

SKRIPSI

**Perbandingan Pendekatan *Classical Test Theory* dan Pendekatan *Item Response Theory*
dalam analisis Karakteristik Psikometrik *The Scale of Ethnic Experience***

**Untuk Memenuhi Persyaratan Memperoleh Gelar Sarjana Psikologi pada Fakultas
Ilmu Sosial dan Ilmu Politik Universitas Brawijaya**



**Disusun oleh :
Syafira Nur Sabilla
145120301111067**

**PROGRAM STUDI S1 PSIKOLOGI
FAKULTAS ILMU SOSIAL DAN ILMU POLITIK
UNIVERSITAS BRAWIJAYA
MALANG
2018**

LEMBAR PENGESAHAN


Perbandingan Pendekatan *Classical Test Theory* dan Pendekatan *Item Response Theory* dalam analisis Karakteristik Psikometrik *The Scale of Ethnic Experience*

SKRIPSI

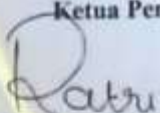
Disusun Oleh
Syaffira Nur Sabilla
NIM. 145120301111067

Telah disetujui dan dinyatakan lulus dalam ujian sarjana
Pada tanggal 13 April 2018


Tim Penguji
Ketua Majelis Sidang Penguji,


Sukaesi Marianti, S.Psi., M.Si., Ph.D.
NIP. 198007312008122002

Ketua Penguji


Ratri Nurwanti, S.Psi., M.Psi.
NIK. 2013048712142001

Anggota Penguji,


Dr. Lusv Asa Akhrani, S.Psi., M.Psi.T
NIK. 20100878052322001

Malang, 28 Mei 2018
Fakultas Ilmu Sosial dan Ilmu Politik
Dekan



Prof. Dr. Ulfy Ludigdo, Ak
NIP. 19680814 199402 1 001

KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT karena rahmat dan karunia-Nya, sehingga penulis masih diberi kesempatan untuk dapat menyelesaikan skripsi *Perbandingan Pendekatan Classical Test Theory dan Pendekatan Item Response Theory* dalam analisis Karakteristik Psikometrik *The Scale of Ethnic Experience*

Atas keberhasilan penyusunan skripsi tersebut, oleh karena itu pada kesempatan ini dengan segala kerendahan dan ketulusan hati, penulis mengucapkan terima kasih kepada :

1. Ibu Sukaesi Marianti, Ph.D. selaku dosen pembimbing yang telah senantiasa membimbing penulis sejak awal penyusunan skripsi ini hingga akhir, tak terkecuali motivasi, nasihat, dan masukan beliau yang telah diberikan kepada penulis selama ini.
2. Ibu Ratri Nurwanti, S.Psi., M.Psi., Psikolog dan Dr. Lusy Asa Akhrani, S.Psi., M.Psi.T selaku dosen penguji yang telah membantu dalam menyempurnakan skripsi ini.
3. Ibu Cleoputri Al Yusainy. Ph.D. selaku Ketua Jurusan Psikologi Universitas Brawijaya dan juga Ibu Ulifa Rahma, S.Psi., M.Psi, Psikolog selaku dosen Pembimbing Akademik penulis.
4. Seluruh Dosen dan Karyawan Jurusan Psikologi Fakultas Ilmu Sosial dan Ilmu Politik Universitas Brawijaya Malang yang telah membantu kelancaran skripsi ini.
5. Kedua orang tua dan adik-adik saya yang telah senantiasa memberikan dukungan tiada akhir kepada saya dalam menyelesaikan skripsi ini.

6. Rekan ó rekan seperjuangan Payung Penelitian seperti Nunung, Dera, Dhia, Dita, Ajeng, Vida, dan Putri yang tak pernah lelah membantu dan mendukung penulis.
7. Serta teman-teman saya yaitu Adimas, Fauridha, Karina, Wafiyyah, Zalfa, dan Indah yang selalu memberikan sugesti-sugesti positif kepada saya demi kelancaran dalam proses penyelesaian skripsi ini.
8. Teman-teman seperjuangan angkatan 2014 Psikologi Universitas Brawijaya yang senantiasa membantu dan memberikan semangat bagi penulis.
9. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang senantiasa memberi dukungan demi kelancaran penyelesaian skripsi ini.

Penulis menyadari sepenuhnya bahwa penyusunan laporan kegiatan ini masih banyak kekurangan. Oleh karena itu, kritik dan saran yang bersifat membangun dari semua pihak sangat diharapkan, harapan penulis semoga laporan magang ini dapat berguna untuk menambah ilmu pengetahuan bagi penulis dan pembaca lainnya.

Malang, Mei 2018

Penulis

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERNYATAAN ORISINALITAS.....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR ISTILAH.....	xi
ABSTRAK.....	xiii
<i>ABSTRACT</i>	xiv
BAB I PENDAHULUAN.....	1
A. Latar Belakang.....	1
B. Rumusan Masalah.....	5
C. Tujuan Penelitian.....	5
D. Manfaat Penelitian.....	6
E. Penelitian Terdahulu.....	6
BAB II TINJAUAN PUSTAKA.....	10
A. <i>Ethnic Experience</i>	10
1. Definisi <i>Ethnic Experience</i>	10
2. Dimensi <i>Ethnic Experience</i>	12
B. Karakteristik Psikometrik.....	13
1. Lokasi Parameter.....	13

2. <i>Item Discrimination</i>	17
3. Reliabilitas.....	18
4. <i>Standard Error of Measurement</i>	21
C. <i>The Scale of Ethnic Experience (SEE)</i>	22
D. Kerangka Pemikiran.....	23
BAB III METODE PENELITIAN	25
A. Desain Penelitian.....	25
B. Fokus Penelitian.....	26
C. Definisi Operasional.....	26
1. Lokasi Parameter.....	26
2. <i>Item Discrimination</i>	26
3. Reliabilitas.....	27
4. <i>Standard Error of Measurement</i>	27
D. Subjek Penelitian.....	27
E. Tahap Pelaksanaan Penelitian.....	28
1. Tahap persiapan.....	28
2. Tahap pengumpulan data.....	30
3. Tahap analisis data.....	30
F. Teknik Analisis Data.....	30
1. <i>Corrected item total correlation</i>	31
2. <i>Cronbach Alpha</i>	31
3. <i>Generalized Partial Credit Model</i>	31
4. <i>Information Function</i>	32

5. <i>Standard Error of Measurement</i>	32
BAB IV HASIL DAN PEMBAHASAN	33
A. Gambaran Umum Subjek Penelitian.....	33
B. Hasil Penelitian.....	34
C. Pembahasan.....	40
D. Keterbatasan Penelitian.....	42
BAB V PENUTUP	43
A. Kesimpulan.....	43
B. Saran.....	44
DAFTAR PUSTAKA	45
Lampiran A. The Scale of Ethnic Experience.....	48
Lampiran B. Lembar Jawab.....	52
Lampiran C. Output corrected item total correlation.....	53
Lampiran D. Output Generalized Partial Credit Model.....	56
Lampiran E. Item Operational Characteristic Curves.....	57
Lampiran F. Kartu Kendali Skripsi.....	59

DAFTAR TABEL

Tabel 1 Skor respon jawaban pada skala <i>ethnic experience</i>	22
Tabel 2 <i>Btusprint</i> skala <i>ethnic experience</i>	23
Tabel 3 Kategorisasi subjek penelitian.....	33
Tabel 4 Perbandingan hasil analisis item pendekatan CTT dan IRT pada SEE...	35

DAFTAR GAMBAR

Gambar 1. Kategori skor pada empat kategori respon.....	15
Gambar 2. Skor <i>category response function</i> pada PCM.....	16
Gambar 3. Desain penelitian perbandingan CTT dan IRT.....	25
Gambar 4. Alur pelaksanaan penelitian	28
Gambar 5. Hubungan SEM dan TIF pada SEE.....	39

DAFTAR ISTILAH

Classical Test Theory (CTT) Teori dasar psikometri yang sudah dikembangkan dan diaplikasikan sejak lama, dan tetap bertahan hingga sekarang.

Corrected item total correlation Koefisien yang dihasilkan dari korelasi antara item dengan total.

Cronbach Alpha Salah satu teknik yang digunakan untuk menguji reliabilitas alat ukur

Generalized Partial Credit Model (GPCM) Salah satu teknik dalam IRT yang merupakan pengembangan dari model PCM.

Graded Response Model (GRM) Salah satu teknik dalam IRT yang merupakan perluasan dari model 1PL.

Information Function Hasil dari pendekatan IRT yang memberikan informasi berupa item dan tes.

Item Characteristic Curve (ICC) Kurva yang dihasilkan dari pendekatan IRT pada masing-masing item.

Item Discrimination kemampuan suatu item dalam membedakan kemampuan subjek pada pendekatan CTT.

Item Information Function Hasil dari pendekatan IRT yang memberikan informasi berupa item.

Item Level Information Informasi yang dihasilkan berasal dari masing-masing item.

Item Response Theory (IRT) Teori dasar psikometri yang muncul untuk menjawab kekurangan dari pendekatan CTT.

Observed Score Skor tampak atau skor amatan.

Partial Credit Model (PCM) Salah satu teknik dalam IRT yang merupakan perluasan dari model 2PL.

Item Total Correlation Teknik yang digunakan untuk melihat daya diskriminasi antar item.

Standard Error of Measurement Teknik yang digunakan untuk mengetahui kesalahan dalam pengukuran.

Test Information Function Hasil dari pendekatan IRT yang memberikan informasi berupa tes.

Test Level Information Informasi yang dihasilkan berasal dari satu tes saja.

The Scale of Ethnic Experience (SEE) Skala yang mengukur komponen kognitif pada akulturasi seseorang yang diciptakan oleh Malcarne, Chavira, Fernandez, dan Liu (2006).

True Score Skor sebenarnya yang diperoleh dari suatu pengukuran.

ABSTRAK

Perbandingan Pendekatan *Classical Test Theory* dan Pendekatan *Item Response Theory* dalam analisis Karakteristik Psikometrik *The Scale of Ethnic Experience*

Syafira Nur Sabilla
145120301111067

Tujuan utama dari penelitian ini ialah melakukan perbandingan antara pendekatan *Classical Test Theory* (CTT) dan pendekatan *Item Response Theory* (IRT). Teknik analisis data yang digunakan dalam pendekatan CTT ialah *corrected item total correlation*, *Cronbach Alpha* dan *Standard Error of Measurement* (SEM), sedangkan pada pendekatan IRT dengan menggunakan *Generalized Partial Credit Model* (GPCM), *Information Function* (IF), dan *Standard Error of Measurement* (SEM). Alat ukur yang digunakan untuk diuji karakteristik psikometriknya ialah *The Scale of Ethnic Experience* (SEE). SEE merupakan alat ukur yang dibuat untuk menilai konstruk yang berkaitan dengan komponen kognitif pada akulturasi seseorang. Dalam penelitian ini juga dilakukan transadaptasi pada skala SEE. Subjek yang digunakan sebanyak 400 mahasiswa yang bukan berasal dari Malang. Hasil dari penelitian ini menunjukkan bahwa pendekatan IRT lebih sesuai untuk digunakan dalam menganalisis SEE. Selain itu, hasil yang diberikan mengatakan bahwa pendekatan IRT lebih sensitif dan lebih mendalam dalam menganalisis item.

Kata kunci : *Classical Test Theory*, *Item Response Theory*, *The Scale of Ethnic Experience*

ABSTRACT

Comparison between Classical Test Theory Approach and Item Response Theory Approach in analyzing Psychometric Characteristics of The Scale of Ethnic Experience

Syafira Nur Sabilla

145120301111067

The main purpose of this research is to compare the Classical Test Theory approach (CTT) and Item Response Theory approach (IRT). Data analysis techniques used in the CTT approach are corrected total correlation items, Cronbach Alpha and Standard Error of Measurement (SEM), while in IRT approach using Generalized Partial Credit Model (GPCM), Information Function (IF), and Standard Error of Measurement (SEM). The measuring instrument used to test its psychometric characteristics is The Scale of Ethnic Experience (SEE). SEE is a measuring tool designed to assess constructs related to the cognitive component of a person's acculturation. In this study also performed transadaptation on the SEE scale. Subjects used as many as 400 students who are not from Malang. The results of this study indicate that the IRT approach is more suitable for use in analyzing SEE. In addition, the results provided say that the IRT approach is more sensitive and deeper in analyzing items.

Keywords : Classical Test Theory, Item Response Theory, The Scale of Ethnic Experience

BAB I

PENDAHULUAN

A. Latar Belakang

Hasil pengukuran yang dapat dipercaya, diperoleh dari suatu alat ukur yang memiliki kualitas yang dapat dipercaya. Sehingga memastikan bahwa alat ukur yang digunakan memiliki kualitas yang baik merupakan suatu keharusan dengan menguji karakteristik psikometriknya (Rahmawati, 2014). Pendekatan yang digunakan dalam menguji karakteristik psikometrik suatu alat ukur juga akan menentukan hasil pengujian (Rahmawati, 2014). Jika dalam dunia psikometri, terdapat dua pendekatan besar dalam menguji karakteristik psikometrik yaitu *Classical Test Theory* (CTT) dan *Item Response Theory* (IRT). CTT lebih menekankan pada tes sehingga hanya menggunakan satu reliabilitas dan satu *Standard Error of Measurement* (SEm) (Cappelleri, Lundy, & Hays, 2014). Sedangkan pendekatan IRT dibuat untuk mengatasi kelemahan dari pendekatan CTT, dimana yang dihasilkan bukan hanya berasal dari satu tes saja (*test-level information*) tetapi juga pada masing-masing *item* (*item-level information*) dan pada masing-masing *level latent traits* (Güler, W{cp,m." ("Vgmt."4236+.

Para ahli psikometri sering membandingkan performa CTT dan IRT dalam menguji karakteristik psikometrik, salah satunya ialah penelitian yang dilakukan oleh I Ängt." W{cp,m." (" Vgmt" *2014) yang

bertujuan untuk menganalisis persamaan dan perbedaan dari pendekatan CTT dan IRT secara teoritis. Kedua pendekatan tersebut sering dibandingkan karena memiliki kelebihan dan kekurangannya masing-masing. Salah satu kelebihan dari CTT ialah menggunakan konsep yang sederhana dalam menentukan reliabilitas (Suwanto, 2004). Selain itu, CTT juga dapat digunakan pada jumlah sampel yang kecil, namun besarnya koefisien reliabilitas tes bergantung pada sampel. Padahal pada pendekatan IRT tidak lagi bergantung pada sampel, namun membutuhkan sampel yang besar agar mampu menghasilkan parameter yang stabil (Suhariyono, Sriyono, & Ngazizah, 2014).

Parameter yang digunakan untuk membandingkan performa pendekatan CTT dan IRT ialah karakteristik psikometriknya yang terdiri dari lokasi parameter, *item discrimination*, reliabilitas, dan *Standard Error of measurement* (SEm). Lokasi parameter atau biasa disebut sebagai *item difficulty* (indeks kesukaran butir) adalah proporsi jumlah subjek yang menjawab benar pada suatu *item* berbanding jumlah subjek yang menjawab pada *item* tersebut (Azwar, 2007). *Item discrimination* (daya beda) ialah kemampuan *item* dalam membedakan antara individu yang memiliki atribut psikologis yang diukur dengan individu yang tidak memiliki atribut psikologis yang diukur, dalam hal ini daya beda *item* pada SEE diartikan sebagai kemampuan *item* dalam membedakan individu yang memiliki konstruk *ethnic experience* dan yang tidak (Suwanto, 2004). Reliabilitas ialah merupakan derajat keajegan (*consistency*) di antara dua

buah hasil pengukuran pada objek yang sama. Lalu yang terakhir, SEM adalah dapat digunakan untuk mamahami kesalahan yang bersifat acak yang mempengaruhi skor subjek dalam pelaksanaan tes (Allen & Yen, 1979). Keempat hal tersebut akan diuji menggunakan dua pendekatan yang berbeda untuk mengetahui bagaimana performa dari masing-masing pendekatan dalam menguji karakteristik psikometrik tersebut.

Corrected item total correlation dan *Cronbach's Alpha* digunakan sebagai teknik analisis data pada pendekatan CTT dalam penelitian ini. Teknik *Cronbach's Alpha* sendiri digunakan karena teknik uji reliabilitas ini dianggap paling mudah dan lebih efisien jika dibandingkan dengan teknik lainnya, terutama pada *item* yang berjumlah banyak (Trochim, 2006). Kemudian pada IRT teknik yang digunakan adalah *Generalized Partial Credit Model* (GPCM) yang termasuk ke dalam model 2PL. Hal ini didasarkan dari hasil penelitian yang dilakukan oleh Ridlo (2016) yang membahas perbandingan antara teknik GPCM dan *Graded Response Model* (GRM). Berdasarkan hasil penelitian tersebut mengatakan bahwa teknik GPCM lebih tepat dalam memprediksi kemampuan subjek dan dalam menyeleksi kemampuan subjek dibandingkan GRM.

Baik penggunaan pendekatan CTT maupun pendekatan IRT tersebut pada awalnya dikembangkan untuk melakukan analisis pada alat ukur yang bersifat kognitif. Seiring berjalannya waktu, kedua pendekatan tersebut juga telah dilakukan pada alat ukur yang bersifat non kognitif. Seperti penelitian yang dilakukan oleh Widhiarso (2016), dengan

menggunakan analisis Rasch, yang pada hakekatnya merupakan analisis IRT model satu parameter, untuk mengidentifikasi proporsi responden yang memiliki gaya respons ekstrim saat mengisi skala harga diri (Nurchahyo, 2016). Oleh karena itu, dalam penelitian ini juga akan menggunakan alat ukur yang bersifat non kognitif yaitu *The Scale of Ethnic Experience* (SEE).

SEE merupakan alat ukur yang dibuat untuk menilai konstruk yang berkaitan dengan komponen kognitif pada akulturasi seseorang. Skala ini terdiri dari 32 *item* dan memiliki empat dimensi, antara lain: *ethnic identity*, *perceived discrimination*, *social affiliation*, dan *mainstream comfort* (Malcarne, Chavira, Fernandez, & Liu, 2006). Setiap dimensi memiliki indikator yang berkaitan dengan persepsi seseorang tentang bagaimana etnisitas yang dimiliki dapat mempengaruhi fungsi sosialnya dalam masyarakat (Thomas, Bardwell, Malcarne, & Dimsdale, 2004). Pada penelitian sebelumnya, tidak dijelaskan pendekatan mana yang sesuai apabila digunakan untuk menganalisis SEE, namun hanya dijelaskan mengenai bagaimana prosedur pembuatan SEE hingga menjadi suatu alat ukur yang valid dan reliabel (Malcarne, Chavira, Fernandez, & Liu, 2006). Sehingga perbandingan CTT dan IRT dilakukan pada SEE untuk mengetahui pendekatan manakah yang akan memberikan hasil yang lebih baik dalam uji karakteristik psikometrik SEE, agar memudahkan untuk penelitian selanjutnya apabila akan menggunakan alat ukur SEE. Indikator-indikator yang dimaksud dengan lebih baik ialah suatu

pendekatan yang mampu memberikan tingkat keakuratan yang lebih tinggi dengan memberikan informasi yang lebih banyak dan mendetail mengenai suatu alat ukur. Selain itu juga suatu pendekatan yang bersifat *independent*, sehingga hasil yang diberikan lebih dapat dipercaya ketika alat ukur tersebut digunakan kembali beberapa kali (Wiberg, 2004). Fokus dalam SEE ialah masyarakat dengan etnis yang beragam, untuk itu SEE sangat sesuai apabila dikembangkan di Indonesia karena memiliki masyarakat yang tergolong dalam masyarakat multikultural dan terdiri dari berbagai etnis dan kebudayaan yang tersebar di seluruh daerah (Hafid, 2016).

B. Rumusan Masalah

Berdasarkan latar belakang tersebut, tampak rumusan masalah yang akan digunakan dalam penelitian ini yaitu apakah terdapat perbedaan pada pendekatan *Classical Test Theory* (CTT) dan *Item Response Theory* (IRT) dalam menganalisis properti psikometrik *The Scale of Ethnic Experience*?

C. Tujuan Penelitian

Untuk mengetahui perbedaan pada pendekatan *Classical Test Theory* (CTT) dan *Item Response Theory* (IRT) dalam menganalisis properti psikometrik *The Scale of Ethnic Experience*?

D. Manfaat Penelitian

1. Manfaat Teoritis

- a. Diharapkan dapat memiliki peran serta dalam pengembangan ilmu pengetahuan pada alat ukur khususnya yang berkaitan dengan *The Scale of Ethnic Experience*.
- b. Untuk memperbanyak kajian ilmu psikologi khususnya mengenai pendekatan *Classical Test Theory* (CTT) dan *Item Response Theory* (IRT) yang biasanya digunakan untuk menguji kualitas psikometrik suatu alat ukur.

2. Manfaat Praktis

- a. Dapat digunakan sebagai alat ukur yang dapat dipercaya terutama dalam mengukur variabel yang sama yaitu *ethnic experience* pada remaja.

E. Penelitian Terdahulu

1. Malcarne, V. L., Chavira, D. A., Fernandez, S., & Liu, P.-J. (2006). The Scale of Ethnic Experience: Development and Psychometric Properties. *Journal Of Personality Assessment*, 86, 150-161.

Penelitian ini memiliki tujuan yaitu untuk mengembangkan alat ukur berupa *The Scale of Ethnic Experience* (SEE), dengan menampilkan karakteristik psikometrinya. Dalam penelitian tersebut menggunakan teknik analisis antara lain reliabilitas *test-retest*, konsistensi internal, validitas kriteria, dan validitas konstruk. Dengan menggunakan sampel yang berasal

dari 4 kelompok etnis yaitu: *African Americans*, *Caucasian Americans*, *Filipino Americans*, dan *Mexican Americans*. Hasil yang diperoleh dari penelitian ini memberikan dukungan terhadap reliabilitas dan validitas pada SEE sebagai alat ukur *multidimensional* yang mengukur aspek kognitif seseorang yang memiliki kelompok etnis ganda dan dapat digunakan di beberapa kelompok etnis di Amerika.

2. Thomas, K. S., Bardwell, W. A., Malcarne, V., & Dimsdale, J. E. (2004). The Effect of Demographic Variables on Ethnic Experience. *Journal of Applied Biobehavioral Research*, 65-79.

Studi ini bertujuan untuk mengungkapkan perbedaan etnis dalam identitas etnis, diskriminasi yang diterima, dan kenyamanan dari masyarakat umum. Selain itu juga meneliti apakah variabel demografis (umur, jenis kelamin, dan status sosial-ekonomi) berhubungan dengan dimensi-dimensi tersebut. Sampel yang digunakan terdiri dari 85 orang yang diambil dari San Diego dan komunitas di sekitarnya. Sampel merupakan *African American* dan *Caucasian American*. Dengan menggunakan *A multivariate analysis of variance* (MANOVA) sebagai teknik analisis datanya. Hasil dari penelitian ini menunjukkan bahwa etnisitas terdiri dari beberapa komponen, yang mungkin dipengaruhi oleh faktor demografis pada *African American*. Ada kemungkinan juga bahwa setiap aspek etnisitas berhubungan dengan hasil kesehatan.

3. I. A. G. T. (2014). Comparison of classical test theory and item response theory in terms of item parameters. *European Journal of Research on Education*, 2, 1-6.

Penelitian ini bertujuan secara empiris untuk menguji persamaan dan perbedaan parameter yang dianalisis dengan menggunakan dua pendekatan yaitu *Classical Test Theory* (CTT) dan *Item Responses Theory* (IRT). Sampel yang digunakan berjumlah 1250 siswa. Dengan teknik pengumpulan data berupa subtes Turkish pada HSEE tahun 2003 yang terdiri dari 25 *item*. Hasil penelitian menunjukkan bahwa model yang paling sesuai dengan data yaitu model IRT 3PL.

4. Suhariyono, Sriyono, & Ngazizah, N. (2014, September). Akurasi pendekatan Classical Test Theory dan pendekatan Item Response Theory dalam menganalisis soal UAS Fisika semester genap kelas 1 SMA Negeri di Purworejo Tahun Pelajaran 2013/2014. *Radiasi*, 5, 75-79.

Tujuan dari penelitian ini ialah untuk mengetahui reliabilitas, tingkat kesukaran dan indeks daya beda berdasarkan *Classical Test Theory* (CTT); *goodness of fit*, *thrushole* dan indeks diskriminasi berdasarkan *Item Response Theory* (IRT); dan *Standart Error of Measurement* (SEM) pada CTT dan IRT. Pada penelitian ini menggunakan metode dokumentasi dengan mengumpulkan lembar jawab UAS Fisika dari SMA Negeri di Kabupaten Purworejo sebagai bagian dari teknik pengumpulan data. Dalam menganalisis data dengan menggunakan program ITEMAN dan program BIGSTEP. Dari hasil yang didapat dapat disimpulkan bahwa

tingkat akurasi analisis menggunakan pendekatan *Item Response Theory* lebih tinggi daripada dengan pendekatan *Classical Test Theory*.

BAB II

TINJAUAN PUSTAKA

A. *Ethnic Experience*

1. Definisi *Ethnic Experience*

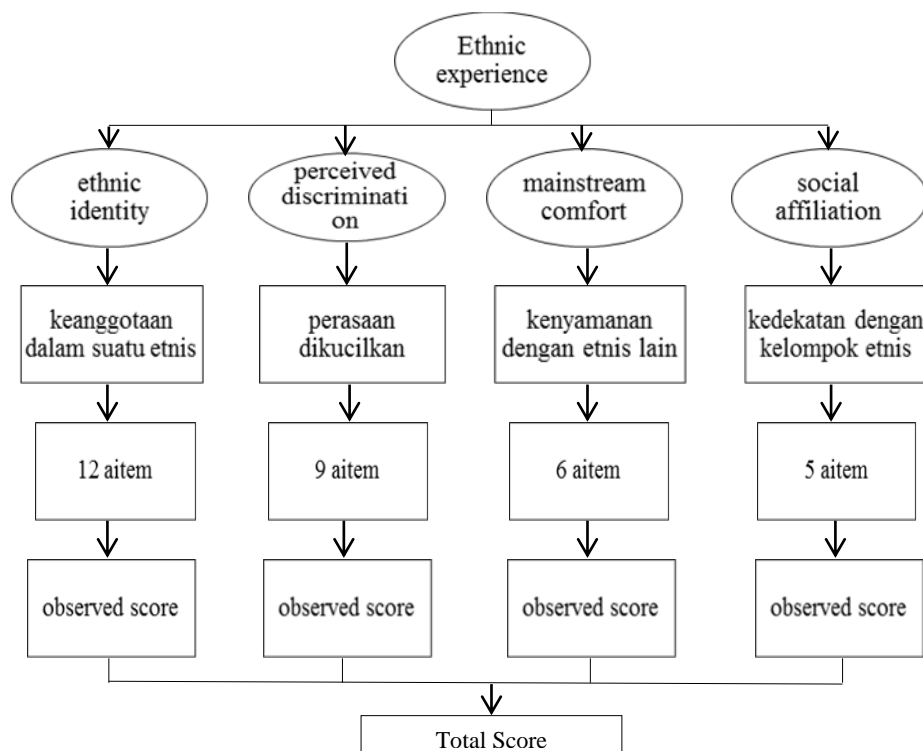
Ethnic Experience adalah pengalaman etnis seseorang pada saat mereka berada di dalam suatu kelompok etnis tertentu, baik itu kelompok etnis mereka sendiri maupun kelompok etnis lain. *Ethnic Experience* termasuk ke dalam komponen kognitif pada teori akulturasi. Akulturasi ialah suatu proses yang melibatkan perubahan secara psikologis dan sosiokultural yang disebabkan karena adanya kontak antar setidaknya dua kelompok etnis yang berbeda (Malcarne, Chavira, Fernandez, & Liu, 2006).

Akulturasi dibagi menjadi dua komponen, yaitu: komponen perilaku dan komponen kognitif. Pada komponen perilaku, memungkinkan individu untuk mempertahankan budaya lamanya ataupun menerapkan budaya baru dalam berbicara, berpakaian, makan, ataupun merayakan suatu tradisi. Sebagai contoh, ketika individu mempertahankan tradisi kelompok etnisnya seperti merayakan hari libur, berbicara bahasa yang berasal dari kelompok etnisnya sendiri, dan mengabaikan keberadaan kelompok etnis lain.

Sedangkan pada komponen kognitif melibatkan preferensi, sikap, dan kepercayaan terhadap kelompok etnis sendiri dan kelompok etnis lain.

Sebagai contoh, ketika individu memilih untuk menciptakan suatu hubungan afiliasi sosial baik dengan kelompok etnis sendiri maupun kelompok etnis lain. Serta adanya persepsi bahwa dirinya dapat diterima oleh kelompok etnis lain. Sehingga dalam mendeskripsikan komponen kognitif pada akulturasi tersebut dibantu oleh konstruk *ethnic experience* (Malcarne, Chavira, Fernandez, & Liu, 2006).

Seperti yang dapat dilihat pada gambar 1, *Ethnic experience* memiliki empat dimensi yaitu *ethnic identity*, *perceived discrimination*, *social affiliation*, dan *mainstream comfort*. Keempat dimensi tersebut masing-masing memiliki indikator perilaku yang kemudian dijadikan sebagai *item*. Keempat dimensi tersebut bersifat *multidimensional with correlated dimensions*, sehingga hanya menghasilkan satu skor saja.



Gambar 1 Konstruk *Ethnic experience*

2. Dimensi *Ethnic Experience*

Menurut Malcarne, Chavira, Fernandez, & Liu (2006), terdapat empat dimensi dalam *ethnic experience*, antara lain:

a. *Ethnic Identity*

Ethnic identity menjelaskan mengenai sikap yang ditunjukkan oleh individu ketika menjadi anggota dalam suatu kelompok etnis tertentu. Sikap tersebut meliputi kebanggaan seseorang terhadap kelompok etnisnya sendiri dan keikutsertaan individu dalam kegiatan yang diadakan oleh kelompok etnis tersebut. Dimensi ini juga menjelaskan bagaimana seseorang merasa menjadi bagian dalam kelompok etnisnya sendiri dan juga ia sangat menghargai nilai-nilai dalam kelompok etnisnya.

b. *Perceived Discrimination*

Perceived discrimination merupakan persepsi individu mengenai bagaimana perlakuan yang didapatkannya dari kelompok etnis lain. Individu percaya bahwa kelompok etnisnya mengalami diskriminasi atau merasa dikucilkan.

c. *Mainstream Comfort*

Mainstream comfort dijelaskan sebagai persepsi seseorang bahwa mereka merasa nyaman hidup bersama dengan kelompok etnis dari masyarakat umum di tempat dimana ia tinggal. Ketika individu bukan berasal dari kelompok etnis tempat dimana ia tinggal sekarang, namun

individu tersebut merasa memiliki karakteristik dari kelompok etnis tersebut.

d. *Social Affiliation*

Social affiliation ialah ketika seseorang memilih untuk menjalin hubungan dengan anggota kelompok etnis sendiri maupun kelompok etnis lain. Hubungan tersebut dapat berupa hubungan pertemanan ataupun hubungan romantis. Selain itu, juga menjelaskan tentang perasaan nyaman yang didapat ketika seseorang tersebut menjalin hubungan dengan kelompok etnis sendiri maupun kelompok etnis lain.

B. Karakteristik Psikometrik

1. Lokasi Parameter

Dinamakan lokasi parameter (atau jika pada pendekatan CTT disebut dengan *item difficulty*), karena dalam pendekatan IRT lokasi parameter mampu menggambarkan suatu *item* yang juga berfungsi di sepanjang skala kemampuan. Sebagai contoh, sebuah *item* yang mudah akan cenderung berada di antara peserta berkemampuan rendah dan sebuah *item* yang sulit akan cenderung berada di antara peserta berkemampuan tinggi (Baker, 2001). Lokasi parameter merupakan salah satu parameter butir soal yang sangat berguna dalam penganalisisan suatu tes. Hal ini disebabkan karena dengan melihat parameter butir ini, akan diketahui seberapa baiknya kualitas suatu butir soal. Jika pada pendekatan IRT hasil dari lokasi parameter digambarkan melalui kurva karakteristik *item* (*item*

characteristic curve), sehingga dapat disimpulkan bahwa dalam setiap *item* harus dipresentasikan melalui satu *Item Characteristic Curve* (ICC) * IÃngt."W{cp,m." ("Vgmt."4234). ICC pada *item* yang lebih sulit berada di sebelah kanan atau semakin tinggi kemampuan yang dibutuhkan untuk menjawab *item* dengan benar, letak ICC semakin ke kanan. Semakin mudah *item* atau semakin rendah kemampuan letak ICC semakin ke kiri (Rahmawati, 2014).

Jika pada pendekatan CTT, lokasi parameter dapat dilihat melalui indeks *item difficulty* yang dihasilkan pada setiap *item*. Besarnya indeks *item difficulty* antara .00 sampai dengan 1.00, dimana *item* dengan indeks *item difficulty* .00 menunjukkan bahwa *item* itu terlalu sulit, sebaliknya indeks 1.00 menunjukkan bahwa *item* terlalu mudah. Adapun perhitungan indeks *item difficulty* dapat dilihat pada Persamaan 1.

$$\boxed{\frac{F_{cp}}{N}} \quad (1)$$

menunjukkan indeks *item difficulty*, yang berasal dari singkatan $\frac{F_{cp}}{N}$ menunjukkan banyaknya subjek yang menjawab *item* tersebut dengan benar. Serta N ialah jumlah seluruh subjek (Solichin, 2017).

Sedangkan pada pendekatan IRT, lokasi parameter didapatkan dari teknik *Generalized Partial Credit Model* (GPCM) dengan melihat parameter θ . GPCM merupakan perluasan dari model 2PL dan pengembangan dari model *Partial Credit Model* (PCM). Model Rasch pada PCM digunakan pada setiap pasang kategori yang berdekatan.

Misalnya, pada *item* yang memiliki empat kategori respon, akan memiliki tiga pasang kategori skor yang berdekatan yaitu: 0 dan 1, 1 dan 2, 2 dan 3. Agar memudahkan dalam memahami konsep dari PCM dapat dilihat pada gambar 1.



Gambar 1. Kategori skor pada empat kategori respon

Oleh karena itu, PCM memberikan probabilitas memperoleh skor x daripada $x - 1$. Hal ini berbeda dengan GRM, yang memberikan probabilitas memperoleh skor x atau di atasnya. PCM memberikan probabilitas mendapatkan skor 1 daripada skor 0 yang diberikan oleh beberapa tingkat kemampuan (), sehingga akan terjadi peningkatan

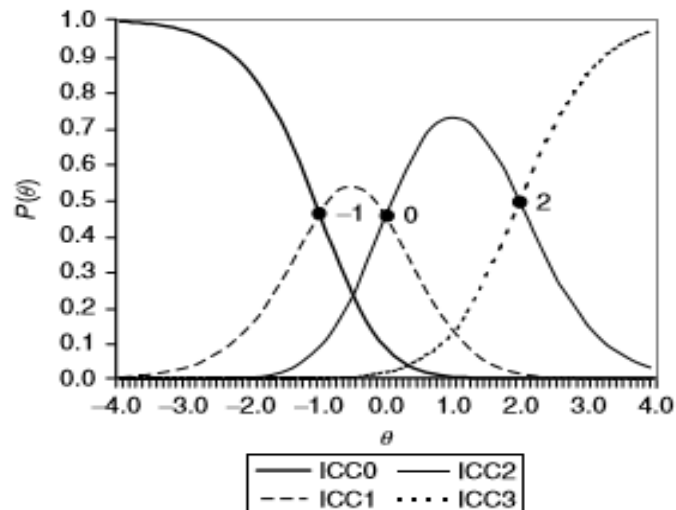
ugectc" o qpqvqp" fgpicp" ogpikmwk" " rfc"ugvkr"kvgo0" Cuwouk" rfc"REO"

menganggap bahwa ketika individu mendapatkan skor 3, maka ia dianggap telah melewati skor 1 dan skor 2 (Keller, 2005). Bagi subjek dengan tingkat kemampuan () tertentu dan untuk item tertentu, memiliki probabilitas menjawab benar seperti yang telah disajikan pada persamaan 2.

$$\frac{e^{-\theta_i - \beta_0}}{e^{-\theta_i - \beta_0} + e^{-\theta_i - \beta_1}} \quad (2)$$

merupakan level trait individu atau lokasi trait individu pada kontinum trait laten. merupakan parameter lokasi *item* yang menunjukkan memperoleh skor 0 dan 1 sama (Widhiarso, 2010). Selain itu

juga terdapat gambar yang mempresentasikan kategori skor, seperti pada gambar 3 untuk 4 kategori respon dengan



Gambar 2. Skor *category response function* pada PCM

Sedangkan perbedaan antara PCM dan GPCM terletak pada daya diskriminasi, dimana pada PCM tidak memiliki daya diskriminasi atau dianggap sama. Oleh karena itu, probabilitas ini diberikan oleh fungsi yang berbeda, yaitu fungsi 2PL yang disajikan dalam Persamaan 3.

$$\boxed{\text{[Empty box for Equation 3]}} \quad (3)$$

merupakan level trait, merupakan lereng pada *item* (*i*), dan merupakan perpotongan antar kategori pada tiap *item* (Widhiarso, 2010).

Suatu pendekatan akan dikatakan lebih baik dalam menganalisis *item difficulty*, ialah ketika pendekatan tersebut mampu memberikan hasil yang bersifat *independent* dimana setiap *item* akan menghasilkan nilai indeks *item difficulty* yang relatif sama meskipun diberikan kepada subjek dengan kemampuan berbeda (Nasir, 2014).

2. *Item Discrimination*

Item discrimination (daya beda) adalah kemampuan *item* dalam membedakan kemampuan subjek, yaitu antara subjek yang memiliki kemampuan tinggi dan subjek yang memiliki kemampuan rendah. *Item discrimination* dapat diketahui dengan melihat besar kecilnya indeks diskriminasi atau angka yang menunjukkan besar kecilnya daya beda. Fungsi dari daya pembeda tersebut adalah untuk mengetahui secara detail pada perbedaan kemampuan individu. Penentuan daya beda *item* yang paling sering digunakan adalah dengan melihat indeks korelasi (Lababa, 2008).

Teknik *corrected item total correlation* digunakan pada pendekatan CTT, dimana korelasi antara skor *item* dan skor total dilakukan dengan tanpa memasukkan skor *item* itu sendiri. Fungsi dari *corrected item total correlation* seperti yang diekspresikan dalam persamaan 4.

$$\frac{r_{it} - r_{it}^2}{\sqrt{1 - r_{it}^2}} \quad (4)$$

merupakan koefisien dari korelasi *item* total setelah dikoreksi (*corrected item total correlation*), merupakan koefisien dari korelasi *item* total sebelum dikoreksi (*item total correlation*), merupakan standar deviasi skor total, dan merupakan standar deviasi dari *item* (Kamila, 2014).

Sedangkan pada pendekatan IRT juga menggunakan teknik GPCM dengan melihat parameter α . Parameter α ini juga menunjukkan sejauh mana respon kategori bervariasi antar item ketika level trait berubah. Nilai item di bawah 1.0 akan membuat kurva menjadi lebih landai sebaliknya nilai lereng di atas 1.0 akan membuat garis kurva menjadi lebih curam. Sama seperti PCM, persamaan tersebut mampu dijelaskan berdasarkan jumlah kategori dalam item (Widhiarso, 2010).

Suatu pendekatan akan dikatakan lebih baik dalam menganalisis *item discrimination*, ialah ketika suatu pendekatan tersebut memiliki suatu parameter dalam model penggambarannya karena *item discrimination* dianggap tidak sama. Akan lebih baik lagi apabila juga memiliki parameter tebakan, karena beberapa subjek dengan kemampuan rendah cenderung menebak jawaban yang benar pada *item* yang paling sulit (Abedalaziz & Leng, 2017).

3. Reliabilitas

Reliabilitas artinya sejauh mana hasil dari suatu alat ukur dapat dipercaya. Alat ukur yang dikatakan reliabel ialah ketika alat ukur tersebut digunakan dalam kelompok subjek yang sama beberapa kali, maka akan memberikan hasil yang relatif sama. Apabila hasil dari reliabilitas semakin mendekati angka 1 maka semakin reliabel, dan begitu juga sebaliknya (Kamila, 2014). Teknik uji reliabilitas ini dengan menggunakan *Etqpdcejøu" Cnrjc* pada pendekatan CTT, karena dianggap paling mudah dan lebih efisien jika dibandingkan dengan teknik lainnya, terutama pada

item yang berjumlah banyak (Trochim, 2006). Fungsi dari $E_{tqpdcej\theta u}$ α seperti yang diekspresikan dalam persamaan 5.

$$\frac{1}{n} \sum_{i=1}^n \frac{\sigma_i^2}{\sigma^2} \quad (5)$$

merupakan koefisien reliabilitas, n merupakan banyaknya *item* pada soal, σ_i^2 merupakan varians pada *item*, dan σ^2 merupakan varians skor total (Kamila, 2014). Sedangkan pada pendekatan IRT dengan menggunakan *information* yang dianalogikan sebagai reliabilitas. Informasi yang dihasilkan berupa *item information function* dan *test information function*. *Item Information Function* (IIF) menggambarkan tentang sejauh mana masing-masing model yang digunakan pada IRT (1PL, 2PL, atau 3PL) dalam memberikan informasi. Informasi tersebut berkaitan dengan estimasi *traits-level* sepanjang skala *latent-traits*. Fungsi yang berkebalikan dengan *Information Function* (IF) dinamakan *Standard Error of Measurement* (SEM), dimana semakin tinggi IF maka semakin rendah SEM dan begitu juga sebaliknya (Ridho, 2006).

IIF memberikan penilaian secara akurat pada *item* dalam membedakan subjek yang memiliki tingkat kemampuan yang berbeda-beda berdasarkan konstruk yang sedang diukur. Semakin tinggi nilai informasi maka semakin akurat. *Item* informasi bergantung pada parameter *item*. Kurva *item* informasi akan runcing apabila parameter () atau parameter diskriminasinya tinggi, dan apabila parameter ()-nya rendah maka akan menghasilkan kurva yang datar. Melalui pengukuran model Rasch, *the person-separation index* digunakan sebagai indeks

reliabilitas yang menggambarkan seberapa akurat suatu *item* dalam membedakan kemampuan subjek (Cappelleri, Lundy, & Hays, 2014). Dikarenakan GPCM merupakan perluasan dari 2PL, untuk itu IIF pada model 2PL diekspresikan dalam Persamaan 6 dan 7.

$$\boxed{} \quad (6)$$

$$\boxed{} \quad (7)$$

Melalui persamaan-persamaan tersebut dijelaskan bahwa merupakan parameter kesulitan *item*, merupakan parameter diskriminasi *item*, merupakan tingkat kemampuan subjek, merupakan nilai konstan dari 2.718, dan merupakan .

Lalu juga dihasilkan *information* pada tes yang disebut dengan *Test Information Function* (TIF). Salah satu fungsi TIF yang tidak dimiliki oleh CTT ialah bahwa setiap *item* berkontribusi pada total informasi secara independen. Kontribusi *item* tes pada subjek dapat dijelaskan tanpa mengetahui nilai *item* lain pada tes. TIF juga digunakan untuk menjelaskan satu set *item* tes pada setiap titik dalam skala. Jumlah informasi dipengaruhi oleh kualitas dan jumlah *item* tes. Fungsi dari TIF diekspresikan dalam Persamaan 8.

$$\boxed{} \quad (8)$$

Dijelaskan juga bahwa semakin curam lereng yang dihasilkan maka akan semakin besar informasi. Lalu, semakin kecil varians pada *item* maka akan semakin besar informasinya. Jumlah informasi yang diberikan oleh serangkaian *item* tes pada tingkat kemampuan berbanding terbalik

C. *The Scale of Ethnic Experience (SEE)*

SEE merupakan alat ukur yang dibuat untuk menilai konstruk yang berkaitan dengan komponen kognitif pada akulturasi seseorang. SEE pertama kali dibuat oleh Malcarne, dkk pada tahun 2006, yang terdiri dari 32 item, dan memiliki *item favorable* dan *item unfavorable*. Serta memiliki empat dimensi, antara lain: *ethnic identity*, *perceived discrimination*, *social affiliation*, dan *mainstream comfort*. SEE berbentuk kuesioner *self report* dengan menggunakan Likert dan memiliki jumlah respon sebanyak lima kategori, seperti yang disajikan pada Tabel 1.

Tabel 1

Skor respon jawaban pada skala *ethnic experience*

<i>Item Favorable</i>	Skor	<i>Item Unfavorable</i>	Skor
Sangat tidak Setuju	1	Sangat tidak Setuju	5
Tidak Setuju	2	Tidak Setuju	4
Agak Setuju	3	Agak Setuju	3
Setuju	4	Setuju	2
Sangat Setuju	5	Sangat Setuju	1

Sebelumnya telah dilakukan uji validitas pada SEE, dengan menggunakan sampel dari beberapa mahasiswa yang berasal dari kelompok etnis yang berbeda-beda, yaitu pada mahasiswa *African American*, *Caucasian American*, *Filipino American*, dan *Mexican American*. Berdasarkan hasil uji validitas dengan menggunakan validitas konkuren diperoleh hasil bahwa SEE memiliki nilai korelasi yang tinggi. Korelasi dilakukan antara SEE dengan alat ukur lain yang juga mengukur dimensi yang sama seperti *ethnic identity* dan dimensi pada akulturasi. Kemudian dari uji reliabilitas telah ditemukan bahwa SEE memiliki karakteristik psikometrik yang bagus, dengan koefisien konsistensi

internal pada dimensi *ethnic identity* berkisar antara .81 hingga .90, pada dimensi *perceived discrimination* berkisar antara .76 hingga .91, pada dimensi *social affiliation* berkisar antara .81 hingga .84, dan pada dimensi *mainstream comfort* berkisar antara .76 hingga .87 (Malcarne, Chavira, Fernandez, & Liu, 2006).

Kemudian *blueprint* yang digunakan yaitu dengan mengikuti *blueprint* dari SEE, sehingga tidak terdapat perubahan pada susunan *itemnya*. Berikut adalah *blueprint* pada SEE:

Tabel 2

Blueprint skala *ethnic experience*

Dimensi	F	UF	Jumlah	Nomor Item	Presentase
<i>Ethnic Identity</i>	6	6	12	*1, *4, *7, *8, *12, 14, 16, 20, 23, 25, *27, 30	37,5%
<i>Perceived Discrimination</i>	5	4	9	*2, *3, 13, 19, *21, 24, 26, 29, *31	28,125%
<i>Mainstream Comfort</i>	4	2	6	6, 9, *11, 17, 28, *32	18,75%
<i>Social Affiliation</i>	5	-	5	5, 10, 15, 18, 22	15,625%
Jumlah			32	32	100%

Catatan. *) = *item unfavorable*

D. Kerangka Pemikiran

Berdasarkan dari uraian penjelasan sebelumnya, kerangka pemikiran dalam penelitian ini berasal dari dua pendekatan yang digunakan yaitu CTT dan IRT. CTT dan IRT merupakan dua pendekatan yang paling umum digunakan dalam perhitungan analisis karakteristik psikometrik suatu alat ukur. Jika pada penelitian ini, pengujian tersebut

dilakukan karakteristik psikometrik dari SEE yang terdiri dari lokasi parameter, *item discrimination*, reliabilitas, dan SEM. Terdapat beberapa indikator yang dapat menentukan pendekatan mana yang dianggap lebih baik dalam membandingkan kedua pendekatan tersebut, diantaranya ialah suatu pendekatan yang mampu memberikan tingkat keakuratan yang lebih tinggi dengan memberikan informasi yang lebih banyak dan mendetail mengenai suatu alat ukur. Selain itu juga suatu pendekatan yang bersifat *independent*, sehingga hasil yang diberikan lebih dapat dipercaya ketika alat ukur tersebut digunakan kembali beberapa kali.

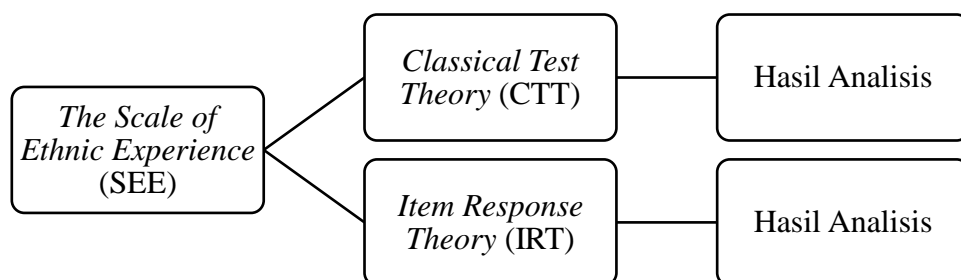
Ketika menyeleksi *item*, pada pendekatan CTT dengan menguji *item discrimination* menggunakan teknik *corrected item total correlation*, serta menguji reliabilitas dengan menggunakan *Cronbach Alpha*. Sedangkan pada pendekatan IRT dengan menguji IF yang berupa IIF dan TIF menggunakan teknik GPCM. Setelah didapatkan hasilnya kemudian akan dibandingkan, agar dapat diketahui *item* mana yang dieliminasi. Apakah nomor *item* yang dieliminasi antara pendekatan CTT dan pendekatan IRT sama atau berbeda. Serta ditentukan pendekatan mana yang lebih sesuai digunakan untuk menganalisis SEE. Sehingga dalam penelitian ini akan diketahui bagaimana pendekatan CTT dan IRT jika digunakan untuk menganalisis karakteristik psikometrik SEE.

BAB III

METODE PENELITIAN

A. Desain Penelitian

Metode kuantitatif psikometrik digunakan pada penelitian ini, dimana pada metode tersebut akan dilakukan perbandingan dua pendekatan antara dua pendekatan yaitu pendekatan CTT dan pendekatan IRT. Serta menggunakan skala adaptasi dari SEE sebagai alat ukur yang akan diuji dalam penelitian ini. Kemudian dalam menganalisis data dengan menggunakan teknik *corrected item total correlation*, Cronbach Alpha, dan SEM jika pada pendekatan CTT, sedangkan pada IRT dengan menggunakan GPCM, IF, dan SEM. Desain penelitian ditampilkan pada Gambar 4.



Gambar 1. Desain penelitian perbandingan CTT dan IRT

B. Fokus Penelitian

Penelitian ini berfokus untuk melihat perbandingan pendekatan CTT dan IRT dalam menganalisis karakteristik psikometrik SEE atau karakteristik tesnya.

C. Definisi Operasional

Definisi operasional digunakan untuk mengatasi kesalahan dalam melakukan interpretasi pada variabel penelitian. Berikut merupakan definisi operasional dari variabel penelitian yang digunakan pada penelitian ini:

1. Lokasi Parameter

Lokasi parameter dapat ditentukan dengan melihat indeks kesulitan *item* jika pada CTT. Sedangkan pada IRT, dapat dengan melihat parameter yang diperoleh dari penggunaan teknik GPCM.

2. Item Discrimination

Item discrimination menunjukkan kemampuan suatu *item* dalam mengungkap kemampuan subjek. Jika pada pendekatan CTT digunakan teknik *corrected item total correlation* yang mengkorelasikan skor *item* dan skor total tanpa memasukkan skor *item* itu sendiri. Sedangkan pada IRT dengan melihat parameter yang juga dihasilkan dari teknik GPCM.

3. Reliabilitas

Reliabilitas digunakan untuk menguji seberapa konsisten suatu alat ukur jika digunakan pada subjek yang sama. Jika pada pendekatan CTT digunakan teknik *Cronbach Alpha* untuk menguji reliabilitas, sedangkan pada pendekatan IRT menggunakan *information* yang terdiri dari *item information function* dan *test information function*.

4. *Standard Error of Measurement*

SEm digunakan untuk melihat seberapa besar kemungkinan *error* yang terjadi pada suatu alat ukur. Dimana jika pada CTT hanya menghasilkan satu SEM pada tes, sedangkan pada IRT selain menghasilkan SEM pada tes juga menghasilkan SEM pada masing-masing *item*.

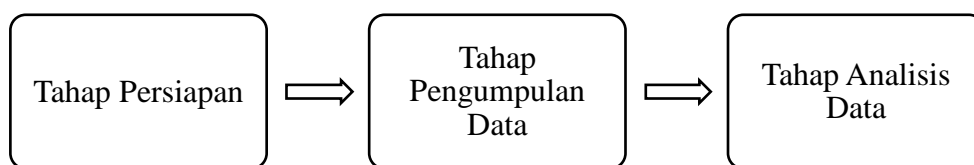
D. Subjek Penelitian

Populasi dalam penelitian ini adalah mahasiswa laki-laki dan perempuan di Fakultas Ilmu Sosial dan Ilmu Politik Universitas Brawijaya Malang yang berusia 18-25 tahun dan bukan berasal dari Malang. Sedangkan sampel yang digunakan sebanyak 400 mahasiswa. Penentuan jumlah sampel berdasarkan Baldwin (2006), yaitu jumlah sampel untuk pendekatan IRT minimal 400 orang. Teknik sampling yang digunakan ialah *accidental sampling*, yaitu pengambilan responden sebagai sampel berdasarkan jumlah sampel yang tersedia di suatu tempat dan yang sesuai

dengan kriteria subjek yang telah ditetapkan berdasarkan tujuan penelitian (Sugiyono, 2004).

E. Tahap Pelaksanaan Penelitian

Terdapat tiga tahapan yang wajib dilakukan dalam penelitian ini. Tahapan tersebut terdiri dari tahap persiapan, tahap pengumpulan data, dan tahap analisis data. Alur pelaksanaan penelitian akan dijelaskan pada Gambar 5.



Gambar 2 Alur pelaksanaan penelitian

1. Tahap persiapan

Tahap persiapan merupakan tahapan yang pertama kali dilakukan pada penelitian ini, dimana pada tahapan tersebut terdiri dari:

- a. Peneliti menerjemahkan skala dari alat ukur SEE milik Malcarne, Chavira, Fernandez, & Liu (2006). Menurut Beaton, dkk (2000) menyarankan bahwa terdapat langkah-langkah yang wajib diikuti ketika menerjemahkan skala, antara lain:

1) *Translation*

Pertama, Skala diterjemahkan dari bahasa aslinya yaitu bahasa Inggris ke bahasa yang diinginkan yaitu Bahasa Indonesia.

Digunakan dua translator ahli yang berbeda dalam proses menerjemahkan skala.

2) *Synthesis*

Kemudian hasil terjemahan dari dua translator yang berbeda tersebut dinamakan T1 dan T2. Lalu, antara T1 dan T2 disatukan dengan mengambil kesimpulannya, yang diebut dengan T-12.

3) *Back Translation*

Lalu T-12 dilakukan *back translation* dari bahasa Indonesia ke dalam bahasa Inggris. Digunakan dua translator yang berbeda dari sebelumnya dalam proses *back translation*. Hasil dari *back translation* disebut dengan BT1 dan BT2.

4) *Expert Comittee Review*

Setelah itu, dilakukan review pada skala yang digunakan, T1, T2, T12, BT1, dan BT2 oleh dengan bantuan ahli (*expert judgement*).

5) *Pretesting*

Setelah skala sudah selesai disusun, kemudian dilakukan uji coba pada sekelompok orang yang berjumlah sekitar 30-40 orang.

- b. Menyiapkan waktu dan tempat untuk mengambil data dari sampel. Kemudian mengadministrasikan skala kepada subjek secara langsung.

2. Tahap pengumpulan data

Digunakan *The Scale of Ethnic Experience* (SEE) sebagai alat pengumpulan data pada tahap ini. Tahap pengumpulan data pada penelitian ini akan dilakukan di Universitas Brawijaya.

3. Tahap analisis data

Tahap analisis data merupakan tahapan terakhir yang akan dilakukan oleh peneliti, yang terdiri dari :

- a) Memeriksa kembali data-data yang telah terkumpul tersebut, guna memisahkan antara data mana saja yang sesuai dan dapat dianalisis dan yang tidak.
- b) Lalu memberikan skor pada jawaban yang telah diisi subjek pada SEE.
- c) Kemudian menghitung dan menganalisis sesuai dengan teknik yang akan digunakan.
- d) Setelah itu, membahas hasil analisis data dan diinterpretasi. Baru kemudian diberikan kesimpulan dari hasil penelitian tersebut.

F. Teknik Analisis Data

Teknik analisis data yang digunakan merupakan beberapa teknik perhitungan yang termasuk ke dalam pendekatan CTT dan IRT. Kemudian dalam menganalisis data dibantu oleh R versi 3.4.0 dengan package *psych* dan *ltm*, antara lain:

1. *Corrected item total correlation*

Teknik ini digunakan untuk menguji validitas dengan mengkorelasikan antar skor tiap *item* yang didapatkan dengan skor total subjek, dimana skor *item* itu sendiri tidak dimasukkan ke dalam perhitungan. Bila koefisien korelasi yang dihasilkan baik pada SEE, maka *item* dalam SEE dapat diterima dalam pengukuran. Sehingga akan didapatkan beberapa *item* dalam skala yang dapat digunakan untuk mengukur konstruk yang sedang diukur. Apabila ingin mengetahui rumus dari *corrected item total correlation* dapat dilihat pada Persamaan 4.

2. *Cronbach Alpha*

Selain menguji validitas suatu alat ukur, diperlukan juga untuk mengestimasi reliabilitas suatu alat ukur guna melihat kualitas dari alat ukur tersebut. Teknik dari pendekatan CTT yang digunakan untuk mengetahui reliabilitas ialah Cronbach Alpha. Apabila ingin mengetahui perhitungan Cronbach Alpha dapat dilihat pada Persamaan 5.

3. *Generalized Partial Credit Model*

Fungsi dari teknik GPCM hampir sama dengan teknik PCM, perbedaannya terletak pada penambahan parameter lereng atau parameter diskriminasi (). Dimana pada parameter diskriminasi menjelaskan sejauh mana *item* mampu membedakan tingkat kemampuan individu yang berbeda-beda. Hasil dari GPCM dapat dilihat pada IIF. IIF memberikan hasil berupa *item discrimination* hingga *item difficulty* pada setiap *item*

(Ridlo, 2016). Apabila ingin mengetahui perhitungan GPCM dapat dilihat pada Persamaan 3.

4. *Information Function*

IF yang digunakan berupa IIF dan TIF. IIF memberikan informasi pada *item*, sedangkan TIF memberikan informasi pada tes. Apabila ingin mengetahui perhitungan IIF pada model 2PL dapat dilihat pada Persamaan 6 dan 7. Sedangkan TIF dapat dilihat pada Persamaan 8.

5. *Standard Error of Measurement*

SEm memberikan perkiraan adanya *error* pada skor setiap subjek. Fungsi yang berkebalikan dengan SEm adalah IF, dimana semakin rendah SEm maka semakin tinggi IF dan begitu juga sebaliknya (Ridho, 2006). Untuk mengetahui perhitungan SEm dapat dilihat pada Persamaan 9.

BAB IV

HASIL DAN PEMBAHASAN

A. Gambaran Umum Subjek Penelitian

Subjek yang digunakan dalam penelitian ini memiliki karakteristik antara lain mahasiswa Universitas Brawijaya Malang, berada pada rentang usia 18-25 tahun, dan bukan berasal dari Malang. Subjek yang digunakan berjumlah sebanyak 400 orang. Subjek tersebut dibagi dalam beberapa kategorisasi berdasarkan etnis, usia, dan jenis kelamin seperti yang disajikan pada tabel 3.

Tabel 1
Kategorisasi subjek penelitian

Kategorisasi Subjek Penelitian	Jenis Kategori	Jumlah
Etnis	Ambon	2
	Bali	2
	Banjar	4
	Batak	22
	Betawi	14
	Bugis	4
	Dayak	4
	Flores	1
	Jawa	267
	Komering	1
	Kutai	2
	Madura	4
	Makassar	2
	Mbojo	1
	Melayu	9
	Minang	9
	Nias	4
	Padang	3
	Palembang	2
	Pamona	1
Sasak	3	
Sunda	27	
Tionghoa	2	
Toraja	4	

(lanjut)

Tabel 3

Kategorisasi subjek penelitian

Kategorisasi Subjek Penelitian	Jenis Kategori	Jumlah
Usia	17-20 tahun	273
	21-25 tahun	127
Jenis Kelamin	Laki-Laki	162
	Perempuan	238

Seperti yang dijelaskan pada Tabel 3 bahwa subjek penelitian yang digunakan terdiri dari beberapa macam etnis dengan etnis yang paling dominan yaitu berasal dari etnis Jawa yang berjumlah 267 orang. Serta subjek yang paling dominan berada pada rentang usia 17-20 tahun.

B. Hasil Penelitian

Perbandingan pendekatan *Classical Test Theory* (CTT) dan *Item Response Theory* (IRT) dalam menganalisis karakteristik psikometrik *The Scale of Ethnic Experience* (SEE) menjadi tujuan utama pada penelitian ini. SEE merupakan skala yang pertama kali dikembangkan oleh Malcarne, Chavira, Fernandez, dan Liu (2006). Terdapat empat dimensi pada SEE diantaranya adalah *Ethnic Identity*, *Perceived Discrimination*, *Mainstream Comfort*, dan *Social Affiliation*.. Keempat dimensi tersebut bersifat *multidimensional with correlated dimensions*, sehingga hasil yang diberikan tidak pada masing-masing dimensi akan tetapi dilihat secara keseluruhan. Berikut adalah perbandingan hasil dari pendekatan CTT dan IRT yang disajikan pada Tabel 4.

Tabel 2

Perbandingan hasil analisis item pendekatan CTT dan IRT pada SEE

Item	CTT	IRT				A
		GPCM				
1	.32	-3.34	-1.34	.66	1.43	.49
2	-.19	-6.43	1.00	8.70	4.27	.10
3	-.06	-3.86	7.58	5.71	.09	.11
4	.30	-1.57	-.22	.26	2.50	.43
5	.29	-2.94	.73	1.48	3.20	.35
6	.22	-4.01	-3.19	-.61	3.62	.40
7	.33	-1.42	.38	-.41	3.63	.49
8	.41	-2.29	-1.53	-1.16	1.40	.99
9	.19	-5.68	-1.8	-.45	4.11	.38
10	.41	-2.69	.54	1.30	1.66	.61
11	.15	-5.43	-2.62	-.95	3.83	.37
12	.23	-2.67	-1.64	-1.84	1.31	.44
13	-.03	-4.51	6.17	4.45	1.39	.10
14	.33	-1.85	-2.17	-1.44	1.42	.79
15	.32	-.19	2.39	2.81	3.56	.36
16	.18	-5.23	-3.31	-1.75	1.93	.43
17	.33	-3.44	-1.86	0.32	2.60	.54
18	.43	-2.51	.24	1.74	2.77	.66
19	.29	-2.26	-1.64	1.27	2.92	.43
20	.45	-2.42	-1.83	.06	1.91	1.15
21	.08	-3.66	4.98	4.79	1.44	.15
22	.28	-2.65	1.13	3.45	4.46	.40
23	.32	-2.92	-1.46	-.99	1.74	.72
24	.03	-6.13	5.75	4.85	3.33	.15
25	.41	-2.78	-1.67	.57	1.90	.80
26	-.00	-5.85	4.85	7.21	3.72	.14
27	.40	-3.01	-.70	.05	2.32	.82
28	.01	-5.18	-4.56	-4.44	3.15	.29
29	.08	-5.33	2.81	-.18	7.61	.16
30	.46	-2.42	-1.51	.27	2.19	1.00
31	.09	-6.17	4.02	2.04	5.15	.16
32	.26	-2.30	-3.05	.28	3.04	.45

Etqpdcejou" Cnr jc"092

SEM: 5.59

TIF: 57.94

GPCM Log.lik: -17048.96

Catatan: Hasil yang dicetak tebal merupakan item yang dieliminasi. CTT = Classical Test Theory; IRT = Item Response Theory; r_{ctt} = corrected item total correlation; b , d , f = item difficulty; a = item discrimination; Etqpdcejou" Cnr jc" ?" tgnkedknkvu" rcf" EVV=" UGO" ?" Standard Error of Measurement; TIF = Total Information Function; Log.lik = Log Likelihood

Berdasarkan analisis dari pendekatan CTT memberikan hasil yang akan dijelaskan sebagai berikut:

- a. Nilai lokasi parameter pada pendekatan CTT dapat dilihat pada indeks *item difficulty*, dimana nilai yang dihasilkan pada setiap *item* berkisar antara -.1 hingga .53. Menurut Allen & Yen (1979) nilai yang disarankan berkisar antara .30 hingga .70, karena nilai di bawah itu .30 akan dianggap sebagai *item* yang sulit sedangkan *item* dengan nilai di atas .70 akan dianggap sebagai *item* yang mudah.
- b. *Item discrimination* pada pendekatan CTT diuji melalui teknik *corrected item total correlation* dengan menghitung yang menggunakan *cut-off point* sebesar .20. Sehingga *item* yang dianggap baik dan dapat diterima apabila *item* memiliki $\times"042."$ ugfcpi MCP" *item* yang dianggap buruk ialah *item* yang memiliki $< .20$ dan akan dieliminasi. Seperti yang disajikan pada Tabel 4, diperoleh hasil bahwa *item* yang diterima pada SEE berjumlah 20 *item*.
- c. Teknik analisis *Etqpdcejøu" Cnrjc"* pada pendekatan CTT untuk melihat suatu reliabilitas, dimana reliabilitas yang dihasilkan hanya 1 yaitu berupa reliabilitas pada tes. Reliabilitas yang dihasilkan dari teknik *Etqpdcejøu" Cnrjc"* pada SEE yaitu sebesar .70.
- d. Berdasarkan teknik analisis yang telah digunakan pada pendekatan CTT, diperoleh hasil nilai SEM pada SEE sebesar 5.59, dimana nilai tersebut termasuk angka yang besar karena menurut Hambleton &

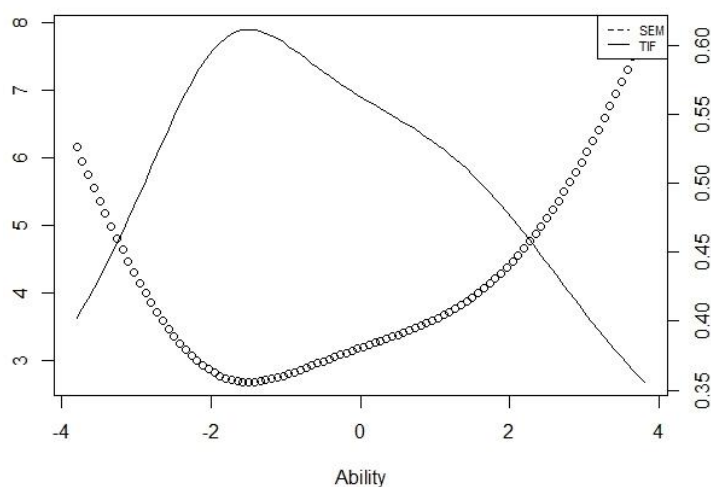
Swaminathan (1985) batas nilai *error* yang masih bisa ditoleransi yaitu sebesar .40.

Sedangkan berdasarkan analisis dari pendekatan IRT memberikan hasil yang akan dijelaskan sebagai berikut:

- a. Lokasi parameter pada pendekatan IRT dilihat melalui parameter dengan menggunakan teknik GPCM. Melalui teknik GPCM menghasilkan parameter dari μ , σ , dan δ pada setiap item. Item yang dianggap baik ialah item yang memiliki parameter yang berurutan (*in order*) dari μ , σ , hingga δ dan tidak memiliki nilai yang ekstrem. Parameter yang disarankan oleh Hambleton & Swaminathan (1985) yaitu berada pada rentang -6 "Ö" " Ö"-6. Sehingga untuk item 2, 3, 7, 9, 11, 12, 13, 14, 16, 21, 24, 26, 28, 29, 31, 32 dianggap sebagai item dengan kualitas yang rendah karena cenderung memiliki nilai parameter yang tidak berurutan dan ekstrim. Hasil tersebut dapat dilihat pada Tabel 4.
- b. Parameter juga dihasilkan melalui teknik GPCM, dimana pada teknik ini memiliki probabilitas seseorang memilih daripada . Berdasarkan dari analisis yang telah dilakukan, menghasilkan parameter yang memiliki fungsi sebagai daya pembeda pada besarnya varian item antara subjek yang memiliki *latent trait* dengan subjek yang tidak memiliki *latent trait*. Menghasilkan interpretasi dcjyc"ugocmkp"vkpiik"rctcogvgt" "omc" ckan semakin baik item tersebut. Seperti yang dapat dilihat pada Tabel 4, bahwa hasil yang

diberikan berada pada kategori rendah dan sedang. Karena menurut Baker (2001), item yang berada pada rentang nilai .35-.64 dikatakan memiliki parameter dalam kategori rendah, dan item yang berada pada rentang nilai .65-.1.34 dikatakan memiliki parameter dalam kategori sedang.

- c. Sedangkan pada pendekatan IRT, reliabilitas dianalogikan sebagai *information*. Sehingga untuk melakukan pengukuran reliabilitas dengan melihat *Test Information Function* (TIF), dimana TIF sendiri merupakan kumpulan dari seluruh IIF yang diperoleh pada setiap item dalam suatu alat ukur. Berdasarkan teknik analisis yang telah dilakukan, nilai TIF pada SEE yang dihasilkan yaitu sebesar 57.94. Seperti yang telah ditunjukkan pada Gambar 7, bahwa puncak tertinggi nilai TIF terletak pada subjek dengan rentang *ability* -2 hingga 0 dan rendah pada subjek dengan rentang *ability* 2 hingga 4, dan berkebalikan dengan nilai SEM dimana informasi tertinggi terletak pada rentang *ability* 2 hingga 4 dan informasi terendah terletak pada *ability* -2 hingga 0.



Gambar 1. Hubungan SEM dan TIF pada SEE

Selain itu juga dihasilkan *information* pada setiap *item* yang disebut dengan *Item Information Function* (IIF). Dimana nilai yang dihasilkan pada setiap *item* berada pada rentang antara .24 hingga 4.02.

- d. Jika pada pendekatan IRT, SEM dapat digunakan untuk semua subjek sehingga masing-masing subjek memiliki nilai *error* yang berbeda-beda. Nilai SEM pada IRT memiliki hubungan yang terbalik dengan nilai TIF, dimana semakin tinggi nilai TIF maka akan semakin rendah nilai SEM dan begitu juga sebaliknya. Nilai SEM pada IRT, dihitung sesuai dengan Persamaan 8 dan menghasilkan nilai sebesar .13. dimana nilai tersebut termasuk angka yang kecil karena menurut Hambleton & Swaminathan (1985) batas nilai *error* yang masih bisa ditoleransi yaitu sebesar .40.

C. Pembahasan

Berdasarkan dari hasil analisis yang telah dijelaskan sebelumnya, dapat diketahui bahwa dalam menganalisis *item difficulty* pada SEE lebih baik digunakan pendekatan IRT karena seperti yang dikatakan oleh Nasir (2014) bahwa suatu pendekatan akan dikatakan baik ketika pendekatan tersebut mampu memberikan hasil yang bersifat *independent*. Hal ini dapat dilihat pada hasil yang diberikan pada pendekatan IRT melalui parameter yang bersifat *person-free* atau tidak bergantung pada sampel yaitu dimana nilai yang dihasilkan tidak akan berubah ketika diberikan kepada subjek yang memiliki kemampuan berbeda. Sedangkan hasil yang diberikan pada pendekatan CTT dilihat melalui indeks *item difficulty* yang cenderung bergantung pada sampel (Nasir, 2014).

Kemudian dalam menganalisis *item discrimination* pada SEE lebih baik digunakan pendekatan IRT karena seperti yang dikatakan oleh Abedalaziz & Leng (2017) bahwa suatu pendekatan akan dikatakan baik ketika suatu pendekatan tersebut memiliki suatu parameter dalam model penggambarannya. Hal ini sesuai dengan hasil yang diberikan oleh pendekatan IRT yang digambarkan melalui parameter yaitu parameter α . Sedangkan pada pendekatan CTT hanya dibatasi oleh *cut off point* dari teknik *corrected item total correlation*. Selain itu pendekatan IRT juga lebih sensitif dan mendalam. Hal ini ditunjukkan yang diberikan dari hasil perbandingan pendekatan CTT dan IRT bahwa terdapat beberapa *item*

yang lolos atau dianggap baik apabila dianalisis dengan menggunakan CTT namun tidak bisa lolos atau dianggap buruk pada pendekatan IRT.

Lalu dalam menganalisis reliabilitas pada SEE lebih baik digunakan pendekatan IRT karena dikatakan oleh Nasir (2014) bahwa suatu pendekatan akan dikatakan baik ketika pendekatan tersebut memberikan hasil yang bersifat informatif atau lebih banyak informasi yang diberikan. Hal ini sesuai dengan hasil yang diberikan oleh pendekatan IRT dimana nilai reliabilitas yang dihasilkan tidak hanya pada tes saja melalui TIF, akan tetapi masing-masing *item* juga menghasilkan nilai reliabilitasnya sendiri-sendiri melalui IIF. Sedangkan pada pendekatan CTT hanya menghasilkan nilai reliabilitas pada tes saja melalui teknik *Etqpdcejøu"Cnrjcl*

Kemudian yang terakhir, dalam menganalisis SEM pada SEE lebih baik digunakan pendekatan IRT karena menurut Nasir (2014) suatu pendekatan yang menghasilkan nilai SEM yang lebih rendah dibandingkan dengan nilai SEM yang dihasilkan oleh pendekatan lainnya dianggap lebih akurat. Hal ini dikarenakan fungsi SEM sendiri dapat mempengaruhi tingkat keakuratan pada suatu alat ukur (Magno, 2009). Dari hasil yang diberikan dikatakan bahwa pada pendekatan IRT menghasilkan nilai SEM sebesar .13 yang tergolong dalam angka yang kecil, sedangkan pada pendekatan CTT menghasilkan nilai SEM sebesar 5.59. Dimana menurut Hambleton & Swaminathan (1985) batas nilai *error* yang masih bisa ditoleransi yaitu sebesar .40.

Implikasi dari perbandingan CTT dan IRT pada SEE ini ialah untuk mengetahui pendekatan manakah yang akan lebih sesuai jika digunakan untuk menganalisis SEE. Dimana menurut hasil dari pengujian keempat karakteristik psikometrik SEE, bahwa pendekatan IRT lebih sesuai untuk digunakan pada SEE. Hal ini dikarenakan pada pendekatan IRT lebih banyak memberikan informasi mengenai SEE baik informasi pada tes maupun informasi pada masing-masing *item*, sehingga mampu memberikan hasil yang lebih akurat. Serta memberikan hasil yang bersifat *independent*, sehingga apabila dalam penelitian selanjutnya menggunakan alat ukur SEE ini akan memberikan hasil lebih dapat dipercaya karena karakteristik psikometrik yang diuji cenderung tidak berubah apabila diberikan kepada subjek dengan kemampuan berbeda.

D. Keterbatasan Penelitian

Keterbatasan dalam penelitian ini ialah hanya menggunakan sampel pada satu lingkup saja yaitu Fakultas Ilmu Sosial dan Ilmu Politik Universitas Brawijaya. Sehingga keberagaman etnis pada sampel dirasa kurang. Hal ini terbukti bahwa etnis Jawa lebih dominan pada kelompok sampel yang digunakan. Sehingga seharusnya populasi yang digunakan diperluas, tidak hanya fokus pada satu Fakultas saja tetapi bisa pada beberapa Fakultas yang tersebar di Universitas Brawijaya. Agar mendapatkan sampel yang lebih beragam etnisnya.

BAB V

PENUTUP

A. Kesimpulan

Hasil penelitian yang telah dijelaskan sebelumnya, dapat disimpulkan bahwa pendekatan IRT lebih sesuai untuk digunakan pada SEE. Hal ini dikarenakan pada pendekatan IRT lebih banyak memberikan informasi mengenai SEE baik informasi pada tes maupun informasi pada masing-masing *item*, sehingga mampu memberikan hasil yang lebih akurat. Serta memberikan hasil yang bersifat *independent*, sehingga apabila dalam penelitian selanjutnya menggunakan alat ukur SEE ini akan memberikan hasil lebih dapat dipercaya karena karakteristik psikometrik yang diuji cenderung tidak berubah apabila diberikan kepada subjek dengan kemampuan berbeda.

Selain itu, hasil yang diberikan mengatakan bahwa pendekatan IRT lebih sensitif dan lebih mendalam dalam menganalisis *item*. Karena terdapat beberapa *item* yang lolos atau dianggap baik apabila dianalisis dengan menggunakan CTT namun tidak bisa lolos atau dianggap buruk pada pendekatan IRT.

B. Saran

Dikarenakan dalam penelitian ini memiliki beberapa keterbatasan, oleh karena itu terdapat sejumlah saran yang sekiranya dapat dilakukan untuk penelitian selanjutnya, antara lain:

1. Sebaiknya untuk penelitian selanjutnya menggunakan pendekatan IRT dalam menganalisis SEE, karena hal ini didasarkan pada hasil analisis yang telah dilakukan bahwa pendekatan IRT lebih baik daripada pendekatan CTT dalam menganalisis SEE.
2. Sebaiknya untuk penelitian selanjutnya digunakan sampel yang memiliki etnis yang lebih heterogen utamanya pada subjek yang sedang tinggal di lingkungan yang tidak sama dengan etnisnya.

DAFTAR PUSTAKA

- Abedalaziz, N., & Leng, C. H. (2017). The relationship between ctt and irt approaches in analyzing item characteristics. *The Malaysian Online Journal of Educational Science*, 1(1), 52-58.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Azwar, S. (2007). *Metode penelitian*. Yogyakarta: Pustaka Belajar.
- Baker, F. B. (2001). *The basics of item response theory*. United States of America: ERIC Clearinghouse on Assessment and Evaluation.
- Baldwin, P. (2006). *A modified irt model intended to improve parameter estimates under small sample conditions*. San Fransisco: University of Massachusetts.
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *SPINE*, 3186-3191.
- Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, 1-15.
- DeMars, C. (2010). *Item response theory (understanding statistics measurement)*. New York: Oxford University Press.
- Güler, N., Uyanık, G. K., & Teker, G. T. (2014). Comparison of classical test theory and item response theory in terms of item parameters. *European Journal of Research on Education*, 2, 1-6.
- Hafid, A. (2016). Hubungan sosial masyarakat multietnik di kabupaten luwu sulawesi selatan. *Jurnal "Al-Qalam"*, 22, 256-268.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory (principles and application)*. New York: Springer Science+Business Media, LLC .
- Kamila, N. (2014). Hubungan antara nilai tugas (task value) mata kuliah psikodiagnostik dengan orientasi tujuan mahasiswa psikologi fakultas ilmu pendidikan universitas pendidikan indonesia. *Universitas Pendidikan Indonesia*, 1-60.

- Keller, L. A. (2005). Item response theory (irt) models for polytomous response data. *Encyclopedia of Statistics in Behavioral Science*, 990-995.
- Lababa, D. (2008). Analisis butir soal dengan teori tes klasik: sebuah pengantar. *STAIN Manado*, 29-37.
- Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1-11.
- Malcarne, V. L., Chavira, D. A., Fernandez, S., & Liu, P.-J. (2006). The scale of ethnic experience: development and psychometric properties. *Journal of Personality Assessment*, 86, 150-161.
- Nasir, M. (2014). Application of classical test theory and item response theory to analyze multiple choice questions. *UNIVERSITY OF CALGARY*, 1-194.
- Nurchahyo, F. A. (2016). Aplikasi IRT dalam analisis aitem tes kognitif. *Buletin Psikologi*, 24, 64-75.
- Rahmawati, E. (2014). Evaluasi karakteristik psikometri Intelligenz Struktur Test (IST). *Seminar Nasional Psikometri*, 270-282.
- Ridho, A. (2006). Karakteristik psikometrik tes berdasarkan pendekatan teori tes klasik dan teori respon aitem. *Fakultas Psikologi UIN Malang*.
- Ridlo, S. (2016). Pengembangan tes pengetahuan praktikum biologi berdasarkan Graded Response dan Generalized Partial Credit. *Jurnal Penelitian dan Evaluasi Pendidikan*, 166-182.
- Solichin, M. (2017, Juni). Analisis daya beda soal, taraf kesukaran, validitas butir tes, interpretasi hasil tes dan validitas ramalan dalam evaluasi pendidikan. *Jurnal Manajemen & Pendidikan Islam*, 2, 192-213.
- Sugiyono. (2004). *Metode penelitian*. Bandung: Alfabeta.
- Suhariyono., Sriyono, & Ngazizah, N. (2014). Akurasi pendekatan Classical Test Theory dan pendekatan Item Response Theory dalam menganalisis soal UAS Fisika semester genap kelas X SMA Negeri di Purworejo Tahun Pelajaran 2013/2014. *Radiasi*, 5, 75-79.
- Suwarto. (2004, November). Analisis item dalam pembuatan tes. *Jurnal Pendidikan*, 289-300.

- Thomas, K. S., Bardwell, W. A., Malcarne, V., & Dimsdale, J. E. (2004). The effect of demographic variables on ethnic experience. *Journal of Applied Biobehavioral Research*, 65-79.
- Trochim, W. M. (2006). Introduction to validity. *Social Research Methods*.
- Wiberg, M. (2004). *Classical test theory vs. item response theory (An evaluation of the theory test in the Swedish driving-license test)*. Swedish: University UMEA.
- Widhiarso, W. (2010). Model politomi dalam teori respons butir. *Fakultas Psikologi UGM*, 1-14.