan. Usia tersebut merupakan tahap formal-operasional yang ditandai adanya kemampuan berpikir abstrak sehingga partisipan diharapkan mampu melakukan evaluasi dan penilaian terhadap hidupnya (Papalia, 2004). Individu

yang mencapai tahap ini dianggap memiliki kemampuan mengintegrasikan *self* agar terjadi koherensi antara nilai-nilai moral yang dipahami, *beliefs*, *tindakan*, *komitmen* serta perkataannya. Tahapan usia tersebut juga menjadikan individu lebih mampu menjalani proses-proses untuk tercapainya integritas seperti yang dikemukakan oleh Petrick dan Quinn (2000). Partisipan penelitian mencakup 1210 pekerja yang berdomisili di wilayah Pulau Sumatera dan Jawa yang merupakan pulau dengan tingkat kepadatan penduduk serta jumlah perusahaan terbesar di Indonesia. Partisipan dari Pulau Sumatera berasal dari Sumatera Utara (Medan, Binjai, Rantau Prapat, Pematang Siantar dan Tebing Tinggi), Sumatera Barat (Padang), Sumatera Selatan (Palembang), Pekanbaru (Riau), Lampung, Jambi dan Bengkulu. Partisipan dari Pulau Jawa berasal dari Jakarta, Jawa Barat (Bandung), Jawa Tengah (Semarang), Jawa Timur (Surabaya) dan Yogyakarta.

Gambaran demografis penyebaran partisipan penelitian dihitung berdasarkan jenis kelamin, status pekerjaan dan jabatan. Berikut gambaran demografis berdasarkan distribusi frekuensi.

Berdasarkan tabel diketahui bahwa partisipan perempuan dalam penelitian ini berjumlah 500 orang (50%) Sementara partisipan laki-laki dalam penelitian ini berjumlah 500

Tabel 1. Penyebaran Partisipan Berdasarkan Jenis Kelamin, Status Pekerjaan dan Jabatan

Jenis Kelamin	Status Pekerjaan	Jabatan	Frekuensi	Persentase (%)
Laki-Laki	PNS	Manager	119	11.9
		Supervisor/Staf	113	11.3
	Swasta	Manager	168	16.8
		Supervisor/Staf	100	10
Perempuan	PNS	Manager	148	14.8
		Supervisor/Staf	120	12
	Swasta	Manager	115	11.5
		Supervisor/Staf	117	11.7
Jumlah			1000	100

Tabel 2. Penyebaran Partisipan Berdasarkan Usia

Usia	Frekuensi	Persentase (%)
> 50 tahun	99	9.9
40 – 50 tahun	351	35.1
30 – 39 tahun	415	41.5
< 30 tahun	135	13.5
Jumlah	1000	100

orang (50%).

Partisipan yang memiliki status pekerjaan sebagai PNS berjumlah 500 orang (50%), sementara partisipan yang memiliki status pekerjaan sebagai swasta berjumlah 500 orang (50%).

Partisipan yang memiliki jabatan sebagai manager berjumlah 550 orang (55%), sementara partisipan yang memiliki jabatan sebagai supervisor/staf berjumlah 450 orang (45%).

Berdasarkan tabel diketahui bahwa penelitian ini melibatkan 135

(13.5%) partisipan yang berusia di bawah 30 tahun, 415 (41.5%) partisipan yang berusia antara 30-39 tahun, 351 (35.1%) partisipan yang berusia antara 40-50 tahun, serta 99 (9.9%) partisipan yang berusia di atas 50 tahun.

Analisis Data

Peneliti melakukan kajian literatur mengenai integritas serta melakukan elisitasi untuk mendapatkan gambaran spesifik indikator-indikator integritas. Selanjutnya dilakukan pengujian

validitas, analisis *item* dan uji asumsi unidimensionalitas menggunakan *confirmatory factor analysis* (CFA). Jika hasil CFA menunjukkan bahwa alat ukur memenuhi asumsi unidimensionalitas, maka dapat dilakukan perhitungan IRT untuk menentukan apakah model sudah *fit* dan terdiri dari *item-item* yang *fit* mengukur integritas. Uji coba alat ukur integritas diawali dengan melibatkan *expert judgement* yang terdiri dari praktisi beberapa perusahaan serta akademisi beberapa universitas di Pulau Sumatera dan Pulau Jawa dan dilanjutkan dengan pengujian validitas konstruk.

Validitas konstruk diuji menggunakan metode analisis faktor yaitu dengan *confirmatory factor analysis* (CFA) yang dilakukan untuk menguji dimensionalitas sebagai acuan menguji asumsi unidimensionalitas alat ukur. CFA juga memberi informasi apakah model *fit* untuk mengukur integritas menggunakan alat ukur ini. Pada tahap ini, analisis dilakukan menggunakan program LISREL versi 8.7 (Joreskog & Sorbom, 1996).

Logika dasar dari CFA adalah sebagai berikut (Umar, 2011) :

1. Sebuah konsep atau trait berupa kemampuan yang didefinisikan secara operasional sehingga dapat disusun pertanyaan atau pernyataan untuk mengukurnya. Kemampuan ini disebut faktor. Sedangkan pengukuran terhadap faktor dilakukan melalui

analisis terhadap respon atau jawaban atas setiap *item*.

2. Setiap *item* dikategorikan hanya mengukur atau memberi informasi tentang satu faktor tertentu saja, baik *item* maupun subtes bersifat uni-dimensional.

3. Berdasarkan penjelasan sebelumnya, dapat disusun sehimpunan persamaan matematis yang dapat digunakan untuk memprediksi matriks korelasi (berdasarkan data) antar *item* yang seharusnya akan diperoleh jika teori tersebut benar-benar bersifat unidimensi. Matriks korelasi tersebut dinamakan sigma (Σ). Kemudian matriks tersebut dibandingkan dengan matriks korelasi yang diperoleh secara empiris dari data (dinamakan matriks S). Jika teori tersebut benar, atau dengan kata lain terbukti unidimensi, maka seharusnya tidak ada perbedaan yang signifikan antara elemen matriks Σ dengan elemen matriks S. Secara matematis dapat dituliskan: $S - \Sigma = 0$.

4. Pernyataan matematis sebagaimana dikemukakan di atas dijadikan hipotesis nihil yang akan dianalisis menggunakan CFA. Dalam hal ini dilakukan uji signifikansi dengan *Chi-Square*. Jika *Chi-Square* yang dihasilkan tidak signifikan, maka dapat disimpulkan bahwa hipotesis nihil yang menyatakan “tidak ada perbedaan antara elemen matriks S dan Σ ” tidak ditolak. Artinya teori yang menyatakan bahwa semua *item* mengukur hal yang sama, terbukti

sesuai dengan data. Sebaliknya jika nilai *Chi-Square* yang diperoleh signifikan, maka hipotesis nihil $S - \sum = 0$ ditolak. Artinya teori tersebut tidak didukung oleh data.

5. Jika teori diterima (model fit), langkah selanjutnya adalah menguji hipotesis tentang bagaimana signifikansi masing-masing item dalam mengukur apa yang hendak diukur. Uji hipotesis ini dilakukan menggunakan *t-test*. Jika nilai t signifikan, berarti *item* yang bersangkutan signifikan dalam mengukur apa yang hendak diukur. Dengan cara seperti ini, dapat dinilai butir item mana yang valid dan yang tidak valid di dalam konteks validitas konstruk. Dengan kata lain, CFA dalam hal ini merupakan pengujian terhadap hipotesis nihil (H_0): $S - \sum = 0$. Artinya tidak ada perbedaan antara matriks korelasi yang diharapkan oleh teori dengan matriks korelasi yang diperoleh dari hasil observasi.

6. Persamaan matematis pada poin-poin di atas adalah persamaan regresi untuk setiap butir soal dalam hubungannya dengan faktor yang diukur, dengan rumus yaitu:

$$X_i = \lambda_i F + \delta_i$$

Dimana:

X_i = Skor yang diperoleh pada item i
 F = Konstruk yang hendak diukur (faktor)

λ_i = Koefisien regresi item i dalam mengukur F (koefisien muatan faktor)

δ_i = Segala hal yang mempengaruhi Varians X_i (selain F), disebut

juga “kesalahan pengukuran”.

Persamaan-persamaan tersebut akan menjadi dasar untuk membuat persamaan-persamaan untuk setiap elemen dari matriks \sum , sehingga setiap elemen tersebut dapat diperoleh nilai e s t i m a s i n y a u n t u k e m u d i a n dibandingkan dengan setiap elemen yang sama dalam matriks S . Kesesuaian model dengan data, seperti disebutkan pada poin 4, diuji secara statistik, yaitu dengan tes *Chi-Square*.

Untuk menilai apakah model pengukuran benar-benar *fit* dengan data, perlu diperhatikan nilai indeks fit. Penelitian ini menggunakan kriteria indeks fit berupa *Chi-Square* dan Root Mean Square Error of Approximation (RMSEA). Nilai *chi-square* merupakan ukuran mengenai buruknya *fit* suatu model. Probabilitas *chi-square* u a r e d a l a m p e n g u j i a n diharapkan tidak signifikan ($p > 0.05$) agar model fit dengan data. Indeks *fit* kedua yang digunakan adalah *Root Mean Square Error of Approximation* (RMSEA). Nilai $RMSEA < 0.05$ mengindikasikan model fit dan nilai $RMSEA$ yang berkisar antara 0.08 menyatakan bahwa model memiliki perkiraan kesalahan yang *reasonable*. Sementara nilai $RMSEA$ yang berkisar antara 0.08 sampai dengan 0.10 menunjukkan model memiliki *fit* yang cukup. Sedangkan nilai $RMSEA > 0.1$ menunjukkan model *fit* yang sangat buruk. Kesimpulannya, model dikatakan fit dengan data ketika nilai $RMSEA < 0.10$.

1. Pengujian *First Order CFA* Integritas 40 item pada 100

partisipan

Pengujian *first order* CFA dilakukan dengan membuat model pengukuran untuk menggambarkan sebaik apa indikator-indikator dalam alat ukur dapat digunakan sebagai instrumen pengukuran variabel laten. Dalam pengujian ini, peneliti membuat model pengukuran dengan satu variabel laten (integritas) dan 40 *item*nya dijadikan sebagai indikator yang diukur secara langsung. Pengujian ini membuktikan bahwa

seluruh item mengukur satu variabel laten yaitu integritas.

Hasil pengujian terhadap model *first order* CFA menghasilkan nilai p -value = 0.016 ($p < 0.05$) dan RMSEA = 0.034 (RMSEA < 0.1). Hanya kriteria RMSEA yang berhasil terpenuhi sehingga peneliti memutuskan model ini tidak fit dengan data. Berikut ini tabel hasil pengujianya :

Peneliti lalu mengeliminasi

Tabel 3. Indeks Kecocokan First Order CFA 40 Item

	Indeks Kecocokan	Keterangan
<i>Chi-Square</i>	825.07; df=740; p=0.01582	Model tidak <i>fit</i>
RMSEA	0.034	Model <i>fit</i>

beberapa item yang memberi sumbangan kecil pada alat ukur sehingga total diperoleh 26 *item* integritas. Uji coba alat ukur integritas 26 *item* kembali dilakukan dengan melibatkan 100 pekerja lain yang masih representatif terhadap partisipan penelitian.

2. Pengujian kembali first order

CFA

Pengujian model pengukuran *first order* CFA 26 item menunjukkan bahwa p -value dan RMSEA berhasil terpenuhi sehingga disimpulkan model benar-benar *fit* dengan data. Berikut ini tabel hasil pengujianya :

Hasil pengujian menunjukkan

Tabel 4. Indeks Kecocokan First Order CFA 26 Item

	Indeks Kecocokan	Keterangan
<i>Chi-Square</i>	299.39; df=299; p=0.48	Model <i>fit</i>
RMSEA	0.00	Model <i>closed fit</i>

bahwa 26 *item* tersebut merupakan indikator yang valid bagi pengukuran konstruk integritas. Hasil juga menunjukkan bahwa 26 *item* tersebut mengukur satu variabel laten, yaitu integritas. Maka alat ukur integritas ini dapat dikatakan telah memenuhi

asumsi unidimensionalitas sehingga penerapan model dengan pendekatan IRT bisa dilakukan. Hasil *path* diagram dari pengujian dapat dilihat pada gambar berikut.

Selain mengetahui model *fit* dari

Chi-Square=299.39, df=299, P-value=0.48285, RMSEA=0.004

Gambar 1. Path Diagram First Order CFA 26 item

suatu pengukuran, CFA juga memberi informasi mengenai indeks *fit* dari setiap item. Berikut adalah tabel *fit* *item first order* CFA alat ukur integritas. Tabel menunjukkan bahwa

Tabel 5. Hasil First Order CFA Alat Ukur Integritas 26 Item

<i>Item</i>	<i>Loading Factor</i>	<i>t-Value</i>	<i>R²</i>	<i>Keterangan</i>
Faktor: JUJUR				
<i>Item 1</i>	0.091	8.08	0.51	<i>Item Fit</i>
<i>Item 5</i>	0.089	7.52	0.46	<i>Item Fit</i>
<i>Item 9</i>	0.080	7.19	0.43	<i>Item Fit</i>
<i>Item 13</i>	0.10	8.06	0.51	<i>Item Fit</i>
<i>Item 21</i>	0.094	6.90	0.40	<i>Item Fit</i>
<i>Item 24</i>	0.080	7.17	0.42	<i>Item Fit</i>
Faktor: TEGUH				
<i>Item 2</i>	0.084	7.78	0.48	<i>Item Fit</i>
<i>Item 6</i>	0.093	6.95	0.40	<i>Item Fit</i>
<i>Item 18</i>	0.090	7.03	0.41	<i>Item Fit</i>
<i>Item 22</i>	0.093	7.24	0.43	<i>Item Fit</i>
<i>Item 25</i>	0.098	7.68	0.47	<i>Item Fit</i>
<i>Item 28</i>	0.092	7.45	0.45	<i>Item Fit</i>
<i>Item 30</i>	0.10	7.08	0.42	<i>Item Fit</i>
<i>Item 31</i>	0.093	6.88	0.40	<i>Item Fit</i>
<i>Item 33</i>	0.10	6.95	0.40	<i>Item Fit</i>
<i>Item 34</i>	0.096	7.06	0.41	<i>Item Fit</i>
<i>Item 36</i>	0.10	7.61	0.47	<i>Item Fit</i>
<i>Item 38</i>	0.10	7.26	0.43	<i>Item Fit</i>
<i>Item 39</i>	0.093	7.60	0.46	<i>Item Fit</i>
Faktor: SELF CONTROL				
<i>Item 3</i>	0.076	7.99	0.50	<i>Item Fit</i>
<i>Item 7</i>	0.10	7.31	0.44	<i>Item Fit</i>
<i>Item 19</i>	0.095	7.27	0.43	<i>Item Fit</i>
<i>Item 23</i>	0.094	7.29	0.44	<i>Item Fit</i>
Faktor: SELF ESTEEM				
<i>Item 4</i>	0.100	7.53	0.46	<i>Item Fit</i>
<i>Item 12</i>	0.095	7.32	0.44	<i>Item Fit</i>
<i>Item 20</i>	0.086	6.93	0.40	<i>Item Fit</i>

seluruh item *fit* dalam mengukur integritas. Hal ini terlihat dari indeks *t-value* yang lebih besar dari 1.96. Item yang memberikan kontribusi terbesar untuk alat ukur integritas adalah item 4, 7, 13, 30, 33, 36 dan 38 dengan muatan faktor sebesar 0.10 sementara item yang memberikan kontribusi terkecil adalah item 3 dengan muatan faktor sebesar 0.076.

3. Pengujian *second order* CFA

Hasil pengujian yang dilakukan terhadap model pengukuran *second order* CFA pada 26 item menghasilkan nilai $p\text{-value} = 0.47$ ($p > 0.05$), dan $RMSEA = 0.00$ ($RMSEA < 0.1$). Berdasarkan data, $p\text{-value}$ dan

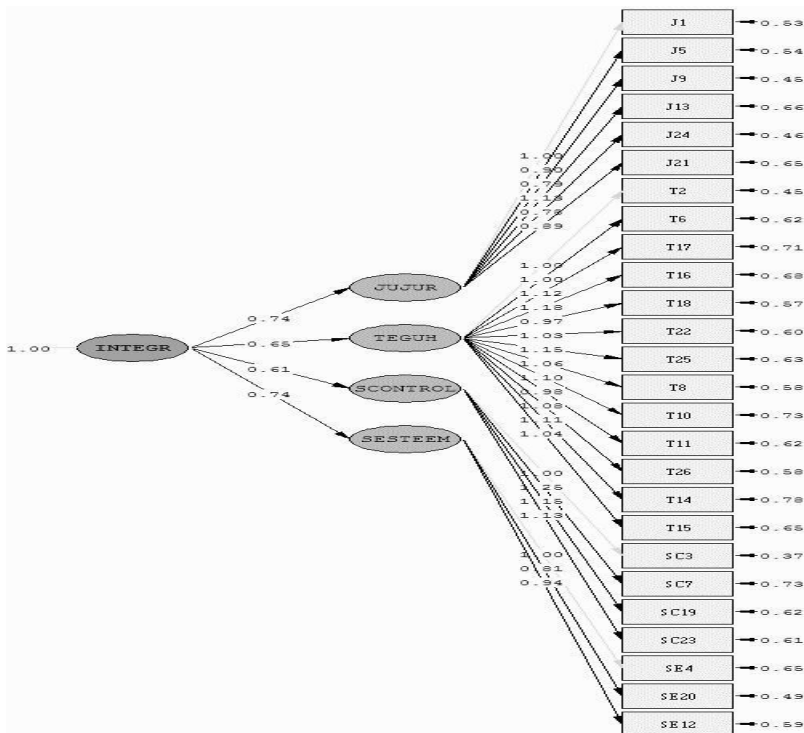
RMSEA berhasil terpenuhi sehingga dapat disimpulkan model ini benar-benar *fit* dengan data. Nilai RMSEA sebesar 0.00 menunjukkan bahwa model ini *closed fit* atau sangat *fit*. Dengan kata lain 26 item tersebut

Tabel 6. Indeks Kecocokan Second Order CFA 26 Item

	Indeks Kecocokan	Keterangan
<i>Chi-Square</i>	297.36; df=296; p=0.47	Model <i>fit</i>
RMSEA	0.00	Model <i>closed fit</i>

merupakan indikator yang valid bagi pengukuran konstruk integritas. Hasil juga menunjukkan bahwa 26 item tersebut mengukur satu variabel laten, yaitu integritas. Maka disimpulkan bahwa alat ukur integritas ini

memenuhi asumsi unidimensionalitas sehingga penerapan model dengan pendekatan IRT bisa dilakukan. *Path* diagram dari pengujian adalah:



Chi-Square=297.36, df=296, P-value=0.46691, RMSEA=0.007

Gambar 2. Path Diagram Second order CFA 26 item

Berikut adalah tabel *fit* seluruh item dari hasil pengujian *second order* CFA alat ukur integritas.

Tabel 7. Hasil *Second Order* CFA Alat Ukur Integritas 26 Item

<i>Item</i>	<i>Loading Factor</i>	<i>t-value</i>	<i>R²</i>	<i>Keterangan</i>
Faktor: JUJUR				
<i>Item 1</i>	-	-	0.50	<i>Reference Variable</i>
<i>Item 5</i>	0.14	6.53	0.44	<i>Item Fit</i>
<i>Item 9</i>	0.12	6.35	0.42	<i>Item Fit</i>
<i>Item 13</i>	0.16	6.95	0.50	<i>Item Fit</i>
<i>Item 21</i>	0.14	6.12	0.39	<i>Item Fit</i>
<i>Item 24</i>	0.12	6.30	0.41	<i>Item Fit</i>
Faktor: TEGUH				
<i>Item 2</i>	-	-	0.48	<i>Reference Variable</i>
<i>Item 6</i>	0.16	6.09	0.40	<i>Item Fit</i>
<i>Item 18</i>	0.14	7.27	0.41	<i>Item Fit</i>
<i>Item 22</i>	0.17	6.26	0.43	<i>Item Fit</i>
<i>Item 25</i>	0.18	6.54	0.47	<i>Item Fit</i>
<i>Item 28</i>	0.16	6.41	0.45	<i>Item Fit</i>
<i>Item 30</i>	0.18	6.16	0.41	<i>Item Fit</i>
<i>Item 31</i>	0.16	6.06	0.40	<i>Item Fit</i>
<i>Item 33</i>	0.18	6.07	0.40	<i>Item Fit</i>
<i>Item 34</i>	0.17	6.15	0.41	<i>Item Fit</i>
<i>Item 36</i>	0.18	6.51	0.46	<i>Item Fit</i>
<i>Item 38</i>	0.18	6.27	0.43	<i>Item Fit</i>
<i>Item 39</i>	0.17	6.50	0.46	<i>Item Fit</i>
Faktor: SELF CONTROL				
<i>Item 3</i>	-	-	0.50	<i>Reference Variable</i>
<i>Item 7</i>	0.19	6.41	0.44	<i>Item Fit</i>
<i>Item 19</i>	0.18	6.42	0.44	<i>Item Fit</i>
<i>Item 23</i>	0.18	6.39	0.44	<i>Item Fit</i>
Faktor: SELF ESTEEM				
<i>Item 4</i>	-	-	0.47	<i>Reference Variable</i>
<i>Item 12</i>	0.15	6.24	0.46	<i>Item Fit</i>
<i>Item 20</i>	0.13	6.09	0.44	<i>Item Fit</i>

Berdasarkan *t - value* dari pengujian *second order* CFA diketahui seluruh item *fit* mengukur integritas karena seluruh *t-value* lebih besar dari 1.96. Dari tabel diketahui juga bahwa item 7 memberikan kontribusi terbesar untuk alat ukur dengan muatan faktor sebesar 0.19, sementara item 9 dan 24 memberikan kontribusi terkecil dengan muatan faktor sebesar 0.12.

Hasil

Dalam pengujian diperoleh hasil pengujian *case fit*, pengujian terhadap model *fit*, item *fit* dan estimasi parameter setiap *item*. Berikut hasil pengujian *model fit*, item *fit* dan estimasi parameter setiap *item*. Pengujian *model fit* dalam IRT terdiri dari 2 bentuk yaitu pengujian *person fit* dan pengujian item *fit*.

Pengujian Person fit

Dari 1000 partisipan, seluruh kasus bisa digunakan karena tidak ada kasus yang memiliki nilai salah semua (*zero*) atau benar semua (*perfect*). Selain itu analisis terhadap nilai *infit mean square* pada estimasi kasus menunjukkan bahwa model dari seluruh kasus *fit* dengan data, yaitu sebesar 0.96 dan berada pada rentang yang dapat diterima. Peneliti juga melakukan pengujian *model fit*. Dalam pengujian *case fit* menggunakan QUEST, terdapat kriteria yang menunjukkan seberapa baik model tersebut dapat menjelaskan hasil tes yang diperoleh. Data dianggap sesuai dengan model apabila nilai *mean*

square mendekati 1,0 (baik *infit* maupun *outfit*) dan nilai uji *t* mendekati 0 (baik *infit* maupun *outfit*). Adams dan Khoo (1993) menyarankan penggunaan *fit mean square* karena lebih berguna untuk melihat kesesuaian antara model dengan data, dibandingkan uji *t* yang sangat peka dengan jumlah sampel. *Infit mean square* dianggap lebih baik karena berkaitan dengan diskriminasi item, dibandingkan dengan *outfit mean square*, karena nilai *outfit sensitivity* terhadap data yang menyimpang (*outlier*). Selain itu, nilai *infit* lebih kuat dan stabil dibanding nilai *outfit*. Meskipun biasanya nilai *infit mean square* dan *outfit mean square* tidak jauh berbeda.

Kriteria pengujian item *fit* dalam pengujian alat ukur integritas ini adalah nilai *infit mean square*. Menurut Adams dan Khoo (1993), model dikatakan *fit* dengan data jika nilai *infit mean square* berada pada rentang 0,77 dan 1,30. Hasil pengujian item *fit* pada alat ukur integritas memperoleh nilai *infit mean square* sebesar 0.96. Nilai tersebut berada dalam rentang *infit mean square* yang dapat diterima, sehingga dapat disimpulkan bahwa alat ukur integritas memiliki model yang *fit* dengan data.

Pengujian Item Fit

Suatu item dikatakan *fit* jika nilai *infit mean square* berada pada rentang 0,77 dan 1,30 (Adams dan Khoo, 1993). Pengujian *item fit* memberi

informasi mengenai *item-item* mana yang tidak *fit*, sehingga sebaiknya tidak digunakan dalam analisis lebih

lanjut. Berikut hasil pengujian *item fit* terhadap 26 *item* integritas.

Tabel 8. Nilai Infit Mean Square 26 Item

Item	Infit Mean Square	Keterangan
<i>Item 1</i>	0.83	<i>Item Fit</i>
<i>Item 2</i>	0.85	<i>Item Fit</i>
<i>Item 3</i>	0.80	<i>Item Fit</i>
<i>Item 4</i>	0.89	<i>Item Fit</i>
<i>Item 5</i>	1.18	<i>Item Fit</i>
<i>Item 6</i>	0.97	<i>Item Fit</i>
<i>Item 7</i>	0.89	<i>Item Fit</i>
<i>Item 9</i>	0.92	<i>Item Fit</i>
<i>Item 12</i>	1.10	<i>Item Fit</i>
<i>Item 13</i>	1.13	<i>Item Fit</i>
<i>Item 18</i>	1.04	<i>Item Fit</i>
<i>Item 19</i>	0.87	<i>Item Fit</i>
<i>Item 20</i>	0.98	<i>Item Fit</i>
<i>Item 21</i>	1.00	<i>Item Fit</i>
Item 22	0.71	Item Tidak Fit
<i>Item 23</i>	1.02	<i>Item Fit</i>
<i>Item 24</i>	1.23	<i>Item Fit</i>
<i>Item 25</i>	0.81	<i>Item Fit</i>
<i>Item 28</i>	0.81	<i>Item Fit</i>
<i>Item 30</i>	0.88	<i>Item Fit</i>
<i>Item 31</i>	1.05	<i>Item Fit</i>
<i>Item 33</i>	1.00	<i>Item Fit</i>
<i>Item 34</i>	0.96	<i>Item Fit</i>
<i>Item 36</i>	0.88	<i>Item Fit</i>
<i>Item 38</i>	1.03	<i>Item Fit</i>
<i>Item 39</i>	1.15	<i>Item Fit</i>

Berdasarkan tabel diketahui terdapat satu *item* yang tidak *fit* karena berada di luar rentang *infit mean square* yang dapat diterima, yaitu item 22 yang berbunyi “Sebesar apapun resiko yang saya hadapi, saya tidak akan melakukan kecurangan”. Maka dapat disimpulkan bahwa dari 26 *item*, terdapat 25 *item* yang *fit* mengukur integritas.

Pengujian Kembali Person Fit dan Item Fit

Berdasarkan pengujian *item fit* diketahui terdapat satu *item* yang tidak *fit* yaitu *item* 22. Langkah selanjutnya adalah mengeliminasi *item* tersebut dan menguji kembali *person fit* dan *item fit* untuk *item* yang tersisa.

Pengujian Kembali Person Fit

Setelah melakukan pengujian

kembali menguji model *fit* dengan mengeluarkan *item* 22, diperoleh nilai *infit mean square* untuk alat ukur integritas sebesar 0,97. Berdasarkan penjelasan sebelumnya telah diketahui bahwa model dikatakan *fit* apabila memiliki nilai *infit mean square* pada rentang 0,77 dan 1,30 (Adams dan Khoo, 1993). Untuk itu, dapat dikatakan bahwa alat ukur integritas dengan 25 *item* memiliki model yang *fit* dengan data.

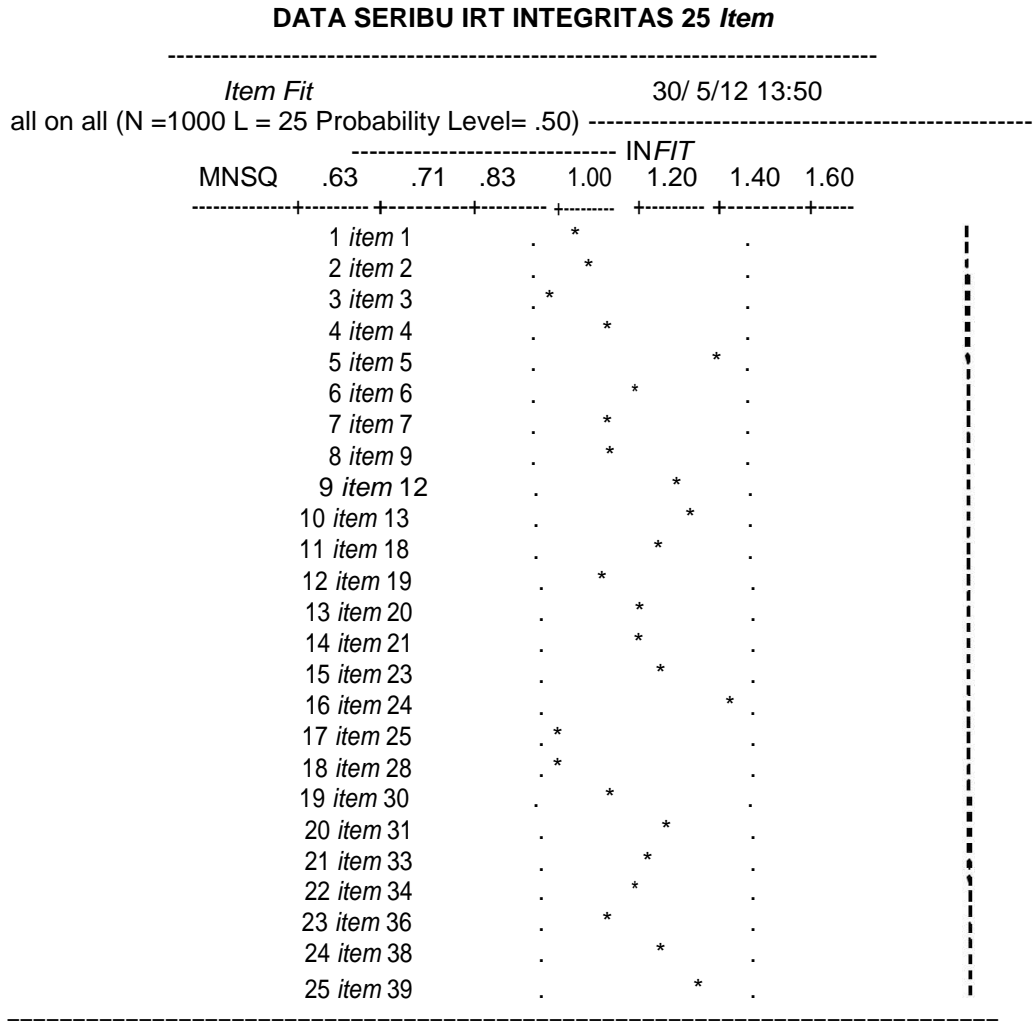
Pengujian Kembali Item Fit

Setelah menguji model *fit* untuk 25 *item*, peneliti melakukan pengujian *item fit* untuk setiap *item* alat ukur integritas. Item dikatakan *fit* jika nilai *infit mean square* berada pada rentang 0,77 dan 1,30 (Adams & Khoo, 1993). Berikut tabel yang menggambarkan pengujian *item fit* terhadap 25 *item* integritas.

Tabel 9. Nilai Infit Mean Square 25 Item

<i>Item</i>	<i>Infit Mean Square</i>	Keterangan
<i>Item</i> 1	0.82	<i>Item Fit</i>
<i>Item</i> 2	0.85	<i>Item Fit</i>
<i>Item</i> 3	0.79	<i>Item Fit</i>
<i>Item</i> 4	0.89	<i>Item Fit</i>
<i>Item</i> 5	1.17	<i>Item Fit</i>
<i>Item</i> 6	0.96	<i>Item Fit</i>
<i>Item</i> 7	0.89	<i>Item Fit</i>
<i>Item</i> 9	0.92	<i>Item Fit</i>
<i>Item</i> 18	1.03	<i>Item Fit</i>
<i>Item</i> 19	0.87	<i>Item Fit</i>
<i>Item</i> 20	0.97	<i>Item Fit</i>
<i>Item</i> 21	0.99	<i>Item Fit</i>
<i>Item</i> 23	1.01	<i>Item Fit</i>
<i>Item</i> 24	1.23	<i>Item Fit</i>
<i>Item</i> 25	0.81	<i>Item Fit</i>
<i>Item</i> 28	0.81	<i>Item Fit</i>
<i>Item</i> 30	0.88	<i>Item Fit</i>
<i>Item</i> 31	1.04	<i>Item Fit</i>
<i>Item</i> 12	1.08	<i>Item Fit</i>
<i>Item</i> 13	1.11	<i>Item Fit</i>
<i>Item</i> 33	0.99	<i>Item Fit</i>
<i>Item</i> 34	0.95	<i>Item Fit</i>
<i>Item</i> 36	0.88	<i>Item Fit</i>
<i>Item</i> 38	1.03	<i>Item Fit</i>
<i>Item</i> 39	1.14	<i>Item Fit</i>

Berdasarkan tabel diketahui disimpulkan bahwa seluruh *item fit* bahwa setelah mengeluarkan *item 22*, dalam mengukur integritas. Berikut seluruh *item* alat ukur integritas berada gambaran plot *infit mean square* dari dalam rentang *infit mean square* yang setiap *item*. dapat diterima sehingga dapat

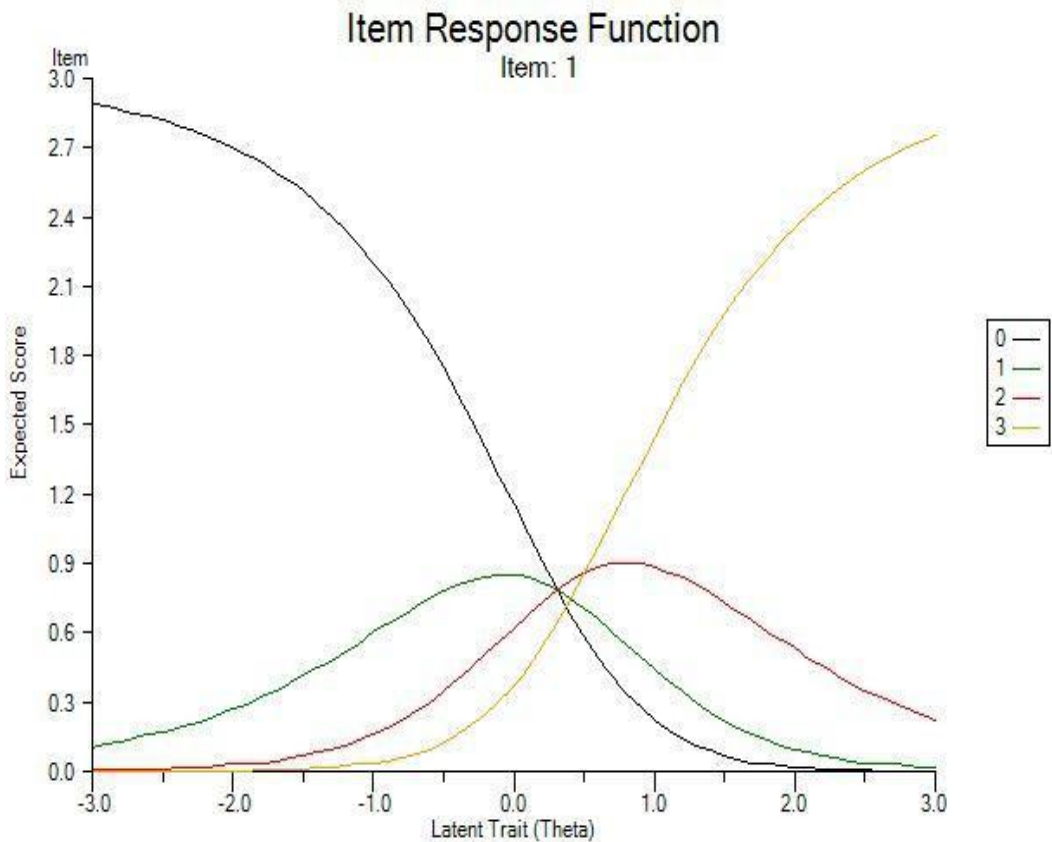


Gambar 3. Plot Item Fit (25 item)

Item Characteristic Curves dan Item Information Function

Item characteristic curves didasarkan pada asumsi bahwa setiap respon yang dipilih oleh penempuh tes mewakili kemampuan - kemampuan yang ia miliki. Dengan demikian diasumsikan bahwa setiap individu yang menempuh tes memiliki nilai *numerik*, *berupa skor*, yang menentukan posisinya dalam skala kemampuan. Skor kemampuan tersebut dilambangkan dengan *theta* (θ).

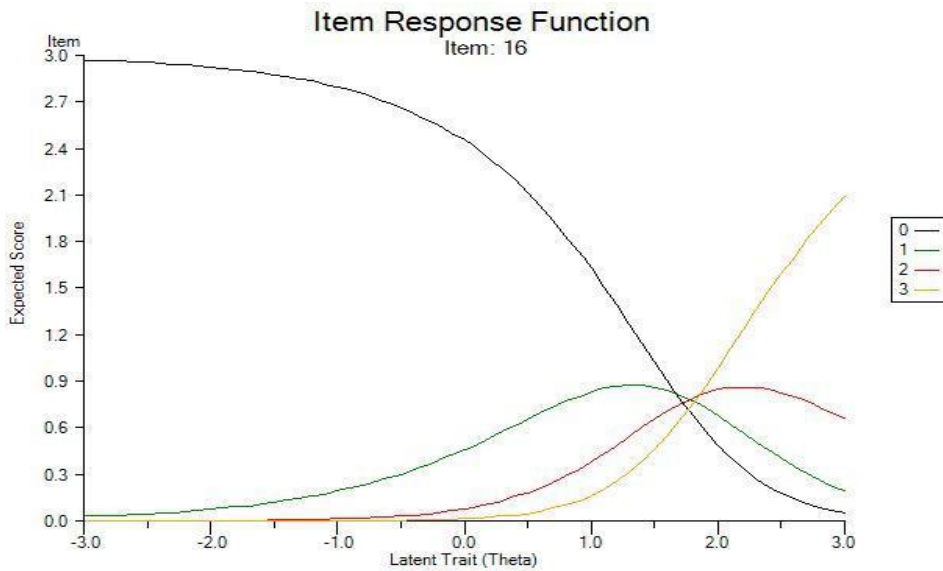
Item-item dalam alat ukur integritas ini memiliki empat kategori respon (*Sangat Setuju (SS) Setuju (S) - Tidak Setuju (TS) - Sangat Tidak Setuju (STS)*) dan empat buah kurva. Kurva pertama adalah tingkat kesukaran dan tiga kurva lainnya adalah *threshold*. *Threshold* merupakan probabilitas untuk tidak memilih. Jika peserta tes memiliki *theta* yang lebih besar atau lebih kecil dari suatu *threshold*, peserta tes tidak memilih opsi yang diwakili *threshold* tersebut.



Gambar 4. Contoh Item Characteristic Curve (Item 1)

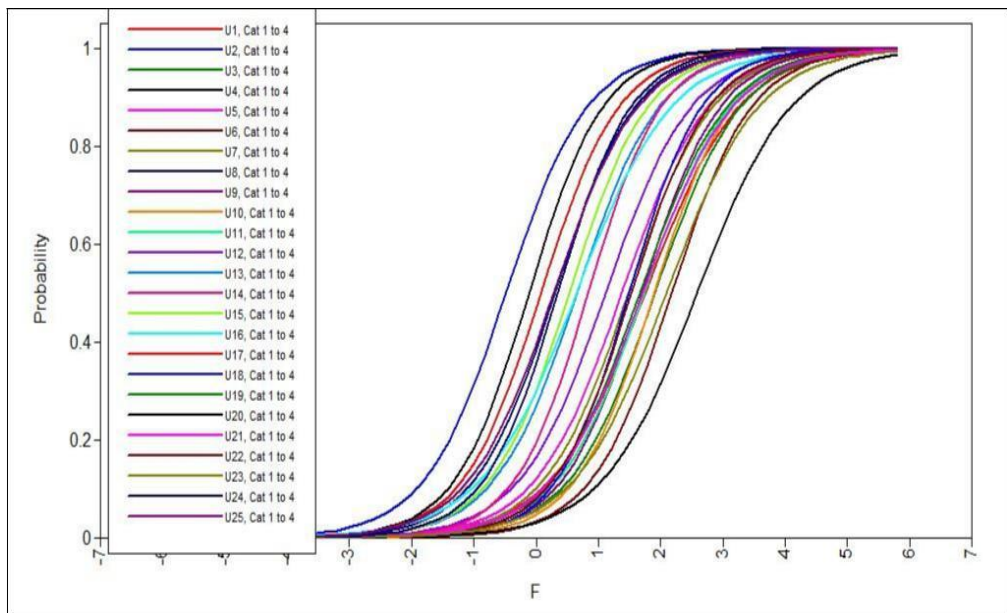
Kurva pertama (berwarna hitam) adalah probabilitas menjawab salah. Semakin rendah θ maka semakin tinggi peluang menjawab salah sehingga kurva turun (S terbalik = *monotonic decreasing*), karena semakin ke kanan maka semakin kecil peluang untuk menjawab STS. Kurva memiliki 3 *threshold*. Jika penempuh tes memiliki θ lebih rendah dari *threshold* pertama, maka opsi yang dipilih adalah STS. Semakin tinggi θ maka semakin rendah kemungkinan memilih opsi STS karena θ kemungkinan berada pada *threshold* berikutnya, sehingga semakin ke kanan semakin rendah peluang memilih STS. *Threshold* kedua adalah probabilitas memilih opsi TS. Kalau θ sudah melewati *threshold* pertama, semakin tinggi θ maka semakin besar peluang menjawab TS. Ketika θ mendekati batas *threshold* kedua (mendekati batas S) opsi yang dipilih menjadi S. Bentuk kurva naik lalu turun. Dalam hal ini peluangnya menjadi lebih kecil (menurun) tetapi tidak serendah kurva *threshold* pertama. Selanjutnya *threshold* ketiga adalah seluruh penempuh tes yang memiliki θ lebih besar dari *threshold* kedua. Semakin tinggi θ maka semakin besar peluang menjawab S (diatas TS), tetapi karena di atasnya masih ada *threshold* ketiga

maka mendekati *threshold* ketiga kurva akan turun lagi. Ketika θ melewati *threshold* ketiga, opsi yang dipilih menjadi SS (kurva naik tetapi turun ketika mendekati *threshold* ketiga). Sedangkan kurva terakhir (berwarna kuning) sama halnya dengan tes pilihan ganda (dikotomi). Penempuh tes yang memiliki θ lebih besar dari *threshold* ketiga, memilih opsi SS. Penempuh tes yang memiliki θ lebih kecil dari *threshold* ketiga mencakup semua penempuh tes yang memilih opsi selain SS (opsi S, TS dan STS) sehingga kurva berbentuk dikotomi (tidak naik dan turun). *Item* dianggap bagus jika kurva *threshold* mulai dari sisi sebelah kanan *threshold* sebelumnya, baik naik maupun turun. Kurva yang naik dari sisi sebelah kanan tapi turunnya tajam memotong di sebelah kiri tergolong kurva *item* yang tidak terlalu bagus walaupun masih memenuhi unidimensionalitas. Jika ada kurva yang terputus, kurva masih bagus tetapi jika informasinya rendah. Kurva yang paling bagus jatuhnya seimbang antara bagian kiri dan kanan. Hal ini berkaitan dengan distribusi θ penempuh tes.



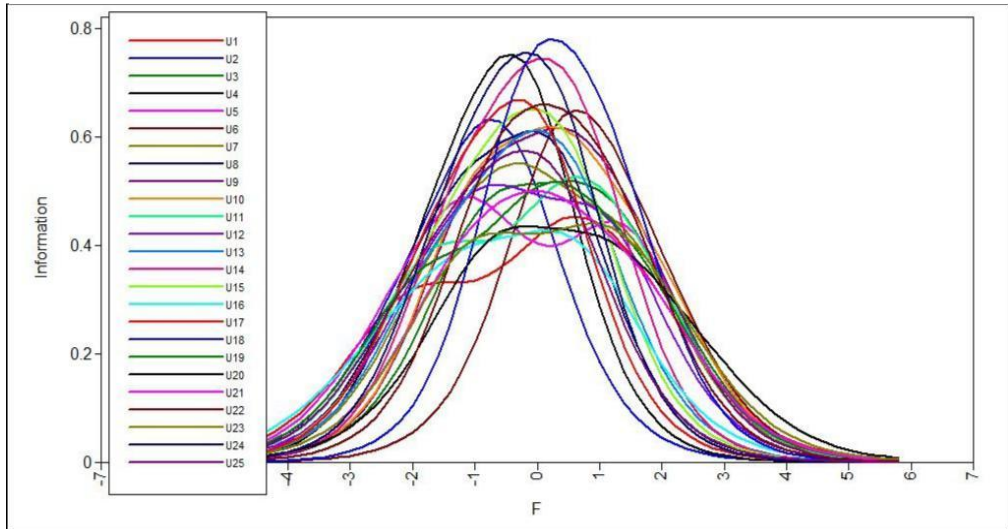
Gambar 5. Contoh Item Characteristic Curve (Item 16)

Kurva yang ditampilkan pada tabel berikut adalah kumpulan treshold ketiga seluruh item.



Gambar 6. Total Item Characteristic Curves

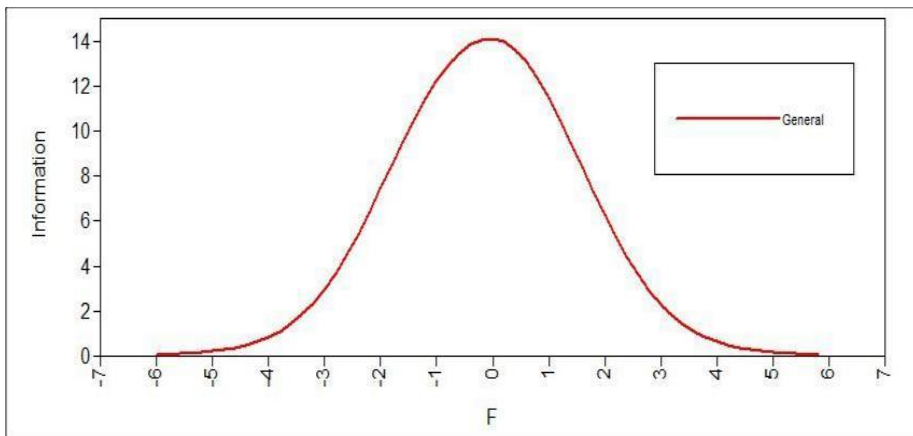
Dari 25 *item* alat ukur integritas yang dirancang dalam penelitian ini, semuanya memiliki kurva yang ideal. Semua naik di sebelah kanan dan turun juga di sebelah kanan kurva sebelumnya.



Gambar 7. Item Information Function

Seluruh kurva *item information function* ini tergolong bagus. Perbedaan bentuk puncak kurva terjadi karena ada ketidaksesuaian antara tingkat kesukaran soal dengan kemampuan individu yang menempuh tes. Ketika banyak kelompok seperti itu yang menempuh sebuah *item*,

puncak kurva menjadi lebih rendah. Hal ini biasa terjadi karena item terlalu mudah atau terlalu susah bagi individu penempuh tes. Secara umum alat ukur integritas ini merupakan alat ukur yang baik berdasarkan kurva *test information function* sebagai berikut.



Gambar 8. Test Information Function

Analisis *Differential Item Functioning* (DIF)

Peneliti melakukan analisis DIF untuk mendeteksi kemungkinan terjadinya bias respon terhadap *item-item* dalam alat ukur integritas. Penelitian ini mendeteksi dan DIF dilakukan menggunakan program QUEST. *Item* dikatakan terdeteksi DIF apabila

memiliki tingkat kecenderungan persetujuan yang berbeda secara signifikan antara kelompok yang diperbandingkan (Adams dan Khoo, 1993). Pengujian DIF terhadap *item-item* alat ukur integritas dalam penelitian ini membandingkan kelompok berdasarkan jenis kelamin. Kelompok berdasarkan jenis kelamin terdiri dari laki-laki dan perempuan.

Tabel 10. Hipotesis Pengujian

Hipotesis	
Ho :	Tidak ada perbedaan kecenderungan persetujuan antar kelompok
Ha :	Ada perbedaan kecenderungan persetujuan antar kelompok

Tabel 11. Perbandingan Estimasi Item Berdasarkan Jenis Kelamin

Item	delta		difference L-P	chi-sq	p	Keterangan
	Laki-laki	Perempuan				
Item1	0.10	0.13	-0.03	0.10	0.75	Tidak DIF
Item2	0.01	0.01	0.00	0.00	0.99	Tidak DIF
Item3	0.11	-0.02	0.13	1.91	0.17	Tidak DIF
Item4	0.76	0.68	0.08	0.78	0.38	Tidak DIF
Item5	-1.38	-1.24	-0.14	2.31	0.13	Tidak DIF
Item6	-0.87	-0.80	-0.07	0.57	0.45	Tidak DIF
Item7	1.09	1.05	0.04	0.18	0.67	Tidak DIF
Item9	0.71	0.75	-0.04	0.19	0.66	Tidak DIF
Item12	-0.47	-0.52	0.05	0.33	0.56	Tidak DIF
Item13	-0.89	-0.98	0.09	1.08	0.30	Tidak DIF
Item18	-0.90	-0.89	0.00	0.00	0.99	Tidak DIF
Item19	0.56	0.47	0.09	1.13	0.29	Tidak DIF
Item20	0.40	0.28	0.13	2.24	0.13	Tidak DIF
Item21	-0.86	-0.86	0.00	0.00	0.96	Tidak DIF
Item23	0.54	0.62	-0.08	1.03	0.31	Tidak DIF
Item24	1.28	1.43	-0.15	3.04	0.08	Tidak DIF
Item25	0.51	0.65	-0.14	2.51	0.11	Tidak DIF
Item28	0.33	0.16	0.17	3.60	0.06	Tidak DIF
Item30	-0.33	-0.37	0.04	0.20	0.65	Tidak DIF
Item31	-0.34	-0.39	0.05	0.36	0.55	Tidak DIF
Item33	-0.14	-0.09	-0.04	0.29	0.59	Tidak DIF
Item34	-0.71	-0.50	-0.20	5.77	0.02	DIF
Item36	-0.87	-0.99	0.11	1.45	0.23	Tidak DIF
Item38	0.57	0.57	0.00	0.00	0.98	Tidak DIF
Item39	0.77	0.85	-0.08	1.15	0.28	Tidak DIF

Statistik dalam pengujian ini adalah *chi-square* dengan hipotesis *null* tidak terdapat perbedaan kecenderungan persetujuan antar kelompok dan hipotesis alternatif terdapat perbedaan kecenderungan persetujuan antar kelompok.

Analisis DIF terhadap kelompok jenis kelamin dilakukan berdasarkan hasil penelitian Ones (1993) yang melaporkan terjadinya kemungkinan bias respon antara laki-laki dan perempuan dalam menjawab *item-item* yang mengukur integritas. Perbandingan estimasi kecenderungan persetujuan untuk kelompok responden laki-laki dan perempuan disajikan pada tabel 11 di atas.

Dari tabel diketahui bahwa item 34 yang memiliki nilai *p-value* < 0.05 (signifikan) terdeteksi DIF, sehingga hipotesis yang menyatakan tidak terdapat perbedaan kecenderungan persetujuan antara kelompok laki-laki dan perempuan ditolak dan hipotesis alternatif diterima. Dengan kata lain, *item* tersebut memiliki kecenderungan persetujuan yang berbeda untuk kelompok perempuan dan laki-laki. Artinya *item* 34 teridentifikasi mengandung bias respon terkait dengan jenis kelamin. *Item* 34 yang teridentifikasi mengandung DIF tersebut, yang berbunyi “*Saya merasa lebih produktif mengerjakan tugas-tugas ketika mendekati deadline*”, memiliki tingkat kecenderungan persetujuan yang lebih tinggi untuk kelompok laki-laki. Dengan kata lain *item* tersebut cenderung disetujui oleh

kelompok laki-laki dibandingkan perempuan.

Pengujian DIF ini memberikan gambaran bahwa *item-item* pada alat ukur integritas pada awalnya masih mengandung DIF sehingga responden dari kelompok jenis kelamin berbeda memiliki kecenderungan memberi respon yang berbeda terhadap *item* dikarenakan atribusi yang melekat pada dirinya. *Item* yang mengandung DIF tersebut dieliminasi dari alat ukur untuk mengurangi bias pada hasil pengukuran.

Deskripsi Karakteristik *Band-Scale* Alat Ukur Integritas

Berdasarkan hasil pengujian *item fit* dan analisis DIF terhadap seluruh *item* pada alat ukur integritas, peneliti memutuskan menghilangkan *item* 34 sehingga total *item* alat ukur integritas menjadi 24 *item*. Dari 24 *item* tersebut, peneliti membuat deskripsi karakteristik setiap *band-scale* alat ukur integritas. Tidak seperti di dalam teori tes klasik dimana skor tes diinterpretasi dengan membandingkan posisinya dalam kelompok normatif, dalam IRT skor tes diinterpretasi berdasarkan *item*. Pembuatan deskripsi karakteristik *band-scale* alat ukur didasarkan pada *domain reference* atau *criterion reference*. Penyekorantes menggunakan *criterion reference* dilakukan setelah melalui proses kalibrasi indeks *threshold* untuk setiap *item*. Kemudian diestimasi pula skala *true score* berbasiskan *item* yang

selanjutnya dibuat rentang *true-score* untuk menjadi skala penafsiran skor. Untuk menafsirkan setiap skor tes, diidentifikasi *item-item* yang mewakili rentang *true-score* tertentu dan dibuat deskripsi tentang karakteristik yang direpresentasikan oleh himpunan *item* pada setiap rentang *true score* tersebut. Dalam

pembuatan deskripsi karakteristik *band-scale* alat ukur integritas ini, peneliti mengalikan setiap rentang *true score* dengan SD sebesar 10 dan menjumlahkannya dengan mean 50. Berdasarkan tahapan atau proses tersebut, peneliti membuat deskripsi karakteristik *band-scale* seperti dalam tabel berikut.

Tabel 12. Potongan Deskripsi Karakteristik Band-Scale Alat Ukur Integritas

<i>Rentang True Score</i>	<i>Scaled Score</i>	Karakteristik
1 s.d. 1.5	55-59	<p>Individu dalam kategori ini sangat menekankan kejujuran tetapi mampu mempertimbangkan cara menyampaikan kejujuran tanpa menyakiti pihak-pihak tertentu.</p> <p>Individu dalam kategori ini, memiliki sisi keteguhan yang didukung kestabilan <i>self</i>. Prinsip-prinsip sudah terinternalisasi sehingga tahan menghadapi pertentangan dan ketidakpercayaan. Individu dalam kategori ini memiliki <i>self-control</i> yang kuat, sehingga mampu dengan cepat mengatasi rasa cemas yang ia hadapi.</p> <p>Individu dalam kategori ini memiliki sisi <i>self-esteem</i> yang stabil. Individu mampu menjalankan prinsip-prinsip moral yang dianut dan memiliki keyakinan bahwa dirinya bermoral tanpa memerlukan pujian atau penghargaan dari pihak luar.</p>

Deskripsi karakteristik *band-scale* alat ukur integritas yang dipaparkan pada tabel, penggunaannya masih terbatas pada tingkat integritas menengah. *Band-scale* belum mampu mendeskripsikan karakteristik-karakteristik integritas yang berada pada tingkat ekstrim, yaitu individu yang sangat

memiliki integritas dan individu yang sama sekali tidak memiliki integritas.

Kesimpulan, Diskusi, dan Saran

Kesimpulan

Penelitian ini merancang alat ukur integritas berdasarkan teori yang dikemukakan oleh Rogers (1961)

karena dianggap lebih mampu menggambarkan integritas sebagai variabel psikologis. Namun masih perlu dilakukan pengujian untuk membuktikan bahwa teori integritas tersebut memiliki konstruk yang valid. Proses ujicoba alat ukur integritas memberikan hasil perhitungan validitas alat ukur serta pengujian unidimensionalitas berdasarkan metode *confirmatory factor analysis* (CFA) yang menegaskan bahwa konstruk alat ukur integritas yang dikemukakan oleh Rogers (1961) terbukti valid.

Penelitian ini juga melakukan pengujian menggunakan metode *first order* dan *second order* CFA untuk menguji asumsi unidimensionalitas. Pengujian *first order* CFA terhadap 40 item alat ukur integritas memberikan hasil yang *fit* pada RMSEA, tetapi *chi-square* tidak *fit*. Hal ini bisa dianggap *fit* mengingat *chi-square* memiliki sensitifitas terhadap jumlah sampel dan normalitas data. Namun umumnya indeks *chi-square* lebih di yakini sehingga peneliti memutuskan kembali melakukan pengujian *first order* CFA dengan mengeliminasi *item-item* yang memberi kontribusi kecil. Eliminasi menghasilkan total 26 item integritas yang akan menjalani kembali pengujian *first order* CFA. 26 *item* tersebut juga melalui pengujian empat faktor *second order* CFA. Hasil pengujian *first order* dan empat faktor *second order* CFA menunjukkan bahwa model yang digunakan *fit* dengan data sehingga dapat dipastikan

bahwa *item-item* dalam alat ukur yang dirancang dalam penelitian ini *fit* untuk mengukur integritas. Hasil pengujian *first order* dan empat faktor *second order* CFA membuktikan bahwa alat ukur integritas yang dirancang dalam penelitian ini memenuhi asumsi unidimensionalitas. Kesimpulannya, hasil pengujian alat ukur integritas asmengunakan pendekatan IRT menunjukkan bahwa model *fit* dengan data.

Terkait model polytomous IRT, yaitu *Rating Scale Model* (RSM), yang peneliti terapkan pada alat ukur integritas dalam penelitian ini, hasil pengujian dari program QUEST menunjukkan bahwa dari 26 *item* alat ukur integritas, terdapat satu *item* (*item* 22) yang tidak *fit* dengan model. Namun secara keseluruhan, rata-rata nilai *infit mean square* yang berada pada $0,77$ dan $1,30$ mengindikasikan bahwa model RSM yang diterapkan *fit* dengan data.

Pengujian differential *item functioning* (DIF) yang dilakukan dengan membuat klasifikasi partisipan berdasarkan jenis kelamin, ternyata mendeteksi DIF pada *item* 34. *Item* yang berbunyi 'Saya merasa lebih produktif mengerjakan tugas-tugas ketika mendekati deadline' tersebut memiliki tingkat kecenderungan persetujuan yang lebih tinggi untuk kelompok laki-laki. Ini berarti *item* tersebut cenderung disetujui oleh kelompok laki-laki dibandingkan perempuan. *Item* yang terdeteksi DIF disinyalir dapat mempengaruhi hasil pengukuran terhadap integritas.

Partisipan dengan level *trait (ability)* yang sama dalam kelompok yang berbeda akan memiliki peluang (probability) yang berbeda untuk memilih kategori tertentu (“Sangat Sesuai”, “Sesuai”, “Tidak Sesuai” atau “Sangat Tidak Sesuai”) pada *item* yang terdeteksi DIF, maka *item* tersebut dieliminasi. Berdasarkan seluruh analisis dan pengujian yang dilakukan terhadap 26 *item* alat ukur integritas yang dirancang dalam penelitian ini, satu *item* (*item* 22) didrop karena tidak *fit* dan satu *item* lain (*item* 34) didrop karena terdeteksi DIF. Maka terdapat total 24 *item* yang digunakan sebagai alat ukur integritas.

Peneliti kemudian merancang deskripsi karakteristik *band-scale* alat ukur integritas berdasarkan *domain referenced*, yaitu melakukan penyekoran tes setelah melalui proses kalibrasi dan estimasi skala *true score* (θ). *Item-item* yang mewakili setiap rentang *true score* diidentifikasi dan dibuat deskripsi karakteristiknya, sehingga dapat disimpulkan bahwa alat ukur yang dikembangkan dalam penelitian ini merupakan alat ukur integritas yang reliabel, valid, memiliki model dan *item-item* yang *fit* serta kontekstual dengan pekerja di Indonesia.

Diskusi

Pengembangan alat ukur integritas dalam penelitian ini melibatkan dua pengujian *first order* CFA dan satu pengujian empat faktor *second order* CFA. Pengujian *first*

order CFA pertama melibatkan 40 *item* integritas namun hasil pengujian hanya memenuhi kriteria *fit* pada RMSEA. Hal ini bisa dianggap *fit*. Namun indeks *chi-square* umumnya lebih diyakini sehingga peneliti memutuskan kembali melakukan pengujian *first order* CFA dengan mengeliminasi *item-item* integritas yang memberi kontribusi kecil. Proses eliminasi menghasilkan total 26 *item* integritas yang akan menjalani kembali pengujian *first order* CFA. 26 *item* integritas tersebut juga mendapat pengujian empat faktor *second order* CFA. Seluruh hasil pengujian kembali menunjukkan bahwa model *closed fit* dengan data serta disusun oleh indikator yang valid untuk mengukur konstruk integritas. Hasil juga menunjukkan bahwa 26 *item* tersebut mengukur satu variabel laten, yaitu integritas. Sehingga alat ukur integritas yang dirancang dalam penelitian ini terbukti mampu memenuhi asumsi unidimensionalitas.

Dari 26 *item* yang digunakan sebagai alat ukur integritas, satu *item* didrop karena tidak *fit* dan satu *item* lain didrop karena terdeteksi DIF. Maka diperoleh total 24 *item* yang dipertahankan dan digunakan sebagai alat ukur integritas. Selanjutnya peneliti mengembangkan deskripsi karakteristik setiap *band-scale* untuk alat ukur integritas berdasarkan *domain referenced*, namun deskripsi karakteristik tersebut perlu dievaluasi lebih lanjut. Dibutuhkan keterlibatan *content specialist* serta psikolog yang

berpengalaman untuk merancang deskripsi kemampuan atau karakteristik yang direpresentasikan oleh himpunan *item* pada setiap *band-scale* agar kemampuan tiap individu yang dites dapat dideskripsikan dengan merujuk kepada *band-scale* tersebut.

Terkait jumlah *item*, tidak ada standar yang menyebutkan batasan jumlah *item* yang dapat dikatakan banyak atau sedikit sehingga sulit mengkategorikan 24 *item* alat ukur integritas yang dirancang dalam penelitian ini panjang atau pendek. Selain itu, dalam pendekatan modern (IRT) dijelaskan bahwa alat ukur yang memiliki sedikit *item* bisa lebih reliabel dibandingkan alat ukur yang memiliki *item* lebih banyak (Embretson dan Reise, 2000). Tetapi ada kemungkinan bahwa *item* dalam alat ukur integritas ini, baik ketika berjumlah 40 *item* maupun 26 *item*, terlalu sedikit sehingga kurang mampu menggali integritas yang dimiliki oleh partisipan. Padahal jumlah 40 *item* tersebut ditetapkan berdasarkan pendapat ahli dan proses elisitasi, karena ketika diberikan alat ukur di atas 40 *item* partisipan memperlihatkan kejenuhan dan tidak berkonsentrasi. Sedangkan jumlah 26 *item* diperoleh berdasarkan hasil pengujian model *fit*.

Langkah selanjutnya yang dilakukan dalam penelitian ini bertujuan untuk mendeteksi bias respon dalam *item-item* alat ukur integritas, peneliti menggunakan analisis DIF dengan metode *Item*

Parameter Comparison. Pengujian terhadap jenis kelamin menemukan satu *item terdeteksi DIF*. *Item yang berbunyi "Saya merasa lebih produktif mengerjakan tugas-tugas ketika mendekati deadline"* tersebut cenderung disetujui oleh kelompok laki-laki dibandingkan perempuan. Berkaitan dengan hal ini, Harter (dalam Peterson dan Seligman, 2004) menjelaskan bahwa laki-laki memiliki kecenderungan lebih terbuka dalam menyuarakan pendapat dibandingkan perempuan. Namun tidak berarti setiap perempuan pasti lebih tertutup menyuarakan pendapatnya dibandingkan laki-laki.

Terkait penggunaan IRT dalam perancangan alat ukur, Hambleton, Swaminathan dan Rogers (1991) menyatakan kemungkinan suatu data *fit* ketika menerapkan model IRT tertentu tetapi tidak *fit* ketika menerapkan model IRT yang lain. Hal ini diistilahkan sebagai *falsifiable model*. Berkaitan dengan penjelasan tersebut, penerapan polytomous IRT menggunakan *rating scale* model (RSM) dalam perancangan alat ukur integritas pada penelitian ini dinilai tepat karena alat ukur memiliki model yang *fit* dengan data.

Poin penting selanjutnya terkait penggunaan IRT dalam perancangan alat ukur adalah rentang *true score* (θ) yang bervariasi. Alat ukur integritas dalam penelitian ini memiliki *item-item* yang mampu mengukur variasi θ dari rentang -1.45 sampai dengan 1.56. Namun variasi rentang *item* lebih kecil

daripada variasi kemampuan partisipan yang berada pada rentang - 3.84 sampai dengan 3.79. Hal ini terjadi karena *item-item* alat ukur integritas dalam penelitian ini disinyalir kurang mampu menggali situasi yang paling memunculkan integritas partisipan. Situasi-situasi tersebut seharusnya dapat tergal melalui proses elisitasi dan *focus group discussion* (FGD) yang dilakukan peneliti. Tetapi hal ini kemungkinan terjadi karena peneliti dapat dikatakan masih kurang familiar dengan konteks pekerja dan perusahaan di Indonesia, meskipun peneliti telah mendiskusikan topik ini dengan beberapa praktisi dan akademisi terkait.

Saran

Tuntutan terhadap alat ukur integritas dapat dipastikan semakin luas. Pengembangan alat diperlukan untuk mendapatkan alat ukur integritas yang berkualitas, khususnya menggunakan pendekatan *item response theory* (IRT). Alat ukur integritas yang dirancang dalam penelitian ini dapat digunakan karena terbukti valid dan *fit* untuk mengukur integritas secara kontekstual terkait kondisi pekerja dan perusahaan di Indonesia. Hasil-hasil pengujian yang telah dipaparkan juga membuktikan bahwa alat ukur integritas yang dikembangkan menggunakan pendekatan modern (IRT) lebih baik dan lebih informatif daripada alat ukur yang dikembangkan menggunakan pendekatan klasik. Tetapi alat ukur

integritas yang dirancang dalam penelitian ini memiliki kelemahan-kelemahan, sehingga penelitian selanjutnya diharapkan melakukan beberapa perbaikan.

Penelitian ini merancang 24 *item* integritas dengan melibatkan 1210 pekerja sebagai partisipan. Dari awal penelitian berupaya melibatkan partisipan yang memiliki latar belakang bervariasi. Namun ternyata rentang item masih lebih kecil daripada rentang kemampuan partisipan. Rentang yang kecil (skor homogen) kemungkinan terjadi karena kelompok partisipan terlalu homogen sehingga varians yang diperoleh kecil. Hal ini menyebabkan analisis apapun yang dilakukan dengan alat ukur ini akan memberi korelasi yang rendah, padahal dasar dari IRT dan Lisrel menggunakan data korelasi. Peneliti selanjutnya sebaiknya melibatkan kelompok sampel yang dipastikan lebih heterogen agar diperoleh varians yang lebih tinggi.

Kelemahan lain dari alat ukur yang dirancang dalam penelitian ini adalah jumlah *item* yang disinyalir masih terlalu sedikit. Penelitian selanjutnya mengembangkan integritas juga disarankan menggunakan *item* dengan jumlah yang lebih banyak serta lebih bervariasi agar lebih mampu mengukur integritas yang dimiliki partisipan serta diperoleh deskripsi karakteristik *band-scale* yang mampu mendekati kriteria-kriteria ekstrim dari integritas.

Alat ukur integritas yang dikonstruksi dalam penelitian ini bersifat kontekstual dan ditujukan khusus bagi kepentingan promosi pekerjaan pada level manajer. Mengingat integritas merupakan karakter positif yang memiliki sifat universal, sangat baik apabila penelitian berikutnya mengembangkan alat ukur yang mampu mengukur integritas secara umum.

Integritas secara sederhana terlihat ketika individu harus membuat keputusan di antara pilihan situasi yang kontras. Gambaran situasi tersebut diperoleh melalui proses elisitasi dan *focus group discussion* (FGD). Proses ini sangat penting dan memiliki pengaruh sangat signifikan terhadap alat ukur yang dirancang. Peneliti yang tertarik mengembangkan alat ukur integritas harus melaksanakan proses elisitasi yang mampu menggali situasi kontras terkait konteks integritas yang diharapkan. Jika memungkinkan FGD sebaiknya melibatkan kelompok-kelompok yang berbeda agar memperoleh masukan yang lebih bervariasi.

Selain itu, alat ukur integritas yang dikonstruksi dalam penelitian ini menggunakan pilihan respon yang meminta partisipan menentukan seberapa pasti kesesuaian atau ketidaksesuaian dirinya dengan pernyataan pada *item* alat ukur. Peneliti lain yang tertarik mengembangkan alat ukur integritas dapat mempertimbangkan

penggunaan pilihan respon yang bersifat netral. Tentu sangat menarik mencermati pengaruh pilihan opsi tertentu terhadap alat ukur integritas yang dikembangkan.

Terakhir, peneliti juga menyarankan agar ada penelitian yang melakukan proses *equating* terhadap item politomi agar dapat diperoleh bank soal untuk *item* integritas. Dengan adanya bank soal *item-item* integritas, tentu dapat dikembangkan alat ukur integritas yang berbentuk *computerized adaptive testing* (CAT). Alat ukur integritas yang terstandarisasi akan sangat bermanfaat secara praktis dan ekonomis.

Daftar Pustaka

- Adams, R.J., & Khoo, S. (1993). *Quest: The interactive test analysis system*. Australia: The Australian Council for Educational Research.
- Barrett, P. (2001). *Pre-employment integrity testing: Current methods, problems and solutions*. Paper presented at: British Computer Society: Information Security Specialist Group March 29th-30th, 2001, Milton Hill, Oxford.
- Barry, C.M., Sackett, P.S., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, 60, 271-301.
- Blasi, A. (2004). Moral functioning: Moral understanding and personality. In D.K. Lapsley & D.

- Narvaez, *Moral development, self and identity* (pp. 335-348). New Jersey: Lawrence Erlbaum Associates.
- Carter, S. (1996). *Integrity*. New York: Basic Books, A Division Of Harper Collins Publishers.
- Deci, E.L., & Ryan, R.M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 4, 227-268.
- Egberink, I.J.L., & Veldkamp, B.P. (2007). The development of a computerized adaptive test for integrity. In D.J. Weiss (Ed.), *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*. Retrieved February, 14th 2012 from www.psych.umn.edu/psylabs/CATCentral/
- Embretson, S.E. & Reise, S.P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Erhard, W., Jensen, M.C. and Zaffron, S. (2011). "Integridad: Un Modelo Positivo Que Incorpora Fenomenos Normativos de Moral, Etica y Legalidad Abreviado (Integrity: A Positive Model that Incorporates the Normative Phenomena of Morality, Ethics, and Legality - Abridged) (March 18, 2011)". *Harvard Business School NOM Unit Working Paper No. 10-061*; *Barbados Group Working Paper No. 10 - 01*; *Simon School Working Paper No. 10-07*. Available at SSRN : <http://ssrn.com/abstract=1756285>.
- Hambleton, R.K., Swaminathan H., & Rogers H.J. (1991). *MMSS fundamentals of item response theory (volume 2)*. California: Sage Publications.
- Harris, W.G. (1987). A components analysis of a pre-employment integrity measure: A replicated study. *Psychological Reports*, 60, 1051-1055.
- Hunter, R. (In preparation). *Criterion-related validity of integrity tests in South Africa*.
- Impelman, K. (2006). *Predicting counter-productive workplace behavior: Item level analysis of an integrity test*. Ann Arbor: Proquest Information and Learning Company.
- Iwao, S. (1997). Consistency orientation and models of social behavior: Is it not time for West to meet East? *Japanese Psychological Research*, 39, 323-332.
- Joreskog, K.G., & D. Sorbom. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International, Inc.
- Khalil, E. (2004). *Integrity, shame and self-rationalization*. Vassar College Department of Economics Working Paper Series 55. Vassar College Department of Economics.

- Marchus, B., & Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal Of Applied Psychology, Vol. 89, No.4*, 647-660.
- Markus, H.R., Kitayama, S., & Heiman, R. (1996). Culture and basic psychological principles. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social Psychology: Handbook of basic principles* (pp. 857-913). New York: Guilford Press.
- Martelli, T. A. (1988, August). *Preemployment screening for honesty: The construct validity, criterion related validity and test retest reliability of a written integrity test*. Unpublished Doctoral Dissertation, Ohio University.
- Miller, M.L., & Schlenker, B.R. (2011). Integrity and identity: Triangulating private and public perceptions of moral identity. *European Journal of Personality, 25*, 2-15.
- Mount, M., Ilies, R., & Johnson, E. (2006). Relationship of personality traits and counterproductive work behaviors: The mediating effects of job satisfaction. *Personnel Psychology, 59*, 591-622.
- Mumford, M.D., Zaccaro, S.J., Harding, F.D., Jacobs, T.O., & Fleishman, E. A. (2000). Leadership skills for a changing world: Solving complex social problems. *Leadership Quarterly, 11*, 11-35.
- Ones, D.S. (1993). *Establishing construct validity for integrity tests*. Doctoral Dissertation, University of Iowa, Iowa City, IA.
- Ones, D.S., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal Of Selection And Assessment, 9*, 31-38.
- Ones, D.S., Viswesvaran, C., & Schmidt, F. (1995). Integrity tests: Overlooked facts, resolved issues and remaining questions. *American Psychologist, 50*, 456-4567.
- Ones, D.S., Viswesvaran, C., & Schmidt, F. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal Of Personality, 17*, S19-S38.
- Palanski, M.E., & Yammarino, F.J. (2007). Integrity and leadership: Clearing the conceptual confusion. *European Management Journal 25*(3), 171-184.
- Papalia, D.E., Olds, W.S., & Feldman. (2004). *Human development (9th ed)*. New York: McGraw Hill.
- Permatasari, M. (2011). *Pengaruh gaya berpikir, integritas dan usia dalam perilaku kerja kontraproduktif*. Tesis. Tidak Diterbitkan. Depok: Fakultas Psikologi Universitas Indonesia.
- Peterson, C., & Seligman, M.E. (2004). Integrity. In *Character*

- strengths and virtues: A handbook and classification* (pp. 249-271). Oxford: Oxford University Press.
- Petrick, J.A., & Quinn, J.F. (2000). The integrity capacity construct and moral progress in business. *Journal Of Business Ethics, 34* (3-4), 331-343.
- Rogers, C.R. (1961). *On becoming a person: A therapist's view of psychotherapy*. Boston: Houghton Mifflin.
- Sackett, P.R., Burris, L.R., & Callahan, C. (1989). Integrity testing for personnel selection: A n u p d a t e . *Personnel Psychology, 42*, 491-529.
- Schlenker, B.R., Miller, M.L., & Johnson, R.M. (2009). Moral identity, integrity and personal responsibility. In D. Narvaez & D.K. Lapsey, *Personality, identity and character* (pp. 316-340). Cambridge: Cambridge University Press.
- Suh, E.M. (2001). *Culture, identity consistency and subjective well-being*. Unpublish Manuscript, University of California - Riverside.
- Umar, Jahja. (2011). *Bahan kuliah psikometri*. Tidak diterbitkan.
- U S C o n g r e s s i o n a l O f f i c e o f Technology Assessment. (1990). *The use of integrity tests for pre-employment screening, OTA-SET-442*. Washington, DC: US Government Printing Office.
- Wanek, J.E., Sackett, P.R., & Ones, D. S . (2 0 0 3) . T o w a r d s a n understanding of integrity test similarities and difference: An item-level analysis of seven tests. *Personnel Psychology, 56*, 873-894.

