

Uji Validitas Konstruk *General Aptitude Test Battery* (GATB) dengan Metode *Confirmatory Factor Analysis* (CFA)

Afifah

Fakultas Psikologi UIN Jakarta

Abstrak

Tujuan dari penelitian ini adalah untuk menguji validitas konstruk dari General Aptitude Test Battery (GATB) yang sudah lama digunakan sebagai salah satu alat tes yang berhubungan dengan jabatan yang berorientasi pada beberapa tes bakat baterai yang mengukur sembilan bakat dalam delapan tes tulis serta empat perangkat tes. Dalam penelitian ini digunakan empat subtes dalam GATB yang mengukur bakat skolastik yaitu computation, three dimensional space, vocabulary, dan arithmetic reasoning dengan jumlah total 175 item. Data yang digunakan merupakan hasil dari rekrutmen sebuah perusahaan. Pelaksanaan tes dilakukan pada tahun 2009 dan ditempuh oleh 3257 orang. Metode analisis faktor yang digunakan dalam penelitian ini adalah Confirmatory Factor Analysis (CFA) dengan program lisrel 8.70. Berdasarkan perhitungan dengan metode CFA dapat disimpulkan bahwa semua subtes fit (sesuai) mengukur model satu faktor, namun untuk subtes three dimensional space serta arithmetic reasoning diperlukan modifikasi model pengukuran untuk dapat memperoleh nilai fit. Hasil pengujian melalui analisis faktor dua tingkat (second order confirmatory factor analysis) menghasilkan bahwa terdapat tiga dari empat subtes GATB yang signifikan dalam mengukur inteligensi umum, yaitu subtes computation, three dimensional space, dan vocabulary.

Kata kunci:

validitas konstruk, *general aptitude test battery*, *confirmatory factor analysis*

Tes-tes inteligensi pada umumnya di desain untuk mengukur *general ability*, tetapi orang merasakan bahwa kemampuan-kemampuan yang terukur oleh tes inteligensi tidak meliputi kemampuan kemampuan atau fungsi-fungsi khusus yang dibutuhkan dalam pekerjaan. Sejak Perang Dunia I para

psikolog mulai membuat tes-tes *aptitude* yang diperuntukan khusus untuk kebutuhan konseling pekerjaan (*vocational counseling*) yang juga dapat melengkapi tes-tes inteligensi umum.

Salah satu alat tes yang biasa dipakai tersebut adalah *General Apti-*

tude Test Battery (GATB). GATB adalah suatu alat tes yang berhubungan dengan jabatan yang berorientasi pada beberapa tes bakat baterai yang mengukur sembilan bakat dalam delapan tes tulis serta empat perangkat tes yaitu *name comparison, computation, three dimensional space, vocabulary, tool match-ing, arithmetic reason, form matching, mark making, place, turn, assemble* dan *disassemble*.

Penelitian ini menggunakan empat subtes dari GATB yaitu *computation, three dimensional space, vocabulary, dan arithmetic reasoning* dengan jumlah total 175 item. Keempat subtes tersebut digunakan karena subtes ini merupakan tes bakat dalam penilaian skolastik. Tes bakat skolastik adalah sebuah tes yang bertujuan untuk mengetahui bakat dan kemampuan seseorang di bidang keilmuan. Tes ini juga dapat mencerminkan tingkat kecerdasan intelektual (IQ) seseorang.

Dalam penelitian ini GATB digunakan karena tes tersebut mengukur kemampuan-kemampuan umum pada individu dan dapat dipakai untuk semua kalangan. GATB merupakan salah satu alat tes bakat dan inteligensi yang telah lama diciptakan dan digunakan dalam pengetesan psikologi. GATB dapat digunakan secara individual, klasikal maupun berkelompok, serta dapat memberikan gambaran atau profil seseorang mengenai kelemahan maupun kekuatan yang dimilikinya berdasarkan berbagai aspek yang terkait dengan fungsi inteligensinya.

Keuntungan dari penggunaan alat tes GATB sebagai salah satu tes untuk mengukur inteligensi adalah pengadministrasian tes lebih mudah, waktu pengerjaan tes yang relatif singkat dapat menghemat waktu dalam pengerjaan. Dalam penghitungan skor IQ juga jauh lebih mudah apabila dibandingkan dengan menggunakan alat tes inteligensi lainnya sehingga tidak perlu lagi menggunakan tenaga profesional yang berdampak pada penghematan biaya tes. Namun, alat tes GATB ini telah lama digunakan karena termasuk salah satu alat tes yang tertua. Belum ada pengujian validitas pada item subtes GATB yang menyebabkan item subtes GATB belum memuaskan. Oleh karena itu, peneliti akan meneliti dengan teori modern dalam rangka menguji validitas konstruk dari GATB. Dikarenakan peneliti belum mempelajari secara khusus statistik tentang analisis faktor, maka peneliti hanya akan mempraktekkannya saja dengan *software* yang sudah ada, yaitu Lisrel kemudian menafsirkan hasil analisis faktor terhadap data hasil tes.

Gambaran Umum GATB

General Aptitude Test Battery (GATB) diciptakan oleh Charles E. Odell dari *United States Employes Services* dan mulai dikembangkan pada tahun 1940. *General Aptitude Test Battery* (GATB) telah digunakan sejak 1947 oleh *State Employment service* yang bergabung dengan *United States Employment Service*

untuk memenuhi kebutuhan tes yang bisa dipergunakan untuk berbagai tujuan (*multipurposes*). Sejak masa itu, GATB telah dimasukkan ke dalam program penelitian yang berkelanjutan untuk menjadikannya tes yang akurat terhadap kesuksesan pada berbagai pekerjaan yang berbeda. Karena dasar risetnya yang luas, GATB dikenal sebagai sejumlah tes bakat ganda yang akurat dalam pengadaanya untuk digunakan dalam bimbingan jurusan dan menilai kecerdasan umum dari seseorang.

Dewasa ini GATB meliputi 12 tes; 4 tes membutuhkan alat sederhana, sementara 8 tes yang lainnya hanya menggunakan kertas dan pensil. Keseluruhan kumpulan tes dapat diselenggarakan dalam waktu kurang lebih 2,5 jam (Anastasi, 1997). Namun dalam penelitian ini, GATB dijadikan *short form* dengan empat subtes yang memiliki jumlah item sebanyak 175, dikarenakan ke-empat subtes tersebut mengukur bakat skolastik atau kognisi dari seseorang dan dari hasil penilaian tes tersebut dapat menghasilkan skor IQ. Keempat subtes tersebut adalah (Jigau, 2007):

1. Subtes kemampuan numeric (*computation*). Waktu yang ditentukan 6 menit, skor maksimal 50.

Tes ini berisi suatu latihan bilangan aritmatik yang membutuhkan penjumlahan, pengurangan, perkalian, atau pembagian seluruh bilangan. Subtes ini mengukur kemampuan numerical (*numerical aptitude*)

2. Subtes tiga dimensi ruang (*three dimensional space*). Waktu 6 menit, skor maksimal 40.

Tes ini berisi empat rangkaian gambar dengan objek tiga dimensi. Gambar perangsang disajikan sebagai suatu potongan-potongan metal yang datar dimana salah satu berputar-putar (berguling-guling) atau keduanya. Garis-garis mengindikasikan di mana salah satu dari empat gambar dengan objek tiga dimensi bisa dibentuk dari gambar perangsang. Subtes ini mengukur kemampuan pengetahuan umum dan kemampuan ruang (*general learning ability dan spatial aptitude*).

3. Subtest kemampuan verbal (*Vocabulary*). Waktu 6 menit, skor maksimal 60.

Tes ini berisi empat set kata-kata. Peserta tes menunjukkan dua kata-kata yang sama atau arti kata yang berlawanan. Subtes ini mengukur kemampuan pengetahuan umum dan kemampuan verbal (*general learning ability dan verbal ability*).

4. Subtes penalaran aritmatik (*arithmetic reason*). Waktu 7 menit, skor maksimal 25.

Tes ini berisi masalah angka aritmatik yang diekspresikan secara verbal. Subtes ini mengukur kemampuan pengetahuan umum dan kemampuan numerical (*general learning ability & numerical aptitude*).

Metode

Subjek Penelitian

Untuk menguji validitas tersebut digunakan pendekatan uji validitas konstruk yang akan menentukan apakah setiap subtes dalam GATB mengukur komponen yang dapat mengukur *general intelligence*. Dalam penelitian ini peneliti menggunakan data mentah yang disediakan oleh lembaga asesmen dalam rekrutmen sebuah perusahaan. Pelaksanaan tes dilakukan pada tahun 2009 dan ditempuh oleh 3257 orang.

Analisis Data

Metode analisis yang digunakan dalam penelitian ini adalah metode analisis faktor. Pada dasarnya terdapat dua jenis pandangan mengenai analisis faktor, yaitu: *Exploratory Factor Analysis* (EFA) dan *Confirmatory Factor Analysis* (CFA). Analisis faktor pada awalnya dikemukakan oleh Spearman (1940) yang pada saat ini dikenal dengan *Exploratory Factor Analysis* (EFA). Dalam EFA, peneliti tidak memiliki ekspektasi tertentu mengenai jumlah atau sifat faktor yang mendasari konstruk. Metode analisis faktor yang lebih modern adalah *Confirmatory Factor Analysis* (CFA). Dalam CFA, peneliti harus memiliki gambaran yang spesifik mengenai (a) jumlah faktor, (b) variabel yang mencerminkan suatu faktor, dan (c) faktor-faktor yang saling berkorelasi (Thompson, 2004).

Dalam rangka penelitian mengenai studi validitas konstruk *General Aptitude Test Battery* (GATB) maka penulis menggunakan metode *Confirmatory Factor Analysis* (CFA) dengan program lisrel 8.70. Alasan penulis menggunakan CFA sebagai metode dalam penelitian karena dengan menggunakan CFA maka setiap dimensi dapat diuji satu persatu. Validitas dari masing-masing item juga dapat diuji dan digambarkan dalam matriks korelasi CFA.

Adapun 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 ini dilakukan melalui analisis terhadap respon (jawaban) atas item-itemnya.
2. Setiap item diteorikan hanya mengukur atau memberi informasi tentang satu faktor tertentu saja. Artinya baik item maupun subtes bersifat unidimensional.
3. Berdasarkan teori yang dipaparkan di atas, dapat disusun sehimpunan persamaan matematis. Persamaan tersebut dapat digunakan untuk memprediksi (dengan menggunakan data yang tersedia) matriks korelasi antar-item yang seharusnya akan diperoleh jika teori tersebut benar (unidimensional). Matriks korelasi ini dinamakan sigma (Σ).

Kemudian matriks ini akan dibandingkan dengan matriks korelasi yang diperoleh secara empiris dari data (disebut matriks S). Jika teori tersebut benar (unidimensional), maka seharusnya tidak ada perbedaan yang signifikan antara elemen matriks Σ dengan elemen matriks S. Secara matematis dapat dituliskan: $S - \Sigma = 0$.

4. Pernyataan matematik inilah yang dijadikan hipotesis nihil yang akan dianalisis menggunakan CFA. Dalam hal ini dilakukan uji signifikansi dengan Chi Square. Jika Chi Square yang dihasilkan tidak signifikan (nilai $p > 0,05$), maka dapat disimpulkan, bahwa hipotesis nihil yang menyatakan: "tidak ada perbedaan antara matriks S dan Σ " tidak ditolak. Artinya teori yang menyatakan bahwa semua item mengukur hal yang sama, dapat diterima kebenarannya (didukung oleh data). Sebaliknya, jika nilai Chi Square yang diperoleh signifikan, maka hipotesis nihil $S - \Sigma = 0$ ditolak. Artinya teori tersebut tidak didukung data (ditolak).
5. Jika teori diterima (model fit), langkah selanjutnya, adalah menguji hipotesis tentang signifikan tidaknya masing-masing item dalam mengukur apa yang hendak diukur (kemampuan ber-pikir analogis). Uji hipotesis ini dilakukan dengan *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 yang

mana yang valid dan yang tidak valid di dalam konteks validitas konstruk. Dengan kata lain, analisis faktor konfirmatori dalam hal ini adalah pengujian terhadap hipotesis nihil (H_0): $S - \Sigma = 0$. Artinya, tidak ada perbedaan antara matriks korelasi yang diharapkan oleh teori dengan matriks korelasi yang diperoleh dari hasil observasi.

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

$$X_1 = \lambda_1 F + \delta_1$$

Dimana:

- X_1 = Skor yang diperoleh pada item no.1
- = Konstruk yang hendak diukur (faktor)
- λ_1 = Koefisien Regresi untuk item no.1 dalam mengukur F, disebut juga sebagai "koefisien muatan faktor"
- δ_1 = Segala hal yang mempengaruhi Varians X_1 (selain F), disebut juga "kesalahan pengukuran".

Persamaan-persamaan inilah yang akan dijadikan dasar untuk membuat persamaan-persamaan untuk setiap elemen dari matrik Σ , sehingga setiap elemen tersebut dapat diperoleh nilai *estimate*-nya untuk kemudian dibandingkan dengan setiap elemen yang sama dalam matrik S. Kesesuaian

model dengan data, seperti disebutkan pada butir-4 di atas, diuji secara statistik, yaitu dengan test Chi Square.

Hasil

Validitas Konstruk Subtes *Computation*

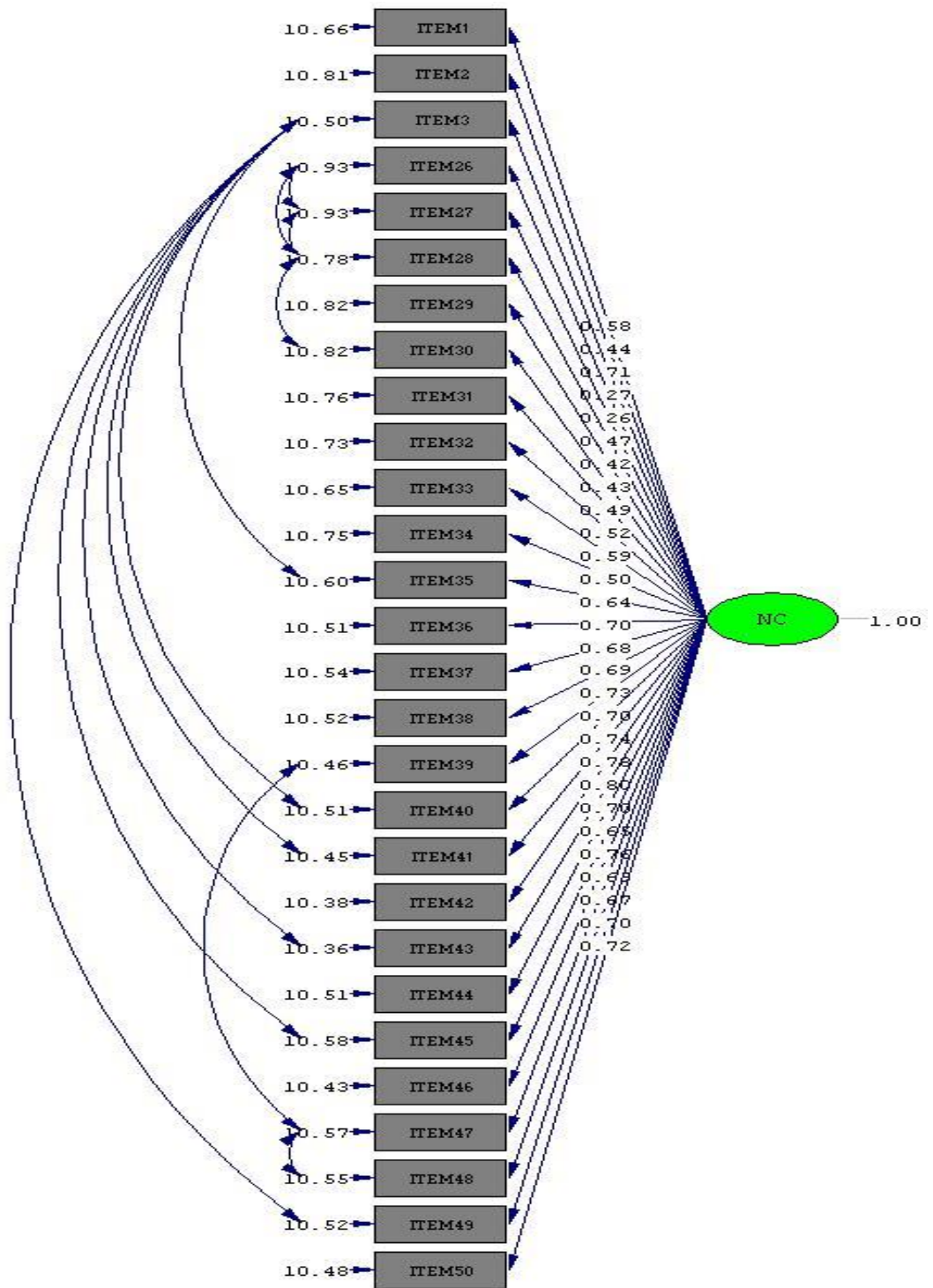
Hasil perhitungan awal untuk subtes *computation* dengan model satu faktor (unidimensional) tidak fit, karena didapatkan Chi - Square = 1557.36, df = 1175, P-value = 0.00000 RMSEA = 0.010. Setelah dilakukanlah modifikasi terhadap model dengan cara membebaskan atau memperbolehkan kesalahan pada setiap item saling berkorelasi satu dengan lainnya, maka didapatkanlah model fit dengan $P > 0,05$ (tidak signifikan). Dengan demikian model dengan hanya satu faktor dapat diterima, yang berarti bahwa seluruh item terbukti mengukur satu hal saja, yaitu *computation*. Namun karena pada model ini ada beberapa item yang kesalahannya saling berkorelasi, dapat disimpulkan bahwa beberapa item tersebut sebenarnya bersifat multidimensional.

Pada awalnya df berjumlah 1175, namun setelah mencapai model fit, df tersisa menjadi 1152. Ini berarti dalam perhitungan awal terdapat 1175-1152 = 23 korelasi antar error yang dibebaskan. Sesuai dengan teori yang telah dikemukakan sebelumnya, apabila ada item yang kesalahan pengukurannya saling berkorelasi

maka item tersebut bersifat multidimensio-nal. Artinya, selain mengukur apa yang hendak diukur oleh subtes yang bersangkutan, ada hal yang lain yang diukur oleh item tersebut. Semakin banyak kesalahan pada sebuah item saling berkorelasi dengan kesalahan pengukuran pada item lainnya, maka item tersebut menjadi rendah atau tidak ideal kualitasnya.

Dari 50 item yang mengukur subtes *computation*, terdapat 22 item yang tidak signifikan, karena nilai t lebih kecil dari 1,96 (*absolute*) dan bernilai negatif. Kedua puluh dua item inilah yang harus di-*drop*, yaitu item nomor 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, dan 25.

Untuk mendapatkan hasil dengan item yang benar-benar murni, maka dilakukan analisis ulang setelah seluruh item yang tidak signifikan dan bernilai negatif dibuang. Hasil yang diperoleh untuk perhitungan kedua pada subtes *computation* dengan model satu faktor (unidimensional) dan item negatif yang sudah di *drop* adalah tidak fit, didapatkan Chi Square = 515.28, df = 350, P-value = 0.00000 RMSEA = 0.012. Kemudian dilakukan mo-difikasi kembali terhadap model dengan cara membebas-kan atau memperbolehkan kesalahan pada setiap item saling berkorelasi satu dengan lainnya, maka didapatkanlah model fit seperti pada gambar 1 berikut ini:



Chi-Square=374.86, df=338, P-value=0.08157, RMSEA=0.006

Gambar 1. Analisis Faktor Konfirmatorik untuk Subtes *Computation*

Tabel 1. Muatan Faktor Item GATB subtes *Computation*

Item	Koefisien	Standar Error	T-Value	Sig
1	0.58	0.08	7.20	V
2	0.44	0.08	5.41	V
3	0.71	0.08	7.91	V
26	0.27	0.08	3.31	V
27	0.26	0.08	3.27	V
28	0.47	0.08	5.75	V
29	0.42	0.08	5.25	V
30	0.43	0.08	5.29	V
31	0.49	0.08	6.14	V
32	0.52	0.08	6.44	V
33	0.59	0.08	7.37	V
34	0.50	0.08	6.19	V
35	0.64	0.08	7.72	V
36	0.70	0.08	8.73	V
37	0.68	0.08	8.47	V
38	0.69	0.08	8.57	V
39	0.73	0.08	8.98	V
40	0.70	0.08	8.51	V
41	0.74	0.08	9.03	V
42	0.78	0.08	9.75	V
43	0.80	0.08	9.67	V
44	0.70	0.08	8.72	V
45	0.65	0.08	7.90	V
46	0.76	0.08	9.40	V
47	0.65	0.08	7.89	V
48	0.67	0.08	8.23	V
49	0.70	0.08	8.46	V
50	0.72	0.08	8.94	V

Keterangan: V= Signifikan (t-values > 1,96);
X = Tidak Signifikan

Terlihat dari gambar 1, bahwa nilai Chi Square menghasilkan $p > 0,05$ (tidak signifikan). Dengan demikian, model dengan hanya satu faktor dapat diterima, yang berarti bahwa seluruh item terbukti mengukur satu hal saja, yaitu *computation*. Pada model ini, terdapat 12 pasang korelasi antar-item yaitu item 3 yang berkorelasi dengan

item 35, 40, 41, 43, 45, dan 49. Item 27 yang berkorelasi dengan item 26. Item 28 berkorelasi dengan item 26 dan 27. Item 30 berkorelasi dengan item 28, item 47 yang berkorelasi dengan item 39 dan item 48 yang berkorelasi dengan item 47.

Selanjutnya kualitas item dapat dilihat dari signifikan atau tidaknya item tersebut menghasilkan informasi tentang apa yang hendak diukur. Dalam hal ini, yang diuji adalah hipotesis nihil dari muatan faktor. Pengujiannya dilakukan dengan melihat nilai t bagi setiap koefisien muatan faktor seperti pada tabel 1 berikut ini.

Dari 28 item yang mengukur subtes *computation*, semua item dinyatakan signifikan karena nilai $t > 1,96$ (*absolute*) dan bernilai positif. Item yang paling baik sesuai urutannya: 42, 43, 46, 41, 39, 50, 36, 44, 38, 40, 37, 49, 48, 3, 45, 47, 35, 33, 1, 32, 34, 31, 28, 2, 30, 29, 26, dan 27.

Validitas Konstruk Subtes *Three Dimensional Space*

Hasil perhitungan awal yang di-peroleh untuk subtes *three dimensional space* dengan model satu faktor (unidimensional) tidak fit, didapatkan Chi Square = 71620.74, df = 740, P-value = 0.00000 RMSEA = 0.172. Setelah dilakukan modifikasi terhadap model dengan cara membebaskan atau memperbolehkan kesalahan pada setiap item saling berkorelasi satu dengan lainnya, maka didapatkanlah model fit dengan $P > 0,05$ (tidak signifikan) yang memiliki 490 korelasi yang dibebaskan.

Tabel 2. Muatan Faktor Item GATB subtes *Three Dimensional Space*

Ite m	Koefi sien	Standar Error	T- Value	Sig
1	0.37	0.02	19.84	V
2	0.48	0.02	27.39	V
3	0.62	0.02	34.72	V
4	0.55	0.02	30.72	V
5	0.59	0.02	33.93	V
6	0.22	0.02	11.35	V
7	0.55	0.02	30.86	V
8	0.59	0.02	34.28	V
9	0.15	0.02	7.97	V
10	0.52	0.02	27.83	V
11	0.49	0.02	27.72	V
12	0.70	0.02	41.70	V
13	0.31	0.02	16.72	V
14	0.62	0.02	36.72	V
15	0.56	0.02	31.83	V
16	0.48	0.02	26.57	V
17	0.53	0.02	28.27	V
18	0.36	0.02	19.77	V
19	0.60	0.02	35.32	V
20	0.51	0.02	28.66	V
21	0.51	0.02	28.92	V
22	0.18	0.02	9.48	V
23	0.45	0.02	25.16	V
24	0.29	0.02	15.28	V
25	0.33	0.02	18.05	V
26	0.38	0.02	20.64	V
27	0.23	0.02	11.94	V
28	0.11	0.02	5.69	V
29	0.10	0.02	5.18	V
30	0.24	0.02	12.28	V
31	0.11	0.02	5.93	V
33	0.29	0.02	15.13	V
39	0.14	0.02	7.29	V

Keterangan: V = signifikan (t-values > 1,96) ; X = Tidak signifikan

Dari 40 item yang mengukur subtes *three dimensional space*, terdapat tujuh item yang tidak signifikan, karena nilai t lebih kecil dari 1, 96 (*absolute*) dan signifikan namun

bernilai negatif. Ke-tujuh item inilah yang harus di-*drop*, yaitu item nomor 32, 34, 35, 36, 37, 38, dan 40.

Selanjutnya dilakukan analisis kembali dengan cara membuang semua item negatif dan item yang tidak signifikan pada perhitungan. Hasil yang diperoleh untuk perhitungan kedua pada subtes *three dimensional space* dengan model satu faktor (uni-dimensional) dan item negatif yang sudah di *drop* tidak fit, didapatkan Chi Square = 25586.12, df = 495, P-value = 0.00000 RMSEA = 0.125. Kemudian dilakukan modifikasi kembali terhadap model dengan cara membebaskan atau memperbolehkan kesalahan pada setiap item saling berkolerasi satu dengan lainnya yang akhirnya didapatkanlah model fit dengan Chi Square = 247.29, df = 214, P-Value = 0.05887, RMSEA = 0.007.

Dari 33 item yang mengukur subtes *three dimensional space*, semua item dinyatakan signifikan karena nilai t lebih besar dari 1,96 (*absolute*) dan semua item dinyatakan bernilai positif. Item yang paling baik sesuai urutannya: 12, 14, 19, 3, 8, 5, 15, 7, 4, 2, 20, 17, 10, 11, 2, 16, 23, 26, 1, 18, 25, 13, 24, 33, 30, 27, 6, 22, 9, 39, 31, 28, dan 29.

Validitas Konstruk Subtes *Vocabulary*

Hasil perhitungan awal yang diperoleh untuk subtes *Vocabulary* dengan model satu faktor (uni-dimensional) tidak fit, didapatkan Chi Square = 2163.52, df = 1710, P-value

= 0.00000 RMSEA = 0.009. Setelah dilakukan modifikasi terhadap model dengan cara membebaskan atau memperbolehkan kesalahan pada setiap item saling berkorelasi satu dengan lainnya, maka didapatkan model fit dengan $P > 0,05$ (tidak signifikan) yang memiliki 19 korelasi antar-error yang dibebaskan.

Tabel 3. Muatan Faktor Item GATB subtes *Vocabulary*

Item	Koefisien	Standar Error	T-Value	Sig
1	0.19	0.07	2.66	V
2	0.24	0.07	3.33	V
3	0.28	0.07	3.95	V
4	0.17	0.07	2.40	V
5	0.21	0.07	3.00	V
6	0.19	0.07	2.63	V
7	0.22	0.07	3.14	V
8	0.27	0.07	3.71	V
10	0.17	0.07	2.33	V
11	0.17	0.07	2.43	V
12	0.12	0.07	1.75	X
13	0.06	0.07	0.85	X
14	0.12	0.07	1.65	X
15	0.24	0.07	3.29	V
16	0.03	0.07	0.41	X
17	0.33	0.07	4.56	V
18	0.19	0.07	2.58	V
19	0.24	0.07	3.39	V
20	0.14	0.07	1.95	X
21	0.27	0.07	3.76	V
22	0.28	0.07	3.91	V
23	0.36	0.07	5.01	V
24	0.26	0.07	3.64	V
25	0.51	0.07	7.11	V
26	0.29	0.07	4.08	V
27	0.34	0.07	4.81	V
28	0.46	0.07	6.41	V
29	0.62	0.07	8.67	V
30	0.60	0.07	8.42	V
31	0.62	0.07	8.71	V

Lanjutan

Tabel 3. Muatan Faktor Item GATB subtes *Vocabulary*

Item	Koefisien	Standar Error	T-Value	Sig
32	0.73	0.07	10.34	V
33	0.66	0.07	9.33	V
34	0.58	0.07	8.09	V
35	0.67	0.07	9.41	V
36	0.88	0.07	12.41	V
37	0.87	0.07	12.26	V
38	0.74	0.07	10.42	V
39	0.80	0.07	11.36	V
40	0.89	0.07	12.65	V
41	0.85	0.07	12.09	V
42	0.73	0.07	10.25	V
43	0.88	0.07	12.49	V
44	0.87	0.07	12.27	V
45	0.88	0.07	12.48	V
46	0.85	0.07	12.01	V
47	0.84	0.07	11.84	V
48	0.83	0.07	11.76	V
49	0.87	0.07	12.36	V
50	0.84	0.07	11.96	V
51	0.77	0.07	10.91	V
52	0.78	0.07	11.00	V
53	0.70	0.07	9.93	V
54	0.74	0.07	10.47	V
55	0.70	0.07	9.93	V
56	0.66	0.07	9.31	V
57	0.70	0.07	9.80	V
58	0.62	0.07	8.72	V
59	0.61	0.07	8.53	V
60	0.64	0.07	8.96	V

Keterangan: V= Signifikan (t-values > 1,96);
X = Tidak Signifikan

Dari hasil tersebut, ternyata didapatkan hanya ada satu item yang baik serta bernilai positif dari 60 item yang ada dalam subtes *vocabulary* yaitu item nomor 9 dan ke 59 item sisa yang ada bernilai negatif.

Dikarenakan hasil dari analisis awal yang hanya menghasilkan satu item dengan nilai yang positif, diprediksi subtes ini memiliki kesalahan dalam penyusunan skalanya, maka dilakukan pengecekan kembali terhadap kunci jawaban. Setelah melakukan pengecekan kunci jawaban dan tidak terjadi kesalahan di dalamnya, 1 item yang bernilai positif tersebut dinyatakan berlawanan arah dan dilakukan analisis kembali dengan cara membuang item yang bernilai positif yaitu item nomor 9.

Hasil yang diperoleh untuk subtes *vocabulary* dengan model satu faktor (unidimensional) tidak fit, didapatkan Chi Square = 1804.71, df = 1652, P-value = 0.00479 RMSEA = 0.005. Kemudian dilakukan modifikasi kembali terhadap model dengan cara membebaskan atau memperbolehkan kesalahan pada setiap item saling berkolerasi satu dengan lainnya, maka didapatkanlah model fit dengan Chi Square = 1714.36, df = 1650, P-Value = 0.13186, RMSEA = 0.003.

Dari 59 item yang mengukur subtes *vocabulary*, ternyata ada lima item yang tidak signifikan, karena nilai t lebih kecil dari 1,96 (*absolute*). Kelima item inilah yang harus di *drop*, yaitu item 12, 13, 14, 16, dan 20. Dari hasil tersebut semua item dinyatakan bernilai positif, item yang paling baik sesuai dengan urutannya adalah item nomor 40, 43, 45, 36, 49, 37, 44, 41, 46, 50, 47, 48, 39, 52, 51, 54, 38, 32, 42, 53, 55, 57, 35, 33, 56, 60, 58, 31, 29, 59, 30, 34, 25, 28, 23, 27, 17, 26, 3, 22, 21, 8, 24, 19, 2, 15, 7, 5, 1, 6, 18, 11, 4, dan 10.

Validitas Konstruk Subtes Arithmetic Reasoning (AR)

Dari hasil analisis kedua untuk subtes *arithmetic reasoning* dengan model satu faktor (unidimensional) dan item negatif yang sudah di-*drop* tidak fit, didapatkan Chi Square = 6100.55, df = 252, P-value = 0.00000 RMSEA = 0.084. Kemudian dilakukan modifikasi kembali terhadap model dengan cara membebaskan atau memperbolehkan kesalahan pada setiap item saling berkolerasi satu dengan lainnya.

Tabel 4. Muatan Faktor Item GATB subtes *Arithmetic Reasoning*

Item	Koefisien	Standar Error	T-Value	Sig
1	0.28	0.03	8.71	V
2	0.50	0.03	15.39	V
3	0.49	0.03	14.57	V
4	0.57	0.03	16.65	V
6	0.35	0.03	9.31	V
7	0.59	0.03	18.59	V
8	0.48	0.03	14.38	V
10	0.52	0.03	16.08	V
11	0.53	0.03	16.37	V
12	0.41	0.03	12.98	V
13	0.50	0.03	15.27	V
14	0.18	0.03	5.17	V
15	0.36	0.03	10.96	V
17	0.31	0.03	9.16	V
18	0.37	0.03	10.81	V
19	0.25	0.03	7.54	V
20	0.18	0.03	5.45	V
21	0.27	0.03	8.14	V
24	0.17	0.03	4.75	V
25	0.19	0.03	5.21	V

Keterangan: V= signifikan (t-values > 1,96)
X = Tidak signifikan

Pada model dengan item negatif yang sudah di-*drop* ternyata masih terdapat item yang berkorelasi dengan kesalahan pengukuran pada item lainnya. Hal ini menandakan bahwa beberapa item tersebut sebenarnya memang bersifat multidimensional. Setelah melihat kualitas item dengan menguji hipotesis nihil dari muatan faktor maka didapatkan bahwa dari 24 item yang mengukur subtes *arithmetic reasoning* (AR), ternyata ada empat item yang tidak signifikan, karena nilai t lebih kecil dari 1,96 (*absolute*). Ke-empat item inilah yang harus di *drop*, yaitu item nomor 9, 16, 22, dan 23.

Selanjutnya, bahkan pada hasil yang diperoleh untuk analisis ketiga pada subtes *arithmetic reasoning* dengan model satu faktor (unidimensional) dan item negatif yang sudah di-*drop* tidak fit, didapatkan Chi Square = 4289.60, $df=170$, P-Value = 0.00000 RMSEA=0.086. Kemudian dilakukan modifikasi kembali terhadap model dengan cara membebaskan atau memperbolehkan kesalahan pada setiap item saling berkorelasi satu dengan lainnya, maka didapatkanlah model fit, Chi Square = 116.05, $df = 95$, P-value = 0.07015 RMSEA = 0.008.

Dari 20 item yang mengukur subtes *arithmetic reasoning* (AR), semua item dinyatakan signifikan, karena nilai t lebih besar dari 1,96 (*absolute*) dan juga bernilai positif. Item yang paling baik sesuai dengan urutannya adalah item nomor 7, 4, 11, 10, 2, 13, 3, 8, 12, 15, 18, 6, 17, 1, 21, 19, 20, 25, 14, dan 24.

Validitas Konstruk Seluruh Subtes GATB dalam Mengukur Satu Kontruk Bersifat Umum (*General Intelligence*)

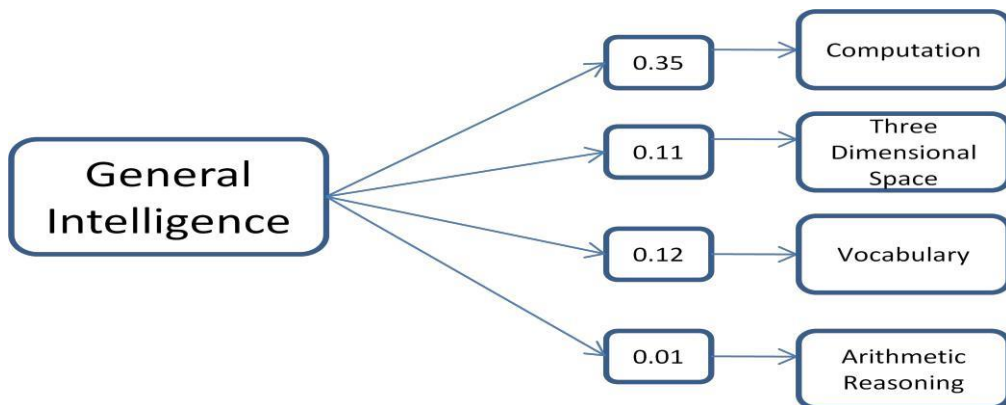
Langkah selanjutnya yang harus dilakukan setelah menganalisis faktor konfirmatorik dari setiap subtes adalah melakukan analisis faktor konfirmatori untuk konstruk umum yang hendak diteliti yaitu inteligensi. Dalam konteks ini, matriks korelasi antar-faktor (subtes) digunakan sebagai input. Namun apabila menggunakan Lisrel, kedua tingkatan analisis faktor ini dapat dilakukan secara simultan (satu kali analisis). Analisis seperti ini disebut juga dengan nama “analisis faktor konfirmatorik orde kedua” (*second order factor analysis*). Dalam hal ini item merupakan indikator dari masing-masing subtes (faktor tingkat satu) dan pada saat yang sama subtes merupakan indikator dari faktor tingkat kedua (inteligensi/*general factor*). Peneliti menggunakan metode ini, karena lebih efisien (hanya satu kali analisis secara simultan) dan dari sudut statistik, analisis seperti ini lebih terpercaya (Joreskog, dan Sorbom, 2006).

Hasil yang diperoleh untuk model dengan *second order* faktor analisis, nilai Chi - Square = 5692.46, $df = 8906$, P-value = 1.00000 atau $P>0,05$, dan RMSEA = 0.000. Dapat disimpulkan bahwa model dengan dua tingkatan faktor (*second order* CFA) fit dengan data. Artinya, teori yang mengatakan bahwa item-item mengukur empat subtes dan keempat

subtes mengukur inteligensi umum pada tabel 5 dan gambar 2. dapat diterima. Hasilnya dapat dilihat

Tabel 5. Koefisien Muatan Faktor Untuk *General Intelligence*

Subtes	Koefisien	Standar Error	T-Value	Signifikan
<i>Computation</i>	0.35	0.06	6.20	V
<i>Three Dimensional Space</i>	0.11	0.03	3.70	V
<i>Vocabulary</i>	0.12	0.05	2.23	V
<i>Arithmetic Reasoning</i>	0.01	0.07	0.17	X



Gambar 2. Koefisien Muatan Faktor Untuk *General Intelligence*

Dari tabel dan gambar di atas ditemukan bahwa, subtes yang muatan faktornya signifikan dalam mengukur inteligensi adalah *computation*, *three dimensional space* dan *vocabulary*. Sedangkan yang tidak signifikan adalah *arithmetic reasoning*. Berdasarkan hasil ini, terdapat kemungkinan bahwa pengukuran terhadap inteligensi umum cukup de-

ngan menggunakan subtes yang signifikan saja, yaitu subtes *computation*, *three dimensional space* dan *vocabulary*.

Kesimpulan, Diskusi dan Saran

Hasil penelitian menunjukkan bahwa semua subtes dalam GATB fit (sesuai) dengan model satu faktor,

yaitu mengukur hanya satu kemampuan yang didefinisikan pada subtes tersebut. Dari empat subtes dalam GATB terdapat dua subtes yang untuk mencapai model fit hanya memerlukan modifikasi yang singkat, namun dua subtes lainnya untuk mencapai model fit harus dilakukan beberapa kali modifikasi yang lebih mendalam karena banyaknya item yang saling berkorelasi.

Terdapat dua subtes yang dalam mencapai model fit hanya memerlukan modifikasi yang singkat karena memiliki item yang baik setelah item-item yang tidak signifikan dan bernilai negatif dibuang. Subtes tersebut adalah *computation* dan *vocabulary*, keduanya dianggap bernilai baik karena tidak memiliki korelasi antar-item yang terlalu banyak sehingga setiap item terbukti memang mengukur konstruk yang dimaksud. Kedua subtes tersebut memiliki karakteristik yang baik dikarenakan memiliki jenis item yang setara dan mengukur satu jenis kemampuan saja sehingga mudah dipahami dalam pengerjaannya.

Subtes yang banyak melakukan modifikasi untuk mencapai model fit adalah *three dimensional space* dan *arithmetic reasoning*. Jenis item dalam subtes ini memiliki karakteristik yang tidak baik karena memiliki terlalu banyak multidimensionalitas, dalam *arithmetic reasoning* hal ini dikarenakan dalam setiap item selain mengukur kemampuan numerik namun juga memerlukan kemampuan penalaran verbal. Item mengukur kemampuan numerik, namun persoal-

an dalam setiap item dibalut dengan soal berbentuk cerita yang terstruktur. Apabila subjek tidak mengerti perintah dalam setiap item pada subtes *arithmetic reasoning*, maka subjek tidak akan mampu untuk menjawab persoalan.

Setelah melakukan analisis faktor terhadap empat subtes dalam GATB yang mengukur inteligensi, menunjukkan bahwa alat tes GATB masih dapat dan layak digunakan sebagai salah satu alat tes inteligensi namun perlu dilakukan perbaikan dan pembaharuan terhadap item-item yang memiliki multidimensionalitas yang terlalu banyak.

Dari hasil pengujian menggunakan CFA menunjukkan bahwa terdapat banyak korelasi antar *measurement error* pada setiap item di subtes GATB. Hal ini menunjukkan bahwa banyak item dalam tes GATB yang selain mengukur apa yang hendak diukur, ternyata juga mengukur hal yang lain (multidimensional).

Melihat hasil analisis yang menyatakan bahwa terdapat subtes yang bersifat multidimensional karena ternyata secara teoritis, tiga subtes dari empat subtes yang diteliti dalam GATB memang mengukur lebih dari 1 faktor. Subtes yang murni mengukur 1 faktor hanyalah *computation* yaitu mengukur kemampuan numerik, sedangkan untuk subtes *three dimensional space* mengukur kemampuan *spatial* dan *general learning ability*, subtes *vocabulary* selain mengukur kemampuan verbal ternyata juga mengukur *general learning ability* dan subtes *arithmetic reasoning* yang

mengukur kemampuan numerik juga *general learning ability*.

Berdasarkan kesimpulan dan diskusi maka dapat disarankan, bahwa:

1. Untuk pengembangan uji validitas kedepannya, dapat menggunakan seluruh subtes dalam GATB pada pengujian validitas konstruk. Dengan menggunakan seluruh subtes dalam GATB maka akan didapatkan konstruk mengenai bakat yang dimiliki seseorang, jadi tidak terbatas hanya inteligensi saja.
2. Terdapat banyak item yang bersifat multidimensional. Hal ini dapat disebabkan oleh kerangka berpikir teori GATB dimana dijelaskan setiap subtes sebenarnya masih terdiri dari beberapa sub faktor, yang seharusnya dapat diwujudkan dalam bentuk faktor tersendiri yang berbeda tingkatan (analisis faktor tiga tingkat). Jadi, akan lebih baik, bila analisis faktor dilakukan 3 tingkat (*third order CFA*). Namun demikian, perlu dilakukan identifikasi terlebih dahulu untuk melihat item yang mengukur sub faktor di dalam masing-masing subtes tersebut.
3. Selanjutnya, bagi institusi psikologi pengguna alat tes GATB, cukup menggunakan 3 subtes saja dalam GATB yaitu *computation, three dimensional space*, dan *vocabulary* untuk mengukur inteligensi. Namun perlu ada penelitian lebih lanjut untuk

membuktikan mengenai skor inteligensi yang dihasilkan.

Daftar Pustaka

- Aiken, Lewis. R. (1997). *Psychological testing and assessment*. Boston: Allyn and Bacon.
- Anastasi, Anne. (1997). *Psychological testing, seventh edition*. New Jersey: Prentice-Hall, Inc.
- Azwar, Saifuddin. (2004). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guildford Press. Buckingham, Marcus and Donald O.
- Clifton. (2001). *Now, discover your strengths*. New York: The Free Press.
- Chaplin, JP. (2006). *Kamus lengkap psikologi*. Jakarta: PT Raja Grafindo Pustaka.
- Coyle, Daniel. (2009). *Talent code*. New York: Bantam Dell.
- Davis, Tony. (2009). *Talent assessment*. Jakarta: PPM.
- Djaali, H & Pudji Muljono. (2007). *Pengukuran dalam bidang psikologi*. Jakarta: Gramedia Pustaka
- Gardner, Howard. (1985). *Frames of mind, the theory of multiple intelligences*. New York: Basic Book, Inc.
- Greogry, Robert J. (2007). *Psychological testing: istory, principles, and application 5th Edition*. New Jersey: Pearson Education, Inc.

- Jigau, Mihai. (2007). *Career counseling, compendium of methods and techniques*. Bucharest : AFIR.
- Jöreskog, K.G. & Sörbom, D. (2006). *LISREL 8.70 for windows (computer software)*. Lincoln-wood, IL: Scientific Software International, Inc.
- Kaplan, Robert M. & Dennis P. Saccuzzo. (2009). *Psychological testing: Principles, application & issues 7th edition*. Belmont: Wadsworth.
- Kerlinger, F.N. (2006). *Asas-asas penelitian behavioral*. Yogyakarta: Gadjah Mada University Press.
- Loewenthal, Kate Miriam. (1996). *An introduction to psychological test and scales*. London: UCL Press.
- Munandar, Ashar Sunyoto. (2001). *Psikologi industri dan organisasi*. Jakarta: Universitas Indonesia.
- Murphy, Kevin R. (1994). *Psychological testing: Principles and application*. London: Prentice-Hall Int.
- Thompson, Bruce. (2004). *Explanatory and confirmatory factor analysis*. Washington DC: American Psychological Association.
- Sadli, Sapparinah. (1991). *Inteligensi bakat dan test IQ*. Jakarta: Gaya Favorit Pres.
- Shaleh, A.R & Wahab, M.A. (2004). *Psikologi suatu pengantar*, Jakarta: Prenada Media.
- Sukardi, Dewa Ketut. (2009). *Analisis tes psikologi teori dan praktik*. Jakarta: Rineka Cipta.
- Umar, Jahja. (2011). *Bahan kuliah psikometri*. Tidak diterbitkan.
- Van Ornum, William. (2008). *Psychological testing across the life span*. New Jersey: Pearson Education, Inc.

