

BERKENALAN DENGAN TEORI RESPONS AITEM

Saifuddin Azwar
Universitas Gadjah Mada

Latar Belakang

Sejak beberapa dekade yang lalu, teori tes klasik (*classical test theory*) telah mendominasi dan banyak berjasa dalam dunia pengukuran. Di antara konsep-konsep yang berdasarkan teori tes klasik yang sangat terkenal dan sangat berguna adalah formula reliabilitas Spearman-Brown, formula-formula Kuder-Richardson, formula eror standar dalam pengukuran (*standard error of measurement*), formula koreksi terhadap atenuasi, dll. Bahkan hampir keseluruhan formula reliabilitas dan validitas yang kita kenal sekarang ini dikembangkan atas konsep teori tes klasik.

Sayangnya, ternyata bahwa dalam teori tes klasik terdapat keterbatasan-keterbatasan yang kemudian menjadi permasalahan dalam aplikasi serta pengembangan tes dan instrumen pengukuran.

Sebagai contoh, parameter-parameter aitem dalam teori klasik, yaitu indeks kesukaran dan indeks diskriminasi, ternyata merupakan karakteristik aitem yang tergantung pada kelompok sampel yang digunakan untuk menghitungnya (*group-dependent*). Angka-angka statistik tersebut besarnya tergantung pada kelompok mana yang digunakan atau yang dikenai tes. Bila kelompok yang dikenai tes adalah kelompok yang kemampuannya tinggi, maka aitem dalam tes akan tampak mudah dan memiliki indeks kesukaran besar (persentase subjek yang dapat menjawab aitem besar). Bila kelompok yang dikenai tes yang sama adalah kelompok yang berkemampuan rendah, maka aitem dalam tes itu akan tampak sulit dan memiliki indeks kesukaran kecil (persentase subjek yang dapat menjawab aitem kecil). Begitu juga, sebagaimana telah diketahui, koefisien reliabilitas dan validitas tes akan menjadi lebih tinggi apabila kelompok subjek merupakan kelompok yang kemampuannya heterogen (bervariasi besar) dan tes yang sama akan mempunyai koefisien yang lebih rendah apabila dikenakan pada kelompok yang kemampuannya relatif lebih homogen. Dependensi terhadap kelompok subjek ini tentu akan mengurangi manfaat parameter aitem dalam berbagai aplikasinya.

Kelemahan lain, dalam teori tes klasik diperlukan pula asumsi kesetaraan eror pengukuran bagi semua subjek yang dikenai tes. Keberatan terhadap asumsi ini adalah kurangnya dukungan yang dapat memperkuatnya dikarenakan pada tes yang sulit, eror pengukuran bagi subjek yang berkemampuan rendah akan berbeda dari eror bagi subjek yang berkemampuan sedang dan tinggi. Di samping itu, definisi tes paralel sebagaimana dimaksudkan oleh teori klasik sangatlah sulit untuk dipenuhi dalam praktek.

Keterbatasan dan kelemahan dalam teori klasik menghambat usaha-usaha aplikasi teori ini dalam pengembangan bank aitem, dalam penyetaraan aitem, dalam komparasi subjek dan lain-lain.

Apa yang diinginkan dalam dunia pengukuran modern adalah statistik aitem yang tidak tergantung pada kelompok subjeknya, skor tes yang dapat menggambarkan profisiensi subjek dan tidak tergantung pada taraf kesulitan tes, model tes yang dapat memberikan dasar pencocokan antara aitem tes dan level kemampuan, model tes yang asumsi-asumsinya punya dukungan kuat, dan model tes yang tidak memerlukan asumsi paralel dalam pengujian reliabilitasnya.

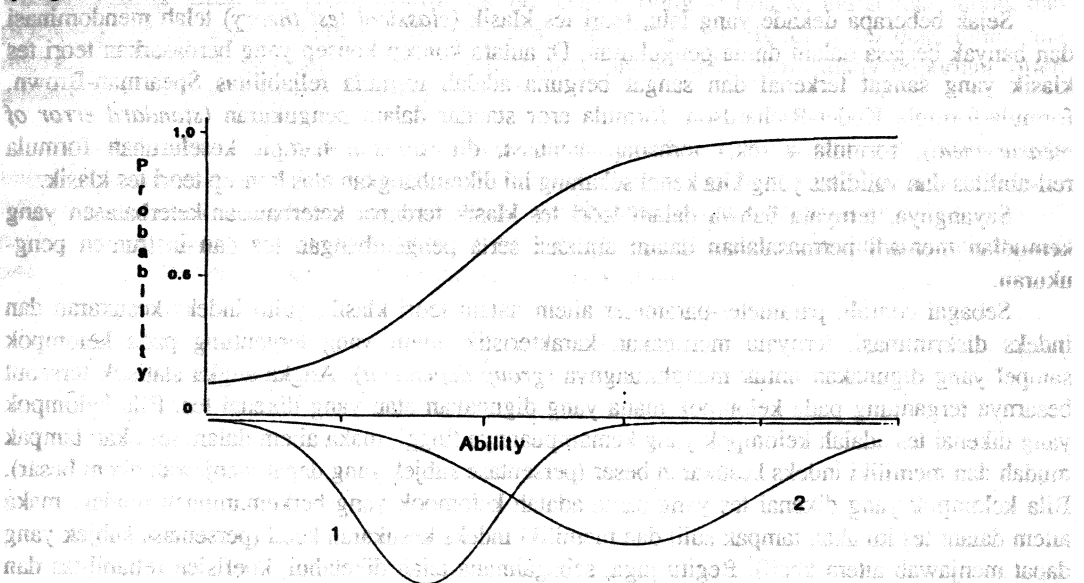
Bukti-bukti menunjukkan bahwa keinginan tersebut dapat dipenuhi oleh suatu teori pengukuran yang disebut sebagai teori respons aitem (*item response theory - IRT*).

Konsep Dasar

Teori respons aitem (selanjutnya disebut IRT) didasari oleh dua postulat.

Pertama, performansi seorang subjek pada suatu aitem dapat diprediksikan oleh seperangkat faktor yang disebut *traits*, *latent traits* atau kemampuan.

Kedua, hubungan antara performansi subjek pada suatu aitem dan perangkat kemampuan (abilitas) yang mendasarinya dapat digambarkan oleh suatu fungsi yang menaik secara monotonis yang disebut *item characteristic function* atau *item characteristic curve* (ICC).



Gambar 1. Kurva Karakteristik Aitem dan Distribusi Kemampuan bagi Dua Kelompok Subjek

Dalam Gambar 1 diperlihatkan satu bentuk ICC dengan hanya satu *trait* yang mendasari performansi dua kelompok yang memiliki kemampuan berbeda. Perhatikan bahwa subjek yang memiliki nilai tinggi pada *trait* memiliki pula probabilitas yang lebih tinggi untuk menjawab aitem dengan benar dibandingkan subjek yang rendah nilai *trait*nya, tanpa peduli dari kelompok mana ia berasal.

Model-model ICC tergantung pada bentuk matematika fungsi karakteristik aitemnya dan pada banyaknya parameter yang dilibatkan dalam model yang bersangkutan. Suatu model repons aitem mungkin cocok pada perangkat data tes tertentu dan dapat tidak cocok pada perangkat data tes yang lain. Oleh karena itu, langkah pertama dalam analisis IRT adalah menentukan kecocokan antara model dengan perangkat data tes yang hendak dianalisis.

Bila model yang cocok untuk perangkat data tes yang hendak dianalisis telah ditemukan, maka akan diperoleh keunggulan-keunggulan tertentu. Keunggulan itu adalah bahwa estimasi kemampuan subjek tidak lagi tergantung pada tes dan estimasi indeks aitem tidak lagi tergantung pada kelompok. Dengan kata lain, estimasi kemampuan yang diperoleh dari perangkat aitem yang berbeda akan selalu sama (di luar error pengukuran) dan parameter aitem yang diperoleh dari kelompok subjek yang berbeda juga akan selalu sama (di luar error pengukuran). Inilah yang disebut dengan sifat parameter aitem dan parameter kemampuan yang *invariant*.

Asumsi-asumsi

Model matematis dalam IRT mengatakan bahwa probabilitas subjek untuk menjawab suatu aitem dengan benar tergantung pada kemampuan subjek dan karakteristik aitem. Untuk itu dalam IRT ada asumsi-asumsi pendukung yang secara tidak langsung dapat diukur dan dibuktikan adanya.

Suatu asumsi yang paling umum adalah bahwa dalam satu tes, kemampuan yang diukur oleh perangkat aitem-aitemnya hanya ada satu. Asumsi ini disebut asumsi *unidimensionality*. Asumsi ini tentu tidak dapat dipenuhi secara ketat dikarenakan adanya faktor-faktor kognitif, kepribadian, dan administrasi tes seperti kecemasan, motivasi, tendensi untuk menebak, dsb. Yang paling penting dalam asumsi ini adalah adanya satu komponen yang "dominan" yang mempengaruhi performansi subjek. Faktor dominan inilah merupakan kemampuan yang diukur oleh tes.

Di samping asumsi unidimensionalitas, dikenal pula asumsi *local independence* yang maksudnya adalah apabila kemampuan-kemampuan yang mempengaruhi performansi dijadikan konstan maka repons subjek terhadap pasangan aitem manapun juga akan independen secara statistik satu sama lain. Jadi, di luar kemampuan subjek, tidak ada hubungan apa-apa antara repons subjek pada suatu aitem dengan reponsnya pada aitem lain. Hal ini sama dengan mengatakan bahwa kemampuan-kemampuan yang dispesifikasikan dalam model merupakan faktor satu-satunya yang mempengaruhi repons subjek. Kemampuan-kemampuan ini merupakan *complete latent space*.

Bila asumsi unidimensionalitas terpenuhi maka *complete latent space* hanya berisi satu kemampuan saja. Akan tetapi independensi lokal dapat diperoleh sekalipun perangkat data tidak bersifat unidimensional. Independensi lokal akan diperoleh bilamana *complete latent space* telah ditetapkan.

Beberapa Model yang Populer

Di antara beberapa model IRT yang populer, perbedaannya terdapat pada banyaknya parameter yang menggambarkan karakteristik aitem yang bersangkutan.

Parameter-parameter aitem adalah:

b_i = indeks kesukaran aitem,

a_i = indeks diskriminasi aitem, dan

c_i = parameter probabilitas tebakan semu.

Bila estimasi terhadap parameter aitem telah diperoleh maka hubungan antara kemampuan dan probabilitas untuk menjawab dengan benar dapat digambarkan dalam diagram ICC.

Model mana yang akan dipilih terserah pada pihak pemakai akan tetapi tergantung juga pada asumsi yang cocok bagi perangkat data yang akan dianalisis. Kecocokan ini dapat dibuktikan kemudian dengan menunjukkan seberapa baiknya model yang bersangkutan dapat menjelaskan hasil tes yang diperoleh.

Berikut akan diperkenalkan selintas tiga macam model IRT unidimensional, yaitu model logistik satu, dua, dan tiga parameter. Model-model ini digunakan pada data jawaban aitem tes yang dikotomi.

Model Logistik Satu-parameter

Model ini merupakan model yang sangat populer. Disebut sebagai model satu-parameter dikarenakan dalam model ini karakteristik aitem hanya ditunjukkan oleh statistik b_i yang merupakan parameter tingkat kesukaran aitem.

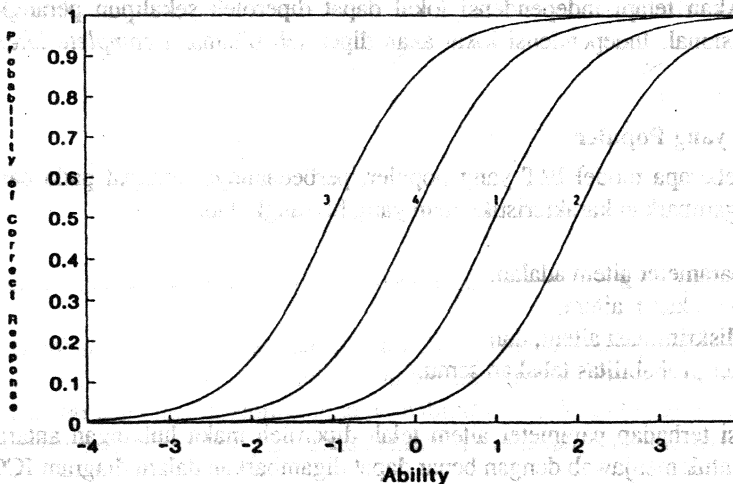
Persamaan matematisnya adalah sebagai berikut:

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

- $P_i(\theta)$ = probabilitas seorang subjek yang memiliki abilitas untuk menjawab aitem i dengan benar
 (θ) = tingkat kemampuan (abilitas)
 b_i = parameter tingkat kesukaran aitem i
 e = angka transendental yang bernilai 2.718

Parameter b_i merupakan suatu titik pada skala kemampuan dimana probabilitas untuk menjawab benar sebesar 0.50. Semakin besar nilai parameter b_i , maka akan semakin besar pula kemampuan yang dituntut dari seorang subjek untuk memiliki 50% kemungkinan menjawab dengan benar. Artinya aitem tersebut semakin sukar. Kurva aitem-aitem yang sukar selalu terletak di ujung sebelah kanan dan kurva aitem-aitem yang mudah berada di ujung sebelah kiri pada skala abilitas.

Beberapa contoh ICC untuk model satu-parameter disajikan pada Gambar 2.



Gambar 2. ICC Satu-parameter untuk 4 Aitem

Untuk aitem 1 $b_1 = 1.0$, untuk aitem 2 $b_2 = 2.0$, untuk aitem 3 $b_3 = -1.0$, dan untuk aitem 4 $b_4 = 0.0$.

Perhatikan bahwa perbedaan keempat kurva tersebut hanya dalam hal letaknya pada skala abilitas. Dalam model satu-parameter diasumsikan bahwa tingkat kesukaran aitem merupakan satu-satunya karakteristik yang mempengaruhi performansi subjek. Hal ini sama saja dengan mengatakan bahwa semua aitem memiliki daya beda yang sama. Perhatikan pula bahwa asimtot-bawah pada ICC adalah nol yang berarti bahwa subjek yang kemampuannya sangat rendah akan memiliki probabilitas nihil untuk dapat menjawab aitem dengan benar.

Model logistik satu-parameter sering juga disebut model Rasch.

Model Logistik Dua-parameter

Model dua-parameter pertama kali dikembangkan oleh Lord berdasar atas distribusi normal kumulatif (*normal ogive*). Kemudian Birnaum (1968) mengganti fungsi ogiv normal untuk model dua-parameter ini dengan fungsi logistik yang lebih mudah dianalisis.

Bentuk matematis persamaan fungsi logistik untuk dua-parameter adalah sebagai berikut :

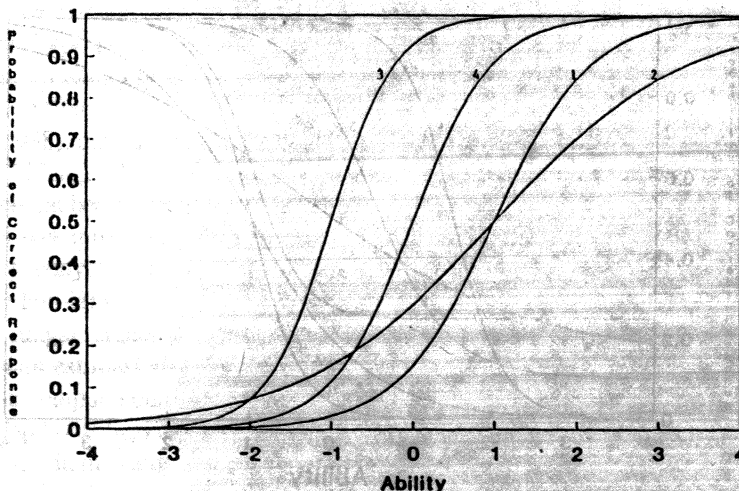
$$P_i(\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \quad i = 1, 2, \dots, n$$

D adalah faktor penskalaan yang diikutkan untuk menjadikan fungsi logistik semirip mungkin dengan fungsi ogiv normal. Apabila $D = 1.7$ ternyata bahwa nilai $P_i(\theta)$ bagi kedua fungsi ogiv normal dan fungsi logistik perbedaannya kurang dari 0.01 untuk semua nilai θ .

Parameter a_i adalah parameter daya diskriminasi aitem. Parameter ini proporsional terhadap slop ICC di titik b_i pada skala abilitas. Aitem-aitem yang memiliki slop yang curam akan memiliki daya diskriminasi lebih tinggi dan lebih mampu memisahkan subjek menurut tingkat kemampuan mereka daripada aitem-aitem yang memiliki slop landai.

Secara teoretis, parameter diskriminasi ditetapkan pada skala $(-\infty, +\infty)$. Dalam prakteknya, parameter negatif menghendaki agar aitem yang bersangkutan dibuang sedangkan parameter yang lebih besar daripada 2 sangat jarang terjadi, sehingga biasanya yang dilihat hanyalah parameter a_i yang besarnya antara 0 sampai dengan 2.

Beberapa contoh model ICC untuk dua-parameter disajikan pada Gambar 3.



Gambar 3. ICC Dua-parameter untuk 4 Aitem

Dalam Gambar 3 untuk aitem 1 $b_1 = 1.0$ dan $a_1 = 1.0$, untuk aitem 2 $b_2 = 1.0$ dan $a_2 = 0.5$, untuk aitem 3 $b_3 = -1.0$ dan $a_3 = 1.5$, untuk aitem 4 $b_4 = 0.0$ dan $a_4 = 1.2$. Keempat ICCnya tidaklah paralel sebagaimana dalam model satu-parameter. Masing-masing memiliki slop yang berbeda

menandakan bahwa daya diskriminasi masing-masing tidak sama.

Juga harus diperhatikan bahwa asimtot-bawah pada semua ICC adalah nol dikarenakan pada model dua-parameter pun diasumsikan bahwa mereka yang sangat rendah kemampuan atau abilitasnya memiliki probabilitas nihil untuk dapat menjawab aitem dengan benar.

Model Logistik Tiga-parameter

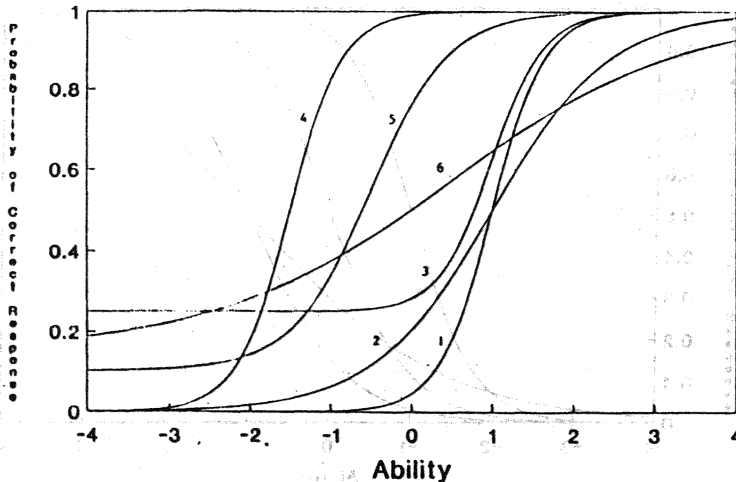
Pada model tiga-parameter dimasukkan satu parameter karakteristik aitem lagi yaitu parameter probabilitas menjawab benar dengan menebak yang lebih dikenal dengan nama parameter *pseudo-chance level* dan disimbolkan oleh huruf c_i . Jadi, dalam model ini, asimtot-bawah ICC dapat saja tidak sama dengan nol yang berarti terdapat asumsi bahwa subjek dengan kemampuan sangat rendah masih mungkin menjawab aitem dengan benar. Hal ini berlaku terutama pada tes dengan format aitem pilihan ganda atau aitem yang menuntut rekognisi jawaban.

Biasanya, harga c_i diasumsikan akan lebih kecil daripada harga yang akan diperoleh bila subjek menjawab aitem dengan tebakan secara random.

Persamaan matematis fungsi logistik dengan tiga para-meter adalah sebagai berikut:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{D_{ai}(\theta - b_i)}}{1 + e^{D_{ai}(\theta - b_i)}} \quad i = 1, 2, \dots, n$$

Contoh model ICC logistik tiga-parameter disajikan pada Gambar 4 sedangkan parameter aitem masing-masing disajikan pada Tabel 1.



Gambar 4. ICC Tiga-parameter untuk 6 Aitem

Tabel 1.
Parameter untuk Aitem dalam Gambar 4

Nomor item	b_i	parameter aitem	
		a_i	c_i
1	1.00	1.80	0.00
2	1.00	0.80	0.00
3	1.00	1.80	0.25
4	-1.50	1.80	0.00
5	-0.50	1.20	0.10
6	0.50	1.40	0.15

Dari Gambar 4 dan Tabel 1 tampaklah bahwa bentuk dan letak ICC ditentukan tidak saja oleh parameter b_i dan a_i akan tetapi ditentukan pula oleh harga c_i .

Beberapa Bidang Aplikasi

Sebagaimana telah dikemukakan, dasar yang paling pokok dalam teori repons aitem adalah sifat *invariant* yang dimiliki oleh parameter aitem dan parameter kemampuan. Sifat *invariant* ini hanyalah akan diperoleh apabila terdapat kecocokan antara model yang digunakan dengan perangkat data tes yang akan dianalisis. Oleh karena itu terdapat prosedur-prosedur pengujian, apakah model yang hendak digunakan ternyata memang cocok dengan data, seperti prosedur analisis *goodness-of-fit*.

Dalam prosedur aplikasi IRT dilakukan langkah estimasi parameter aitem dan parameter abilitas. Estimasi ini dilakukan dengan perhitungan-perhitungan tertentu, misalnya dengan menggunakan analisis *maximum likelihood* atau menggunakan metoda Bayesian. Untuk itu telah tersedia program-program komputer yang memudahkan dan memungkinkan pekerjaan dengan perangkat data yang besar. Di antara program termaksud antara lain adalah BICAL (1979), RASCAL (1988), MICROSCALE (1986), ANCILLES (1978), ASCAL (1988), LOGIST (1982), BILOG (1984), MULTLOG (1988), dan lain-lain.

Berbagai bidang dalam konstruksi dan pengembangan tes kemampuan memperoleh manfaat aplikasi yang sangat besar dari teori repons aitem. Di antara bidang aplikasi tersebut adalah analisis fungsi efisiensi dan informasi aitem dan tes, identifikasi aitem yang potensial untuk bias (*potentially bias item*), penyetaraan (*equating*) skor tes, pengembangan bank aitem, *computerized adaptive testing*, dan lain-lain.

Analisis fungsi efisiensi dan informasi akan memperlihatkan seberapa besar kontribusi aitem-aitem dalam estimasi abilitas. Kontribusi ini sangat tergantung pada daya diskriminasi aitem sedangkan letak tempat kontribusi itu dirasakan akan tergantung pada tingkat kesukaran aitem. Estimasi terhadap kontribusi aitem akan menghasilkan aitem-aitem yang berfungsi secara efisien.

Identifikasi aitem yang berpotensi untuk bias lebih dikenal dengan nama analisis DIF (*differential item functioning*). DIF didefinisikan sebagai aitem yang apabila dijawab oleh individu-individu yang memiliki kemampuan setara tapi berasal dari kelompok yang berbeda akan menyebabkan perbedaan probabilitas untuk menjawab dengan benar. Analisis DIF akan memberikan jalan untuk melakukan evaluasi dan revisi aitem yang pada gilirannya akan membawa pada *fairness* dalam pengambilan keputusan menyangkut hasil tes yang bersangkutan. Berbagai pendekatan DIF telah ditegakkan, antara lain adalah metoda area dan metoda Mantel-Haenszel.

Prosedur penyetaraan skor tes tidak saja penting dalam pengembangan bank aitem akan tetapi memungkinkan pula dilakukannya komparasi hasil dari tes yang berbeda.

Demikianlah beberapa contoh bidang aplikasi yang menggunakan teori repons aitem.

Penutup		
<p>Uraian selintas ini dapatlah memperkenalkan pembaca umumnya dan peminat psikometri khususnya akan teori repons aitem yang dewasa ini sedang mendapat perhatian sangat besar dari para psikometrisian di dunia. Ratusan penelitian telah dilakukan dan pengembangan model-model serta program komputer yang mendukungnya terus diadakan di luar negeri.</p> <p>Langkanya para ahli psikometri di Indonesia menyebabkan kita tidak dapat segera mengikuti perkembangan tersebut. Apalagi dukungan <i>software</i> untuk analisis prosedur IRT sangatlah mahal bagi ukuran keuangan kita. Akan tetapi hal itu tidak menjadi alasan bagi kita untuk tidak segera belajar mengenal teori modern ini. Tulisan ini adalah salah-satu usaha ke arah tersebut.</p>		

Referensi

Hambleton, R.K. (1989) *Principles and Selected Applications of Item Response Theory*. In R.L. Linn (Ed), *Educational Measurement* (3rd ed.) (pp.147-200). New York: Macmillan.

Hambleton, R.K., Swaminathan, H. & Rogers, H.J. (1991) *Fundamentals of Item-Response Theory*. Newbury Park: Sage.

Yogyakarta, 3 Juli 1993