

## **ANALISIS KUALITAS AITEM INTELLIGENCE STRUCTURE TEST (IST) MELALUI METODE ITEM RESPONSE THEORY (IRT)**

**Dwi Agustin Nuriani Sirodj**

Fakultas MIPA, Universitas Islam Bandung, Jl. Tamansari No.1 Bandung  
dwi.agustinns@gmail.com

### **Abstrak**

Intelligence Structure Test (IST) merupakan salah satu tes psikologi untuk mengukur kemampuan intelegensi seseorang. Di Universitas "X" penggunaan IST dijadikan sebagai acuan untuk seleksi masuk mahasiswa. Sebagai salah satu prediktor keberhasilan mahasiswa dalam menempuh pendidikan sarjana, penting untuk memastikan apakah kualitas dari tes IST masih baik untuk saat ini. Mengingat bahwa alat tes IST yang ada di Indonesia pada saat ini masih hasil adaptasi dari IST-70 dan belum ada revisi hingga saat ini. Seiring dengan berkembangnya ilmu pengetahuan maka berkembang pula salah satu metode yaitu item respon theory (IRT). IRT merupakan teori tentang bagaimana variabel kemampuan seseorang dan variabel aitem menentukan data respons ketika seseorang menjawab aitem tersebut. Kualitas tes IST apakah baik atau tidak ditentukan oleh ketepatan alat tes tersebut dalam mengukur kecerdasan individu (validitas), konsistensi dari alat tes tersebut (reliabilitas), derajat kesukaran aitem yang bergerak dari mudah ke sulit, daya pembeda aitem yang mampu membedakan individu yang mempunyai kemampuan tinggi dan rendah, serta index guessing agar aitem tidak mudah untuk ditebak jawabannya. Atas dasar tersebut peneliti ingin melakukan penelitian mengenai analisis kualitas aitem IST melalui metode IRT pada proses seleksi penerimaan mahasiswa baru Universitas "X" pada periode 2017- 2018. Dari hasil penelitian didapat bahwa secara keseluruhan IST masih layak digunakan sebagai alat test yang mengukur kecerdasan seseorang dan dipakai sebagai prediktor keberhasilan calon mahasiswa baru dalam menempuh pembelajaran di universitas, hanya saja perlu adanya perbaikan dalam penempatan aitem yang disesuaikan dengan index facility nya. Selain itu terdapat delapan aitem tes yang tidak layak digunakan sehingga perlu diganti.

Kata kunci : Intelligence Structure Test, Item Respon Theory, derajat kesukaran aitem, index guessing

### **Abstract**

*Intelligence Structure Test (IST) is one of the psychological tests to measure a person's intelligence capabilities. At the University "X", IST is used as a reference for student admission. As one of the predictors of student success in taking undergraduate education, it is important to ascertain the quality of the IST test is still good or not. The IST test in Indonesia are still adapted from the IST-70 and there has been no revision until now. Along with the development of science, one method is developing, namely the item response theory (IRT). IRT is a theory of how a person's ability variable and item variable determine response data when someone answers the item. The quality of the IST test is determined by the accuracy of the test tool in measuring individual intelligence (validity), consistency of the test instrument (reliability), facility index of items that move from easy to difficult, discrimination index that can individuals who have the ability high and low, and a guessing index so that the answer is not easy to guess. So, this research wanted to conduct*

*the analysis of IST item quality through the IRT method in the selection process for new "X" University admissions in the period 2017- 2018. From the results of the study, overall IST is still feasible to be used as a test tool that measures one's intelligence and used as a predictor of the success of prospective new students in taking university learning, however there are still need some improvement in the placement of items adjusted to the index facility. In addition, there are eight test items that are not suitable for use, so these items need to be replaced.*

*Keywords: Intelligence Structure Test, Item Respon Theory, Index Facility, Index Guessing*

## **Pendahuluan**

Alat tes psikologi telah cukup lama dikenal oleh masyarakat Indonesia seiring dengan masuknya ilmu Psikologi pada awal tahun 1950. Pemerintah Indonesia melalui Balai Psychototechnick dari Kementerian Pendidikan Pengajaran & Kebudayaan RI yang mengadakan seleksi siswa untuk masuk ke sekolah menengah kejuruan teknik serta pengukuran psikometris untuk keperluan penjurusan sekolah. Perkembangan penggunaan alat tes Psikologi di Indonesia cukup pesat dimana penggunaannya tidak hanya dalam bidang pendidikan, namun juga dalam bidang militer, hukum, industri, organisasi maupun individual.

Dalam tes Psikologi, secara umum dibagi menjadi dua jenis, yaitu tes psikologi yang mengukur kepribadian seseorang seperti emosi, sikap, motivasi, serta minat dan tes psikologi yang mengukur kemampuan seseorang seperti bakat dan kemampuan intelegensi. Berbagai alat tes psikologi dikembangkan dan dikaji secara terus menerus guna memenuhi fungsinya dalam menggambarkan kondisi psikologis individu, mengingat manusia adalah makhluk dinamis yang terus berubah dan berkembang sesuai dengan tuntutan jaman dan budaya yang mempengaruhi kehidupannya.

Penggunaan alat tes psikologi yang populer adalah alat tes psikologi yang berkaitan dengan kegiatan seleksi untuk memilih kandidat yang terbaik untuk dapat diterima menjadi anggota suatu organisasi baik perusahaan, sosial, pemerintahan maupun pendidikan. Alat tes yang digunakan dalam kegiatan seleksi pun ada yang bersifat projektif maupun objektif seperti tes grafis, tes pauli, dan tes kecerdasan. Berkaitan dengan tes intelegensi, ada sebuah alat tes psikologi yang cukup populer digunakan di Indonesia untuk menggambarkan tingkat kecerdasan seseorang. Alat tes ini cukup populer karena dapat diberikan kepada banyak orang dalam waktu yang bersamaan (klasikal) dan mengukur bukan hanya tingkat kecerdasan saja namun juga memetakan aspek kecerdasan yang dimiliki seseorang.

Intelligence Structure Test (IST) merupakan salah satu tes psikologi untuk mengukur tingkat intelegensi seseorang. Menurut Polhaupessy (2002), IST adalah tes intelegensi yang dikembangkan oleh Rudolf Amthauer di Jerman pada tahun 1953 (Adinugroho, 2016). IST yang digunakan di Indonesia adalah IST hasil adaptasi Fakultas Psikologi Universitas Padjajaran Bandung, yaitu IST-70. Tes ini pertama kali digunakan oleh Psikolog Angkatan Darat Bandung, Jawa Barat. IST menjadi salah satu

tes yang paling penting saat ini, karena dijadikan sebagai salah satu prediktor untuk keberhasilan individu dalam mengerjakan suatu tugas. Dalam bidang pendidikan, IST dapat dijadikan sebagai salah satu prediktor keberhasilan mahasiswa untuk dapat menempuh pendidikan sarjana dalam waktu yang telah ditentukan. Banyaknya penggunaan IST sebagai salah satu syarat seleksi dalam penerimaan mahasiswa baru di berbagai perguruan tinggi memicu munculnya buku-buku psikotes yang banyak dipakai oleh para calon mahasiswa tersebut sebagai latihan dalam mengerjakan psikotes itu sendiri sehingga memungkinkan pengukuran intelegensi menjadi tidak akurat.

Di Universitas "X" penggunaan IST dijadikan sebagai acuan untuk seleksi masuk mahasiswa. Pada awalnya IST hanya digunakan untuk seleksi masuk fakultas Psikologi, namun saat ini digunakan untuk seleksi masuk oleh fakultas-fakultas lain seperti Kedokteran, Ilmu Komunikasi, Matematika dan Ilmu Pengetahuan Alam (MIPA), dan Ekonomi. Sebagai salah satu prediktor keberhasilan mahasiswa dalam menempuh pendidikan sarjana, penting untuk memastikan apakah kualitas dari tes IST masih baik untuk saat ini mengingat bahwa alat tes IST yang ada di Indonesia pada saat ini masih hasil adaptasi dari IST-70 dan belum ada revisi.

Penelitian mengenai evaluasi atau peninjauan kembali mengenai kualitas aitem-aitem yang ada pada IST pernah dilakukan oleh Adinugroho (2016), hanya saja dalam penelitian tersebut dilakukan pada satu subtes IST saja yaitu subtes Form Auswahl (FA), metode yang digunakan menggunakan classical test theory (CTT). Seiring dengan berkembangnya ilmu pengetahuan maka berkembang pula salah satu metode yaitu item respon theory (IRT). IRT merupakan teori tentang bagaimana variabel kemampuan seseorang dan variabel aitem menentukan data respons ketika seseorang menjawab aitem tersebut (Huriaty, 2015). Perbandingan metode IRT dan CTT pun banyak di teliti untuk menentukan metode mana yang paling baik dalam menilai kualitas suatu aitem. Kelebihan metode IRT dibandingkan dengan metode CTT yaitu statistik aitem tidak bergantung pada kelompok, skor tes yang diperoleh dapat menggambarkan kemampuan individu, tidak memerlukan tes yang paralel untuk menghitung koefisien reliabilitas, dan dapat menyediakan ukuran yang tepat untuk setiap skor kemampuan (Huriaty, 2015). Sehingga bisa dikatakan bahwa metode IRT lebih baik dalam menilai kualitas aitem dibandingkan dengan metode CTT.

Penelitian ini dimaksudkan untuk menilai kualitas aitem dari Intelligence Structure Test (IST) menggunakan Item Response Theory (IRT). Teori respon aitem (IRT) merupakan salah satu cara untuk menilai kelayakan aitem dengan membandingkan rata-rata penampilan aitem terhadap tampilan bukti kemampuan kelompok yang diramalkan oleh model. Pengembangan IRT bertujuan untuk mengatasi kelemahan teori tes klasik yang tidak independen terhadap kelompok peserta yang mengerjakan tes maupun terhadap tes yang diujikan. IRT menghubungkan antara probabilitas jawaban benar pada suatu aitem dengan kemampuan subjek ukur. IRT memiliki kelebihan diantaranya: (1) tanpa varian pada parameter aitem, (2) tanpa varian pada parameter abilitas, dan (3) adanya ketepatan pada pengukuran lokal.

Penggunaan IRT dalam menganalisis kualitas aitem dalam tes kemampuan maupun tes prestasi masih terbatas. Penelitian-penelitian yang telah dilakukan mengenai IRT diantaranya penelitian yang dilakukan oleh Ridho (2007) melakukan perbandingan antara pendekatan teori tes klasik dan teori respon aitem dalam mengevaluasi suatu alat ukur untuk variabel achievement. Alat ukur terdiri dari 40 soal multiple choice yang diujicobakan pada 7000 sampel siswa SMA. Hasil penelitian memperlihatkan bahwa pendekatan teori respon aitem 2 parameter logistic (IRT 2 PL) yang lebih baik dan cocok untuk mengevaluasi alat tes. Anggreyani (2009) melakukan perbandingan antara teori uji klasik dan teori respon aitem dalam mengevaluasi aitem “Tingkat Persiapan Bersama” di IPB. Dari hasil penelitian diperoleh model yang menggambarkan karakteristik aitem adalah model teori respon butir 3 parameter.

Tes IST yang ada di Indonesia merupakan adaptasi dari IST-70 yang dilakukan oleh Psikologi UNPAD pada 47 tahun yang lalu. Dari segi usia tes IST-70 sudah berlaku cukup lama sehingga perlu evaluasi mengenai kualitas dari aitemnya apakah masih baik atau perlu dilakukan revisi atau bahkan diganti. Atas dasar tersebut peneliti mencoba menganalisis kualitas aitem dari IST pada delapan dari sembilan subtes yang terdapat di tes IST. Kedelapan subtes memiliki karakteristik yang sama dalam hal skoring yaitu dichotomi (satu dan nol). Skoring bernilai satu jika jawabannya adalah benar dan sebaliknya skoring bernilai nol jika jawabannya adalah salah. Pengujian terhadap kualitas butir IST ini juga masih terbatas. Beberapa penelitian yang telah dilakukan tidak menganalisis seluruh subtes dalam IST tetapi hanya satu sub tes saja seperti yang dilakukan oleh Adinugroho (2016) yaitu pengujian properti psikometrik pada alat tes IST akan tetapi hanya pada subtes Form Auswahl (FA) saja. Metode yang digunakan adalah analisis aitem; validasi convergent – discriminant dan uji reliabilitas dengan split-half. Hasil penelitiannya menunjukkan bukti empiris untuk mempertahankan seluruh aitem. Akan tetapi, analisis distraktor menunjukkan hasil yang berbeda. Ada kebutuhan untuk merevisi pilihan jawaban dalam soal karena ketidakmampuan distraktor untuk mengecoh respon individu. Secara umum, uji validitas dan reliabilitas menunjukkan bahwa subtes FA masih sesuai untuk mengukur kemampuan spasial terhadap orang Indonesia. Analisis aitem yang dilakukan oleh Adinugroho (2016) masih menggunakan teori tes klasik dengan dua parameter. Selain itu evaluasi terhadap aitem-aitem IST pernah dilakukan juga oleh Rahmawati (2014), penelitiannya bertujuan untuk mengevaluasi apakah IST masih memiliki kualitas yang baik dengan menganalisis tingkat kesukaran aitem, daya diskriminasi aitem, dan peluang tebakan semu menggunakan pendekatan item response theory dengan metode marginal maksimum likelihood menggunakan Program R. Hasil analisisnya menunjukkan bahwa 53.125% dari aitem-aitem IST yang dianalisis memiliki karakteristik psikometri yang kurang baik menurut batasan psikometri. Penelitian lain pun dilakukan oleh Suryani (2018) mengevaluasi aitem IST menggunakan model Rasch, hasil penelitiannya menunjukkan pada instrumen tes IST dapat diketahui bahwa secara umum kualitas tes IST termasuk dalam kategori baik. Dari 176 item tes IST hanya ada 1 item yang tidak bagus, yaitu aitem 155 (WU19) sehingga aitem 155 harus

dibuang. Berdasarkan analisis DIF dapat dilihat bahwa ada 28 item yang mendukung satu jenis kelamin saja, sehingga dua puluh delapan item harus direvisi.

Kualitas tes IST apakah baik atau tidak ditentukan oleh ketepatan alat tes tersebut dalam mengukur kecerdasan individu (validitas), konsistensi dari alat tes tersebut (reliabilitas), derajat kesukaran aitem yang bergerak dari mudah ke sulit, daya pembeda, serta tidak mudah untuk ditebak jawabannya. Atas dasar tersebut peneliti ingin melakukan penelitian mengenai analisis kualitas aitem IST melalui metode IRT pada proses seleksi penerimaan mahasiswa baru Universitas “X” pada periode 2017-2018.

## Metode

Analisis penelitian dilakukan dengan metode kuantitatif menggunakan alat test IST yang telah diterjemahkan oleh lembaga psikologi Universitas Padjadjaran dan digunakan oleh Universitas “X” dalam proses seleksi calon mahasiswa baru setiap tahun. Selanjutnya metode IRT akan diaplikasikan untuk menilai apakah kualitas setiap aitem / pertanyaan pada IST masih baik dalam mengukur kecerdasan calon mahasiswa baru atau memang perlu dilakukan revisi atau perbaikan. Seluruh proses analisis akan menggunakan *R-Software 3.3.0*. Penentuan subjek penelitian (*sampel*) dilakukan dengan metode *accidental sampling* yaitu siapapun yang menjadi calon peserta mahasiswa baru yang mengikuti psikotes saringan ujian masuk Universitas “X” pada periode waktu 2017 (Penerimaan Mahasiswa Baru Gelombang 3) sampai 2018 (Penerimaan Mahasiswa Baru Gelombang 1) yang berjumlah 1382 orang.

Model IRT menunjukkan hubungan antara *ability* atau *trait* yang disimbolkan dengan  $\theta$  yang diukur melalui sebuah instrumen dengan respon yang diberikan pada setiap aitem (Mars, 2010). Dalam IRT, model matematisnya mempunyai makna bahwa probabilitas subjek untuk menjawab aitem dengan benar tergantung pada kemampuan subjek dan karakteristik aitem. Ini berarti bahwa peserta tes dengan kemampuan tinggi akan mempunyai probabilitas menjawab benar lebih besar jika dibandingkan dengan peserta yang mempunyai kemampuan rendah. Hambleton & Swaminathan (1985) menyatakan bahwa ada tiga asumsi yang mendasari IRT, yaitu : (a) Unidimensi : setiap aitem tes hanya mengukur satu kemampuan; (b) Independensi local : Jika faktor-faktor yang mempengaruhi prestasi konstan, maka respons subjek terhadap aitem yang manapun akan independen secara statistik satu sama lain; dan (c) Invariansi parameter : karakteristik aitem tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak bergantung dari ciri aitem.

Bentuk persamaan model IRT bisa terdiri dari model IRT 1 *Parameter Logistic* (PL), model IRT 2 PL, dan model IRT 3 PL. berikut adalah model persamaannya :

Model IRT 1 PL :

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}} \quad (1)$$

Model IRT 2 PL :

$$P_i(\theta) = \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}} \quad (2)$$

Model IRT 3 PL :

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}} \quad (3)$$

Dengan :  $i = 1, 2, 3, \dots, n$

$P_i(\theta)$  : Peluang peserta tes yang memiliki kemampuan  $\theta$  dapat menjawab aitem- $i$  dengan benar

$\theta$  : tingkat kemampuan subyek

$b_i$  : *index facility* aitem ke- $i$

$a_i$  : *index discrimination* aitem ke- $i$

$c_i$  : *index guessing* aitem ke- $i$

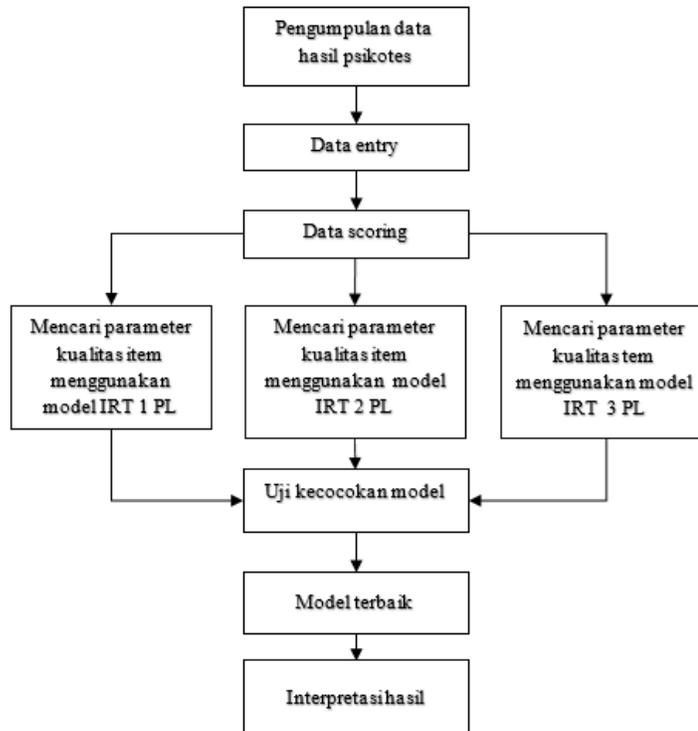
$n$  : banyaknya aitem

*Index discrimination* aitem menunjukkan seberapa baik suatu aitem dapat membedakan antara peserta yang memiliki kemampuan yang tinggi dan peserta yang memiliki kemampuan yang rendah. *Index discrimination* aitem bergerak dari 0 sampai dengan 2 atau  $0 \leq a_i \leq 2$  (Hambleton, 1991; Retnawati, 2014).

*Index facility* aitem merupakan parameter untuk menunjukkan tingkat kesulitan aitem untuk dijawab oleh peserta / responden. Bila nilai  $b_i$  lebih besar, lebih besar pula kemampuan yang diperlukan peserta untuk dapat menjawab aitem dengan benar, yang berarti aitem lebih sulit. Menurut (Hambleton dan Swaminathan, 1985; Retnawati, 2014) nilai  $b_i$  bergerak dari -2 sampai dengan +2 atau  $-2 \leq b_i \leq 2$ .

*Index guessing* menggambarkan peluang peserta yang memiliki kemampuan rendah dapat menjawab aitem sulit dengan benar. Secara umum disebut juga dengan parameter tebakan karena diasumsikan peserta dapat menjawab aitem dengan benar dengan cara menebak. Secara teoritis, nilai  $c$  bergerak dari 0 sampai dengan 1 atau  $0 \leq c_i \leq 1$ . Tetapi menurut (Baker, 2001; Retnawati, 2014) nilai  $c$  di atas 0,35 tidak dapat diterima.

Adapun tahapan analisis data akan mengikuti alur berikut :



Gambar 1. Tahapan Penelitian

### Hasil Dan Pembahasan

Sebelum membahas mengenai parameter-parameter dalam model IRT, maka yang dilakukan adalah memilih model mana yang cocok untuk setiap subtes. Pemilihan model didasarkan pada nilai *Akaike Information Criterion* (AIC) terkecil di antara tiga model yang diuji cobakan. Berikut adalah nilai AIC untuk keseluruhan model IRT :

Tabel 1. Rekapitulasi Nilai AIC Model IRT

SUBTES	AIC			MODEL IRT YANG TERPILIH
	IRT 1 PL	IRT 2PL	IRT 3PL	
IST				
SE	29139.33	28934.01	<b>28641.99</b>	3 PL
WA	30630.34	30390.45	<b>30287.92</b>	3 PL
AN	27315.68	26966.53	<b>26727.42</b>	3 PL
RA	17086.88	16534.9	<b>16465.08</b>	3 PL
ZR	23445.08	23122.15	<b>23088.65</b>	3 PL
FA	30179.56	30107.34	<b>30001.19</b>	3 PL

<b>WU</b>	27009.59	26899.97	<b>26254.28</b>	3 PL
<b>ME</b>	28530.77	27908.44	<b>27847.46</b>	3 PL

Berdasarkan pada tabel 1 terlihat bahwa model IRT 3 PL untuk keseluruhan subtes mempunyai nilai AIC terkecil, sehingga dapat disimpulkan bahwa untuk menilai kualitas keseluruhan aitem IST berdasarkan parameter-parameternya akan mengacu pada parameter pada model IRT 3 PL. Parameter-parameter yang ditaksir pada model IRT 3 PL antara lain : *index guessing* ( $c_i$ ), *index facility* ( $b_i$ ), dan *index discrimination* ( $a_i$ ). Adapun rekapitulasi perhitungan setiap parameter pada setiap aitem di seluruh subtes IST yang telah disesuaikan dengan acuan / kriteria yang ditetapkan adalah sebagai berikut :

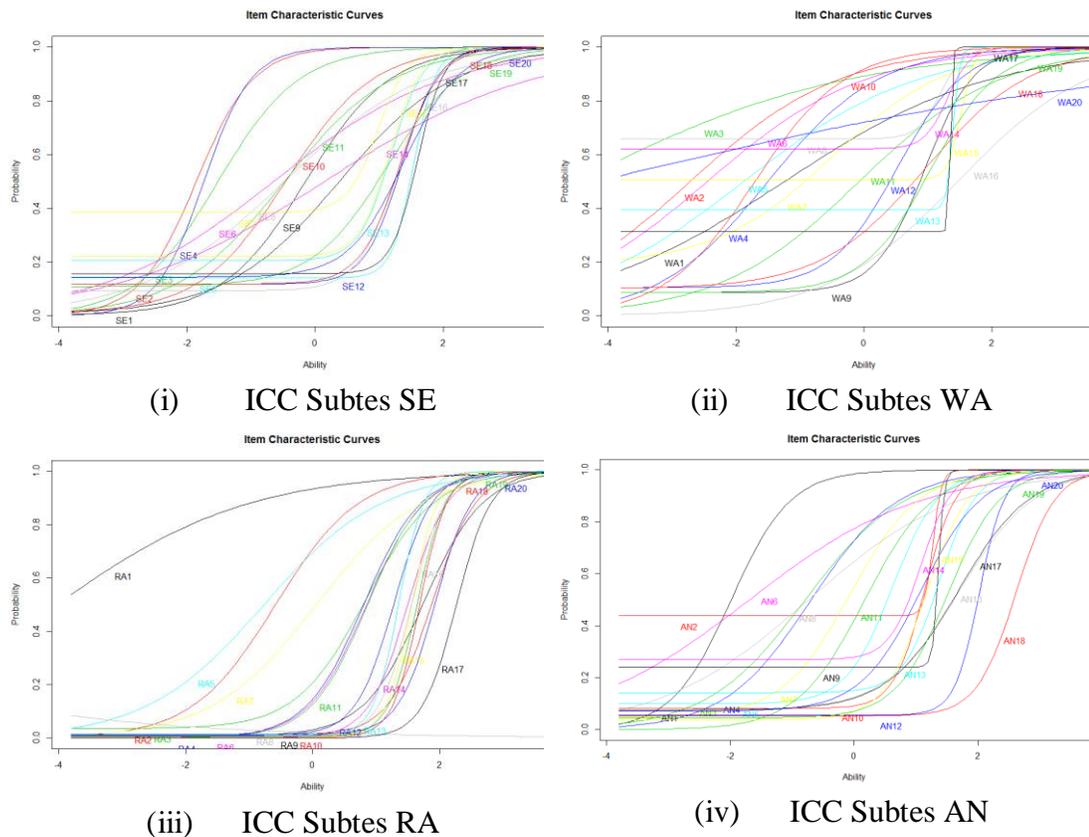
**Tabel 2. Rekapitulasi Kualitas Item Berdasarkan Parameter- Parameter Model IRT 3PL**

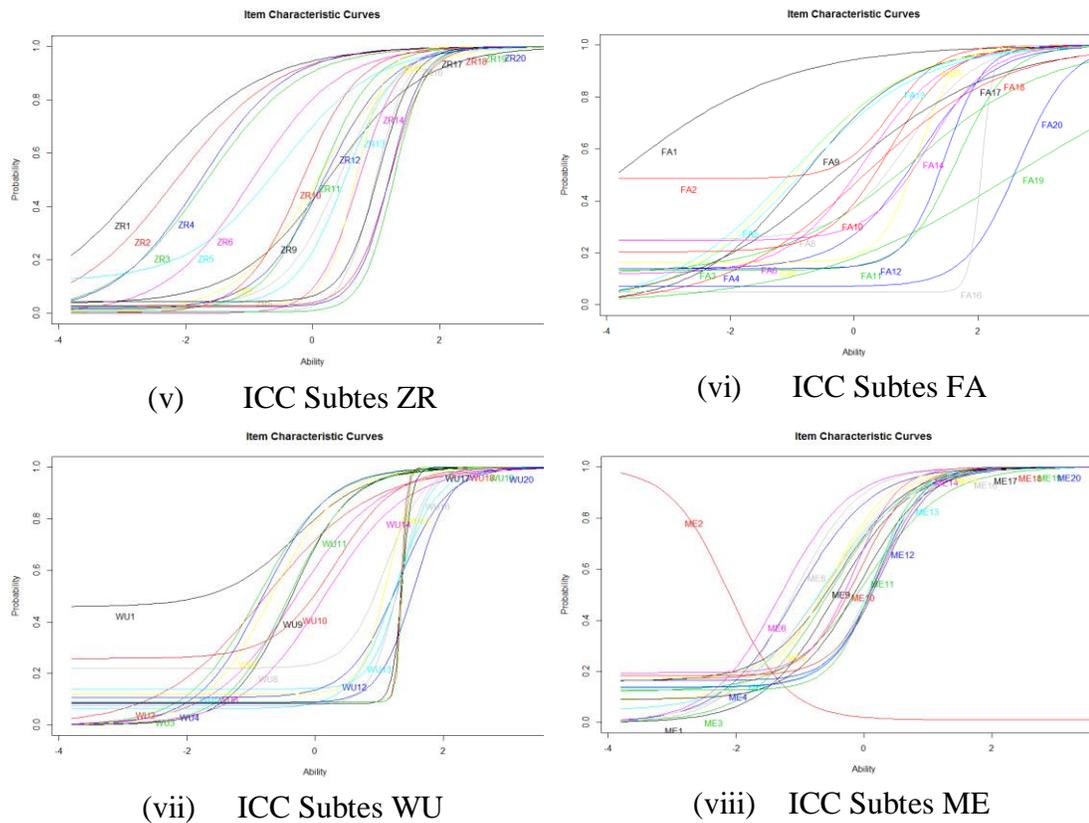
SUBTES IST	REKAPITULASI JUMLAH ITEM					
	<i>index guessing</i>		<i>index facility</i>		<i>index discrimination</i>	
	Baik	Perlu Perbaikan	Baik	Perlu Perbaikan	Baik	Perlu Perbaikan
<b>SE</b>	95%	5%	<b>100%</b>	0%	45%	55%
<b>WA</b>	80%	<b>20%</b>	80%	<b>20%</b>	<b>70%</b>	30%
<b>AN</b>	95%	5%	90%	10%	40%	60%
<b>RA</b>	<b>100%</b>	0%	85%	15%	25%	<b>75%</b>
<b>ZR</b>	<b>100%</b>	0%	90%	10%	45%	55%
<b>FA</b>	95%	5%	80%	<b>20%</b>	60%	40%
<b>WU</b>	95%	5%	<b>100%</b>	0%	35%	65%
<b>ME</b>	<b>100%</b>	0%	95%	5%	55%	45%

Berdasarkan pada tabel 2 terlihat bahwa untuk parameter *index guessing*, aitem-aitem di subtes WA memiliki persentase terbesar yaitu 20% aitem perlu perbaikan karena *index guessing* berada di luar batas kriteria yang ditetapkan yaitu  $0 \leq c_i \leq 0.35$ , berbeda dengan subtes RA, ZR, dan ME yang keseluruhan aitemnya baik jika dilihat dari *index guessing*nya. Hal ini menandakan bahwa jawaban keseluruhan aitem pada subtes tersebut tersebut tidak mudah ditebak oleh peserta yang memiliki

kemampuan rendah sekalipun. Jika dilihat dari parameter *index facility*, aitem-aitem di subtes WA dan FA memiliki persentase terbesar yaitu 20% aitem perlu perbaikan karena *index facility* berada di luar batas kriteria yang ditetapkan yaitu  $-2 \leq b_i \leq 2$ , berbeda dengan subtes SE, dan WU yang keseluruhan aitem nya baik jika dilihat dari *index facility* nya. Hal ini menandakan bahwa tingkat kesulitan dalam menjawab keseluruhan aitem pada subtes tersebut sudah cukup baik, tidak terlalu sukar dan tidak terlalu mudah. Jika dilihat dari parameter *index discrimination*, aitem-aitem di subtes RA memiliki persentase terbesar yaitu 75% aitem perlu perbaikan karena *index discrimination* berada di luar batas kriteria yang ditetapkan yaitu  $0 \leq a_i \leq 2$  begitupun dengan subtes lain nya yang mempunyai prosentase relative besar (di atas 50%), berbeda dengan subtes WA yang 70 % aitem nya baik. Sehingga secara keseluruhan perlu ada nya perbaikan aitem di sebagian besar subtes agar aitem dapat membedakan antara peserta yang memiliki kemampuan yang tinggi dan peserta yang memiliki kemampuan yang rendah.

Secara grafik, ketiga parameter model IRT 3PL ini dapat digambarkan melalui *Item Characteristic Curve (ICC)*. Bentuk ICC untuk keseluruhan aitem di seluruh subtes dapat dilihat pada gambar berikut :





Gambar 2. Item Characteristic Curve (ICC) keseluruhan Subtes IST

Berdasarkan gambar 2 dapat disimpulkan hal-hal seperti berikut : (1) Aitem ME 2 menghasilkan pola grafik yang berbeda dengan aitem yang lain, hal ini mengindikasikan bahwa aitem tersebut tidak mampu mengukur apa yang akan diukur dalam subtes tes tersebut; (2) Aitem WA 13, WA 14, dan WA 15 menghasilkan pola grafik yang memiliki tingkat *guessing* yang tinggi sehingga pada aitem tersebut mudah ditebak yang mengakibatkan aitem tersebut tidak bisa membedakan kemampuan dari subjek yang diukur; (3) Aitem RA 1 menghasilkan pola grafik yang lebih tinggi dibanding dengan aitem yang lain, hal ini diakibatkan karena *index facility* yang terlalu rendah atau aitem tersebut terlalu mudah dijawab oleh subjek; (4) Aitem AN 2 menghasilkan pola grafik yang memiliki tingkat *guessing* yang tinggi sehingga pada aitem tersebut mudah ditebak yang mengakibatkan aitem tersebut tidak bisa membedakan kemampuan dari subjek yang diukur; (5) Aitem FA 1 menghasilkan pola grafik yang lebih tinggi dibanding dengan aitem yang lain, hal ini diakibatkan karena *index facility* yang terlalu rendah atau aitem tersebut terlalu mudah dijawab oleh subjek; (6) Aitem FA 2 menghasilkan pola grafik yang memiliki tingkat *guessing* yang tinggi sehingga pada aitem tersebut mudah di tebak yang mengakibatkan aitem tersebut tidak bisa membedakan kemampuan dari subjek yang diukur.

## Simpulan

Berdasarkan hasil analisis dapat disimpulkan bahwa secara keseluruhan IST masih layak digunakan sebagai alat test yang mengukur kecerdasan seseorang dan dipakai sebagai prediktor keberhasilan calon mahasiswa baru dalam menempuh pembelajaran di universitas, hanya saja perlu adanya perbaikan dalam penempatan aitem yang disesuaikan dengan *index facility* nya. Kemudian terdapat delapan aitem tes yaitu ME 2, WA 13, WA 14, WA 15, RA 1, AN 2, FA 1, dan FA 2 yang tidak layak digunakan sehingga perlu diganti.

## Daftar Pustaka

- Adinugroho, I. (2016). Pengujian Properti Psikometrik Intelligenz Struktur Test Subtes Kemampuan Spasial Dua Dimensi (Form Auswahl): Studi Pada Dua Sma Swasta Di Jakarta . Jurnal Ilmiah Psikologi MANASA, Vol. 5, No. 2, 165-180.
- Anggreyani, A. (2009). Penerapan Teori Uji Klasik Dan Teori Respon Butir Dalam Mengevaluasi Butir Soal (Studi Kasus : Soal Ujian Akhir Semester Tingkat Persiapan Bersama Institut Pertanian Bogor Mata Kuliah Fisika Tahun Ajaran 2008/2009). Bogor: Skripsi Departemen Statistika Fakultas Matematika Dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- Hambleton, R. K., & Swaminathan, H. (1985). Item Response Theory : Principles and Applications. New York: Kluwer Nijhoff Publishing.
- Huriaty, D. (2015). Metode Kalibrasi Dan Desain Tes Berdasarkan Teori Respon Butir (IRT). Math Didactic : Jurnal Pendidikan Matematika Vol 1, No 3.
- Mars, C. D. (2010). Item Response Theory : Understanding Statistics Measurement. New York: Oxford University Press, Inc.
- Rahmawati, E. (2014). Evaluasi Karakteristik Psikometri Intelligenz Struktur Test (IST) . Seminar Nasional Psikometri (hal. 270-282). SURakarta: Fakultas Psikologi Universitas Muhammadiyah Surakarta.
- Retnawati, H. (2014). Teori Respon Butir dan Penerapannya. Yogyakarta: Nuha Medika.
- Ridho, A. (2007). Karakteristik Psikometrik Tes Berdasarkan Pendekatan Teori Tes Klasik Dan Teori Respon Aitem. INSAN Media Psikologi.
- Suryani, Y. E. (2018). Aplikasi Rasch Model dalam Mengevaluasi Intelligenz Structure Test (IST). Psikohumaniora: Jurnal Penelitian Psikologi, 73-99.