

PENGARUH BANYAKNYA PESERTA TES, BUTIR, PILIHAN JAWABAN, SERTA INDEKS KESULITAN TERHADAP STATISTIK DAYA PEMBEDA DAN RELIABILITAS

(A “Simulated” Data Driven Research)

ALI HASMY

*Pasca Sarjana IAIN Pontianak
HP. 081391856700, E-mail: ali_hasmy@yahoo.com*

ABSTRAK

This research was conducted to see the effect of the number of testee (sample size), the number of items (test length), the number of options, and index of difficulty to various item discrimination statistics and test reliability. The data used are simulated data and analyzed using the Test Analysis Program (TAP) version 6.65 with a full factorial design. In general, the results show that the number of testee, the number of items, and index of difficulty (except number of options) significantly affect the various item discrimination statistics and test reliability. The statistics are robust to these three factors is only the Mean of Item Discrimination and Spearman Brown's 1-2 Split-Half, while the most sensitive is Split-Half Odd-Even statistic.

1. Pendahuluan

Dalam pengembangan tes yang berbasiskan *Classical Test Theory* (CTT) dengan model,

$$X = T + E,$$

ada beberapa hal yang penting untuk diperhatikan antara lain seperti reliabilitas tes, indeks kesulitan, dan indeks daya pembeda item (butir soal).

Ketiga hal tersebut semestinya terkait dengan banyaknya item yang digunakan (panjang tes), maupun banyaknya *option* jawaban yang diberikan. Selain itu, hasil analisis yang dilakukan terkait dengan ketiga hal dimaksud juga, secara statistik, terkait dengan banyaknya *testee*. Dengan demikian ada beberapa

faktor yang harus diperhatikan yaitu banyaknya *testee* (N), banyaknya item (n), dan banyaknya *option*.

Meskipun diperkirakan ketiga faktor tersebut mempengaruhi daya pembeda dan reliabilitas, namun apakah pengaruhnya signifikan? Jika ya, apakah pengaruh faktor tersebut bersifat individual atau bersifat interaktif? Banyak hal yang dapat dipertanyakan dan sekaligus dikaji dalam kaitannya dengan hal ini, antara lain yaitu:

- a. Apakah banyaknya *testee*, item, *option*, maupun tingkat kesulitan secara individual berpengaruh secara signifikan terhadap besaran statistik untuk daya pembeda dan reliabilitas tes?
- b. Apakah banyaknya *testee*, item, *option*, maupun tingkat kesulitan secara bersama-sama (interaktif) berpengaruh

secara signifikan terhadap besaran statistik untuk daya pembeda dan reliabilitas tes?

- c. Bagaimanakah model yang cocok untuk dipergunakan memprediksi besaran statistik daya pembeda dan reliabilitas tes?
- d. Statistik mana saja yang bersifat kokoh (*robust*) dan yang sensitif terhadap perubahan pada banyaknya *testee*, item, *option*, maupun tingkat kesulitan?

Penelitian ini diharapkan dapat memberikan manfaat antara lain:

- a. Hasilnya dapat digunakan dalam rangka pengembangan tes, berkenaan dengan banyak item, banyak *option*, banyak *testee*, dan indeks kesulitan tertentu yang berhubungan dengan daya pembeda dan reliabilitas yang diharapkan.
- b. Menambah khasanah keilmuan di bidang pengukuran pendidikan yang berbasiskan CTT.
- c. Bagi peneliti, penelitian ini bermanfaat dalam rangka mengaplikasikan teori-teori statistika khususnya *Measurement Theory* dalam pengukuran pendidikan.

Karena beragamnya statistik yang terkait, maka pada penelitian ini dibatasi pada statistik-statistik sebagai berikut:

- a. *Mean of Item Difficulty*
- b. *Mean of Item Discrimination*
- c. *Mean of Biserial Correlation*
- d. *Mean of Point Biserial*
- e. *Mean of Adjusted Point Biserial*
- f. Koefisien KR20
- g. *Standard Error of Measurement* KR20
- h. Koefisien KR21
- i. Koefisien *Split-Half First-Second*
- j. Koefisien *Split-Half First-Second* Spearman Brown
- k. Koefisien *Split-Half Odd-Even*
- l. Koefisien *Split-Half Odd-Even* Spearman Brown.

Data yang digunakan dibatasi pada data hasil simulasi, bukan data lapangan, sedangkan analisisnya tentunya menggunakan CTT dan bukan *Item Respon Theory* (IRT).

2. Metode

Penelitian ini merupakan "*data driven research*" yaitu penelitian yang di-drive oleh data, sebagai komplemen dari "*theory driven research*" yaitu penelitian yang di-drive oleh teori.

Jika *theory driven research* cenderung pada *theory-then-research* yang bersifat *confirmatory* yang menguji teori, maka pada *data driven research* cenderung pada *research-then-theory* yang bersifat *exploratory* untuk menemukan teori.

Data yang digunakan pada penelitian ini adalah bukan data riil yang didapat dari lapangan (*field research*), tetapi merupakan data hasil simulasi, yaitu data yang di-generate dengan menggunakan "*Test Analysis Program*" (TAP) versi 6.65.

Data hasil generate (pembangkitan) dimaksud merupakan *metric data* yang memiliki *level of measurement* interval/rasio (*scale*).

Simulasi untuk meng-generate data ditentukan (*fixed*) variabelnya (*factor*) yaitu:

- a. Banyaknya *testee* sesuai dengan pendapat Hambleton & Cook (1983) serta Segall (2000) tetapi dalam hal ini aplikasinya pada CTT bukan IRT dengan *level* didasarkan pada kepentingan praktis di STAIN Pontianak yaitu:
 - 1) Kelas (45 *testee*)
 - 2) Angkatan (180 *testee*)
 - 3) Seleksi masuk mahasiswa baru (500 *testee*).

Namun hal ini mirip dengan yang dilakukan oleh Jiao & Kamata (2003) yang menggunakan 250 dan 500 *testee*.

- b. Banyaknya item (*Test Length*) sesuai dengan pendapat Mislevy & Bock (1990) yang menyatakan bahwa tes pendek dengan ukuran 11 – 20 *items* serta tes yang panjang dengan ukuran > 20 *items*. Tetapi dalam penelitian ini *level*-nya ditentukan sebagai berikut:

- 1) Paket A (20 *items*)
- 2) Paket B (40 *items*), dan
- 3) Paket C (60 *items*).

- c. Banyaknya *option* yang *level*-nya ditentukan untuk 2 jenis tes yaitu:
- 1) *True-False* (2 *options*)
 - 2) *Multiple Choice* (4 *options*).
- d. *Difficulty Index* dengan *level*-nya ditentukan sebagai berikut:
- 1) *Extremely Easiest* (95%) yang dalam konteks distribusi normal mirip dengan yang dilakukan oleh Jiao & Kamata dengan menggunakan nilai mulai -2,00.
 - 2) *Extremely Difficult/Hardest* (40%).
 - 3) *Moderate* (70%) sebagai titik tengah kategorinya.

Sedangkan untuk replikasinya ditentukan 5, meskipun Harwel, dkk (1996) menyatakan bahwa banyaknya replikasi dapat mempengaruhi statistik yang dihasilkan.

Karena *level-level* variabelnya (*independent variable*) ditentukan oleh peneliti tanpa randomisasi, maka dalam konteks *General Linear Model – Multivariate*, mereka akan ditempatkan sebagai *fixed-factors*.

Data hasil *generate* dianalisis dengan *software* TAP 6.65 dengan. berdasarkan *Classical Test Theory*.

Statistik-statistik hasil analisis dengan TAP beserta variabel yang diteliti pengaruhnya (banyaknya *testee*, panjang tes, banyaknya *option*, serta indeks kesulitan) di-entry sebagai data untuk dianalisis dengan *General Linear Model – Multivariate* beserta pemeriksaan asumsi *Homogeneity of Variance Error* beserta *Lack of Fit* (uji kesesuaian model)-nya dengan menggunakan *software* PASW Statistics 18 (dulu bernama SPSS). Adapun model yang digunakan adalah *full model* yang mencakup *main effects* dan *interaction effects* sebagai berikut:

$$\begin{aligned}
 &J_TESTEE + J_ITEM + DIF_INDEX + J_OPTION + \\
 &J_TESTEE * J_ITEM + J_TESTEE * DIF_INDEX + \\
 &J_ITEM * DIF_INDEX + \\
 &J_TESTEE * J_ITEM * DIF_INDEX + \\
 &J_TESTEE * J_OPTION + J_ITEM * J_OPTION + \\
 &J_TESTEE * J_ITEM * J_OPTION + \\
 &DIF_INDEX * J_OPTION + \\
 &J_TESTEE * DIF_INDEX * J_OPTION + \\
 &J_ITEM * DIF_INDEX * J_OPTION + \\
 &J_TESTEE * J_ITEM * DIF_INDEX * J_OPTION
 \end{aligned}$$

Model yang digunakan tanpa melibatkan *intercept*, $E(Y)$ bila $X = 0$, karena bila banyaknya *testee* = 0 maka $E(Y) = 0$, begitu pula bila banyaknya *item* = 0, maupun banyaknya *option* = 0.

Untuk keringkasan, maka analisis *descriptive* menggunakan tampilan dari *output* Lisrel v.8 begitu pula dengan *Multivariate Normality Test* yang tidak tersedia pengujiannya di PASW Statistics 18.

3. Hasil-hasil

3.1. Pemeriksaan Data

Sebelum dilakukan analisis data lebih lanjut, terlebih dahulu perlu dilakukan pemeriksaan terhadap data yang akan dianalisis. Hasilnya dapat dilihat pada Tabel 1 *output* PASW Statistics 18 di halaman dberikut ini.

Berdasarkan tabel di atas dapat dilihat bahwa:

- a. Banyaknya data untuk *testee* (*examinee*) adalah 90 untuk *level* kelas (45 mahasiswa), 90 untuk *level* angkatan (4 kelas @ 45 mahasiswa = 180 mahasiswa), dan 90 untuk *level* seleksi calon mahasiswa baru (500 peserta), sehingga totalnya = 270.
- b. Banyaknya data untuk item paket tes A (20 item) adalah 90, begitu juga untuk paket tes B (40 item) dan paket tes tes C (60 item), sehingga keseluruhannya juga sebanyak 270.
- c. Banyaknya data untuk indeks kesulitan item (*difficulty index*) adalah 90 untuk indeks kesulitan 40% (*hardest*), 90 untuk indeks kesulitan 70% (*moderate*), dan 90 untuk tingkat kesulitan 95% (*easiest*), sehingga keseluruhannya adalah sebanyak 270.
- d. Banyaknya data untuk 2 *option* jawaban (*true-false*) adalah 135 dan untuk 4 *option* jawaban (*multiple choice*) juga 135, sehingga total banyaknya juga = 270.

Tabel 1. *Between-Subjects Factors*

		Value Label	N
B_TESTEE	45	KELAS	90
	180	ANGKATAN	90
	500	SELEKSI	90
B_ITEM	20	PAKET A	90
	40	PAKET B	90
	60	PAKET C	90
DIF_INDEX	40	HARDEST	90
	70	MODERATE	90
	95	EASIEST	90
B_OPTION	2	T-F	135
	4	M-CHOICE	135

Keempat hal di atas menunjukkan bahwa:

- Keseluruhan data telah di-entry dengan lengkap.
- Tidak ada data yang *missing*, sehingga tidak perlu dilakukan penanganan dan *analysis of missing data*.
- Data memenuhi syarat untuk dianalisis lebih lanjut.

Hasil Analisis Deskriptif

Hasil analisis data dengan menggunakan LISREL v.8 dapat dilihat di Lampiran 1 (hasil analisis dengan PASW Statistics 18 tidak ditampilkan karena tabelnya memuat *output* yang sangat rinci, melibatkan hasil analisis deskriptif tidak hanya untuk tiap faktor tetapi juga untuk tiap kombinasi interaksinya pada tiap-tiap sel, sehingga sangat banyak memakan tempat).

Hasil *output* deskriptif di atas memperlihatkan antara lain untuk,

- Banyaknya *testee*:

Mean-nya 241,667 didapat dari $((90 \times 45 + 90 \times 180 + 90 \times 500) / 270)$ atau karena *weight*-nya sama (90) cukup dihitung dengan cara $(45 + 180 + 500) / 3$. Dengan koefisien *Skewness* < 2,00 dan Kurtosis < 7,00 maka secara univariat dapat dikatakan sebaran data berdasarkan banyaknya *testee* adalah normal (Currant, West & Finch, 1996).

- Banyaknya item:

Mean-nya adalah 40 didapat dari $(20 + 40 + 60) / 3$.

Dengan koefisien *Skewness* < 2,00 dan Kurtosis < 7,00 maka secara univariat dapat dikatakan sebaran data berdasarkan banyaknya item adalah normal.

- Difficulty Index*:

Mean-nya 68,333 didapat dari $(40 + 70 + 95) / 3$.

Dengan koefisien *Skewness* < 2,00 dan Kurtosis < 7,00 maka secara univariat dapat dikatakan juga bahwa sebaran data berdasarkan *Difficulty Index* (istilah *Index* merujuk pada item - individual, sedangkan istilah *Coefficient* merujuk pada tes - aggregate) adalah normal.

- Banyaknya *option*:

Mean-nya adalah 3 didapat dari $(2 + 4) / 2$.

Dengan koefisien *Skewness* < 2,00 dan Kurtosis < 7,00 maka secara univariat dapat dikatakan sebaran data berdasarkan banyaknya *option* adalah normal.

- Mean of Item Difficulty*:

Mean dari *Mean*-nya 0,691 (69,1%) yang ternyata mendekati 70% (*moderate*) didapat dari $((90 \times 45 + 90 \times 180 + 90 \times 500) / 270)$ dan dengan koefisien *Skewness* > 3,00 dan Kurtosis > 21,00 maka secara univariat dapat dikatakan sebaran datanya adalah tidak normal (Currant, West & Finch, 1996).

- Mean of Item Discrimination*:

Mean dari *Mean*-nya adalah 2,037 dan dengan koefisien *Skewness* > 3,00 dan Kurtosis > 21,00 maka secara univariat dapat dikatakan sebaran datanya adalah tidak normal.

- Mean of Biserial Correlation*:

Mean-nya adalah 0,488 > 0,3 sehingga dalam konteks *Discrimination Index* dengan perbandingan antar *testee* dengan perbandingan 2 kelompok)

- termasuk kategori memenuhi syarat, karena *Determination Index*-nya mendekati 10%.
- Dengan koefisien *Skewness* $< 2,00$ dan Kurtosis $< 7,00$ maka secara univariat dapat dikatakan sebaran data berdasarkan *Biserial Correlation* adalah normal.
- h. *Mean of Point Biserial*:
Mean-nya 0,325 sehingga dalam konteks *Discrimination Index* termasuk kategori memenuhi syarat. Selain itu, dengan koefisien *Skewness* $< 2,00$ dan Kurtosis $< 7,00$ maka secara univariat dapat dikatakan sebaran data berdasarkan *Mean of Point Biserial* adalah normal.
 - i. *Mean of Adjusted Point Biserial*:
Mean-nya 0,236 sehingga dalam konteks *Discrimination Index* termasuk kategori tidak memenuhi syarat. Penurunan koefisiennya terjadi karena item yang dianalisis dihubungkan dengan total yang tidak termasuk nilai item yang bersang-kutan.
 Namun koefisien *Skewness* $< 2,00$ dan Kurtosis $< 7,00$ maka secara univariat dapat dikatakan sebarannya adalah normal.
 - j. Koefisien KR20 (Alpha untuk *Dichotomous Score*)
Mean-nya 0,698 lebih besar dari *standard* reliabilitas untuk item dalam pengembangan yaitu 0,6 sehingga dapat dinyatakan reliabilitas relatif cukup memenuhi syarat.
 Dengan koefisien *Skewness* $< 2,00$ dan Kurtosis $< 7,00$ maka secara univariat juga dapat dikatakan sebarannya adalah normal.
 - k. *Standard Error of Measurement* KR20
Mean-nya 2,151 dan dengan koefisien *Skewness* $< 2,00$ dan Kurtosis $< 7,00$ maka secara univariat dapat dikatakan bahwa sebarannya adalah normal.
 - l. Koefisien KR21
Mean-nya 0,673 lebih besar dari *standard* reliabilitas untuk item dalam pengembangan yaitu 0,6 sehingga dapat dinyatakan reliabilitasnya memenuhi syarat.
 - m. Koefisien *Split-Half First-Second*
Mean-nya 0,575 yang $< 0,6$ sehingga dapat dinyatakan bahwa reliabilitasnya berdasarkan korelasi separoh item bagian pertama dengan separoh bagian kedua tidak memenuhi syarat.
 Namun dengan koefisien *Skewness* $< 2,00$ dan Kurtosis $< 7,00$ maka secara univariat dapat dikatakan sebarannya termasuk kategori normal.
 - n. Koefisien *Split-Half First-Second* Spearman Brown
 Dengan menggunakan koreksi dari Spearman Brown, *mean*-nya adalah 0,967 yang jauh di atas *standard* minimal pengembangan tes, bahkan lebih tinggi dari reliabilitas tes *standard* yaitu 0,9.
 Namun koefisien *Skewness*-nya $> 3,00$ dan Kurtosis $> 21,00$ maka secara univariat dapat dikatakan sebarannya tidaklah normal.
 - o. Koefisien *Split-Half Odd-Even*
Mean-nya 0,570 juga tidak memenuhi syarat minimal reliabilitas.
 Namun koefisien *Skewness*-nya $< 2,00$ begitu juga dengan Kurtosisnya $< 7,00$ sehingga secara univariat dapat dikatakan bahwa sebarannya adalah normal.
 - p. Koefisien *Split-Half Odd-Even* Spearman Brown
 Dengan menggunakan koreksi dari Spearman Brown didapat *mean*-nya sebesar 0,702 (memenuhi syarat).
 Dengan koefisien *Skewness* $< 2,00$ dan Kurtosis $< 7,00$ maka secara univariat sebarannya juga dapat dikatakan normal.
 Sehingga dalam konteks reliabilitas belah dua, ini adalah pilihan yang paling menguntungkan.

Hasil Analisis Multivariat

3.3.1. Pemeriksaan Normalitas

Salah satu dari asumsi yang fundamental dalam analisis multivariat adalah normalitas yang terkait dengan bentuk dari suatu distribusi data variabel metrik (Hair, Anderson, Tatham, & Black, 1998).

Dengan menggunakan kriteria yang berbeda, dibandingkan dengan yang telah dicantumkan di bagian hasil analisis deskriptif, maka normalitas univariat dengan menggunakan pengujian berdasarkan statistik Z maupun *Chi-Square* memberikan hasil sebagaimana yang dicantumkan pada Lampiran 4.

Hasil pengujian yang dicantumkan pada lampiran tersebut memperlihatkan bahwa:

- Untuk pengujian individual dengan *Z-Score* untuk *Skewness* yang memenuhi syarat distribusi normal hanya: Banyaknya Item, Indeks Kesulitan, Banyaknya *Option*, Mean of Biserial Correlation, dan SEM of KR20.
- Untuk pengujian individual dengan *Z-Score* untuk *Skewness* yang memenuhi syarat distribusi normal hanya: Koefisien KR20 dan KR21.
- Untuk pengujian bersama-sama dengan *Chi-Square* seluruhnya tidak memenuhi syarat distribusi normal karena nilai signifikansinya $< 0,05$ sehingga H_1 yang menyatakan bahwa data tidak berdistribusi normal harus diterima.

Sedangkan pengujian normalitas multivariat (*Multivariate Normality*) hasilnya dapat dilihat pada Lampiran 5.

Hasil pengujian dimaksud memperlihatkan bahwa baik pengujian *Skewness* dan kurtosis secara individual maupun bersama-sama menunjukkan bahwa data tidak berdistribusi multivariat normal karena nilai signifikansinya $> 0,05$.

3.3.2 Pemeriksaan Homogenitas Variansi Error

Output pengujian homogenitas variansi *error* dapat pada Lampiran 6.

Hasil ini memperlihatkan bahwa secara keseluruhan H_0 yang menyatakan bahwa variansi *error* homogen gagal diterima. Dengan demikian dapat disimpulkan bahwa variansi *error* mengalami problem *heteroscedastisity*. Konsekuensi dari hal ini membuat statistik yang digunakan adalah statistik untuk "*Variance Equal Not Assumed*".

3.3.3. Pemeriksaan *Lack of Fit* (Kesesuaian Model)

Pemeriksaan *Lack of Fit* (Kesesuaian Model) berdasarkan *output* PASW Statistics 18 (tidak ditampilkan karena memerlukan puluhan halaman) seluruhnya tidak signifikan karena nilai probabilitas alphanya $> 0,05$. Dengan demikian dapat dinyatakan bahwa "model sesuai dengan data" atau H_0 yang menyatakan bahwa "*Lack of Fit* tidak signifikan" dinyatakan diterima (*accepted*), hanya asumsi *multivariate normal* dan homogenitasnya tak terpenuhi (*violated*) dan ini mengakibatkan menurunnya akurasi inferensi.

Uji multivariat baik untuk *Main Effect* maupun *Interaction Effect* dengan menggunakan PASW Statistics 18 menghasilkan *output* sebagaimana yang dicantumkan di Lampiran 2.

Output dimaksud memperlihatkan bahwa untuk pengujian secara keseluruhan (*Overall*):

- Seluruh *Main Effect* (kecuali untuk banyaknya *option*) adalah sangat signifikan karena memiliki *p-value* untuk $\text{Alpha} \approx 0,000$ baik dengan menggunakan statistik Pillai's Trace, Wilks' Lambda, Hotelling's Trace, maupun Roy's Largest Root.
- Seluruh *Interaction Effect* yang mengandung faktor *option* tidak signifikan karena memiliki nilai Sig. (*Probability Value*) $> 0,05$ baik dengan

menggunakan statististik Pillai's Trace, Wilks' Lambda, Hotelling's Trace, maupun Roy's Largest Root.

Dengan demikian model (*design*) yang terbentuk berdasarkan data adalah:

$$\begin{aligned} &B_TESTEE + B_ITEM + DIF_INDEX + \\ &B_TESTEE*B_ITEM + \\ &B_TESTEE*DIF_INDEX + \\ &B_ITEM*DIF_INDEX + \\ &B_TESTEE*B_ITEM*DIF_INDEX. \end{aligned}$$

Atau dengan kata lain dapat dinyatakan bahwa daya pembeda dan reliabilitas dipengaruhi oleh banyaknya *testee*, banyaknya item, indeks kesulitan, dan interaksi dari faktor-faktor tersebut.

Untuk pengujian terhadap *response* (*Dependent Variables*) secara individual, hasilnya dapat dilihat pada Lampiran 3.

Pada lampiran tersebut (yang diberi *background* kuning memiliki hasil uji yang tidak signifikan) memperlihatkan bahwa:

- a. Banyaknya *option* tidak berpengaruh signifikan terhadap semua statistik untuk daya pembeda maupun reliabilitas, termasuk *mean of difficulty index* (dalam artian secara individual).
- b. Selain itu banyaknya *testee*, banyaknya item (panjang tes), indeks kesulitan maupun interaksinya tidak berpengaruh secara signifikan terhadap statistik:
 - 1) *Mean of Item Discrimination*.
 - 2) *Split-Half First-Second from Spearman Brown*.

Dengan demikian dapat dinyatakan bahwa statistik ini bersifat "*Robust*".

- c. Statistik yang paling sensitif terhadap banyaknya *testee*, banyaknya item (panjang tes), Indeks Kesulitan maupun interaksinya adalah *Split-Half Odd-Even from Spearman Brown*, karena hanya tidak signifikan terhadap efek interaksi Banyaknya Item*Indeks Kesulitan meskipun hasilnya cenderung memenuhi syarat minimal

reliabilitas (0,6) dan berdistribusi normal.

4. Diskusi

Hasil-hasil di atas memperlihatkan hal-hal sebagai berikut:

- a. *Main Effects* yang signifikan adalah dari:

- 1) Banyaknya *Testee* (ukuran sampel).
- 2) Banyaknya Item dalam satu tes (panjang tes).
- 3) Tingkat Kesulitan.

Sedangkan banyaknya *option* tidak berpengaruh secara signifikan terhadap besaran statistik daya pembeda. Dengan kata lain, banyak atau sedikitnya *option* tidak akan memberikan perbedaan yang berarti antara *testee* yang mampu menjawab item dengan benar dan yang tidak mampu menjawab dengan benar. Dengan demikian memperbanyak *option* dalam hal ini hanya akan merepotkan pengembang tes belaka.

- b. *Interaction Effects* yang signifikan adalah dari:

- 1) Banyaknya *Testee**Banyaknya Item.
- 2) Banyaknya *Testee**Tingkat Kesulitan.
- 3) Banyaknya Item*Tingkat Kesulitan.
- 4) Banyaknya *Testee**Banyaknya Item*Tingkat Kesulitan.

Sementara yang tidak signifikan adalah interaksi antara:

- 1) Banyaknya *Testee**Banyaknya *Option*.
- 2) Banyaknya Item*Banyaknya *Option*.
- 3) Indeks Kesulitan*Banyaknya *Option*.
- 4) Seluruh interaksi pada ordo yang lebih tinggi.

Interaksi yang melibatkan 3 dan 4 (keseluruhan faktor dalam kajian ini) tidak memiliki pengaruh yang signifikan memang masuk di akal, karena kenyataan pada umumnya dalam penelitian eksperimental

memang memperlihatkan hal tersebut. Karena itulah dikembangkan berbagai desain yang memangkasnya seperti keluarga dari *Fractional Factorial Design* (Kirk, 1995).

Sedangkan 3 interaksi pada ordo kedua yang tidak signifikan karena melibatkan banyaknya *option*, yang juga secara individual tidak signifikan pengaruhnya, sehingga hal ini pada dasarnya dapat dimengerti.

- c. Model yang sesuai dengan data untuk memprediksi Daya Pembeda dan Reliabilitas adalah:

$$B_TESTEE + B_ITEM + DIF_INDEX + B_TESTEE*B_ITEM + B_TESTEE*DIF_INDEX + B_ITEM*DIF_INDEX + B_TESTEE*B_ITEM*DIF_INDEX.$$

- d. Model di atas mampu memprediksi:

- 1) *Mean of Item Discrimination* hanya sebesar 0,002 atau 0,2%.
- 2) *Mean of Biserial Correlation* sebesar 0,993 atau 99,3%.
- 3) *Mean of Point Biserial* sebesar 0,993 atau 99,3%.
- 4) *Mean of Adjusted Point Biserial* sebesar 0,979 atau 97,9%.
- 5) Koefisien KR20 (Alpha) sebesar 0,993 atau 99,3%.
- 6) Koefisien KR21 sebesar 0,991 atau 99,1%.
- 7) Koefisien *Split-Half* 1-2 sebesar 0,984 atau 98,4%.
- 8) Koefisien *Split-Half* 1-2 dari Spearman Brown hanya sebesar 0,048 atau 4,8%.
- 9) Koefisien *Split-Half O-E* sebesar 0,982 atau 98,2%.
- 10) Koefisien *Split-Half O-E* Spearman Brown sebesar 0,987 atau 98,7%.

Hasil ini memperlihatkan bahwa, walau secara umum faktor banyaknya *testee*, banyaknya butir, dan indeks kesulitan berpengaruh secara signifikan, namun pengaruh

tersebut umumnya pada statistik-statistik reliabilitas (kecuali *Split-Half* 1-2) namun tidak demikian halnya pada statistik *mean* daya pembeda item.

Dengan kata lain dapat dinyatakan bahwa *Mean of Item Discrimination* dan *Split-Half* 1-2 *Index* dengan koreksi dari Spearman Brown, *robust* terhadap banyaknya *testee*, item, dan indeks kesulitan.

- e. Statistik yang paling sensitif terhadap banyaknya *Testee*, banyaknya Item (panjang tes), Indeks Kesulitan maupun interaksinya adalah *Split-Half Odd-Even* Spearman Brown karena hanya tidak signifikan terhadap efek interaksi Banyaknya Item*Indeks Kesulitan meskipun hasilnya cenderung memenuhi syarat minimal reliabilitas (0,6) dan berdistribusi normal.

Namun yang perlu diperhatikan terkait dengan penelitian ini adalah beberapa statistik yang dikaji sebarannya tidak berdistribusi normal. Dengan demikian untuk kajian yang melibatkan statistik - statistik dimaksud direkomendasikan melakukan transformasi terlebih dahulu agar data yang dianalisis berdistribusi normal sehingga hasil analisis dengan MANOVA lebih akurat.

Selain itu penelitian ini semata-mata menggunakan data simulasi, sehingga perlu diperiksa lebih lanjut mengenai hasilnya bila menggunakan data lapangan.

Selain itu, diharapkan pula untuk melakukan penelitian lebih lanjut terkait dengan kajian statistik-statistik ini berdasarkan:

- a. *Classical Test Theory* (CTT) terutama pada level-level faktor yang belum dilibatkan seperti:

- 1) *Testee* sebanyak 250 seperti pada penelitian Jiao & Kamata atau jumlah *testee* yang lebih besar seperti pada pengujian di *level*

- Daerah Tingkat II, Daerah Tingkat I, dan Nasional.
- 2) Panjang tes 25, 30, 45, 50, 80, 90, 100 item atau yang lainnya.
 - 3) Indeks kesulitan sangat mudah (*very easy*), relatif mudah (*relatively easy*), agak sulit (*fairly difficult*), maupun sangat sulit (*very difficult*).
 - 4) Jumlah *option* 2, 3, 5 atau lebih termasuk *mixed option*.
- b. *Item Response Theory* (IRT) dan membandingkan hasilnya dengan CTT.
- Referensi**
- Coakes, S. J. & Steed, L. G. (1996). *SPSS for windows: Analysis without anguish*. Brisbane: John Wiley & Sons.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Bulletin*, 109: 512 – 519.
- Field, a. (2000). *Discovering statistics using SPSS for windows: Advanced techniques for the beginner*. London: Sage Publications
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998), *Multivariate data analysis*, 5th edition, Prentice Hall Internasional: UK
- Hambleton, R. K., & Cook, L. L. (1983). Robustness of item response models and effects of test length and sample size on the precision of ability estimates. Dalam D. Weiss (Ed.). *New horizon in testing*, 31 – 49. New York: Academic Press
- Harwel, M. R., Stone, C. A., Hsu, T. C., dkk (1996). Monte-Carlo studies in item response theory. *Applied Psychological Measurement*, 20, 101 – 125.
- Jiao, H., & Kamata, A. (2003). Model comparison in the presence of local item dependence, *Paper presented at the annual meeting of the AERA*, Chicago, April 21-25, 2003
- Joreskog, K. G., & Sorbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Sciencetific Software International
- Lewis, M. (2002). *Test analysis program version 4.2.5: User's guide*. <http://oak.cats.ohiu.edu/~brooksg/tap.htm>: Gordon P. Brooks.
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences*. Pacivic Grove, California: Brooks/Cole Publishing Company.
- Mislevy, R. J., & Bock, R. D. (1990). *BILOG 3: Item analysis & test scoring with binary logistic models*. Moorseville: Sciencetific Software Inc.
- Segall, D. O. (2000). General ability measurement: An application of multidimensional item respon theory. *Psychometrika*, 66, 79 – 97.

Lampiran 1

Total Sample Size = 270

Univariate Summary Statistics for Continuous Variables

Variable	Mean	St. Dev.	T-Value	Skewness	Kurtosis	Min	Freq.	Max	Freq.
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
B_TESTEE	241.667	191.157	20.773	0.454	-1.511	45.000	90	500.000	90
B_ITEM	40.000	16.360	40.175	0.000	-1.511	20.000	90	60.000	90
DIF_INDE	68.333	22.526	49.845	-0.111	-1.511	40.000	90	95.000	90
B_OPTION	3.000	1.002	49.204	0.000	-2.015	2.000	135	4.000	135
MI_DIF	0.691	0.303	37.500	4.179	42.740	0.040	1	3.860	1
MI_DIS	2.037	28.708	1.166	16.431	266.986	0.066	2	472.000	1
MBC	0.488	0.074	108.946	-0.218	1.452	0.259	2	0.706	2
MPB	0.325	0.069	77.047	-0.524	-0.873	0.166	2	0.435	4
MAPB	0.236	0.091	42.574	-0.519	-1.189	0.035	2	0.358	3
KR20_ALP	0.698	0.184	62.290	-0.953	0.025	0.103	2	0.907	2
SEM_KR20	2.151	0.830	42.582	-0.011	-1.281	0.681	2	3.423	2
KR21	0.673	0.187	59.064	-0.904	-0.033	0.081	2	0.894	2
SH1_2	0.575	0.204	46.263	-0.511	-0.845	0.069	2	0.871	2
SH1_2SB	0.967	4.282	3.710	16.386	265.998	0.128	2	71.000	1
SHO_E	0.570	0.204	45.981	-0.599	-0.540	0.030	2	0.869	2
SHO_ESB	0.702	0.189	61.021	-1.134	0.804	0.058	2	0.930	2

Lampiran 2

Tabel 2. Multivariate Tests(c)

Effect		Value	F	Hypothesis df	Error df	Sig.
J_TESTEE	Pillai's Trace	.804	11.531	24.000	412.000	.000
	Wilks' Lambda	.265	16.106(a)	24.000	410.000	.000
	Hotelling's Trace	2.516	21.383	24.000	408.000	.000
	Roy's Largest Root	2.408	41.340(b)	12.000	206.000	.000
J_ITEM	Pillai's Trace	1.342	35.010	24.000	412.000	.000
	Wilks' Lambda	.011	149.493(a)	24.000	410.000	.000
	Hotelling's Trace	60.564	514.794	24.000	408.000	.000
	Roy's Largest Root	60.006	1030.095(b)	12.000	206.000	.000
DIF_INDEX	Pillai's Trace	1.213	26.432	24.000	412.000	.000
	Wilks' Lambda	.009	163.486(a)	24.000	410.000	.000
	Hotelling's Trace	85.980	730.832	24.000	408.000	.000
	Roy's Largest Root	85.691	1471.037(b)	12.000	206.000	.000
J_OPTION	Pillai's Trace	.020	.341(a)	12.000	205.000	.981
	Wilks' Lambda	.980	.341(a)	12.000	205.000	.981
	Hotelling's Trace	.020	.341(a)	12.000	205.000	.981
	Roy's Largest Root	.020	.341(a)	12.000	205.000	.981
J_TESTEE * J_ITEM	Pillai's Trace	.428	2.074	48.000	832.000	.000
	Wilks' Lambda	.625	2.141	48.000	791.720	.000
	Hotelling's Trace	.521	2.207	48.000	814.000	.000
	Roy's Largest Root	.324	5.610(b)	12.000	208.000	.000
J_TESTEE * DIF_INDEX	Pillai's Trace	1.077	6.384	48.000	832.000	.000
	Wilks' Lambda	.142	10.882	48.000	791.720	.000
	Hotelling's Trace	4.607	19.530	48.000	814.000	.000
	Roy's Largest Root	4.303	74.583(b)	12.000	208.000	.000
J_ITEM * DIF_INDEX	Pillai's Trace	1.162	7.095	48.000	832.000	.000
	Wilks' Lambda	.079	15.375	48.000	791.720	.000
	Hotelling's Trace	8.796	37.292	48.000	814.000	.000
	Roy's Largest Root	8.486	147.097(b)	12.000	208.000	.000
J_TESTEE * J_ITEM * DIF_INDEX	Pillai's Trace	.664	1.599	96.000	1696.000	.000
	Wilks' Lambda	.486	1.637	96.000	1391.225	.000
	Hotelling's Trace	.786	1.664	96.000	1626.000	.000
	Roy's Largest Root	.277	4.890(b)	12.000	212.000	.000
J_TESTEE * J_OPTION	Pillai's Trace	.039	.339	24.000	412.000	.999
	Wilks' Lambda	.962	.338(a)	24.000	410.000	.999
	Hotelling's Trace	.040	.336	24.000	408.000	.999
	Roy's Largest Root	.026	.444(b)	12.000	206.000	.944
J_ITEM * J_OPTION	Pillai's Trace	.040	.350	24.000	412.000	.998
	Wilks' Lambda	.960	.349(a)	24.000	410.000	.998
	Hotelling's Trace	.041	.348	24.000	408.000	.998
	Roy's Largest Root	.031	.530(b)	12.000	206.000	.894
J_TESTEE * J_ITEM * J_OPTION	Pillai's Trace	.075	.332	48.000	832.000	1.000

DIF_INDEX * J_OPTION	Wilks' Lambda	.926	.331	48.000	791.720	1.000
	Hotelling's Trace	.078	.330	48.000	814.000	1.000
	Roy's Largest Root	.042	.731(b)	12.000	208.000	.720
	Pillai's Trace	.042	.368	24.000	412.000	.998
J_TESTEE * DIF_INDEX * J_OPTION	Wilks' Lambda	.958	.369(a)	24.000	410.000	.998
	Hotelling's Trace	.044	.371	24.000	408.000	.998
	Roy's Largest Root	.041	.710(b)	12.000	206.000	.741
	Pillai's Trace	.082	.361	48.000	832.000	1.000
J_ITEM * DIF_INDEX * J_OPTION	Wilks' Lambda	.920	.359	48.000	791.720	1.000
	Hotelling's Trace	.085	.358	48.000	814.000	1.000
	Roy's Largest Root	.047	.819(b)	12.000	208.000	.631
	Pillai's Trace	.085	.377	48.000	832.000	1.000
J_TESTEE * J_ITEM * DIF_INDEX * J_OPTION	Wilks' Lambda	.917	.377	48.000	791.720	1.000
	Hotelling's Trace	.089	.377	48.000	814.000	1.000
	Roy's Largest Root	.059	1.027(b)	12.000	208.000	.426
	Pillai's Trace	.157	.353	96.000	1696.000	1.000
	Wilks' Lambda	.850	.353	96.000	1391.225	1.000
	Hotelling's Trace	.167	.354	96.000	1626.000	1.000
	Roy's Largest Root	.088	1.548(b)	12.000	212.000	.109

a Exact statistic

b The statistic is an upper bound on F that yields a lower bound on the significance level.

c Design: J_TESTEE+J_ITEM+DIF_INDEX+J_OPTION+J_TESTEE * J_ITEM+J_TESTEE * DIF_INDEX+J_ITEM * DIF_INDEX+J_TESTEE * J_ITEM * DIF_INDEX+J_TESTEE * J_OPTION+J_ITEM * J_OPTION+J_TESTEE * J_ITEM * J_OPTION+DIF_INDEX * J_OPTION+J_TESTEE * DIF_INDEX * J_OPTION+J_ITEM * DIF_INDEX * J_OPTION+J_TESTEE * J_ITEM * DIF_INDEX * J_OPTION

Lampiran 3

Tabel 3. Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	MI_DIF	143.645(a)	54	2.660	58.265	.000
	MI_DIS	44911.822(b)	54	831.700	1.010	.465
	MBC	65.323(c)	54	1.210	670.182	.000
	MPB	29.564(d)	54	.547	737.588	.000
	MAPB	17.006(e)	54	.315	235.843	.000
	KR20_ALP	139.901(c)	54	2.591	667.392	.000
	KR21	130.636(f)	54	2.419	565.439	.000
	SEM_KR20	1432.505(g)	54	26.528	3058.719	.000
	SH1_2	99.117(h)	54	1.835	299.184	.000
	SH1_2SB	1236.576(i)	54	22.900	1.253	.133
	SHO_E	97.586(j)	54	1.807	280.565	.000
	SHO_ESB	141.280(k)	54	2.616	386.590	.000
J_TESTEE	MI_DIF	.073	2	.037	.800	.450
	MI_DIS	1644.782	2	822.391	.999	.370
	MBC	.265	2	.132	73.383	.000
	MPB	.001	2	.000	.632	.532
	MAPB	.010	2	.005	3.728	.026
	KR20_ALP	.120	2	.060	15.409	.000
	KR21	.168	2	.084	19.606	.000
	SEM_KR20	.045	2	.023	2.609	.076
	SH1_2	.025	2	.012	2.033	.133
	SH1_2SB	35.257	2	17.629	.965	.383
	SHO_E	.048	2	.024	3.708	.026
	SHO_ESB	.075	2	.038	5.550	.004
J_ITEM	MI_DIF	.080	2	.040	.876	.418
	MI_DIS	1657.287	2	828.643	1.006	.367
	MBC	.353	2	.177	97.861	.000
	MPB	.120	2	.060	80.870	.000
	MAPB	.053	2	.026	19.711	.000
	KR20_ALP	2.027	2	1.014	261.103	.000
	KR21	2.413	2	1.207	282.021	.000
	SEM_KR20	59.837	2	29.919	3449.681	.000
	SH1_2	2.357	2	1.178	192.069	.000
	SH1_2SB	24.757	2	12.378	.677	.509
	SHO_E	2.600	2	1.300	201.836	.000
	SHO_ESB	1.956	2	.978	144.530	.000
DIF_INDEX	MI_DIF	12.495	2	6.247	136.840	.000
	MI_DIS	1769.595	2	884.797	1.074	.343
	MBC	.004	2	.002	1.181	.309
	MPB	.962	2	.481	647.979	.000
	MAPB	1.853	2	.926	693.713	.000
	KR20_ALP	5.838	2	2.919	751.969	.000
	KR21	5.587	2	2.793	652.928	.000

J_OPTION	SEM_KR20	118.284	2	59.142	6819.194	.000
	SH1_2	7.207	2	3.603	587.337	.000
	SH1_2SB	57.797	2	28.899	1.581	.208
	SHO_E	6.877	2	3.438	533.831	.000
	SHO_ESB	5.727	2	2.864	423.119	.000
	MI_DIF	.036	1	.036	.781	.378
	MI_DIS	822.590	1	822.590	.999	.319
	MBC	3.20E-005	1	3.20E-005	.018	.894
	MPB	.000	1	.000	.000	1.000
	MAPB	3.70E-005	1	3.70E-005	.028	.868
	KR20_ALP	.000	1	.000	.000	1.000
	KR21	3.70E-009	1	3.70E-009	.000	.999
	SEM_KR20	.005	1	.005	.615	.434
	SH1_2	.000	1	.000	.000	1.000
	SH1_2SB	18.299	1	18.299	1.001	.318
	SHO_E	.000	1	.000	.000	1.000
J_TESTEE * J_ITEM	SHO_ESB	3.70E-007	1	3.70E-007	.000	.994
	MI_DIF	.195	4	.049	1.066	.374
	MI_DIS	3305.060	4	826.265	1.003	.407
	MBC	.007	4	.002	.973	.423
	MPB	.007	4	.002	2.334	.057
	MAPB	.010	4	.003	1.884	.114
	KR20_ALP	.032	4	.008	2.074	.085
	KR21	.041	4	.010	2.368	.054
	SEM_KR20	.035	4	.009	1.019	.398
	SH1_2	.098	4	.025	4.007	.004
	SH1_2SB	77.358	4	19.340	1.058	.378
	SHO_E	.063	4	.016	2.457	.047
	SHO_ESB	.083	4	.021	3.084	.017
	MI_DIF	.186	4	.047	1.021	.398
	MI_DIS	3294.895	4	823.724	1.000	.408
J_TESTEE * DIF_INDEX	MBC	.323	4	.081	44.753	.000
	MPB	.023	4	.006	7.862	.000
	MAPB	.010	4	.002	1.787	.132
	KR20_ALP	.110	4	.027	7.081	.000
	KR21	.137	4	.034	7.984	.000
	SEM_KR20	.054	4	.014	1.560	.186
	SH1_2	.136	4	.034	5.541	.000
	SH1_2SB	75.512	4	18.878	1.033	.391
	SHO_E	.087	4	.022	3.363	.011
	SHO_ESB	.117	4	.029	4.337	.002
	MI_DIF	.186	4	.046	1.018	.399
	MI_DIS	3304.423	4	826.106	1.003	.407
	MBC	.103	4	.026	14.255	.000
	MPB	.013	4	.003	4.330	.002
	MAPB	.006	4	.002	1.145	.336
J_ITEM * DIF_INDEX	KR20_ALP	.129	4	.032	8.291	.000
	KR21	.113	4	.028	6.599	.000
	SEM_KR20	4.972	4	1.243	143.315	.000

J_TESTEE * J_ITEM * DIF_INDEX	SH1_2	.055	4	.014	2.222	.068
	SH1_2SB	73.828	4	18.457	1.010	.403
	SHO_E	.022	4	.005	.846	.497
	SHO_ESB	.057	4	.014	2.122	.079
	MI_DIF	.353	8	.044	.966	.464
	MI_DIS	6588.871	8	823.609	1.000	.437
	MBC	.010	8	.001	.695	.696
	MPB	.002	8	.000	.368	.936
	MAPB	.005	8	.001	.426	.905
	KR20_ALP	.030	8	.004	.980	.452
	KR21	.036	8	.004	1.052	.399
	SEM_KR20	.101	8	.013	1.449	.178
	SH1_2	.012	8	.002	.252	.980
	SH1_2SB	145.647	8	18.206	.996	.440
	SHO_E	.084	8	.011	1.632	.117
J_TESTEE * J_OPTION	SHO_ESB	.139	8	.017	2.565	.011
	MI_DIF	.100	2	.050	1.091	.338
	MI_DIS	1647.086	2	823.543	1.000	.370
	MBC	6.41E-005	2	3.20E-005	.018	.982
	MPB	.000	2	.000	.000	1.000
	MAPB	7.41E-005	2	3.70E-005	.028	.973
	KR20_ALP	.000	2	.000	.000	1.000
	KR21	7.41E-009	2	3.70E-009	.000	1.000
	SEM_KR20	.006	2	.003	.359	.699
	SH1_2	.000	2	.000	.000	1.000
	SH1_2SB	36.598	2	18.299	1.001	.369
	SHO_E	.000	2	.000	.000	1.000
	SHO_ESB	7.41E-007	2	3.70E-007	.000	1.000
J_ITEM * J_OPTION	MI_DIF	.100	2	.050	1.094	.337
	MI_DIS	1647.086	2	823.543	1.000	.370
	MBC	6.41E-005	2	3.20E-005	.018	.982
	MPB	.000	2	.000	.000	1.000
	MAPB	7.41E-005	2	3.70E-005	.028	.973
	KR20_ALP	.000	2	.000	.000	1.000
	KR21	7.41E-009	2	3.70E-009	.000	1.000
	SEM_KR20	.006	2	.003	.359	.699
	SH1_2	.000	2	.000	.000	1.000
	SH1_2SB	36.598	2	18.299	1.001	.369
	SHO_E	.000	2	.000	.000	1.000
	SHO_ESB	7.41E-007	2	3.70E-007	.000	1.000
J_TESTEE * J_ITEM * J_OPTION	MI_DIF	.172	4	.043	.939	.442
	MI_DIS	3292.265	4	823.066	.999	.409
	MBC	.000	4	3.20E-005	.018	.999
	MPB	.000	4	.000	.000	1.000
	MAPB	.000	4	3.70E-005	.028	.999
	KR20_ALP	.000	4	.000	.000	1.000
	KR21	1.48E-008	4	3.70E-009	.000	1.000
	SEM_KR20	.017	4	.004	.487	.745

DIF_INDEX * J_OPTION	SH1_2	.000	4	.000	.000	1.000
	SH1_2SB	73.195	4	18.299	1.001	.408
	SHO_E	.000	4	.000	.000	1.000
	SHO_ESB	1.48E-006	4	3.70E-007	.000	1.000
	MI_DIF	.072	2	.036	.786	.457
	MI_DIS	1646.111	2	823.056	.999	.370
	MBC	6.41E-005	2	3.20E-005	.018	.982
	MPB	.000	2	.000	.000	1.000
	MAPB	7.41E-005	2	3.70E-005	.028	.973
	KR20_ALP	.000	2	.000	.000	1.000
	KR21	7.41E-009	2	3.70E-009	.000	1.000
	SEM_KR20	.006	2	.003	.359	.699
	SH1_2	.000	2	.000	.000	1.000
	SH1_2SB	36.598	2	18.299	1.001	.369
	SHO_E	.000	2	.000	.000	1.000
	SHO_ESB	7.41E-007	2	3.70E-007	.000	1.000
J_TESTEE * DIF_INDEX * J_OPTION	MI_DIF	.200	4	.050	1.093	.361
	MI_DIS	3293.239	4	823.310	1.000	.409
	MBC	.000	4	3.20E-005	.018	.999
	MPB	.000	4	.000	.000	1.000
	MAPB	.000	4	3.70E-005	.028	.999
	KR20_ALP	.000	4	.000	.000	1.000
	KR21	1.48E-008	4	3.70E-009	.000	1.000
	SEM_KR20	.017	4	.004	.487	.745
	SH1_2	.000	4	.000	.000	1.000
	SH1_2SB	73.195	4	18.299	1.001	.408
	SHO_E	.000	4	.000	.000	1.000
	SHO_ESB	1.48E-006	4	3.70E-007	.000	1.000
J_ITEM * DIF_INDEX * J_OPTION	MI_DIF	.199	4	.050	1.092	.362
	MI_DIS	3293.239	4	823.310	1.000	.409
	MBC	.000	4	3.20E-005	.018	.999
	MPB	.000	4	.000	.000	1.000
	MAPB	.000	4	3.70E-005	.028	.999
	KR20_ALP	.000	4	.000	.000	1.000
	KR21	1.48E-008	4	3.70E-009	.000	1.000
	SEM_KR20	.017	4	.004	.487	.745
	SH1_2	.000	4	.000	.000	1.000
	SH1_2SB	73.195	4	18.299	1.001	.408
	SHO_E	.000	4	.000	.000	1.000
	SHO_ESB	1.48E-006	4	3.70E-007	.000	1.000
J_TESTEE * J_ITEM * DIF_INDEX * J_OPTION	MI_DIF	.343	8	.043	.938	.486
	MI_DIS	6585.462	8	823.183	.999	.437
	MBC	.000	8	3.20E-005	.018	1.000
	MPB	.000	8	.000	.000	1.000
	MAPB	.000	8	3.70E-005	.028	1.000
	KR20_ALP	.000	8	.000	.000	1.000

Error	KR21	2.96E-008	8	3.70E-009	.000	1.000
	SEM_KR20	.029	8	.004	.423	.907
	SH1_2	.000	8	.000	.000	1.000
	SH1_2SB	146.391	8	18.299	1.001	.436
	SHO_E	.000	8	.000	.000	1.000
	SHO_ESB	2.96E-006	8	3.70E-007	.000	1.000
	MI_DIF	9.861	216	.046		
	MI_DIS	177900.741	216	823.615		
	MBC	.390	216	.002		
	MPB	.160	216	.001		
	MAPB	.288	216	.001		
	KR20_ALP	.838	216	.004		
	KR21	.924	216	.004		
	SEM_KR20	1.873	216	.009		
	SH1_2	1.325	216	.006		
	SH1_2SB	3947.817	216	18.277		
	SHO_E	1.391	216	.006		
	SHO_ESB	1.462	216	.007		
Total	MI_DIF	153.507	270			
	MI_DIS	222812.563	270			
	MBC	65.713	270			
	MPB	29.724	270			
	MAPB	17.295	270			
	KR20_ALP	140.740	270			
	KR21	131.560	270			
	SEM_KR20	1434.378	270			
	SH1_2	100.442	270			
	SH1_2SB	5184.392	270			
	SHO_E	98.977	270			
	SHO_ESB	142.742	270			

a R Squared = .936 (Adjusted R Squared = .920)

b R Squared = .202 (Adjusted R Squared = .002)

c R Squared = .994 (Adjusted R Squared = .993)

d R Squared = .995 (Adjusted R Squared = .993)

e R Squared = .983 (Adjusted R Squared = .979)

f R Squared = .993 (Adjusted R Squared = .991)

g R Squared = .999 (Adjusted R Squared = .998)

h R Squared = .987 (Adjusted R Squared = .984)

i R Squared = .239 (Adjusted R Squared = .048)

j R Squared = .986 (Adjusted R Squared = .982)

k R Squared = .990 (Adjusted R Squared = .987)

Lampiran 4

Test of Univariate Normality for Continuous Variables

	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	Z-Score	P-Value	Z-Score	P-Value	Chi-Square	P-Value
J_TESTEE	2.411	0.016	61.197	0.000	3750.878	0.000
J_ITEM	0.000	1.000	61.196	0.000	3744.910	0.000
DIF_INDE	-1.051	0.293	61.194	0.000	3745.851	0.000
J_OPTION	0.000	1.000	36.098	0.000	1303.055	0.000
MI_DIF	4.792	0.000	10.452	0.000	132.210	0.000
MI_DIS	6.268	0.000	12.638	0.000	199.013	0.000
MBC	-1.658	0.097	3.311	0.001	13.708	0.001
MPB	-2.564	0.010	-5.150	0.000	33.099	0.000
MAPB	-2.553	0.011	-11.174	0.000	131.380	0.000
KR20_ALP	-3.202	0.001	0.295	0.768	10.340	0.006
KR21	-3.145	0.002	0.095	0.924	9.897	0.007
SEM_KR20	-0.117	0.906	-15.114	0.000	228.457	0.000
SH1_2	-2.536	0.011	-4.834	0.000	29.801	0.000
SH1_2SB	6.265	0.000	12.635	0.000	198.893	0.000
SHO_E	-2.705	0.007	-2.301	0.021	12.612	0.002
SHO_ESB	-3.388	0.001	2.263	0.024	16.596	0.000

Lampiran 5

Relative Multivariate Kurtosis = 3.746

Test of Multivariate Normality for Continuous Variables

Skewness		Kurtosis		Skewness and Kurtosis	
Value	Z-Score P-Value	Value	Z-Score P-Value	Chi-Square	P-Value

878.927 160.361 0.000 793.029 30.520 0.000 26647.000 0.000

Lampiran 6

Tabel 4. Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
MI_DIF	6.865	53	216	.000
MI_DIS	7.109	53	216	.000
MBC	2.306	53	216	.000
MPB	2.793	53	216	.000
MAPB	3.121	53	216	.000
KR20_ALP	4.291	53	216	.000
KR21	5.641	53	216	.000
SEM_KR20	4.158	53	216	.000
SH1_2	4.548	53	216	.000
SH1_2SB	7.081	53	216	.000
SHO_E	3.579	53	216	.000
SHO_ESB	5.660	53	216	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: J_TESTEE+J_ITEM + DIF_INDEX+J_OPTION +
J_TESTEE * J_ITEM + J_TESTEE * DIF_INDEX +
J_ITEM * DIF_INDEX +
J_TESTEE * J_ITEM * DIF_INDEX +
J_TESTEE * J_OPTION + J_ITEM * J_OPTION +
J_TESTEE * J_ITEM * J_OPTION +
DIF_INDEX * J_OPTION +
J_TESTEE * DIF_INDEX * J_OPTION +
J_ITEM * DIF_INDEX * J_OPTION +
J_TESTEE * J_ITEM * DIF_INDEX * J_OPTION