

796 / Pengukuran dan Evaluasi Pendidikan

**LAPORAN
PENELITIAN DISERTASI DOKTOR**



**PENSKALAAN TIPE LIKERT DAN THURSTONE DENGAN TEORI
KLASIK DAN MODERN :
STUDI PADA INSTRUMEN *MULTIPLE INTELLIGENCES***

**PENGUSUL
Farida Agus Setiawati, M.Si
0013087201**

**Dibiayai oleh : DIPA UNY dengan Surat Perjanjian Penugasan dalam rangka Penelitian
Program Disertasi Doktor Tahun Anggaran 2013 Nomor: 532a/BOPTN/UN34.21/2013
Tanggal 27 Mei 2013**

**UNIVERSITAS NEGERI YOGYAKARTA
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
TAHUN 2013**

HALAMAN PENGESAHAN

Judul Kegiatan : Penskalaan Tipe Likert dan Thurstone dengan Teori Klasik dan Modern
: Studi pada Instrumen Multiple Intelligences

Peneliti / Pelaksana
Nama Lengkap : FARIDA AGUS SETIAWATI M.Si.
NIDN : 0013087201
Jabatan Fungsional :
Program Studi : Bimbingan Dan Konseling
Nomor HP : 081215536324
Surel (e-mail) : faridaagus@yahoo.co.id.

Institusi Mitra (jika ada)
Nama Institusi Mitra :
Alamat :
Penanggung Jawab :
Tahun Pelaksanaan : Tahun ke 1 dari rencana 1 tahun
Biaya Tahun Berjalan : Rp. 30.000.000,00
Biaya Keseluruhan : Rp. 0,00



Mengetahui
Dekan FIP
(Dr. Haryanto, M.Pd)

NIP/NIK 196009021987021001

Yogyakarta, 28 - 11 - 2013,
Ketua Peneliti,

(FARIDA AGUS SETIAWATI M.Si.)

NIP/NIK197208131998022001



Menyetujui,
Ketua UPPM

(Prof. Dr. Hanik Ghufro)

NIP/NIK 196211111988031001

DAFTAR ISI

| | |
|--|-----------|
| HALAMAN SAMPUL | |
| HALAMAN PENGESAHAN | |
| DAFTAR ISI | |
| ABSTRAK | |
| ABSTRACT | |
| BAB I. PENDAHULUAN | 1 |
| A. Latar Belakang | 1 |
| B. Identifikasi Masalah | 5 |
| C. Tujuan Penelitian | 6 |
| D. Urgensi Penelitian | 6 |
| | |
| BAB II. KAJIAN PUSTAKA | 7 |
| A. Penskalaan | 7 |
| B. Penskalaan Tipe Thurstone | 8 |
| C. Penskalaan Tipe Likert | 9 |
| D. Karakteristik Psikometrik Penskalaan pada Pendekatan Klasik | 10 |
| E. Penskalaan dengan Pendekatan Teori Modern atau IRT | 10 |
| F. Kerangka Berpikir | 11 |
| G. Pertanyaan Penelitian | 14 |
| | |
| BAB III. METODE PENELITIAN | 15 |
| A. Pendekatan Penelitian | 15 |
| B. Subjek Penelitian | 15 |
| C. Alur Penelitian | 16 |
| D. Pengembangan Instrumen | 17 |
| E. Validitas Isi | 17 |
| F. Ujicoba Instrumen | 18 |
| G. Seleksi Butir | 18 |
| H. Proses Penskalaan | 18 |
| I. Penskalaandengan Pendekatan Klasik | 18 |
| J. Penskalaandengan Pendekatan Modern (IRT) | 19 |
| | |
| BAB IV. HASIL PENELITIAN DAN PEMBAHASAN | 20 |
| A. Hasil Penskalaan dengan Pendekatan Klasik | |
| A.1. Penskalaan <i>Paired Comparison</i> | 20 |
| A.2. Penskalaan <i>Summated Rating</i> | 21 |
| B. Diskripsi Data Kemampuan (Θ) Subjek Penelitian sebelum dan setelah Ditransformasi | 22 |
| C. Dimensi Instrumen | 24 |
| D. Analisis Reliabilitas | 24 |
| E. Penskalaan dengan Pendekatan Modern (IRT) | 28 |
| F. Estimasi Parameter | 28 |
| G. Fungsi Informasi Butir | 39 |

| | |
|-------------------|----|
| BAB V. KESIMPULAN | 33 |
| A. Kesimpulan | 33 |
| B. Saran | 34 |
| | |
| DAFTAR PUSTAKA | 35 |

ABSTRAK

Issue tentang penskalaan muncul terkait dengan makna skor hasil pengukuran aspek psikis manusia memiliki karakteristik data 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. Tujuan penelitian ini dapat dirumuskan sebagai berikut : 1). Melakukan proses penskalaan menggunakan pendekatan klasik dan modern pada tipe Thurstone dan Likert. 2) membandingkan karakteristik psikometrik instrumen kedua tipe pada data asli dan data yang diskalakan. 3) Membandingkan karakteristik psikometrik pada kedua tipe data yang sudah diskalakan dengan menggunakan pendekatan klasik dan modern

Secara umum penelitian ini menggunakan pendekatan penelitian kuantitatif yang dalam pelaksanaannya terdiri dari beberapa bagian penelitian yang saling terkait, yaitu penelitian pengembangan instrumen, proses penskalaan pada data hasil ujicoba, analisis karakteristik psikometrik instrumen, dan perbandingan karakteristik psikometrik instrumen. Bagian pertama adalah penelitian pengembangan. Pada penelitian pengembangan dibuat instrumen *multiple intelligences* dengan menggunakan dua tipe, yaitu tipe Thurstone dan tipe Likert. Setelah diuji validitas isi dari ahli selanjutnya diperbaiki. Instrumen yang sudah teruji validitas isinya selanjutnya diujicobakan dilapangan. Bagian berikutnya adalah melakukan proses penskalaan. Proses penskalaan dilakukan pada data hasil ujicoba dilapangan dengan menggunakan pendekatan klasik dan modern. Setelah diskalakan akan dianalisis karakteristik psikometrik kedua tipe instrumen. Bagian terakhir dalam penelitian ini adalah membandingkan karakteristik psikometrik kedua tipe instrument secara diskriptif dan inferensi.

Hasil penskalaaninstrumen MI pada pendekatan klasik dengan metode *paired comparison* pada tipe Thurstone didapatkan urutan skor dari yang rendah adalah logika matematika, musik, linguistik, kinestetik, naturalis , visual, interpersonal, eksistensial dan intrapersonal. Dengan metode summated rating pada instrumen tipe Thurstone dan Likert dihasilkan skor yang bervariasi pada tiap respon. Terdapat perubahan varian, rata-rata, reliabilitas dan kesalahan baku pengukuran (SEM) dari data asli dengan data yang ditransformasi menggunakan skor z pada penskalaan teori klasik. SEM pada data yang ditransformasi mengalami penurunan. Karakteristik psikometrik koefisien reliabilitas instrumen tipe Thurstone lebih rendah dibanding tipe Likert, baik pada keseluruhan data maupun data tiap dimensi, begitu pula dalam SEMnya. SEM pada instrumen tipe Thurstone lebih rendah dibanding tipe Likert. Hasil penskalaan teori IRT dengan menggunakan model respon berjenjang didapatkan hasil daya beda dan indeks kesulitan butir yang bervariasi pada kedua tipe instrumen. Butir-butir pada instrumen tipe Thurstone memiliki daya beda lebih rendah dan index kesulitan butir lebih tinggi dibanding butir-butir pada instrumen tipe Likert. Fungsi informasi butir maupun fungsi informasi tes pada instrumen tipe Likert lebih tinggi dibanding tipe Thurstone, sebaliknya kesalahan baku pengukuran atau SEM pada instrumen tipe Thurstone lebih tinggi dibanding tipe Likert. Dengan demikian instrumen tipe Likert lebih akurat dibanding tipe Thurstone dalam mengukur kecenderungan kemampuan subjek.

Kata Kunci :

Penskalaan, Thurstone, Likert, Multiple Intelligence

ABSTRACT

The study aimed: 1) to reveal the result the scaling of multiple intelligence (MI) instruments of Thurstone and Likert types using the classical and modern approaches, 2) to reveal the characteristics of multiple intelligence instruments of Thurstone and Likert types in the original data and the scaled data, 3) to compare the psychometric characteristics of the two scaled types of data using the classical and modern approaches.

The study generally used the quantitative research approach. The activity consisted of four interrelated parts: developing research instrument, processing the data scaling, analyzing the psychometric characteristics of the instrument, and comparing the characteristics of psychometric instruments. The MI instrument was developed using Thurstone and Likert types in the same constructs. The instrument was validated experts and tried out. The psychometric characteristics of the two types of data of instrument were compared using the classical and modern approaches. The comparison of both types were analyzed by descriptive statistics.

The result of scaling data using the classical approach with paired comparison method of Thurstone type are the sequential scores stimulus, there are logical-mathematic, musical, linguistic, kinesthetic, naturalist, visual-spatial, interpersonal, existential and intrapersonal. The summated rating method of Thurstone and Likert type produce the standartdize scores from lowerst to highest in each response items. There are changes of skor, variant, and standard error of measurement (SEM) on the original data to the scaled data on the classical scaling. The SEM of data instrument that have hight reliability of scaled data is lower than original data. The reliability coefficient and SEM of Thurstone type is lower than that of the Likert type, because it has data variant lower than Likert type. The analyze of the psychometric characteristics of modern scaling using item response theory are index of discrimination, index of difficulty items, item and test information function. The index of discrimination item of Thurstone type is lower than that of Likert type . The index of difficulty item of Thurstone type is higher than that of Likert type. The item information function and test information function on a Likert type are higher than those on Thurstone type and SEM of Thurstone type is higher than that of Likert type. So, the instrument MI of Likert type more precise than that of Thurstone type

Key Word: scaling, multiple intelligence, Thurstone type, Likert type

BAB I PENDAHULUAN

E. Latar Belakang

Pengukuran aspek psikis manusia banyak digunakan dalam pengembangan ilmu pengetahuan melalui berbagai penelitian dan penerapan ilmu di masyarakat. Dalam penggunaannya dibutuhkan instrumen atau alat ukur yang dapat mengungkap aspek atau atribut psikis tersebut. Pengukuran aspek psikis ini dibedakan menjadi kognitif dan nonkognitif. Instrumen yang mengungkap atribut nonkognitif dikenal dengan berbagai istilah, diantaranya : *questionnaire* dan *rating scale* yang diperkenalkan oleh Galton (Aiken, 1997: 5), inventori diperkenalkan oleh Aiken (1997: 9), dan tes performansi tipikal diperkenalkan oleh Cronbach (1984: 32). Cronbach juga menjelaskan bahwa pengukuran performansi tipikal ini memiliki metode yang berbeda dengan pengukuran aspek kognitif atau performansi maksimal. Dalam pengukuran performansi maksimal yang diungkap menunjukkan seberapa besar kemampuan manusia pada performansi atau aspek yang diukur, sedangkan dalam pengukuran performansi maksimal, informasi yang diungkap bukanlah seberapa besar kemampuan yang diungkap melainkan seberapa besar kecenderungan manusia pada performansi yang diungkap yang tampak dari karakteristik sikap, sifat, kepribadian atau karakteristik psikis manusia yang lain.

Salah seorang yang melaporkan pengukuran aspek nonkognitif adalah Thurstone dengan mencoba menerapkan pemikiran psikofisik dalam bidang sosial (1927, 384). Metodenya dikenal dengan *paired comparison* atau perbandingan pasangan. Dalam metode ini diperbandingkan 2 stimulus pada tiap item untuk dipilih yang paling sesuai dengan keadaan subjek. Metode ini selanjutnya menjadi dasar dalam pengukuran minat, sikap dan nilai dalam pendekatan stimulus. Dalam pendekatan ini penskalaan dilakukan pada stimulus. Model instrumen yang dikembangkan dengan metode *paired comparison* dalam penelitian ini selanjutnya disebut dengan instrumen tipe Thurstone.

Instrumen tipe Thurstone merupakan dasar pengukuran *forced choice*. Model instrumen ini meminta responden memilih satu dari sepasang atau lebih stimulus atau pernyataan yang diminta menanggapi. Model instrumen ini juga mengharuskan subjek untuk memilih satu pernyataan dari beberapa pernyataan yang disediakan. Pilihan subjek didasarkan pada kesesuaian dirinya, namun dalam model ini subjek tetap harus memilih

satu pernyataan meskipun pernyataan yang dipilih kurang menggambarkan keadaan dirinya atau semua pernyataan sesuai dengan keadaan subjek. Contoh lain model instrumen tipe ini adalah instrumen tipe ranking yang dalam analisisnya didasari metode perbandingan berpasangan dari Thurstone.

Pemilihan satu dari beberapa pernyataan dalam tipe ini menyebabkan tipe ini memiliki karakteristik yang berbeda dengan tipe instrumen yang menyajikan sebuah pernyataan dengan berbagai respon. Tipe instrumen ini akan memberi respon yang berbeda pada suatu stimulus, sehingga akan didapatkan respon yang bervariasi pada berbagai stimulus yang disajikan. Hal ini sejalan dengan pendapat Olivares & Brown (2009:935) yang menyebutkan bahwa tipe ini terhindar dari jawaban yang sama atau adanya bias dalam memberi respon seperti respon persetujuan yang ekstrim, atau kelemahannya dalam pemberian respon yang tidak bervariasi atau efek “halo”.

Instrumen tipe Likert muncul dalam jurnal *Archives of Psychology* yang berjudul *Technique for Measurement of Attitudes*. Model ini disusun berdasar distribusi respon subjek yang berbeda dengan model Thurstone dalam proses penskalaan. Dalam tipe ini respon subjek diberikan dalam taraf kesetujuan atau ketidaksetujuan, misalkan : sangat setuju atau *strongly approve*, setuju atau *approve*, tidak tentu atau *undecided*, tidak setuju atau *disapprove*, dan sangat tidak setuju atau *strongly disapprove*. Dalam perkembangan selanjutnya respon subjek ini tidak hanya berbentuk taraf kesetujuan, tetapi berkembang dalam berbagai format, misalnya: taraf kesesuaian, taraf keyakinan, frekwensi, ketertarikan dan skala rating lainnya. Jumlah respon dalam tipe ini juga berkembang dan bervariasi, yang semula tipe ini menggunakan 5 respon, berkembang menjadi 6,7 hingga 11 respon. Karena proses penskalaan tipe Likert dikenakan pada respon subjek dengan berbagai variasinya, dan kriteria yang digunakannya adalah respon, maka penskalaan model ini dapat dikatakan sebagai proses penskalaan dengan pendekatan respon.

Secara kuantitatif, pengukuran aspek nonkognitif ditandai dengan memberi skor atau angka pada hasil pengukuran. Skor menunjukkan kualifikasi atribut yang diukur. Makna skor hasil pengukuran dapat dibedakan dari *level of measurement* atau level pengukuran. Ada empat level pengukuran, yaitu nominal, ordinal, interval dan rasio. Data yang berupa angka pada level nominal berfungsi untuk mengkategorikan dan tidak menunjukkan level tertentu. Pada level ordinal angka menunjukkan tingkatan tertentu tetapi tidak ada satuan ukuran tertentu pada masing-masing tingkatan. Pada level interval disamping ada tingkatan

juga ada satu ukuran tertentu, dan pada level rasio disamping memiliki karakteristik sebagaimana data interval juga memiliki nilai nol mutlak.

Pemberian skor pada instrumen tipe Thurstone skor 1 pada pernyataan yang dipilih dan 0 pada pernyataan yang tidak dipilih pada tiap pasangan pernyataan. Skor 1 menunjukkan pernyataan yang dipilih yang menunjukkan ia cenderung pada aspek atau atribut yang dikembangkan dan 0 menunjukkan ketidaksesuaian pada atribut tersebut, sehingga dapat dimaknai skor 1 memiliki makna lebih tinggi dari 0 . Pada tipe Likert, respon subjek diskor dari 1 hingga 5 secara berurutan. Skor 1 lebih rendah dari 2, 2 lebih rendah dari 3, dan seterusnya. Urutan 1,2,3 hingga 5 berdasar pada kesesuaian diri subjek dan tidak ada satuan yang pasti jarak atau rentang antar satuan. Dengan melihat karakteristik tersebut data hasil pengukuran instrumen tipe Thurstone dan Likert merupakan data ordinal.

Data nominal dalam pengukuran merupakan data kategorik yang memiliki keterbatasan dalam operasi matematika, yaitu hanya dapat menggunakan modus dan tidak dapat menggunakan mean, median, maupun deviasi standart (SD). Begitu pula pada data ordinal, hanya dapat menggunakan modus dan median dan tidak dapat menggunakan mean dan SD. Keterbatasan pengukuran dengan menggunakan data tersebut tidak memenuhi asumsi parametrik yang terkait dengan data. Glenberg (1988: 144) menyatakan bahwa asumsi dalam analisis parametrik adalah data yang diperoleh berada pada level interval dan ratio. Dengan demikian data pada level ordinal tidak dapat dianalisis dengan statistik parametrik dan analisis yang dilakukan adalah nonparametrik.

Penggunaan data ordinal dalam analisis data parametrik menimbulkan perdebatan panjang dari para ahli. Hingga saat ini belum mendapatkan kesepakatan. James Carifio & Rocco Perla (2008: 1150) menyatakan hal yang sama terkait dengan adanya perdebatan panjang hingga 50 tahun tersebut tentang data yang didapatkan pada model pengukuran Likert. Perdebatan ini terkait dengan sumber informasi yang masing-masing independen. Jamieson memberi informasi bahwa karakteristik data pada skala likert merupakan karakteristik data ordinal, atau *rank order data*, sehingga harus dianalisis dengan non parametrik, yang kurang sensitif dan kurang powerful dibanding parametrik. Suryabrata, (1998: 146) menyatakan bahwa data yang diperoleh dalam pengukuran psikologi belum benar-benar data interval, tetapi diperlakukan sebagai data interval.

Penskalaan merupakan bagian yang mendasar dalam proses pembentukan teori pengukuran (Lord & Novick, 1968: 22). Teori pengukuran dapat digolongkan menjadi 2, yaitu teori klasik dan modern atau *item response theory* (IRT). Proses penskalaan dalam teori modern disebut dengan *item response theory* atau IRT. Dalam pendekatan IRT penskoran didasarkan pada sebuah model yang mengaitkan kemampuan orang (Θ) dengan probabilitas menjawab benar tiap butir (Hambleton & Swaminathan, 1985:35; Embretson & Rouse, 2000:67, Baker, Ho Kim, Seock, 2004:2, Demars, 2010:1). Model teori ini meletakkan karakteristik instrumen yang tidak terikat pada karakteristik kelompok atau sampel, namun pada tiap orang dan tiap butir. Dengan demikian, pendekatan teori modern ini dapat memberi estimasi kesalahan pada masing-masing orang dan butir (Hambleton et al., 1991:94)

Pollard, Dixon, Dieppe & Johnston, (2009:3) menjelaskan bahwa pendekatan IRT memiliki keuntungan dibanding CTT terkait dengan kesalahan sampel yang menggunakan sampel *invariant* sehingga parameter butir tidak tergantung pada tingkat kemampuan sampel. Pendekatan teori modern dianggap lebih cermat dan *powerful* dalam menilai karakteristik instrumen dibanding pendekatan teori klasik. Meskipun demikian pendekatan teori klasik ini masih banyak digunakan hingga sekarang karena kelebihanannya yang tidak ketat terhadap asumsi sehingga mudah sesuai dengan data lapangan (Lord & Novick, 1968: 25, Hambleton & Jones, 1993: 255).

Model IRT dengan berdasarkan pada bentuk data yang digunakan dapat dibagi menjadi dua yaitu model IRT dikotomus dan IRT politomus. Model dikotomus adalah model respon yang mempunyai 2 kemungkinan jawaban, yaitu benar yang diskor 1 dan salah diskor 0. Model IRT politomus adalah model IRT yang memiliki respon atau kemungkinan jawaban lebih dari dua kategori. Pada skala tipe Thurstone, tiap item disajikan sebagai stimulus yang harus dipilih oleh subjek. Stimulus apapun yang dipilih merupakan pilihan yang sesuai dengan minat, kesukaan, dominansi yang sesuai dengan karakteristik dirinya. Analisis skala ini terkait dengan masing-masing alternatif jawaban atau stimulus yang dipasangkan. Karena alternatif jawaban diberi skor 1 atau 0, maka data tersebut dikatakan sebagai data dikotomus. Dengan demikian model analisis IRT yang dianggap tepat untuk mengungkap parameter pada tiap-tiap alternatif jawaban adalah model IRT dikotomus. Instrumen tipe Likert merupakan respon berjenjang dari 1 hingga 5 yang merupakan respon yang *continue* dan saling mempengaruhi. Model respon berjenjang

atau *graded response model* (GRM) merupakan model yang tepat digunakan pada model instrumen tipe Likert.

Pengembangan instrumen ilmu-ilmu sosial terutama psikologi banyak menggunakan tipe Thurstone maupun Likert. Begitu pula dengan instrument *multiple intelligences* (MI) yang dikembangkan dalam penelitian disertasi ini. Dibeberapa negara alat ukur *multiple intelligences* sudah dibuat dan dikembangkan dalam berbagai penelitian. Phyllis Reardon dan Isabell Dyke dari Canada membuat alat ukur untuk mengungkap kecenderungan *multiple intelligences* yang dimiliki manusia. Dari Virginia, Thomas Armstrong membuat alat ukur untuk mengungkap profil *multiple intelligences* yang dimiliki orang dewasa maupun siswa. Dari Ohio USA, Charles Branton Shearer mengembangkan sebuah alat ukur *multiple intelligences* yang diberi nama dengan *Multiple Intelligences Development Scales* (MIDAS) yang dikembangkan pada anak-anak hingga orang dewasa. Pada tahun 2007, peneliti juga sudah mengembangkan instrument *multiple intelligences*, namun instrumen yang sudah dikembangkan ini masih memiliki beberapa kelemahan sehingga perlu diperbaiki lagi. Disamping itu, dalam analisis instrumen belum dilakukan proses penskalaan. Dengan demikian perlu upaya untuk menyempurnakan instrumen yang sudah dikembangkan dengan proses penskalaannya dan perlu pengkajian kembali karakteristik psikometrik instrumen tersebut.

F. Identifikasi Masalah

Berdasar latar belakang dapat diidentifikasi beberapa masalah berikut :

1. Data yang didapatkan dalam instrumen tipe Thurstone dan Likert berupa data ordinal yang memiliki keterbatasan pada analisis statistik inferensi.
2. Penskalaan dengan pendekatan klasik tidak ketat terhadap asumsi dan hanya mempertimbangkan probabilitas menjawab tiap butir, sehingga hasil yang didapatkan kurang akurat dibandingkan dengan menggunakan pendekatan teori modern.
3. Penskalaan dengan teori modern memerlukan perhitungan yang lebih rumit dan mendetail dibandingkan teori klasik sehingga membutuhkan perangkat analisis yang lebih kompleks dan waktu yang lebih lama.
4. Proses penskalaan dengan teori klasik dan modern sudah dilakukan dengan berbagai cara atau metode, namun penerapannya dalam penelitian dan pengembangan instrumen di Indonesia belum banyak dilakukan.

G. Tujuan Penelitian

Tujuan penelitian ini dapat dirumuskan sebagai berikut :

1. Melakukan proses penskalaan menggunakan pendekatan klasik dengan metode *paired comparison* dan *summated rating*.
2. Mengetahui reliabilitas instrument *multiple intelligences* pada tipe Thurstone dan Likert pada data asli dan data yang diskalakan
3. Mengetahui parameter butir, fungsi informasi butir dan fungsi informasi tes pada instrument *multiple intelligences* dengan model logistik dua parameter untuk data dikotomus dan model *graded response* untuk data politomus.
4. Membandingkan karakteristik psikometrik pada kedua tipe data yang sudah diskalakan dengan menggunakan pendekatan klasik dan modern

H. Urgensi Penelitian

Penelitian ini secara teoritis memiliki makna dalam mengembangkan bidang pengukuran, terutama terkait dengan upaya untuk mengkaji metode penskalaan yang kurang diperhatikan dalam pengembangan instrumen. Hasil penelitian ini diharapkan dapat memberi sumbangan pengembangan instrumen dengan melakukan penskalaan terutama ketika menganalisis data hasil pengukuran dengan kedua tipe instrumen.

Penelitian ini juga bermanfaat dalam mengkaji karakteristik psikometrik instrumen seperti *validitas* dan *reliabilitas*. Hasil *validitas* dan *reliabilitas* instrumen dengan penskalaan klasik maupun tanpa penskalaan. Berbagai informasi yang terkait dengan validitas, reliabilitas dan fungsi informasi butir dan instrumen akan menjadi pertimbangan dalam pemilihan dan penggunaan instrumen. Terutama bagi pengguna tes, hasil penelitian ini juga bermanfaat bagi peneliti dalam pemilihan tipe instrumen yang terkait dengan cara penilaian dan analisis data.

Output penelitian ini berupa instrumen yang secara praktis bermanfaat bagi pengguna tes, guru atau dosen dan mahasiswa. Instrumen dapat digunakan untuk memahami karakteristik kemampuan mahasiswa sehingga dapat menjadi pertimbangan dalam pengembangan diri dan pemilihan karir.

BAB II

KAJIAN PUSTAKA

H. Penskalaan

Pengukuran aspek psikis manusia menghasilkan skor yang dalam interpretasinya terkait proses penskalaan. Proses penskalaan tidak terlepas dari istilah penskalaan, yang berasal dari kata dasar skala. Kerlinger (1990: 492) mendefinisikan skala sebagai himpunan lambang, symbol atau angka yang disusun dengan cara tertentu yang diberikan pada individu sehingga angka tersebut dapat memberi makna sesuai dengan apa yang hendak diukur. Crocker dan Algina (1986: 45) juga memberikan definisi skala sebagai perkembangan suatu cara yang sistematis pada unik-unik yang bermakna dari suatu pengukuran dalam menganalisis data empiris. Cohen & Swerdlik (2005: 63) mengartikan skala sebagai serangkaian simbol atau angka, yang memiliki properti dan sifat atau karakteristik sifat-sifat empiris tertentu pada objek atau angka tersebut. Dari beberapa pengertian diatas, skala diartikan sebagai suatu cara yang digunakan dalam menetapkan suatu himpunan angka atau hasil pengukuran suatu variabel sehingga himpunan angka tersebut memiliki sifat atau karakteristik tertentu.

Dalam penskalaan terdapat upaya untuk membandingkan objek atau data hasil pengukuran. Sebagaimana dikatakan oleh Dun-Runkin, Knezek, Wallace, dan Zhang (2004:3) yang mengartikan skala sebagai pengukuran dan perbandingan objek dalam berbagai cara yang bermakna. Perbandingan suatu objek dengan objek lain akan dapat dilakukan apabila kedua objek tersebut memiliki satuan yang sama, sebagaimana dikatakan Shaw & Wright (1967:20) bahwa upaya membandingkan objek dalam proses penskalaan ini dilakukan dengan membuat kesamaan unit perbandingan. Upaya membandingkan baru akan mendapatkan informasi yang benar apabila unit perbandingnya sama, sebaliknya, bila unit perbandingnya tidak sama bisa mendapat informasi yang keliru. Dengan demikian karakteristik jenis data perlu diperhatikan dalam proses penskalaan.

Dalam statistika, karakteristik data dikaitkan dengan data nominal, ordinal, interval, maupun ratio yang menunjukkan makna kumpulan angka hasil dari pengukuran terhadap suatu variable. Makna kumpulan data ini pada awalnya disebut sebagai skala pengukuran, sebagaimana dijelaskan oleh Nunnaly (1970: 16-19) yang menyebutkan bahwa bahwa ada tipe atau macam skala pengukuran, yang masing-masing memiliki karakteristik yang

berbeda dalam pengukuran psikologi. Karakteristik yang berbeda ini terkait dengan fungsi angka yang dapat digolongkan sebagai :a) label, b) kategori, c) skala ordinal, d). skala interval dan e). skala ratio. Dalam pengembangan ilmu pengukuran istilah skala pengukuran ini disebut juga dengan level pengukuran (Lord & Novick, 1968: 20-21, Allen & Yen, 1979: 7-9, Cohen & Swerdlik, 2009 hal 73-74).

Proses penskalaan terkait dengan upaya menempatkan atribut atau karakteristik pada suatu rentang kontinum, yang didalamnya melibatkan perubahan nilai atau transformasi skor baik berupa *transformasi linear* maupun *non linear* (Brennan, 2006:155). Dalam penelitian ini penskalaan dikaitkan dengan upaya untuk menempatkan atribut psikologi dengan mengubah atau mentransformasi data pada yang semula berbentuk data nominal dan ordinal yang tidak memiliki unit pembanding yang sama menjadi data interval atau rasio yang memiliki satuan pembanding yang sama. Berbagai macam cara atau metode dilakukan dalam upaya untuk mentransformasi data menjadi data interval.

I. Penskalaan Tipe Thurstone

[Louis Leon Thurstone](#) dianggap sebagai pelopor penyusunan instrumen nonkognitif. Ia mengembangkan penskalaan untuk mengukur sikap manusia terhadap berbagai tindakan kriminal. Sejak tahun 1920 ia bersama beberapa teman sejawatnya menerbitkan artikel yang menjelaskan dasar-dasar pengukuran sikap dan menjelaskan metodologi pengukuran sikap. 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 (Shaw & Wright, 1967:21-23, Eagly & Chaiken, 1993: 32, Suryabrata, 2002:201). Metode yang dikembangkan Thurstone berpengaruh besar dalam perkembangan ilmu-ilmu sosial terutama yang terkait dengan pengukuran aspek nonkognitif. Metode yang paling awal dikembangkannya dan hingga sekarang banyak digunakan dalam pengembangan instrumen adalah metode *paired comparison* atau metode perbandingan berpasangan. Dalam pengembangan instrumen, metode *paired comparison* memegang peran penting dalam pengukuran preference atau minat, attitude atau sikap dan value atau nilai. Diantara instrumen kepribadian yang menggunakan metode ini adalah instrumen kepribadian yang dikembangkan oleh Edward pada tahun 1959 yang dikenal dengan skala kepribadian EPPS, dan instrumen pengukuran minat seperti Kuder pada tahun 1969, Holland tahun 1999 dan Lee Thorpe tahun 1943.

J. Penskalaan Tipe Likert

Penskalaan tipe Likert dikembangkan oleh Rensis Linkert pada tahun 1932 dalam mengukur sikap sosial. Pengukuran tipe ini pertama kali dibuat untuk mengukur sikap atau pendapat seseorang terhadap hubungan internasional, antar suku, konflik ekonomi, konflik politik dan agama. Selanjutnya, Likert (1932 : 7) menyatakan bahwa tipe pengukuran ini digunakan tidak hanya untuk mengukur pendapat atau sikap tetapi juga digunakan untuk mengukur persepsi dan berbagai karakteristik psikis atau trait manusia serta dalam pengukuran kepribadian

Tipe Likert disusun berdasar distribusi respon subjek. Respon subjek diberikan dalam taraf persetujuan atau ketidaksetujuan. Format awal tipe ini adalah sangat setuju atau *strongly approve*, setuju atau *approve*, tidak tentu atau *undecided*, tidak setuju atau *disapprove*, sangat tidak setuju atau *strongly disapprove*. Dalam perkembangan selanjutnya respon subjek ini tidak hanya berbentuk taraf persetujuan, 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
- e. kemampuan ; sangat mampu, mampu, ragu-ragu, kurang mampu, tidak mampu

Proses penskalaan tipe Likert dikenal dengan istilah *summated ratings* (Shaw & Wright, 1967: 24, Dun-Rankin, et.al.2004 : 105). Proses penskalaan dimulai dengan menghitung jumlah jawaban atau frekwensi (f) pada masing-masing respon pada tiap item. Skor frekwensi atau f ini selanjutnya dibagi dengan jumlah subjek sehingga didapatkan skor proporsi atau p, $p = f/N$. Langkah berikutnya menghitung proporsi kumulatif (pk), yaitu proporsi pada suatu kategori ditambah dengan proporsai kategori sebelumnya. Karena harga pk ini merupakan batas-batas interval, maka perlu dicari titik tengahnya, yaitu pk tengah. Pk-tengah dihitung dari setengah proporsi dari kategori suatu respon ditambah dengan proporsi kumulatif pada kategori sebelumnya, $pk \text{ tengah} = \frac{1}{2} (pk + pkb)$. Selanjutnya melihat harga skor deviasi atau skor z berdasar harga pktengah. Skor z ini merupakan nilai untuk masing-masing respon. Untuk menghindari skor skala yang negatif, skor z terkecil ditambahkan dengan nilai tertentu sehingga nilai terendah adalah 0.

K. Karakteristik Psikometrik pada Pendekatan Klasik

Istilah karakteristik psikometrik terkait dengan definisi psikometrik. Psikometrik diartikan sebagai prosedur yang digunakan untuk mengestimasi dan mengevaluasi berbagai atribut tentang tes (Furr & Bacharach, 2008:8-9). Dengan demikian karakteristik psikometrik terkait dengan berbagai karakteristik yang digunakan untuk mengetahui dan mengevaluasi berbagai atribut tentang tes. Dalam pendekatan klasik atribut tes dapat diketahui dari validitas dan reliabilitas.

Konsep utama dalam pendekatan klasik yang menyatakan bahwa skor yang didapat (*observed score*) merupakan gabungan dari skor murni atau *true score* dan kesalahan hasil pengukuran atau *error score*. Konsep *true score* sangat terkait dengan konsep reliabilitas (Thissen & Wainer, 2001: 26). Sehingga reliabilitas merupakan ciri khas yang menonjol dalam pendekatan klasik.

L. Penskalaan dengan Pendekatan Teori Modern atau IRT

Pendekatan teori modern lebih dikenal dengan item respon teori (IRT), yang dipopulerkan oleh Hambleton dan Swaminathan pada tahun 1985. Pendekatan IRT ini didasarkan pada variabel laten dan berbagai manifestasinya (Hambleton, Swaminathan & Rogers, 1991:2, De Ayala, 2009:2). Model IRT menggunakan konsep matematika yang menyatakan bahwa probabilitas subjek menjawab item dengan benar tergantung dari kemampuan subjek dan karakteristik item. Dalam hal ini diketahui bahwa seseorang yang memiliki kemampuan atau laten trait yang tinggi akan memberikan respon pada suatu item yang berbeda dengan seseorang yang memiliki kemampuan rendah. Terkait dengan karakteristik item, item atau soal yang sulit akan direspon berbeda oleh subjek dibandingkan item yang mudah.

Model yang akan digunakan dalam penskalaan tipe Thurstone adalah Model logistik 2 parameter. Model ini pertama kali dikembangkan oleh Lord dengan menggunakan fungsi distribusi normal ogive. Selanjutnya fungsi normal ogive diubah oleh Birnbaum (1968) dengan menggunakan fungsi logistik. Model ini ditentukan oleh karakteristik indeks kesukaran butir (b) dan daya beda butir. Embretson (2000 : 127) menyebutkan bahwa kesukaran soal dalam instrumen non kognitif diartikan sebagai taraf penerimaan atau *probability of endorsement* dan Kemampuan atau theta disebut dengan tingkat trait atau *trait level*.

Salah satu model IRT *polytomus* yang dapat digunakan untuk mengukur skala Likert adalah model response berjenjang atau *graded response model* (GRM). Model ini dikembangkan oleh Samejima (1969), digunakan untuk data berjenjang dan digunakan ketika respon peserta terhadap item merupakan skor kategori yang berurutan dan cenderung meningkat. Dalam tipe Likert respon diskor 0,1,2,3,4 yang merupakan skor berjenjang dari yang paling kecil hingga besar secara berurutan.

Item Information Function atau fungsi informasi butir merupakan informasi yang penting dalam model IRT yang merupakan suatu metode atau cara untuk menggambarkan kekuatan suatu butir atau tes (Hambleton, et.al.,1991:91). Weis & Yoes (Hambleton & Zaal, 1991: 79) juga menyatakan bahwa informasi butir merupakan ketepatan pada sesuatu butir yang diestimasi. Kedua informasi tersebut menunjukkan bahwa fungsi informasi butir merupakan kekuatan suatu butir atau tes yang mengungkap ketepatan dalam mengestimasi. Informasi ini juga menunjukkan sumbangan butir dalam mengungkap trait laten atau kemampuan orang pada tes yang diukur.

Fungsi informasi butir digambarkan sebagai sebuah kurve yang semula rendah akan meningkat, dan paling tinggi atau maksimal di tengah-tengah dan akan turun lagi menjauhi titik tengah. Fungsi informasi butir melambangkan ketepatan butir yang mempunyai hubungan yang berlawanan dengan kesalahan baku pengukuran (SE). Semakin besar kesalahan baku (SE) pada suatu butir akan memberikan fungsi informasi butir yang semakin rendah.

M. Kerangka Berpikir

Issue tentang penskalaan muncul terkait dengan makna skor hasil pengukuran berupa level nominal, ordinal, interval dan rasio. Sebagian besar data dalam pengukuran aspek psikis manusia memiliki karakteristik data 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.

Pengembangan instrumen melalui proses penskalaan untuk mengukur karakteristik psikis manusia diperkenalkan oleh Thurstone dengan metodenya yang terkenal dengan *paired comparison*. Penskalaan dimulai dengan mengembangkan instrumen dengan membuat serangkaian stimulus atau pernyataan yang masing-masing saling diperbandingkan. 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.

Instumen tipe *force choice* memiliki kelebihan terkait dengan respon subjek yang cenderung terhindar dari *social desirability* dan *faking* (Mc Donald, 1999:24, Chernyshenko, et. al, 2009:108). *Social desirability* yaitu pernyataan bersifat umum yang memiliki kecenderungan untuk direspon subjek dengan tidak jujur, sehingga sulit untuk mengetahui pendapat subjek sebenarnya. Sedangkan *faking* adalah kecenderungan subjek memilih respon yang tidak sesuai dengan karakteristik dirinya atau dengan sengaja mengubah responnya untuk mendapatkan dirinya pada kelompok yang diharapkan.

Olivares & Böckenholt (2005:285) menjelaskan beberapa kelebihan metode *paired comparison* adalah memaksa minimal *constrain* pada perilaku responden terutama pada pilihan alternative yang sedikit. Metode ini dapat memberikan informasi lebih terkait perbedaan individu dibandingkan dengan metode rating, disamping itu, metode ini juga memiliki kelebihan karena adanya pengecekan konsistensi internal.

Metode *paired comparison* memiliki kelemahan di samping beberapa kelebihan di atas, yaitu beban kerjanya yang banyak. Kelemahan ini disebabkan karena item yang digunakan berjumlah cukup banyak akibat dari setiap item yang harus mendapat kesempatan untuk berpasangan dengan item dari aspek lain (Kwan & Chiu, 2007:433). Kelemahan lain model ini menurut McIver & Carmines (1986: 21) adalah dipengaruhi oleh karakteristik *judgment*. Hal ini disebabkan karena proses penskalaan model ini terkait dengan orang yang akan memberikan *judgment*, maka karakteristik *judgment* juga akan mempengaruhi hasil penskalaan. Berbagai kelemahan dalam tipe ini dapat diatasi dengan model pengukuran yang berbasis respon subjek. Metode yang sering digunakan dan sangat terkenal dalam pengembangan instrumen pada penelitian sosial adalah metode yang digunakan dalam instrumen tipe Likert.

Penskalaan yang dilakukan oleh Likert 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.

Tipe Likert memiliki kelebihan terkait dengan kesederhanaannya dan kemudahannya dalam penyusunan dan interpretasi instrumen (Laerhoven, Zaag-Loonen, Derkx, 2004: 830). Namun kritik terhadap tipe ini adalah adanya keterkaitan atau korelasi yang tinggi antara satu item dengan item-item lain atau antara item dengan total. Korelasi tinggi ini akan berpengaruh pada skor varian dan analisis dimensi instrumen (McIver & Carmines, 1986: 30). Kelemahan lain dari model ini adalah memungkinkan responden untuk melakukan *faking*. Oleh karena itu keberadaan *faking* memungkinkan diperoleh karakteristik psikometrik instrumen yang tidak sesuai dengan kondisi sebenarnya.

Kedua tipe instrumen memiliki kelebihan dan kekurangan terkait dengan adanya *social desirability* dan *faking*. 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 di atas dapat dibuat tabel perbandingan penskalaan tipe Thurstone dan Likert dengan teori klasik dan modern dapat diformulasikan pada tabel 1.

Tabel 1. Perbandingan Penskalaan Tipe Thurstone dan Likert

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

N. Pertanyaan Penelitian

Berdasar uraian pada kerangka berpikir yang dijelaskan diatas, terdapat beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana karakteristik psikometrik pada instrument *multiple intelligences* yang diskalakan menggunakan pendekatan klasik dengan metode *paired comparison* pada instruen tipe Thurstone dan metode *summated rating* pada instrumen tipe Likert?
2. Bagaimana karakteristik psikometrik pada instrument *multiple intelligences* yang diskalakan menggunakan pendekatan modern dengan model respon berjenjang pada instrument *multiple intelligences* tipe Thurstone dan instrumen tipe Likert?
3. Diantara instrument tipe Thurstone dan Likert, tipe instrument yang mana yang memiliki karakteristik psikometrik lebih akurat dengan menggunakan pendekatan klasik dan modern?

BAB III METODE PENELITIAN

A. Pendekatan Penelitian

Secara umum penelitian ini menggunakan pendekatan penelitian kuantitatif yang dalam pelaksanaannya terdiri dari beberapa bagian penelitian yang saling terkait, yaitu penelitian pengembangan instrumen, proses penskalaan pada data hasil ujicoba, analisis karakteristik psikometrik instrumen, dan perbandingan karakteristik psikometrik instrumen. Bagian pertama adalah penelitian pengembangan. Pada penelitian pengembangan dibuat instrumen *multiple intelligences* dengan menggunakan dua tipe, yaitu tipe Thurstone dan tipe Likert. Setelah diuji validitas isi dari ahli selanjutnya diperbaiki. Instrumen yang sudah teruji validitas isinya selanjutnya diujicobakan dilapangan. Bagian berikutnya adalah melakukan proses penskalaan. Proses penskalaan dilakukan pada data hasil ujicoba dilapangan dengan menggunakan pendekatan klasik dan modern. Setelah diskalakan akan dianalisis karakteristik psikometrik kedua tipe instrumen. Bagian terakhir dalam penelitian ini adalah membandingkan karakteristik psikometrik kedua tipe instrument secara diskriptif dan inferensi.

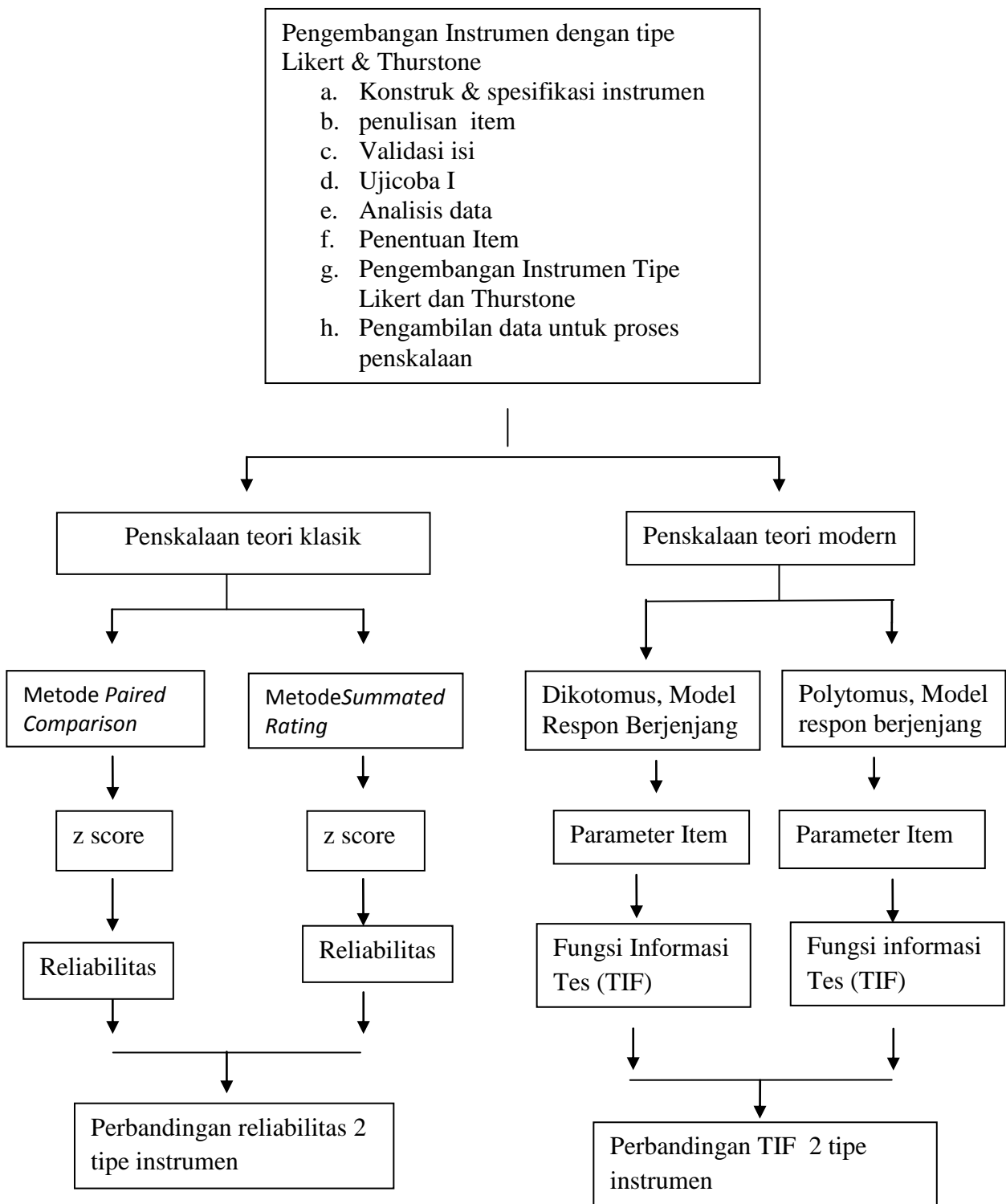
B. Subjek Penelitian

Subjek dalam penelitian ini adalah mahasiswa Universitas Negeri Yogyakarta yang masih aktif kuliah di kampus. Pengambilan data dilakukan dengan *cluster random sampling*, yaitu pengambilan sample yang dilakukan berdasar gugus atau klaster tertentu. Langkah penarikan sampel diawali dengan mengidentifikasi klaster dimana subjek menjadi anggota populasi. Setelah semua klaster teridentifikasi, penarikan responden dilakukan secara acak. Klaster dalam penelitian ini adalah berbagai program studi yang berasal dari beberapa fakultas yang ada di UNY. Tiap fakultas dipilih 2 program studi secara acak. Selanjutnya tiap program studi dipilih satu kelas secara acak juga. Karena sample dipilih secara acak dalam suatu cluster yang bertingkat maka teknik ini disebut juga *multi stage sampling* (Kalton, 1983:28) atau *two stage cluster sampling* (Scheaffer, Mendenhall, Ott Lyman. 1990: 285). Dalam penelitian ini ada dua tingkatan klaster yaitu tingkat fakultas dan tingkat program studi.

C. Alur Penelitian

Alur penelitian merupakan langkah-langkah yang akan dilakukan dalam penelitian ini dari awal hingga akhir. Secara global alur penelitian yang akan dilakukan dapat dilihat pada gambar 1.

Gambar 1. Alur Penelitian



D. Pengembangan Instrumen

Langkah awal penelitian ini adalah pengembangan instrumen. Pengembangan instrumen yang akan dilakukan melalui langkah-langkah sebagai berikut :

1. Penentuan konstruk dan spesifikasi instrumen dengan mengenali teori yang mendasari konstruk
2. Penulisan item
3. Review item dan validasi isi oleh expert judgment
4. Revisi atau perbaikan
5. Uji coba instrumen
6. Seleksi Item
7. Validitas dan reliabilitas 1
8. Penyusunan item dengan tipe Likert dan tipe Thurstone

Instumen *multiple intelligences* dibuat dengan mengacu pada teori Gardner (1993). Spesifikasi instrumen dilakukan dengan menjabarkan teori yang mendasari instrumen yang akan dikembangkan. Masing-masing bentuk inteligensi dianggap sebagai dimensi yang akan dijelaskan lagi kedalam aspek dan indikator. Penyusunan aspek dan indikator dirumuskan dari teori *multiple intelligence* yang dijelaskan oleh Gardner. Indikator-indikator yang sudah penulis buat selanjutnya didiskusikan dengan *expert judgment*.

E. Uji Validitas Isi

Validitas isi merupakan upaya untuk mengetahui karakteristik butir-butir yang dikembangkan sudah sesuai mengukur indikator yang dikembangkan dalam spesifikasi instrumen. Perhitungan validitas isi dengan menggunakan index validitas isi dari Aiken atau disebut dengan V Aiken. Rumus untuk menghitung V Aiken sebagai berikut :

$$V = S / [n*(c-1)], \text{ dimana } S = \sum ni (r-lo),$$

Keterangan :

V = indeks validitas dari Aiken

ni = jumlah expert yang memilih kriteria i,

r = kriteria ke i,

lo = rating terendah,

n = jumlah semua expert,

c = banyaknya rating/kriteria

F. Ujicoba Instrumen

Ujicoba dilakukan untuk mendapatkan butir-butir instrumen yang memiliki karakteristik psikometrik yang bagus. Pelaksanaan ujicoba dilakukan dengan cara meminta mahasiswa mengisi instrumen yang diberikan peneliti dengan terlebih dahulu meminta kesediaan untuk mengisi lembar persetujuan terlibat dalam penelitian disertasi ini. Data yang terkumpul dari empat kelas berjumlah 173 subjek.

G. Seleksi Butir

Seleksi butir dilakukan untuk mendapatkan butir yang memiliki karakteristik psikometris yang tinggi. Seleksi butir dilakukan dengan mempertimbangkan hasil analisis dari validitas isi dari V Aiken, korelasi butir dengan skor total, validitas konstruk melalui loading faktor dan koefisien *chi square* dan signifikannya pada analisis IRT serta keterwakilan item pada masing-masing indikator. Dalam penelitian ini, untuk menyesuaikan format instrumen, jumlah item pada setiap dimensi kecerdasan dibatasi masing-masing 8. Dengan demikian item-item yang sudah dibuat perlu diseleksi dan beberapa item perlu digugurkan. Pengguguran item didasarkan pada beberapa kriteria dilakukan dengan mempertimbangkan beberapa kriteria di atas. Butir-butir yang sudah final selanjutnya disusun kembali sebagai instrumen dengan format atau tipe Thurstone dan Likert. Sehingga kedua bentuk instrumen tersebut memiliki konstruk dan butir-butir yang sama namun model instrumennya yang berbeda.

Kedua bentuk instrumen ini selanjutnya diujicobakan pada mahasiswa UNY. Hasil ujicoba ini selanjutnya dianalisis dengan pendekatan klasik dan modern. Penskalaan dengan pendekatan klasik dilakukan dengan metode *paired comparison* pada instrumen tipe Thurstone dan metode *summated rating* pada instrumen tipe Likert. Sedangkan pendekatan modern dilakukan dengan model respon berjenjang dengan menggunakan program parscale. Setelah dilakukan proses penskalaan dianalisis karakteristik psikometrik dan dibandingkan hasil kedua karakteristik psikometrik keduanya.

H. Proses Penskalaan

1. Penskalaan dengan Pendekatan Klasik

Penskalaan pada pendekatan klasik pada hakekatnya adalah cara mendapatkan data yang semula berada pada level nominal atau ordinal diubah menjadi skor z yang berupa

skor interval. Proses penskalaan dilakukan menggunakan dua metode yaitu : Penskalaan perbandingan berpasangan dan penskalaan *summated rating*. Proses penskalaan dengan metode *pair comparison* pada instrument *multiple intelligence* mengacu pada langkah-langkah yang dilakukan oleh Dunn-Runkin, et.al. (2004 : 94-96). Perhitungan dilakukan secara manual dengan bantuan program Microsoft Exel. Setelah dilakukan proses penskalaan, ditetapkan kembali skor pada tiap butir. Hasil dari proses penskalaan diatas didapatkan skor z pada masing-masing item. Setelah dilakukan proses penskalaan selanjutnya dianalisis karakteristik kedua tipe instrumen berdasar koefisien reliabilitas. Perhitungan reliabilitas dilakukan dengan rumus Alpha sebagaimana ada pada persamaan 8 pada instrumen tipe Likert. Karena dimensi instrumen pada penelitian ini adalah multidimensional maka analisis reliabilitas menggunakan reliabilitas komposit dari alpha pada persamaan. Analisis reliabilitas instrumen dilakukan dengan menggunakan program microsoft excel.

2. Penskalaan dengan Pendekatan Modern (IRT)

Penskalaan dengan pendekatan modern dilakukan melalui beberapa langkah. Langkah pertama mencari parameter daya beda dan index kesukaran soal pada tiap butir. Parameter butir pada instrumen tipe Thurstone maupun instrumen tipe Likert dianalisis dengan menggunakan model respons berjenjang atau *graded response model* (GRM). Analisis model ini dilakukan dengan menggunakan program parscale (Du Toit, M. 2003). Setelah mendapatkan parameter daya beda dan index kesukaran butir, dilakukan estimasi parameters kemampuan atau Θ menggunakan *marginal maximum likelihood*. Analisis estimasi kemampuan dilakukan dengan bantuan komputer menggunakan program Exel.

Setelah didapatkan skor parameter kedua model, selanjutnya dianalisis fungsi informasi butir dan fungsi informasi tes. Perhitungan fungsi informasi butir dan fungsi informasi tes dilakukan dengan program excel dengan menggunakan persamaan 19 dan 21

Perbandingan penskalaan kedua tipe Instrumen dilihat dari karakteristik psikometrik keduanya dengan teori klasik dan modern. Dalam teori klasik dilihat dari reliabilitas dan kesalahan baku pengukuran (SEM) kedua tipe instrumen sedangkan secara modern dilihat berdasarkan fungsi informasi butir (IIF), Fungsi informasi tes (TIF) dan kesalahan baku pengukuran (SEM) pada kedua instrumen. Perbandingan keduanya dilakukan secara diskriptif.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

Penelitian ini memaparkan proses penskalaan yang dimulai dari menyusun instrumen, mengubah skor hingga menganalisis karakteristik psikometrik hasil penskalaan dengan pendekatan klasik dan modern. Penskalaan dilakukan dengan menggunakan instrumen *multiple intelligence* yang butir-butirnya sudah dibuat dan diseleksi sebagaimana dijelaskan pada bab 3, selanjutnya butir-butir yang sudah terseleksi disusun kembali sebagai instrumen dengan format atau tipe Thurstone dan Likert. Sehingga kedua tipe instrumen tersebut memiliki konstruk dan butir-butir yang sama namun tipe instrumennya yang berbeda.

Kedua bentuk instrumen ini selanjutnya diujicobakan pada mahasiswa UNY yang seluruhnya berjumlah 443 orang. Hasil ujicoba ini selanjutnya dianalisis dengan pendekatan klasik dan modern. Penskalaan dengan pendekatan klasik dilakukan dengan metode *paired comparison* pada instrumen tipe Thurstone dan metode *summated rating* pada instrumen tipe Likert. Sedangkan pendekatan modern dilakukan dengan model respon berjenjang dengan menggunakan program *Parscale*. Setelah dilakukan proses penskalaan dianalisis karakteristik psikometrik dan dibandingkan hasil kedua karakteristik psikometrik keduanya.

A. Hasil Penskalaan dengan Pendekatan Klasik

A.1. Penskalaan *Paired Comparison*

Proses penskalaan dilakukan pada skor hasil penelitian sebanyak 443 subjek. Perhitungan dilakukan secara manual dengan bantuan program *Microsoft Excel* dengan mengikuti langkah-langkah sebagai berikut :

- a. Membuat matrik frekuensi (f) yang memuat jumlah frekuensi pilihan subjek terhadap masing-masing pasangan stimulus
- b. Mengurutkan stimulus dari yang terkecil hingga terbesar
- c. Mengubah nilai frekwensi menjadi nilai proporsi (p), $p = f/N$
- d. Mengkonversikan harga p dalam bentuk skor z yang merupakan deviasi dari rata-rata pada distribusi normal
- e. Menghitung rata-rata skor z pada tiap dimensi
- f. Membuat urutan stimulus dari z terkecil hingga besar

Berdasar perhitungan sebagaimana ditunjukkan tabel 16, didapatkan skor logika matematika 0, musik 0,376, linguistik 0,429, kinestetik 0,575, naturalis 0,668, visual 0,714, interpersonal 1,031, eksistensial 1,065 dan intrapersonal 1,541.

A.2. Penskalaan *Summated Rating*

Proses penskalaan instrumen tipe Likert dilakukan dengan metode *summated rating*. Perhitungan penskalaan dilakukan dengan bantuan program Microsoft Excel. Perhitungan penskalaan ini diawali dengan menghitung jumlah frekwensi (f) respon subjek di tiap kriteria pada tiap butir. Skor frekwensi ini selanjutnya diubah menjadi skor proporsi (p) dan proporsi kumulatif. Skor proporsio dihitung dengan cara membagi frekwensi (f) dengan banyaknya responden (N). Proporsi kumulatif (pk) didapatkan dari proporsi pada tiap kategori ditambah proporsi kategori sebelumnya. Proses berikutnya menghitung pk tengah yaitu titik tengah proporsi kumulatif yang dihitung dari setengah proporsi dalam kategori ditambah dengan pk kategori sebelumnya, atau dapat dirumuskan sebagai berikut; $pk\text{-tengah} = \frac{1}{2}p + pk$. Proses selanjutnya menghitung nilai deviasi (z) dengan mengkonversi skor pk-tengah menjadi skor z dengan mengacu pada tabel z kurve normal. Contoh perhitungan penskalaan *summated rating* pada salah satu butir dapat dilihat pada tabel 2.

Tabel 2

Perhitungan penskalaan *summated rating* pada salah satu item

| Butir1 | Respon | | | | |
|-----------|--------|--------|--------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 |
| f | 4 | 36 | 120 | 229 | 54 |
| p | 0,009 | 0,081 | 0,271 | 0,517 | 0,122 |
| pk | 0,009 | 0,090 | 0,361 | 0,878 | 1,000 |
| pk tengah | 0,005 | 0,047 | 0,204 | 0,541 | 0,771 |
| z | -2,611 | -1,671 | -0,826 | 0,103 | 0,741 |
| z+ | 0,000 | 0,940 | 1,785 | 2,714 | 3,352 |

Hasil penskalaan *summated rating* didapatkan skor z tiap respon pada tiap butir. Hasil tersebut menunjukkan bahwa dengan proses penskalaan didapatkan skor respon pada tiap butir berbeda dengan skor respon tanpa penskalaan. Skor hasil penskalaan menunjukkan jarak skor antar respon pada masing-masing butir tidak tetap atau sama

dengan 1. Skor pada respon 1 berubah menjadi -2,611, respon 2 berubah menjadi -1,671, respon 3 berubah menjadi -0,826, respon 4 berubah menjadi 2,714 dan respon 5 berubah menjadi 0,741.

Dengan melihat kedua proses penskalaan instrumen dengan kedua metode yang telah dijelaskan diatas dapat diketahui bahwa penskalaan instrumen tipe Likert dengan metode *summated rating* pada hakekatnya adalah proses penskalaan dengan pendekatan respon, sedangkan penskalaan tipe Thurstone dengan *metode paired comparison* merupakan penskalaan dengan pendekatan stimulus. Dalam aplikasinya instrumen tipe Thurstone ini juga digunakan untuk mengukur karakteristik psikologi seseorang dimana dasar pengukuran dan analisisnya menggunakan respon subjek. Sehingga model instrumen tipe Thurstone ini diskalakan juga dengan pendekatan respon. Dalam penelitian ini upaya untuk membandingkan kedua tipe instrumen dengan menganalisis penskalaan instrumen tipe Thurstone dengan menggunakan pendekatan respon. Pada teori klasik kedua instrumen dianalisis dengan menggunakan metode *summated rating*.

B. Diskripsi Data Kemampuan (Θ) Subjek Penelitian Sebelum dan Setelah Ditransformasi

Diskripsi data hasil penelitian disajikan dalam dua bentuk, yaitu data sebelum diskalakan yang masih berupa skor asli dan data yang sudah diskalakan atau data yang sudah ditransformasi dalam skor z. Diskripsi data yang disajikan pada tiap dimensi atau bentuk kecerdasan. Data hasil penelitian instrumen tipe Thurstone disajikan pada tabel 4. dan instrumen tipe Likert disajikan pada tabel 3.

Tabel 3

Diskripsi Data Penelitian Instrumen Tipe Thurstone

| Dimensi | Thurstone | | | | | | | | | |
|-------------------|-----------|-------|----|-----|-----|-----------------------------|-------|-------|-------|-------|
| | Data Asli | | | | | Data setelah Ditransformasi | | | | |
| | Mean | SD | Md | Min | Max | Mean | SD | Md | Min | Max |
| Linguistik | 3,255 | 1,427 | 3 | 0 | 8 | 3,413 | 1,415 | 3,398 | 0 | 7,6 |
| Logika-matematika | 2,047 | 1,812 | 2 | 0 | 8 | 1,746 | 1,508 | 1,634 | 0 | 6,538 |
| Visual-spasial | 3,966 | 1,625 | 4 | 0 | 8 | 4,371 | 1,616 | 4,482 | 0 | 8,229 |
| Musik | 2,995 | 2,11 | 3 | 0 | 8 | 2,841 | 1,913 | 2,62 | 0 | 7,236 |
| Kinestetik | 3,623 | 1,497 | 4 | 0 | 8 | 3,696 | 1,46 | 3,646 | 0 | 7,775 |
| Intrapersonal | 6,296 | 1,292 | 6 | 2 | 8 | 8,996 | 1,773 | 8,805 | 2,602 | 11,3 |
| Interpersonal | 4,916 | 1,306 | 5 | 0 | 8 | 6,314 | 1,571 | 6,489 | 0 | 9,499 |
| Naturalis | 3,885 | 1,625 | 4 | 0 | 8 | 4,121 | 1,63 | 4,197 | 0 | 8,049 |
| Eksistensial | 5,016 | 1,409 | 5 | 1 | 8 | 6,184 | 1,635 | 6,361 | 1,13 | 9,405 |

Tabel 4

Diskripsi Data Penelitian Instrumen Tipe Likert

| Dimensi | Likert | | | | | | | | | |
|-------------------|-----------|-------|----|-----|-----|-----------------------------|-------|-------|-------|-------|
| | Data Asli | | | | | Data setelah Ditransformasi | | | | |
| | Mean | SD | Md | Min | Max | Mean | SD | Md | Min | Max |
| Linguistik | 29,84 | 3,021 | 30 | 15 | 40 | 30,31 | 2,869 | 30,56 | 15,03 | 37,9 |
| Logika-matematika | 28,36 | 4,605 | 29 | 12 | 40 | 30,16 | 4,068 | 30,85 | 14,75 | 38,47 |
| Visual-spasial | 28,84 | 4,045 | 29 | 16 | 40 | 30,12 | 3,4 | 30,6 | 18,41 | 38,12 |
| Musik | 26 | 5,064 | 26 | 12 | 40 | 30,08 | 3,949 | 30,33 | 18,2 | 39,33 |
| Kinestetik | 29,74 | 3,293 | 30 | 16 | 40 | 30,15 | 2,875 | 30,34 | 17,64 | 37,77 |
| Intrapersonal | 32,332 | 3,437 | 32 | 12 | 40 | 30,26 | 3,18 | 30,52 | 11,7 | 36,53 |
| Interpersonal | 31,418 | 3,168 | 31 | 11 | 40 | 30,27 | 3,039 | 30,42 | 10,12 | 37,08 |
| Naturalis | 29,190 | 3,993 | 29 | 15 | 40 | 30,18 | 3,53 | 30,41 | 16,49 | 38,09 |
| Eksistensial | 32,341 | 3,440 | 32 | 9 | 40 | 30,41 | 3,371 | 30,75 | 9,279 | 36,66 |

C. Dimensi Instrumen

Pengujian dimensi instrumen dilakukan untuk mengetahui banyaknya dimensi atau faktor pada tiap dimensi kecerdasan. Pengetahuan banyaknya dimensi ini digunakan sebagai pertimbangan dalam analisis reliabilitas dan analisis IRT. Salah satu cara mengetahui dimensi suatu instrumen dengan melakukan analisis faktor eksploratori. Pengujian dilakukan menggunakan bantuan program SPSS. Analisis dilakukan secara terpisah pada tiap dimensi atau kecerdasan pada kedua tipe instrumen.

Banyaknya dimensi yang didapat dari hasil analisis faktor dilihat dari varian yang dijelaskan faktor yang terekstraksi dan *scree plot* pada tiap faktor. Hasil penelitian ini menunjukkan besarnya persentase varian yang dijelaskan pada tiap faktor semua dimensi kedua tipe instrumen lebih dari 20%. Pada instrumen tipe Thurstone, faktor pertama menyumbangkan sekitar 20%. Sedangkan pada instrumen tipe Likert sumbangan faktor 1 berkisar antara 33- 56 %. *Scree plot* pada tiap dimensi menunjukkan faktor pertama memiliki nilai eigen paling tinggi dibanding faktor lainnya. Pada instrumen tipe Thurstone dominansi faktor pertama tampak tajam pada dimensi linguistik, logika matematika, visual spasial, musik, kinestetik dan naturalis. Sedangkan dimensi intrapersonal, interpersonal dan eksistensial, dominansinya tidak terlalu tajam karena selisih nilai eigen pada faktor berikutnya tidak terlalu besar. Pada instrumen tipe Likert, faktor 1 memiliki dominansi yang tajam pada semua dimensi. Meskipun beberapa dimensi dalam instrumen ini tidak menunjukkan kecenderungan unidimensi yang tajam, namun karena secara teoritik instrumen ini mengukur satu faktor dan persentase faktor pertama pada semua dimensi sudah lebih dari 20 % dari jumlah faktor yang terekstraksi, maka dalam analisis lebih lanjut digunakan pendekatan unidimensi.

C. Analisis Reliabilitas

Salah satu karakteristik psikometrik yang sering digunakan dalam pendekatan teori klasik adalah reliabilitas. Dalam penelitian ini, perhitungan reliabilitas instrumen pada tipe Thurstone dihitung dengan rumus koefisien reliabilitas internal dari Kendall dan Smith (1940), sementara perhitungan reliabilitas tiap dimensi yang dilakukan dengan rumus *alpha cronbach*. Hasil perhitungan reliabilitas instrumen secara keseluruhandidapatkan koefisien reliabilitas sebesar 0,64. Hasil tersebut menunjukkan instrumen ini cukup reliabel secara internal.

Perhitungan reliabilitas masing masing dimensi pada instrumen tipe Thurstone dihitung dengan mengelompokkan butir-butir yang mengukur dimensi yang sama. Hasil perhitungan reliabilitas tiap dimensi dapat dilihat di pada data asli dan data yang sudah ditransformasi dapat dilihat pada tabel 5.

Reliabilitas instrumen pada tiap dimensi tidak mengalami perbedaan yang bermakna pada data asli maupun data yang sudah ditransformasi, Namun apabila dilihat dari kesalahan baku pengukuran (SEM) pada data yang ditransformasi mengalami penurunan terutama pada dimensi kecerdasan logika matematika dan musik. Pada dimensi linguistik, visual-spasial, kinestetik dan naturalis, terdapat perubahan kesalahan baku pengukuran, namun perubahannya sangat kecil sehingga tidak terlihat jelas dalam grafik. Sedangkan pada dimensi intrapersonal dan eksistensial tampak jelas mengalami peningkatan. Hasil tersebut menunjukkan pada dimensi instrumen yang memiliki reliabilitas cukup bagus SEMnya mengalami penurunan pada data yang ditransformasi, namun penurunan itu tidak terjadi terutama pada data yang memiliki reliabilitas sangat rendah.

Tabel 5
Reliabilitas Tiap Dimensi Instrumen Tipe Thurstone

| No | Dimensi | Data Asli | | Data skor yang ditranformasi | |
|----|-------------------|--------------|-------|------------------------------|-------|
| | | Reliabilitas | SEM | Reliabilitas | SEM |
| 1 | Linguistik | 0,267 | 1,222 | 0,265 | 1,214 |
| 2 | Logika-matematika | 0,651 | 1,071 | 0,638 | 0,908 |
| 3 | Visual-spasial | 0,383 | 1,277 | 0,344 | 1,309 |
| 4 | Musik | 0,7 | 1,156 | 0,687 | 1,070 |
| 5 | Kinestetik | 0,236 | 1,308 | 0,203 | 1,303 |
| 6 | Intrapersonal | 0,25 | 1,12 | 0,256 | 1,529 |
| 7 | Interpersonal | 0,163 | 1,195 | 0,181 | 1,422 |
| 8 | Naturalis | 0,376 | 1,283 | 0,347 | 1,317 |
| 9 | Eksistensial | 0,195 | 1,264 | 0,203 | 1,460 |

Perhitungan reliabilitas instrumen tipe Likert juga dihitung pada tiap dimensi dengan menggunakan rumus *alpha*. Hasil perhitungan reliabilitas tiap dimensi dapat dilihat pada tabel 6.

Tabel 6

Reliabilitas Tiap Dimensi Instrumen Tipe Likert

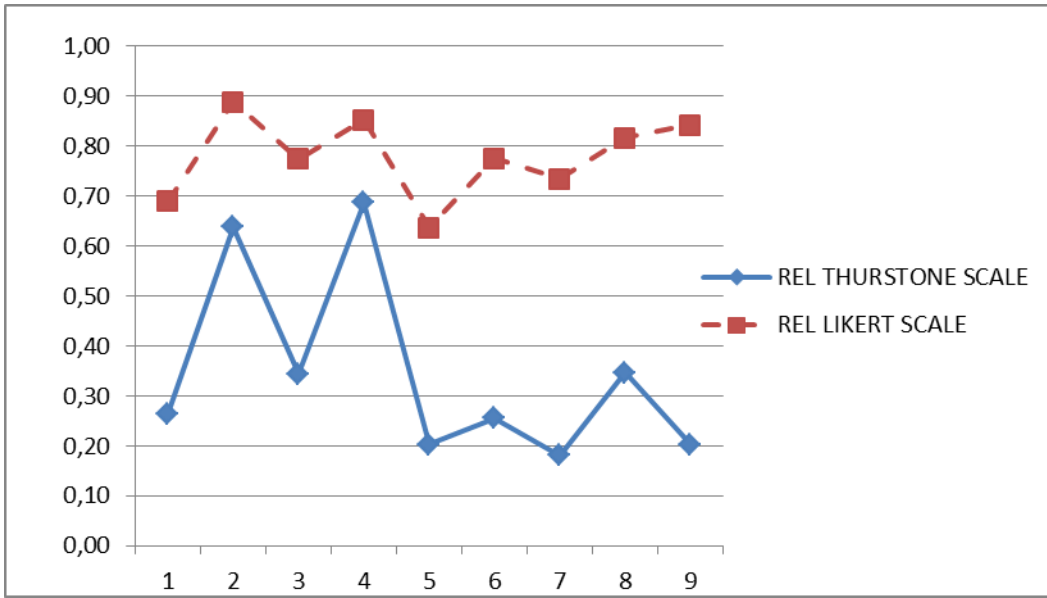
| No | Dimensi | Data Asli | | Data skor yang ditranformasi | |
|----|-------------------|--------------|-------|------------------------------|-------|
| | | Reliabilitas | SEM | Reliabilitas | SEM |
| 1. | Linguistik | 0,700 | 1,654 | 0,690 | 1,598 |
| 2. | Logika-matematika | 0,887 | 1,546 | 0,888 | 1,359 |
| 3. | Visual-spasial | 0,783 | 1,885 | 0,775 | 1,613 |
| 4. | Musik | 0,856 | 1,925 | 0,852 | 1,517 |
| 5. | Kinestetik | 0,639 | 1,980 | 0,636 | 1,734 |
| 6. | Intrapersonal | 0,793 | 1,563 | 0,775 | 1,508 |
| 7. | Interpersonal | 0,749 | 1,588 | 0,735 | 1,565 |
| 8. | Naturalis | 0,825 | 1,670 | 0,817 | 1,510 |
| 9. | Eksistensial | 0,861 | 1,284 | 0,842 | 1,338 |

Reliabilitas tiap dimensi instrumen tipe Likert sebelum dan setelah proses penskalaan juga tidak banyak perubahan, meskipun reliabilitas pada data asli sedikit lebih tinggi dari data yang sudah diskalakan, namun perubahannya sangat kecil dan kurang bermakna. Namun apabila dilihat dari kesalahan baku pengukuran (SEM), ternyata SEM pada data yang sudah diskalakan cenderung lebih rendah, kecuali pada dimensi eksistensial yang mengalami sedikit peningkatan.

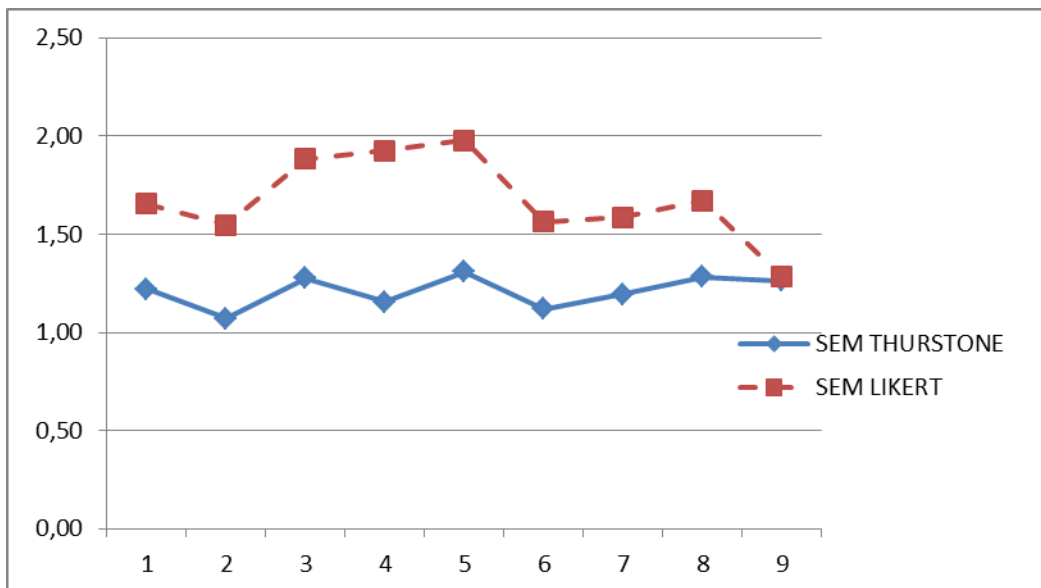
Setelah didapatkan reliabilitas pada tiap dimensi selanjutnya dihitung reliabilitas instrumen tipe Likert dengan menggunakan rumus reliabilitas komposit alpha berstrata (Cronbach, Schoneman, Mc Kie, 1965; 293, Widhiarso, 2009; 42,43) sebagaimana disajikan pada formula . Hasil perhitungan reliabilitas instrumen secara keseluruhan dengan menggunakan rumus tersebut didapatkan koefisien reliabilitas sebesar 0,939 dengan SEM sebesar 8,422 pada data asli dan reliabilitas sebesar 0,796 dengan SEM 4,594 pada data yang sudah ditransformasi ke dalam skor z.

Instrumen tipe Thurstone memiliki karakteristik yang berbeda dengan instrumen tipe Likert. Meskipun keduanya memiliki konstruk yang sama, dengan metode pengukuran yang berbeda memberikan hasil reliabilitas dan SEM yang berbeda pula. Berdasar perhitungan reliabilitas yang sudah disajikan pada tabel 6 dan 7 diketahui bahwa pada semua dimensi reliabilitas instrumen tipe Likert lebih tinggi dari instrumen tipe Thurstone, namun lebih jelasnya data reliabilitas kedua tipe instrumen disajikan pada gambar 2.

Gambar 2 menunjukkan bahwa terdapat perbedaan reliabilitas yang terlihat jelas pada instrumen tipe Thurstone dan Likert. Reliabilitas instrumen tipe Likert jauh lebih tinggi dari reliabilitas instrumen tipe Thurstone. Meskipun demikian apabila dilihat SEM keduanya, tampak adanya perubahan, dimana SEM pada instrumen tipe Thurstone lebih rendah dibanding instrumen tipe Likert. Informasi lebih jelas dapat dilihat pada gambar 3.



Gambar 2
Grafik Perbandingan Reliabilitas Instrumen Tipe Thurstone dan Likert



Gambar 3
Perbandingan SEM Instrumen Tipe Thurstone dan Likert

D. Penskalaan dengan Pendekatan Modern (IRT)

Penskalaan dengan pendekatan modern terkait dengan model pengukuran yang digunakan. Pada penelitian ini model yang digunakan adalah respon berjenjang atau *graded response model* (GRM). Pada model ini ada dua parameter yang turut mempengaruhi transformasi skor dalam proses penskalaan, yaitu parameter daya beda dan index kesulitan butir yang dalam pengukuran nonkognitif sering disebut dengan *index probability of endorsement*, (Embretson, 2000 : 127).

E. Estimasi Parameter

Parameter b dikatakan efektif apabila memiliki nilai tidak terlalu tinggi dan tidak terlalu rendah. Hambleton, et al, (1991 ; 13) menyatakan parameter b akan diterima pada nilai yang berkisar antara -2,0 hingga +2,0. Nilai b kurang dari -2,0 dikatakan item tersebut sangat mudah atau memiliki probabilitas kemendukungan yang sangat tinggi dan diatas +2,0 dikatakan sangat sulit atau probabilitas kemendukungan sangat rendah. Berdasar tabel 27 terdapat beberapa butir yang memiliki index kesulitan rendah adalah butir ke 18, 41, 42, 44, 46, 49, 52, 53, 65, 66. Butir-butir yang memiliki index kesulitan tinggi adalah butir ke 5, 6, 13, 54, 70. Pada instrumen tipe Likert terdapat 4 parameter b, yang bergerak dari tingkat kesulitan rendah hingga tinggi. Parameter b1 pada semua butir memiliki skor dibawah -2, b2 sebagian besar berkisar antara -2 hingga 0, b3 berkisar 0 hingga 2 dan b4 diatas 2. Secara keseluruhan, rata-rata parameter b instrumen tipe Thurstone adalah -0,149. Instrumen tipe Likert memiliki rata-rata b sebesar -0,956, jika rata-rata b antar kategori didapatkan b1 = -3,481, b2 = -1,042, b3 = 1,088, dan b4 = 2,902. Secara umum parameter b instrumen tipe Thurstone lebih tinggi daripada instrumen Tipe Likert.

Daya beda butir menunjukkan seberapa besar suatu butir dapat membedakan seseorang yang memiliki kemampuan atau trait yang tinggi dan rendah. Daya beda butir akan diterima apabila memiliki nilai diatas 0,2. Pada instrumen tipe Thurstone maupun Likert semua butir memiliki daya beda diatas 0,2, sehingga dapat dikatakan bahwa butir butir pada instrumen tipe Thurstone maupun Likert dapat membedakan orang yang mendukung dan tidak mendukung. Rata-rata daya beda butir-butir pada instrumen tipe Thurstone adalah 0,533, sedangkan butir-butir pada instrumen tipe Likert memiliki daya beda lebih tinggi yaitu 0,991.

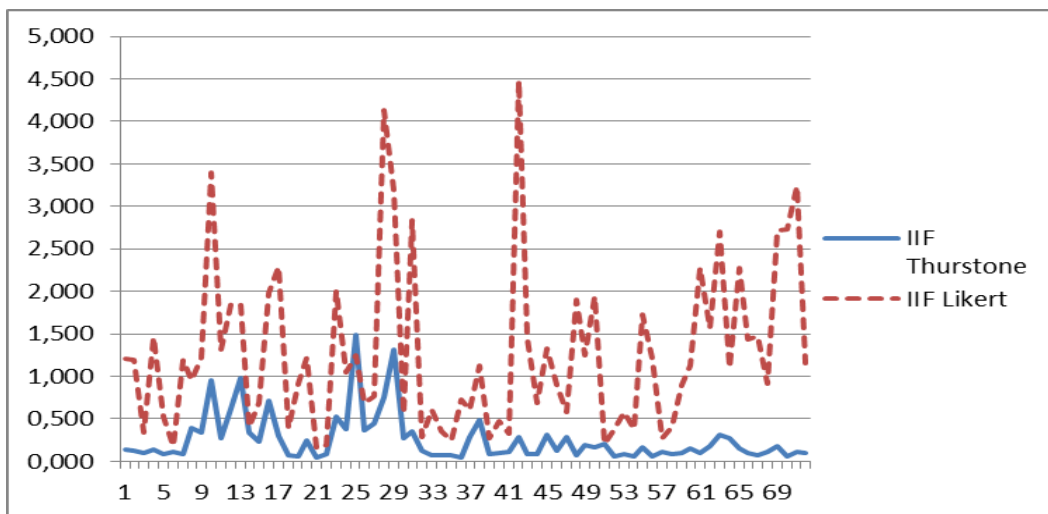
F. Fungsi Informasi Butir

Rata-rata fungsi informasi butir instrumen tipe Thurstone sebesar 0,639 dan pada instrumen tipe Likert sebesar 2,816. Apabila dilihat dari theta atau posisi kemampuan maksimal dimana fungsi informasi butir berperan efektif, maka theta pada butir-butir instrumen tipe Thurstone lebih tinggi dibanding instrumen tipe Likert. Rata-rata theta instrumen tipe Thurstone adalah -0,368 sedangkan pada instrumen tipe Likert adalah -2,185.

Letak kemampuan (Θ) dimana fungsi informasi butir maksimal berada pada item-item instrumen tipe Thurstone maupun Likert bervariasi. Gambar 4 menunjukkan sebagian besar theta pada instrumen tipe Likert lebih tinggi dari theta fungsi informasi butir instrumen tipe Thurstone, meskipun ada sebagian butir di instrumen tipe Thurstone yang lebih tinggi. Theta butir-butir instrumen tipe Likert sebagian besar terletak di atas atau lebih tinggi dari butir-butir pada instrumen tipe Thurstone, meskipun ada yang sebaliknya pada beberapa butir.

Gambar 4

Perbandingan Theta Dimana IIF Berada pada Kedua Tipe Instrumen



Perhitungan fungsi informasi tes dilakukan dengan menjumlahkan fungsi informasi butir pada tiap dimensi. Hasil perhitungan kedua fungsi informasi tes diketahui bahwa dimensi kecerdasan kinestetik memiliki fungsi informasi yang paling tinggi dibanding dimensi lainnya baik pada instrumen tipe Thurstone maupun Likert. Fungsi informasi yang paling rendah terdapat pada dimensi kecerdasan linguistik baik pada instrumen tipe

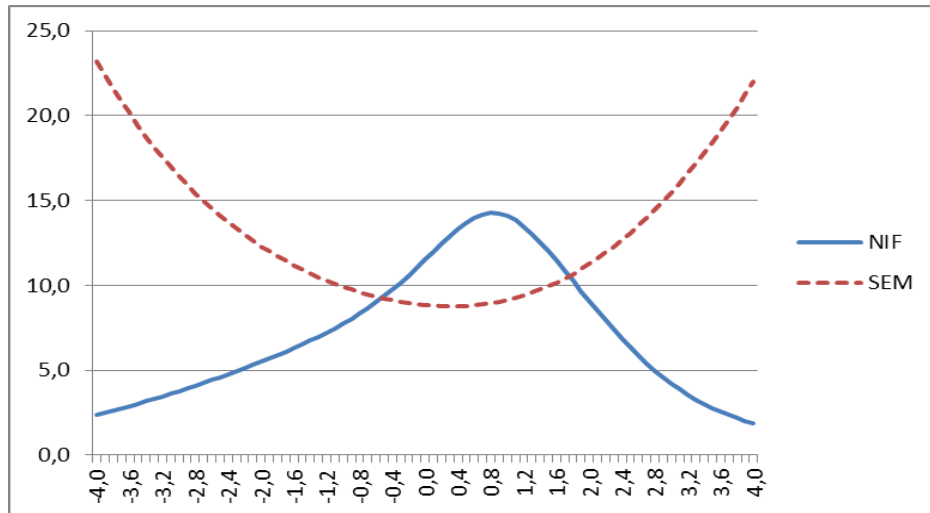
Thurstone maupun instrumen tipe Likert. Dilihat dari rata-ratanya instrumen tipe Thurstone memiliki fungsi informasi sekitar 4 kali lebih rendah dari instrumen tipe Likert.

Tabel 7.

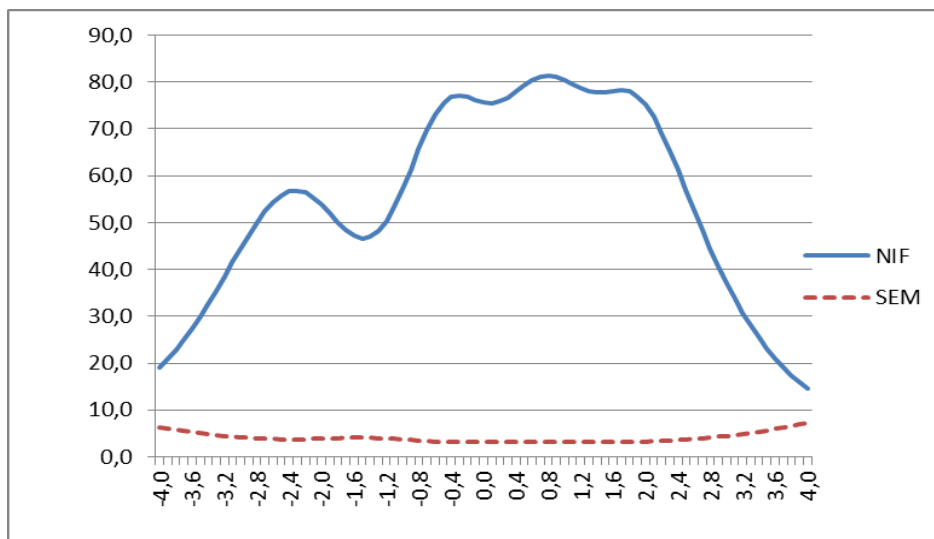
Fungsi Informasi dan Kesalahan Baku Pengukuran pada tiap Dimensi

| No | Jumlah | Thurstone | | Likert | |
|-----------|-------------------|-----------|-------|--------|-------|
| | | TIF | SEM | TIF | SEM |
| 1. | Linguistik | 1,177 | 0,922 | 7,048 | 0,377 |
| 2. | Logika-matematika | 2,214 | 0,672 | 9,259 | 0,329 |
| 3. | Visual-spasial | 2,869 | 0,590 | 10,624 | 0,307 |
| 4. | Musik | 3,987 | 0,501 | 12,183 | 0,286 |
| 5. | Kinestetik | 4,448 | 0,474 | 12,718 | 0,280 |
| 6. | Intrapersonal | 3,519 | 0,533 | 10,780 | 0,305 |
| 7. | Interpersonal | 2,935 | 0,584 | 9,747 | 0,320 |
| 8. | Naturalis | 1,744 | 0,757 | 7,836 | 0,357 |
| 9. | Eksistensial | 1,698 | 0,767 | 8,216 | 0,349 |
| Rata-rata | | 2,732 | 0,644 | 9,824 | 0,323 |

Fungsi informasi tiap dimensi dan keseluruhan item atau tes pada instrumen tipe Thurstone lebih rendah dari instrumen tipe Likert, sehingga kesalahan baku pengukuran pada instrumen tipe Likert lebih rendah dibanding instrumen tipe Thurstone. Perpaduan nilai informasi tes dan kesalahan baku pengukuran pada instrumen tipe Thurstone digambarkan pada gambar 5, sedangkan pada instrumen tipe Likert digambarkan pada gambar 6. Titik potong keduanya merupakan daerah dimana kemampuan peserta dapat diterima. Kemampuan yang dapat diterima pada instrumen tipe Thurstone berkisar antara - 0,7 hingga 1,8, sedangkan pada instrumen tipe Likert memiliki rentang yang lebih lebar dari -4 hingga 4.



Gambar 5
TIF dan SEM instrumen tipe Thurstone



Gambar 6
TIF dan SEM instrumen Tipe Likert

Berdasar gambar 5 dapat ditarik kesimpulan bahwa butir-butir pada instrumen tipe Thurstone hanya efektif pada subjek pengukuran yang memiliki kemampuan disekitar rata-rata dan tidak efektif pada subjek yang memiliki kemampuan tinggi atau rendah. SEM akan tinggi bahkan lebih tinggi dibanding informasi yang didapatkan pada hasil pengukuran subjek yang memiliki kemampuan tinggi dan rendah. Dengan demikian

informasi hasil pengukuran dengan menggunakan instrumen tipe Thurstone pada subjek yang dikatakan tinggi dan rendah kemampuannya kurang tepat karena mengandung kesalahan atau error yang besar. Sedangkan pada instrument tipe Likert didapatkan nilai informasi tes lebih tinggi dari SEMnya dan informasi tes pada instrument tersebut efektif digunakan untuk mengukur subjek dengan berbagai kemampuan. Berdasarkan hasil penelitian ini dapat disimpulkan bahwa instrument multiple intelligence tipe Likert lebih efektif dibanding tipe Thurstone dalam mengukur kecenderungan kemampuan subjek.

BAB V

KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasar hasil penelitian yang disajikan pada bab 4, Ada beberapa hal yang dapat disimpulkan.

1. Penskalaan *multiple intelligence* pada pendekatan klasik didapatkan skor yang telah ditransformasi ke dalam skor z. Dengan menggunakan metode *paired comparison* pada tipe Thurstone didapatkan urutan skor dari yang rendah adalah logika matematika, musik, linguistik, kinestetik, naturalis, visual, interpersonal, eksistensial dan yang paling tinggi adalah intrapersonal. Dengan menggunakan metode summated rating pada instrumen tipe Thurstone dan Likert dihasilkan skor yang bervariasi pada tiap respon, sedangkan pada pendekatan IRT didapatkan estimasi parameter dan kemampuan pada tiap butir.
2. Terdapat perubahan varian, rata-rata, reliabilitas dan kesalahan baku pengukuran (SEM) dari data asli dengan data yang ditransformasi menggunakan skor z pada penskalaan teori klasik. SEM pada data yang ditransformasi mengalami penurunan.
3. Karakteristik psikometrik koefisien reliabilitas instrumen tipe Thurstone lebih rendah dibanding tipe Likert, baik pada keseluruhan data maupun data tiap dimensi, begitu pula dalam SEMnya. SEM pada instrumen tipe Thurstone lebih rendah dibanding tipe Likert,
4. Hasil penskalaan teori IRT dengan menggunakan model respon berjenjang didapatkan hasil daya beda dan indeks kesulitan butir yang bervariasi pada kedua tipe instrumen. Butir-butir pada instrumen tipe Thurstone memiliki daya beda lebih rendah dan index kesulitan butir lebih tinggi dibanding butir-butir pada instrumen tipe Likert.
5. Fungsi informasi butir maupun fungsi informasi tes pada instrumen tipe Likert lebih tinggi dibanding tipe Thurstone, sebaliknya kesalahan baku pengukuran atau SEM pada instrumen tipe Thurstone lebih tinggi dibanding tipe Likert. Dengan demikian instrumen tipe Likert lebih akurat dibanding tipe Thurstone dalam mengukur kecenderungan kemampuan subjek.

B. Saran

1. Proses penskalaan pada teori klasik terbatas menggunakan metode *paired comparison* dan *summated rating*. Penggunaan metode-metode lain seperti metode interval tampak setara, *semantik differential*, dan metode rangking dapat dilakukan untuk pengembangan instrumen yang bervariasi.
2. Reliabilitas yang rendah pada beberapa dimensi instrumen tipe Thurstone perlu dikaji lagi faktor penyebabnya, sehingga dapat diantisipasi pada pengembangan instrumen.
3. Dalam membahas dan membandingkan konsep reliabilitas dan kesalahan baku pengukuran perlu mencantumkan pula varian instrumen tersebut.
4. Dimensi instrumen tipe Thurstone tidak cukup kuat menunjukkan kecenderungan unidimensi, meskipun dalam analisis penelitian ini menggunakan asumsi unidimensi. Diperlukan penelitian lebih lanjut untuk mengkaji analisis dimensi data tipe Thurstone dengan metode analisis yang berbeda.
5. Perlu ada penelitian lebih lanjut untuk menganalisis data hasil pengukuran dengan model lain pada pendekatan IRT, seperti model *unfolding* dan skala *rating*.

Daftar Pustaka

- Aiken L.,R.,. Three Coefficients to Analyzing The Reliability and Validity of Rating. Educational and Psychological Measurements. 1985, 45 : 131-142
- Azwar, S.(1999). *Dasar-Dasar Psikometri*. Yogyakarta : Pustaka Pelajar Offset.
- Azwar, S.(2012). *Penyusunan Skala Psikologis*. Edisi 2. Yogyakarta : Pustaka Pelajar Offset.
- Baker, F.,B., Ho Kim, Seock. (2004). *Item response theory.parameter estimation techniques*. New York ; Marcel Dekker
- Bock RD, Gibbons R., & Muraki E. (1988).Full-information item faktor analysis.*Applied Psychological Measurement*,12, 261-280
- Carifio, J & Perla, R. (2008).Resolving the 50-year debate around using and misusing Likert scales.*Medical Education*.42, 1150–1152 1.
- Chernyshenko O, S., Stark,S, Prewett, M, Gray, A.,A., Stilson, F.R, & Tuttle,M.,D. (2009) Normative scoring of multidimensional pairwise preference personality scales using IRT: Empirical Comparisons With Other Formats. *Human Performance*, 22,105–127
- De Ayala, R., J.(2009). *The theory and practice of item response theory*. New York: The Guilford Press
- Du Toit, M. (2003). IRT from SSI : Bilog-MG, multilog, parscale, testfact. lincolnwood ;*Scientific Software International*.
- Dunn-Runkin, P., Knezek, G., A., Wallace, S.,& Zhang, S. (2004). *Scaling Methods*,(2nd Edition). Mahwah : Lawrence Erlbaum Associates, Inc.
- Embretson S. E., &Reise S P., (2000).Item response theory for psychology. Lawrence New Jersey ; Elbaum Associates, Publisher.
- Gardner, H. (1993). *Multiple intelligences.the theory in practice a reader*. New York : Basic Book, A Subsidiary of Perseus Books, LLC.
- Glenberg, A., M.(1988).*Learning from data, an introduction to statistical reasoning*. San Diego : Harcourt Brace Jovanovich, Publishers.
- Hambleton, R.K., & Swaminathan, H. (1985).*Item response theory : Principles and Application*. Boston : Kluwer
- Hambleton, R.K., & Zaal, J.,N., (1991). *Advance in educational and psychological testing*. Boston: Kluwer Academic Publisher Group.
- Jamieson S. (2004). Likert scales: How to use them. *Medical Education*, 4;38:1212
- Kendall, M.G. & Smith,B.B. (1940). On the method of paired comparisons. *Biometrika* 31:324-345
- Kwan & Chiu. (2007). Modification and siplication of thurstone scalling method and its demonstration with crime seriousness assessment.*The Hong Politecnic University Research Commity*
- Laerhoven H, van der Zaag-Loonen H.J., &Derkx B.H.F. (2004).A comparison of Likert scale and visualanalogue scales as response options in children’s questionnaires. *Acta Pædiatrik*, 93, 830–835.
- Lord, F. M., (1980). Applications of Item Response Theory to Practical Testing Problems. New Jersey : Lawrence Erlbaum Assouciates
- Kwan &Chiu. (2007). Modification and siplication of thurstone scalling method and its demonstration with crime seriousness assessment.*The Hong Politecnic University Research Commity*
- Likert, R., (1932).*Technigue for The Measurement of Attitudes*. New York City : Archives of Psychology.

- Lord, F., M., (1980). *Applications of item response theory to practical testing problems*. Erlbaum : Hillsdale, N.J.
- Lord, F., M., & Novick, M., R., (1974). *Statistical theories of mental test scores*. New York: Addison-Wesley Publishing Company, Inc.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta : Mitra Cendekian Offset
- McDonald R.P., (1999). *Test Theory: A unified treatment*. London : Lawrence Erlbaum Associates
- McIver, J.P., & Carmines, E., G. (1986). *Unidimensional scaling*. London : Sage Publications. Inc
- Netemeyer, R., C., Bearden, W., O., & Sharma, S. (2003). *Scaling procedures, issues and applications*. London : Sage Publications, Inc.
- Olivares, A., M., & Bockenholt, U. (2005). Structural equation modeling of paired-comparison and ranking data. *Psychological Methods*, 10, 3, 285–304
- Pollard, B., Dixon, D., Dieppe, P., & Johnston, M. (2009). Measuring the ICF components of impairment, activity limitation and participation restriction: an item analysis using classical test theory and item response theory. *Health and Quality of Life Outcomes*, 7, 41
- Reckase, M.D., (1979). Unifactor latent trait models applied to multifactor tests : results and implications. *Journal of Educational Statistics*. 4, 3, 207-230
- Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.
- Sheares, B., Midas home page by Branton Sheares. www.agelfire.com/oh/themidas/-10k. Diambil pada tanggal 20 januari 2006
- Shaw, M., E., & Wright, J., M., (1967). *Scales for The Measurement of Attitudes*. New York : McGraw-Hill Book Company
- 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
- Thurstone, L., L. (1927). Method of paired comparison for social values. *Journal of Abnormal and Social Psychology*, 21, 384-400
- Warter, M., K., (1999). Multiple intelligence inventory. diambil dari <http://surfaguarium.com/mi/inventory.htm> pada tanggal 17 januari 2006
- Van der Linden, W.J., & Hambleton, R.K., (1997). *Handbook of Modern Item Response Theory*, New York : Springer-Verlag New York Inc.
- Widhiarso W., (2009). Koefisien reliabilitas pada pengukuran kepribadian yang bersifat multi dimensi. *Psikobuana*, 1, 1, 39-48