

Makalah Seminar Nasional

PERBANDINGAN PENSKALAAN METODE INTERVAL TAMPAK  
SETARA (TIPE THURSTONE) DAN *SUMMATED RATING*  
(TIPE LIKERT)



Oleh :

Farida Agus Setiawati, M.Si

Jurusan Psikologi Pendidikan dan Bimbingan, Fakultas Ilmu Pendidikan

Universitas Negeri Yogyakarta

April 2011

## **PENGANTAR**

### **A. Latar Belakang**

Sebagian besar data dalam pengukuran aspek psikis manusia memiliki karakteristik data nominal dan ordinal, termasuk pengukuran dengan menggunakan instrumen tipe Thurstone dan Likert. Pengukuran pada data ini tidak dapat dianalisis dengan statistik parametrik disebabkan karena tidak memenuhi asumsi analisis statistik parametrik. Sebagaimana yang dinyatakan oleh Glenderg (1988: 144) ada beberapa syarat yang harus dipenuhi agar data dapat dianalisis menggunakan statistic parametrik, yaitu : 1) syarat yang terkait dengan populasi 2) syarat yang terkait dengan sample, 3) dan syarat yang terkait dengan data. Analisis parametrik hanya dapat dianalisis apabila populasi data sample penelitian berdistribusi normal, sample data didapat secara acak dan data yang berupa angka termasuk dalam level interval dan ordinal.

Tidak terpenuhinya berbagai asumsi tersebut menyebabkan data harus dianalisis dengan nonparametrik, padahal power statistik pada data nonparametrik ini lebih rendah daripada parametrik. Data pada level ini terikat pada populasi tertentu, sehingga level dan power error tidak dapat diterapkan. Dengan mengetahui berbagai keterbatasan analisis non parametrik dan melihat pentingnya untuk menggeneralisasi hasil penelitian, para ahli menekankan pentingnya pemenuhan asumsi pada metode statistik parametrik. Dengan demikian proses penskalaan pada data nominal dan ordinal sangat diperlukan untuk mendapatkan data interval.

Penskalaan instrumen untuk mengukur karakteristik psikis manusia sudah dilakukan oleh beberapa ahli pengukuran. Thurstone (1931) mengembangkan penskalaan yang disebut dengan metode interval tampak setara atau methods of equaling interval. Dengan menggunakan aturan tiap stimulus akan berpasangan dengan stimulus

lain akan didapatkan jumlah item tertentu agar semua item saling berpasangan. Cara ini akan membatasi item-item yang dikembangkan dalam tipe Thurstone baik dari jumlah maupun aspek yang diukur, disamping itu juga memungkinkan untuk diketahui konsistensi responden dalam memberi respon pada tiap pernyataan. Meskipun demikian tipe ini akan sulit dikembangkan apabila jumlah stimulus atau aspek yang akan dikembangkan cukup banyak, karena keharusannya untuk berpasangan dengan semua stimulus. Di samping itu responden juga akan merasa bosan dalam memberi respon karena banyaknya item dan cenderung berulang.

Penskalaan lain yang sering digunakan dalam ilmu sosial adalah penskalaan yang dilakukan oleh Likert tahun 1932 yang dikenal dengan *summated rating*. Tipe Likert berbeda dengan Thurstone terkait dengan pernyataan yang dibuat. Tiap pernyataan mewakili indikator yang dibuat dalam kisi kisi. Tidak ada aturan terkait dengan banyaknya item yang harus dibuat. Dengan demikian pembuatan instrumen tipe ini lebih sederhana dibanding tipe Thurstone. Tiap pernyataan disajikan beberapa respon dimana responden harus memilih satu dari beberapa respon yang disajikan. Karena itu pendekatan tipe Likert berbasis respon.

Kedua tipe instrumen memiliki kelebihan dan kekurangan terkait dengan adanya *social desirability* dan *faking* sebagaimana dijelaskan di latar belakang penelitian. Tipe Thurstone tidak sensitif dengan *social desirability* sehingga memiliki kemungkinan kecil untuk dijawab tidak jujur atau *faking*. Hal ini berbeda dengan tipe Likert yang memungkinkan untuk dijawab tidak jujur atau *faking*.

Berdasarkan penjelasan diatas dapat dibuat tabel perbandingan penskalaan tipe Thurstone dan Likert dengan teori klasik dan modern dapat diformulasikan pada tabel 9.

Penskalaan dengan metode *paired comparison* maupun *summated rating* dikatakan berbasis klasik karena dalam penskalaan yang dilakukannya berbasis skor z yang terikat

pada sampel subjek. Disamping itu metode ini tidak mempertimbangkan beberapa asumsi seperti unidimensionalitas, dependensi item dan penyebaran subjek. Penskalaan dengan pendekatan modern atau item respon theory lebih ketat dalam mempertimbangkan berbagai asumsi. Disamping itu penskalaan IRT berkembang dengan berbasis laten trait yang dalam proses penskalaannya dilakukan tidak hanya berbasis stimulus atau respon tetapi juga kemampuan subjek.

Tabel 9. Perbandingan Penskalaan Tipe Thurstone dan Likert

Tipe	Thurstone	Likert
Karakteristik	<ol style="list-style-type: none"> <li>1. Pendekatan stimulus</li> <li>2. Tiap item terdiri 2 pernyataan yang berpasangan</li> <li>3. Responden diminta memilih satu dari dua pernyataan yang disajikan</li> </ol>	<ol style="list-style-type: none"> <li>1. Pendekatan respon</li> <li>2. Tiap item terdiri dari sebuah pernyataan dengan lima alternatif respon</li> <li>3. Responden diminta memberikan respon yang paling sesuai dari beberapa pilihan yang disediakan</li> </ol>
Kelebihan	<ol style="list-style-type: none"> <li>1. Tidak sensitif terhadap <i>social desirability</i> dan <i>faking</i></li> <li>2. Memberikan batasan minimal pada perilaku responden</li> <li>3. Konsistensi internal dapat diketahui</li> </ol>	<ol style="list-style-type: none"> <li>1. Perilaku responden dapat digambarkan secara detail</li> </ol>
Kekurangan	<ol style="list-style-type: none"> <li>1. Beberapa item dapat overload</li> <li>2. Dipengaruhi oleh karakteristik judgment</li> </ol>	<ol style="list-style-type: none"> <li>1. Memungkinkan adanya keterkaitan antar item</li> <li>2. Memungkinkan munculnya <i>social desirability</i> dan <i>faking</i></li> </ol>

Banyak dijumpai penelitian-penelitian untuk mengetahui karakteristik psikometrik suatu instrumen. Namun, sebagian besar dilakukan tanpa proses penskalaan. Karena data dalam proses penskalaan berbeda dari data sebelumnya, maka karakteristik psikometrik dengan data sebelum dan setelah proses penskalaan dimungkinkan akan berbeda. Karakteristik psikometrik pada data setelah diskalakan tentunya akan lebih akurat daripada sebelumnya. Demikian pula dengan perbedaan metode atau cara yang digunakan dalam proses penskalaan akan berakibat pada perbedaan karakteristik psikometrik yang didapatkan.

Bentuk Instrumen yang mengukur aspek psikis manusia banyak digunakan dalam berbagai bidang ilmu terutama dalam bidang psikologi, bimbingan konseling, ekonomi, kesehatan, agama, pendidikan dan berbagai ilmu social lainnya. Bentuk instrumen ini

mengungkap aspek non kognitif yang diungkap dengan pendekatan non tes, yang berbeda dengan pendekatan tes. Pendekatan instrument ini bukan menunjukkan jawaban benar atau salah pada suatu item, tetapi menunjukkan derajat keberpihakan, kecenderungan, kesenangan, kesesuaian akan aspek tertentu. Sebagaimana disampaikan (Cronbach, 1970), Pengukuran aspek psikis manusia dapat dilakukan dengan tes performance tipikal, yang memiliki metode yang berbeda dengan pengukuran aspek kognitif yang dikenal dengan tes performance maksimal.

Perkembangan dalam pengukuran pendidikan menunjukkan perkembangan pula dalam bentuk atau format instrument, metode dan cara analisis instrument. Berbagai penelitian menunjukkan masing-masing bentuk instrument memiliki karakteristik yang berbeda sehingga memiliki kelebihan dan kekurangannya masing-masing. Melalui tulisan ini penulis akan memaparkan cara menganalisis instrument non tes dan membandingkan metode maupun format instrument yang memiliki karakteristik yang berbeda.

Metode penskalaan aspek psikis sudah dilakukan oleh Thurstone dan Likert. Penggunaan skala Likert dan Thurstone ini sering digunakan dalam berbagai penelitian. Respon pada skala Likert ini diskor secara kasar dari 1 hingga 5, sedangkan pada skala Thurstone judgment diminta meletakkan stimulus berdasar skala 1 hingga 7. Perbedaan proses penskalaan ini berakibat pada karakteristik instrument. Makalah ini bertujuan untuk mengembangkan instrument multiple intelligences dengan dua metode dan mamaparkan hasil karakteristik psikometris kedua bentuk instrument

## **B. Berbagai Metode dalam Instrument Non Tes**

Dalam proses penskalaan instrument non tes ada metode yang digunakan :

1. Metode berbasis stimulus
2. Metode berbasis respon
3. Metode berbasis subjek

Dalam Proses penskalaan instrument berbasis stimulus, subjek atau responden diminta untuk memilih berbagai stimulus yang tepat yang paling sesuai dengan karakteristik dirinya. Beberapa contoh instrument yang dibuat dengan metode ini adalah:

- a. Urutkan berbagai Profesi berikut dari yang paling disukai :
1. Guru
  2. Ilmuan
  3. Pekerja sosial
  4. Pedagang
- b. Diantara 2 pernyataan berikut yang paling sesuai dengan diri anda adalah :
1. Belajar matematika
  2. Belajar musik
- c. Beri jawaban yang tepat pada masing-masing pernyataan :
1. Menyukai kegiatan olah raga                    ya                    tidak
  2. Mengerjakan tugas guru                    ya                    tidak

Proses penskalaan berbasis respon dibuat berdasar respon subjek atau respondent. Pernyataan dibuat berdasar respon subjek dan ia diminta memilih salah satu respon yang paling sesuai dengan dirinya. Model instrument yang paling sering digunakan misalnya :

- a. Saya senang pelajaran matematika    STS    TS    N    S    SS

STS            = sangat tidak sesuai

TS             = tidak sesuai

N              = netral

S              = sesuai

SS             = sangat sesuai

- b. Aktif bertanya didalam kelas    TP            J            KD            SR            SL

TP             = tidak pernah

J              = Jarang

KD = Kadang-kadang

SR = Sering

SL = Selalu

Proses Penskalaan berbasis subjek berusaha meletakkan subjek dalam suatu kategori tertentu. Contoh metode ini adalah skala asertifitas :

- a. Teman saya meminjam buku catatan saya yang tidak segera dikembalikan ;
  1. Saya biarkan hingga ada kesadaran ia sendiri untuk mengembalikan
  2. Saya tagih untuk segera mengembalikan
- b. Seorang ibu yang baru datang menerobos antrian pembelian tiket:
  1. Saya biarkan, mungkin ia ada urusan penting
  2. Saya tegur untuk mengantri di belakang

### **C. Penskalaan dengan metode Interval Tampak Setara dari Thurstone dan Sumative Rating dari Likert**

[Louis Leon Thurstone](#) dianggap sebagai “Bapak” penyusunan skala sikap. Sejak tahun 1920 ia dan beberapa teman sejawatnya menerbitkan artikel yang menjelaskan dasar-dasar pengukuran sikap dan menjelaskan metodologi pengukuran sikap. Publikasi Thurstone ini berpengaruh besar dalam perkembangan ilmu-ilmu sosial terutama yang terkait dengan pengukuran aspek non kognitif.

Thurstone mengembangkan tiga macam teknik penyusunan skala , yaitu 1) metode perbandingan pasangan ( *paired comparisons* ), 2). Metode interval tampak sama ( *equal appearing intervals* ), 3) metode interval suksesif ( *successive intervals* ). Ketiga metode ini menggunakan pendapat atau judgment suatu kelompok panel untuk member pendapat mengenai kemendukungan atau kepositifan pernyataan-pernyataan sikap terhadap objek tertentu. Metodenya ini dikenal dengan “The Law of Comparative Judgment”. Harga atau skor kemendukungan pada tiap pernyataan dihitung dari pendapat judgment. Pada metode paired comparison kemendukungan suatu subjek dihitung dengan membandingkan stimulus-stimulus yang dibuat berpasangan. Pada metode interval tampak sama subjek diminta menentukan setiap pernyataan dalam rentang

kategori sangat tidak mendukung hingga sangat mendukung. Sedangkan pada metode interval suksesif proses penskaan ini ditentukan dengan perhitungan statistik.

Sebuah instrument dengan menggunakan metode interval tampak setara disusun dengan membuat serangkaian pernyataan yang menggambarkan seluruh wilayah domain konstruk yang diukur. Setelah pernyataan dibuat, meminta penilai (judgment) untuk meletakkan pernyataan ke dalam kontinum berdasarkan favorabilitasnya. Biasanya kontinum tersebut berisi 7 poin (Mardapi, D. 2008). Setelah penilai meletakkan pernyataan dalam kontinum, selanjutnya merangkum hasil penilaian tersebut dalam skor criteria tertentu. Pada penelitiannya Thurstone merangkum hasil penilaian dengan menggunakan rumus median.

Contoh Instrument :

1. Mengikuti kegiatan bina vocal atau menyanyi

1	2	3	4	5	6	7
---	---	---	---	---	---	---

2. Mengikuti kegiatan bermain musik

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Skala Likert semula disebut *Skala linkert* pertama kali dikembangkan oleh **Rensis Linkert** pada tahun 1932 dalam mengukur sikap masyarakat. Skala ini dibuat untuk mengukur sikap atau pendapat seseorang. Selanjutnya, skala ini digunakan tidak hanya untuk mengukur pendapat, sikap saja tetapi juga digunakan untuk mengukur persepsi dan berbagai karakteristik psikis manusia lainnya.

Model ini disusun berdasar distribusi respon subjek. Respon subjek diberikan dalam taraf kesetujuan atau ketidak setujuan dalam berbagai variasinya, misalkan : sangat setuju (SS), setuju (S), tidak tentu (N), tidak setuju (TS), sangat tidak setuju (STS). Dalam perkembangan selanjutnya respon subjek ini tidak hanya berbentuk taraf kesetujuan, tetapi berkembang dalam berbagai bentuk, misalnya :

- a. taraf kesesuaian : sangat sesuai, sesuai, netral, kurang sesuai, tidak sesuai.
- b. Taraf keyakinan : sangat yakin, yakin, ragu-ragu, kurang yakin, tidak yakin

- c. Frekwensi : selalu, sering, kadang-kadang, jarang , tidak pernah
- d. Ketertarikan : sangat suka, suka, biasa-biasa, kurang suka, tidak suka

Contoh instrument Likert :

No	Item	Respon				
		Selalu	Sering	Kadang -kadang	Jarang	Tidak Pernah
1.	Menuangkan ide, pemikiran, khayalan dalam suatu tulisan					
2.	Berdiskusi untuk memecahkan masalah					

#### D. Analisis Instrumen Non Tes

Kualitas suatu instrumen terkait dengan validitas dan reliabilita instrumen tersebut. Validitas alat ukur/tes menunjukkan bagaimana suatu alat ukur benar-benar dapat mengukur apa yang akan diukur atau diketahui. Misalnya sebuah tes inteligensi atau kecerdasan dikatakan valid apabila tes tersebut benar-benar dapat mengukur kecerdasan seseorang dan dapat mengukur kecerdasan orang yang berbeda-beda. Begitu pula dengan tes kepribadian dikatakan valid apabila dapat mengukur karakteristik kepribadian seseorang yang berbeda-beda antara satu individu dengan individu lain.

Validitas dapat diukur dengan berbagai cara. Pengukuran validitas tergantung pada maksud dan tujuan dalam pembuatan alat ukur. Secara umum validitas dibedakan dalam tiga macam, yaitu : validitas content/isi, validitas kriteria dan validitas konstruk

##### 1. Validitas Isi,

Validitas isi dikembangkan untuk mengetahui bagaimana fakta atribut yang diukur dalam tes sesuai dengan performance yang dikembangkan pada item-item dalam tes. Validitas ini membutuhkan analisis rasional dari seseorang yang ahli dalam bidang yang dikembangkan alat ukur tersebut atau professional judgment.

Validitas isi dibedakan menjadi validitas muka dan validitas logik. *validitas muka* didasarkan pada penampilan tes. Tes yang penampilannya bagus dilihat dari cover, kertas, dan tulisannya dikatakan memiliki validitas muka yang tinggi. Validitas muka ini sering digunakan karena alasan kepraktisannya, misalnya pada tes hasil belajar baik tes semester maupun ulangan-ulangan harian. Tes-tes yang lain juga biasanya dilihat validitas mukanya disamping validitas yang lainnya

*validitas logik* didasarkan pada sejauhmana isi tes dapat menggambarkan domain perilaku atau atribut yang hendak diukur dalam tes tersebut. Untuk mendapatkan validitas logik yang tinggi suatu tes harus dirancang berdasar suatu atribut tertentu yang dijelaskan dalam suatu indikator-indikator yang kemudian dijelaskan dalam item-item yang benar-benar menggambarkan atribut tersebut. Dalam tes prestasi atau hasil belajar validitas isi dilihat dari kesesuaian item dengan pokok-pokok bahasan yang dikembangkan dalam silabus.

Analisis validitas isi dilakukan secara kualitatif dan kuantitatif. Secara kualitatif dilihat dari berbagai coretan, masukan, kalimat-kalimat yang belum tepat diminta memperbaiki. Sedangkan secara kuantitatif dapat dilakukan dengan analisis Content Validity Ratio atau CVR (Lawshe, dalam Shults, 2005).

## **2. Validitas Konstruk,**

Validitas ini menunjukkan sejauhmana tes mengungkap suatu konstruk teori yang diukur atau suatu trait yang dikembangkan dalam pengukuran. Kontruk teori ini ada berbagai macam tergantung pada ahli yang mengembangkannya. Pengujian validitas konstruk membutuhkan teknik analisis statistik yang lebih kompleks daripada pengujian validitas yang lain. Metode yang diakui para ahli untuk mengukur validitas konstruk adalah analisis faktor.

Analisis faktor adalah teknik untuk menyederhanakan atau meringkas data dengan mereduksi jumlah variabel atau dimensi yang diperlukan. Analisis faktor digunakan untuk mengurangi pengukuran dan tes-tes yang beragam supaya menjadi sederhana. Analisis faktor membantu menemukan dan mengidentifikasi keutuhan-keutuhan dan sifat-sifat fundamental yang melandasi pengukuran tersebut. Suatu faktor adalah sebuah konstruk, suatu hipotesis, variabel laten yang dianggap melandasi tes, skala, butir, dan pada semua jenis pengukuran. Tujuan analisis faktor :

- mengidentifikasi faktor-faktor yang tidak langsung tampak, tetapi mendasari variabel yang tampak saling berkaitan.
- mengetahui konstruk suatu instrumen atau kelompok variabel
- menguji hipotesis mengenai eksistensi konstruk (CFA)

Analisis faktor dibedakan menjadi 2 : analisis faktor explanatory dan konfirmatori. Analisis faktor explanatory dilakukan untuk mengetahui atau mencari konstruk suatu instrument, sedangkan analisis faktor konfirmatori digunakan apabila konstruk suatu instrument sudah jelas dan peneliti ingin menguji hipotesis (kebenaran) konstruk yang dibuat. Prosedur analisis faktor lebih efisien dilakukan dengan program komputer, dengan menggunakan SPSS, SAS, SPS, Statistika, tesfact, MPlus dll pada analisis factor exploratory dan menggunakan program Lisrell, Amos, MPlus, untuk analisis factor konfirmatori yang memenuhi asumsi parametric dan PLS, GSCA, MPlus untuk yang non parametric.

### **3. Validitas Kriteria**

Validitas kriteria ini menunjukkan hubungan antara skor tes dengan sebuah kriteria tertentu. Kriteria tersebut merupakan perilaku yang dapat digunakan untuk memprediksi skor tes. Kriteria yang dimaksudkan dapat berupa alat tes lain yang standar yang digunakan mengukur atribut yang sama yang disebut Validitas Konkuren ataupun kriteria prestasi yang akan datang yang menunjukkan performansi yang diukur, misalnya IP, nilai ujian akhir, disebut Validitas Prediktif.

Validitas kriteria dihitung berdasarkan korelasi antara skor tes yang dibuat dengan skor kriteria yang dihubungkan. Semakin tinggi korelasinya menunjukkan keterkaitan antara skor hasil tes dengan kriteria yang diharapkan, sehingga menunjukkan ketepatan alat ukur tersebut dalam mengukur atribut tertentu.

Validitas prediktif sangat penting dilakukan apabila tes yang diukur digunakan sebagai prediktor bagi perilaku atau performance diwaktu yang akan datang. Beberapa tes yang digunakan memprediksikan perilaku seseorang dimasa yang akan datang misalnya tes dalam bimbingan karir, seperti tes bakat dan tes minat, serta tes seleksi masuk sekolah, tes seleksi masuk dunia kerja maupun tes penempatan karyawan.

Pengujian validitas presiktif membutuhkan selangwaktu tertentu, misalnya satu semester atau satu tahun setelah siswa atau karyawan menunjukkan perilaku performansinya yang cenderung menetap. Sebagai contoh validitas prediktif tes bakat

diukur dari korelasi antara hasil tes bakat bahasa dengan hasil prestasi siswa dalam bidang bahasa setelah waktu tertentu misalnya 1 tahun.

### E. Contoh analisis validasi isi oleh expert judgment

Review item dilakukan dengan meminta masukan dari 3 orang ahli/ expert Judgment. Dua orang ahli dalam psikologi yaitu : Ibu Yulia Ayriza (A) dan Ibu Rita Eka Izzaty (B) dan seorang ahli dalam statistika pendidikan yaitu Bapak Ali Hasmy (C). Review item ini dilakukan untuk mengetahui validitas isi instrument. Hasil penilaian dari ahli ini selanjutnya dianalisis secara kuantitatif dan kualitatif. Secara kuantitatif dianalisis dengan content validity rasio atau CVR (Lawshe, dalam Shults, 2005). Dengan menggunakan persamaan sebagai berikut :

$$CVR = \frac{ne - \frac{N}{2}}{\frac{N}{2}}$$

ne = number of subject matter experts (SMEs) rating the item that essential

N = total number of SMEs

Hasil dari penilain dan perhitungan dengan CVR pada tiap item disajikan pada tabel berikut :

No	Pertanyaan	Penilaian			CVR
		A	B	C	
1.	Kesesuaian item 1 dengan indikator membaca	2	2	3	-0.33
2.	Kesesuaian item 2 dengan indikator diskusi	3	2	3	0,33
3.	Kesesuaian item 3 dengan indikator matematika	3	3	3	1
4.	Kesesuaian item 4 dengan indikator penalaran/logika	2	2	3	-0.33
5.	Kesesuaian item 5 dengan indikator melukis	2	3	3	0.33
6.	Kesesuaian item 6 dengan indikator memahat	2	3	3	0.33
7.	Kesesuaian item 7 dengan indikator menyanyi	2	2	3	-0.33
8.	Kesesuaian item 8 dengan indikator bermain music	2	2	3	-0.33
9.	Kesesuaian item 9 dengan indikator olah raga	2	2	3	-0,33
10.	Kesesuaian item 10 dengan indikator menari	2	2	3	-0,33
11.	Kesesuaian item 11 dengan indikator introspeksi	2	3	3	0.33
12.	Kesesuaian item 12 dengan indikator pengembangan pribadi	2	3	3	0.33

13.	Kesesuaian item 13 dengan indikator bergaul	3	3	3	1
14.	Kesesuaian item 14 dengan indikator bekerja sama	2	3	3	0.33
15.	Kesesuaian item 15 dengan indikator berkebun	3	3	3	1
16.	Kesesuaian item 16 dengan indikator beternak	3	3	3	1

1= tidak relevan

2= agak relevan

3= relevan

Nilai CVR bergerak dari -1 hingga 1. Nilai 1 menunjukkan bahwa item tersebut essential (dalam instrument ini relevan) dan dibawah 0 menunjukkan kurang essential (kurang relevan). Item yang memiliki nilai CVR dibawah 0 selanjutnya dievaluasi secara kualitatif berdasar masukan ahli dan diubah menjadi item berdasar masukan tersebut. Disamping hasil diatas, beberapa masukan secara kualitatif dari ahli diantaranya :

- a. Perlu dikaji lagi terkait makna inteligensi dari Gardner dan kesesuaian dengan model instrument yang dikembangkan.
- b. Perlu dikaji apakah mahasiswa yang sering melakukan kegiatan pada salah satu inteligensi menunjukkan bahwa inteligensinya yang tinggi pada bidang tersebut.
- c. Beberapa masukan terkait dengan redaksi kalimat, item yang perlu dibuat lebih umum

Setelah mendapat berbagai masukan dari expert judgment penulis memperbaiki instrumen yang ada dan melakukan ujicoba instrumen

## **F. Ujicoba**

Ujicoba dilakukan pada 115 mahasiswa UNY yang bersal dari Fakultas Ilmu Pendidikan jurusan Bimbingan Konseling dan Teknologi Pendidikan, dan mahasiswa fakultas Bahasa dan Seni jurusan Seni Musik. Ujicoba ini dilakukan pada jam kuliah dengan menekankan pada mahasiswa untuk mengisi instrument yang diberikan dengan seadanya/sejujurnya dan tidak menuliskan identitas diri. Pengambilan data dilakukan pada bulan Nopember 2010.

### G. Analisis Kontruk Intrument

Uji validitas konstruk dilakukan dengan analisis factor confirmatory. Analisis ini dilakukan dengan alasan dalam pengembangan instrument ini penulis menggunakan teori yang sudah cukup terkenal dan sering digunakan yaitu teori Multiple Intelligence dari Howard Gardner. Salah satu program yang semula akan digunakan untuk analisis factor confirmatory adalah program Lisrel. Asumsi mendasar yang harus dipenuhi dalam program Lisrel adalah Normalitas. Hasil uji normalitas kedua model instrument tersebut dapat dilihat pada table berikut :

Type	Skewness			Kurtosis			Skewness and Kurtosis	
	Value	Z-Score	P-Value	Value	Z-Score	P-Value	Chi-Square	P-Value
Thurstone	55.133	4.890	0.000	294.754	2.329	0.020	29.339	0.000
Likert	55.064	3.874	0.000	294.665	2.304	0.021	20.318	0.000

Hasil diatas menunjukkan bahwa uji normalitas tidak terpenuhi baik pada kedua type instrument ( $P < 0.05$ ). Ketidaknormalan data ini kemungkinan disebabkan karena subjek penelitian yang kurang untuk menggunakan program analisis Lisrel (idealnya 5-10 kali parameter yang diujikan). Disamping itu, Curren et al, dalam Ghazali, 2008) memberikan judgment mengenai normalitas data dengan membagi jenis distribusi data menjadi 3 kriteria berdasar skor z sebagai berikut :

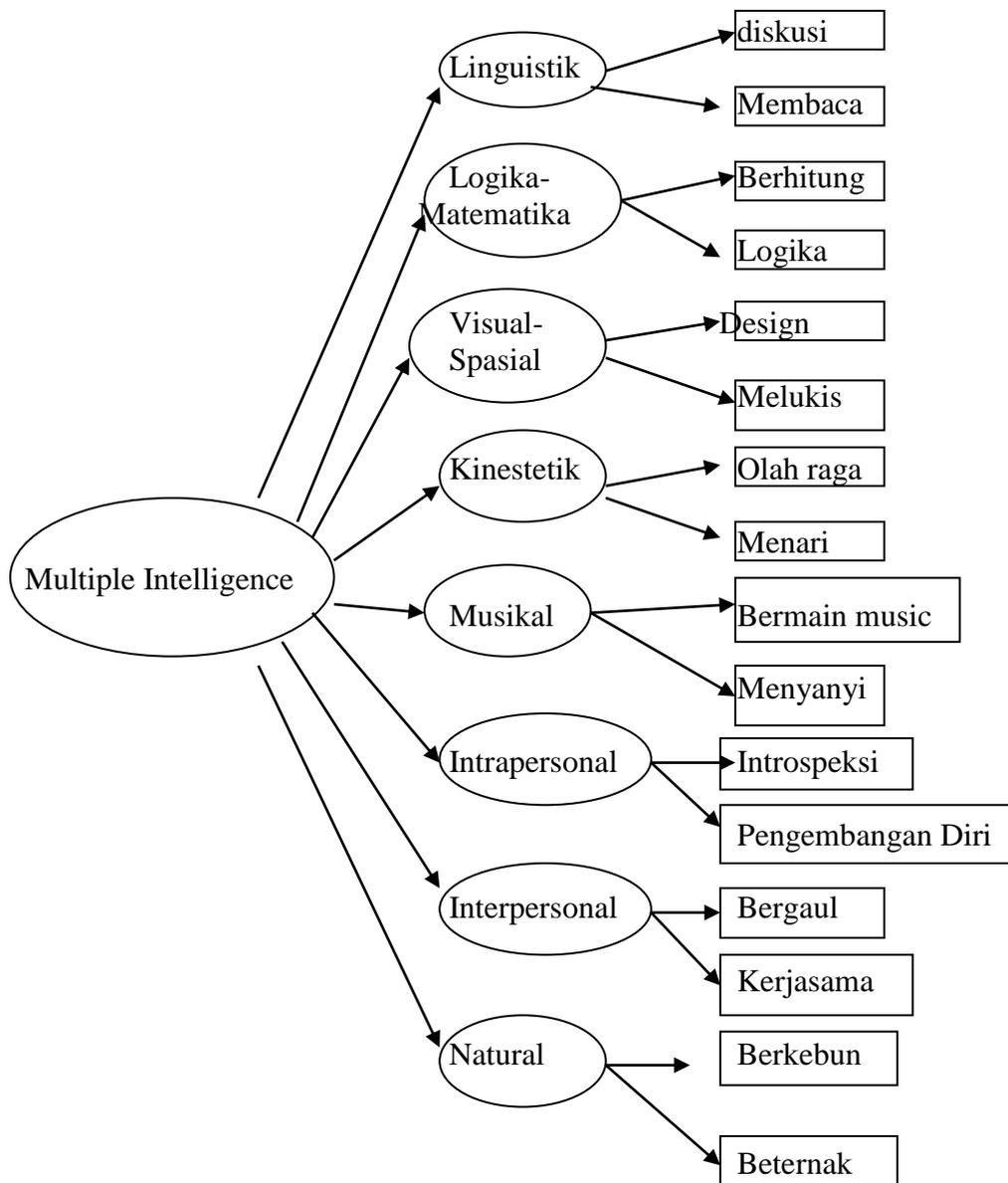
	Skewness	Kurtosis
Normal	< 2	< 7
Moerately Non Normal	2-3	7-21
Extremely Non Normal	> 3	> 21

Dengan melihat criteria dan hasil uji normalitas menunjukkan bahwa skor kurtosis keduanya normal, namun skewness extremely tidak normal. Karena data tidak memenuhi asumsi normalitas multivariate, adalisis konstruk instrument tidak menggunakan Lisrel namun dilakukan dengan menggunakan program SmartPLS versi 1.10. SmartPLS merupakan software yang dikembangkan sebagai proyek di Institute of Operation

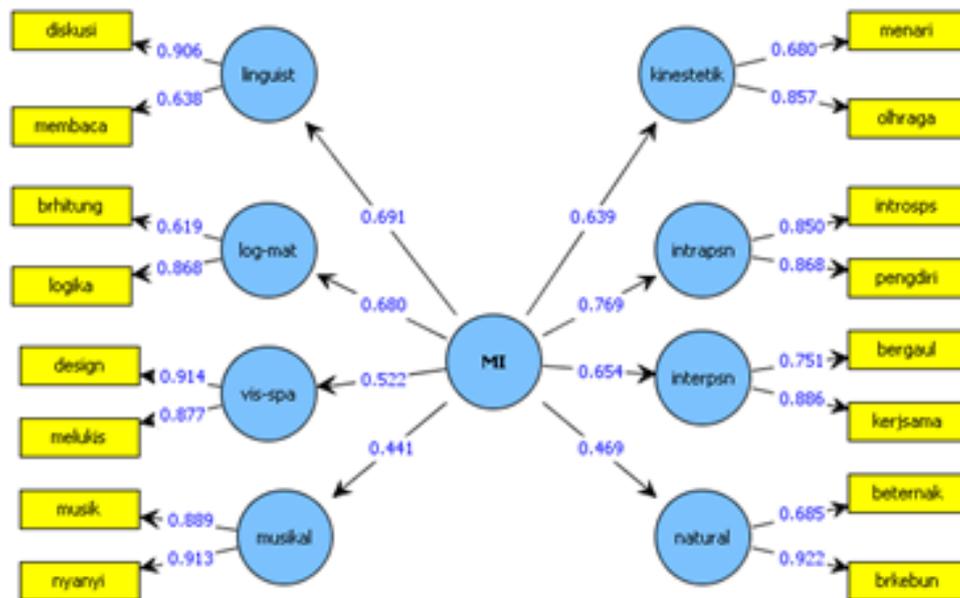
Management and Organization (School of Business) University of Hamburg yang berlokasi di Jerman. Penekatan yang digunakan oleh software ini “distribution free” atau tidak mengasumsikan data berdistribusi tertentu, bias berupa nominal, ordinal, interval dan rasio (Ghozali, I., 2008)

### 1. Kontruk Instrumen type Thurstone

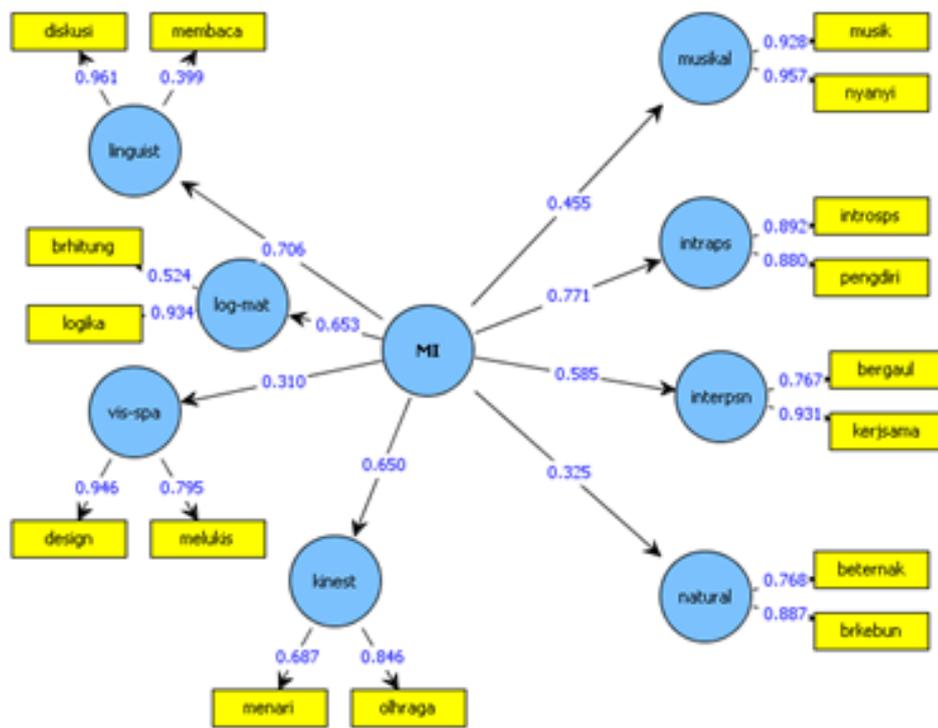
Model diagram jalur yang dikembangkan pada penelitian ini sebagaimana dijelaskan pada penyusunan konstruk dan spesifikasi instrument dapat dirumuskan sebagai berikut :



Berdasar model awal sebagaimana gambar diatas dilakukan analisis dengan program smart PLS dihasilkan model beserta skor loading pada masing-masing variable pada instrument type Thurstone dan type Likert sebagaimana gambar berikut :



**Gambar : Model Second Order Reflektif pada Instrumen type Thurstone**



**Gambar : Model Second Order Reflektif pada Instrumen type Likert**

Hasil diatas menunjukkan bahwa meskipun model awal yang dikembangkan sama namun mendapatkan hasil loading yang bervariasi pada tiap indikator maupun skor variable laten. Hasil loading pada tiap item pada data asli maupun data bootstrapping yang dapat dilihat dari mean of sample pada data instrument type Thurstone disajikan pada tabel berikut :

Hasil Outer loadings [ CSV-Version ] pada data type Thurstone

	original sample estimate	mean of subsamples	Standard deviation	T-Statistic
Diskusi	0.906	0.918	0.045	20.090
Membaca	0.638	0.564	0.221	2.887
Brhitung	0.619	0.617	0.191	3.233
Logika	0.868	0.840	0.118	7.359
Design	0.914	0.912	0.042	21.831
Melukis	0.877	0.866	0.107	8.179
Music	0.889	0.826	0.304	2.928
Nyanyi	0.913	0.876	0.242	3.773
Menari	0.680	0.643	0.211	3.225
Olahraga	0.857	0.845	0.098	8.777

Introsps	0.850	0.839	0.064	13.201
Pengdiri	0.868	0.850	0.071	12.301
Bergaul	0.751	0.709	0.167	4.492
Kerjsama	0.886	0.903	0.042	21.240
Beternak	0.685	0.610	0.320	2.143
Brkebun	0.922	0.902	0.108	8.506

Hasil analisis dengan menggunakan bootstrapping pada instrument type Thurstone diatas menunjukkan pada data asli (original) didapatkan factor loading pada semua item diatas 0.5 (factor loading dibawah 0.5 tidak disarankan digunakan) dengan demikian factor loading untuk first order sudah memenuhi convergent validity karena nilainya semua di atas 0.5.

Hasil for inner weights [ CSV-Version ] pada data type Thurstone

	original sample estimate	mean of subsamples	Standard deviation	T-Statistic
MI -> linguist	0.691	0.707	0.057	12.086
MI -> log-mat	0.680	0.684	0.087	7.863
MI -> vis-spa	0.522	0.532	0.108	4.816
MI -> musikal	0.441	0.443	0.146	3.017
MI -> kinestetik	0.639	0.650	0.085	7.520
MI -> intrapns	0.769	0.747	0.104	7.393
MI -> interpsn	0.654	0.666	0.082	7.988
MI -> natural	0.469	0.492	0.118	3.969

Hasil dari divergent validity diatas menunjukkan skor loading yang tinggi pada hampir semua dimensi, kecuali dimensi musikal dan natural. Apabila dilihat dari hasil analisis uji T menunjukkan semua latent variable yang signifikan atau mendukung karena memiliki nilai T statistik diatas 1.96.

Analisis pada data type Likert dengan menggunakan model yang sama sebagaimana type Thurstone didapatkan hasil output sebagaimana berikut :

Hasil Outer loadings [ CSV-Version ]

	original sample estimate	mean of subsamples	Standard deviation	T-Statistic
Diskusi	0.961	0.947	0.048	19.836
membaca	0.399	0.308	0.363	1.100

Brhitung	0.524	0.398	0.356	1.472
Logika	0.934	0.907	0.146	6.416
Design	0.946	0.724	0.524	1.804
Melukis	0.795	0.528	0.628	1.265
Menari	0.687	0.654	0.191	3.600
Olhraga	0.846	0.842	0.144	5.874
Music	0.928	0.867	0.327	2.837
Nyanyi	0.957	0.905	0.315	3.044
Introsps	0.892	0.901	0.030	29.856
Pengdiri	0.880	0.867	0.070	12.485
Bergaul	0.767	0.747	0.119	6.445
kerjsama	0.931	0.932	0.030	31.432
Beternak	0.768	0.465	0.638	1.204
Brkebun	0.887	0.594	0.579	1.533

Hasil analisis diatas menunjukkan ada satu item yang memiliki factor loading yang rendah yaitu item membaca (0.399). Apabila dilihat dari skor T pada masing-masing item didapatkan 6 item yang memiliki nilai T rendah atau kurang dari 1.96. Dengan demikian konstruk Multiple inteligensi dengan data yang dilakukan bootstrapping tidak didukung oleh item membaca, berhitung, design, melukis, beternak, berkebun.

Hasil analisis pada masing-masing variable laten didapatkan skor loading cukup tinggi, kecuali pada variable visual spatial, musical, dan natural. Meskipun data T statistiknya lebih dari 1.96. Hasil tersebut menunjukkan variable laten mendukung instrument ini ada beberapa yang memiliki nilai loading yang kurang.

Hasil inner weights [ CSV-Version ]

	original sample estimate	mean of subsamples	Standard deviation	T-Statistic
MI -> linguist	0.706	0.702	0.089	7.952
MI -> log-mat	0.653	0.669	0.098	6.661
MI -> vis-spa	0.310	0.362	0.134	2.310
MI -> kinest	0.650	0.647	0.095	6.826
MI -> musikal	0.455	0.505	0.146	3.109
MI -> intraps	0.771	0.758	0.092	8.381
MI -> interpsn	0.585	0.593	0.137	4.260
MI -> natural	0.325	0.349	0.127	2.562

### **R-Square dan Relibilitas pada tiap Indikator**

Instrumen-instrumen non kognitif banyak mengembangkan dimensi yang lebih dari 1. Karena penskalaan pada instrument diatas bersifat multidimensi maka reliabilitasnya dapat dilihat pada masing-masing dimensi. Hasil dari reliabilitas komposit pada tiap dimensi dan R-square (sumbangan pada masing-masing variable atau dimensi terhadap instrument) pada kedua instrument dapat dilihat pada table dibawah ini.

	R-square		Reliabilitas	
	Thurstone	Likert	Thurstone	Likert
linguist	0.478	0.498	0.756	0.669
log-mat	0.463	0.426	0.719	0.714
vis-spa	0.272	0.096	0.890	0.865
musikal	0.194	0.422	0.896	0.743
kinestetik	0.408	0.207	0.746	0.941
intrapsn	0.592	0.594	0.849	0.879
interpsn	0.427	0.343	0.805	0.841
natural	0.220	0.106	0.791	0.815
MI			0.837	0.798

Besarnya reliabilitas komposit tiap dimensi dan R-squarenya menunjukkan hasil yang bervariasi pada kedua type instrument. Namun secara umum menunjukkan skor reliabilitas komposit dan reliabilitas yang lebih tinggi pada instrument type Thurstone dibanding type Likert.

### **H. KESIMPULAN DAN IMPLIKASI**

Instrument multiple inteligensi yang peneliti buat telah memenuhi karakteristik karakteristik psikometrik yang baik pada validitas dan reliabilitas pada metode Thurstone maupun Likert. Pada metode Thurstone menunjukkan karakteristik yang lebih baik daripada Likert. Implikasi hasil tersebut menunjukkan bahwa rentang penskalaan turut mempengaruhi karakteristik instrument. Instrument tipe Thurstone yang memiliki rentang penskalaan lebih banyak daripada Likert mendapatkan hasil karakteristik instrument yang lebih detail.

## SUMBER REFERENSI

- Armstrong, T. 1993. *7 Kinds of Smart : Identifying and Developing Your Many Intelligences*. New York : A Plume Book.
- Azwar, S., 1995. *Sikap Manusia – Teori dan Pengukurannya*. Yogyakarta ; Pustaka Pelajar
- Azwar, S., 1999. *Penyusunan Skala Psikologis*. Yogyakarta ; Pustaka Pelajar
- Allen, M., J., Yen, W., M. 1979. *Introduction to Measurement Theory*. Monterey : Brooks/Cole Publishing Company
- Eagly, A.H., Chaiken, S. 1993. *The Psychology of Attitudes*. Orlando ; Harcourt BBrace Jovanovich, Inc.
- Crocker L, Algina J. 1986. *Introduction to Classical and Modern Tes Theory*. New York : CBS College Publishing.
- Cronbach, L.,J., 1970. *Essentials of Psychological Testing (3<sup>rd</sup> edition)*, New York : Harper &Row Publisher.
- Gardner, H. 2003. *Multiple Intelligences : Kecerdasan Majemuk dalam Praktik* (alih bahasa Sindoro A. Batam : Interaksara.
- Ghazali, I., Fuad. 2008. *Structural Equation Modeling, Teori, Konsep, dan Aplikasi dengan Program Lisrel 8.8*. Semarang; Badan Penerbit Universitas Diponeoro
- Ghazali, I., 2008. *Stuktural Equation Modeling Metode Alternatif dengan Partial Least Square (PLS)*. Semarang; Badan Penerbit Universitas Diponeoro
- Mardapi, D., 2008. *Teknik Penyusunan Instrumen Tes dan NonTes*. Yogyakarta : Mitra Cendekian Offset

Shultz, K.S., Whitney, D.J., 2005. *Measurement Theory in Action, Case Studies and Exercises*. California; Sage Publication, Inc

Suryabrata, S., 2002. *Pengembangan Alat Ukur Psikologis*. Yogyakarta : Penerbit Andi Offset

Thurstone, L.L. 1927. A Law of Comparative Judgment. *Psychological Review*. 34, 273-286

Togerson, W.S., 1958. *Theory and Methods of Scaling*. New York; Wiley.