

ABILITAS KOMPOSIT DALAM TES POTENSI

Saifuddin Azwar & Ali Ridho

2012

ABSTRACT

When a unidimensionality assumption has been actually violated, interpretation of test scores might be jeopardized. It couldn't be overemphasized in case of high-stake exams such as PAPs UGM (academic potentiality test for UGM graduate student candidates) which was supposed to reflect a composite ability.

This study aimed at revealing item characteristics of PAPs A-1 Series based on UIRT and MIRT and discovering dimensionality of the three subtests of the test.

Scores of subject ($n=2035$) on the 3 subtests were analysed and the results showed that 27 items (10 of Verbal, 8 of Kuantitatif, and 9 of Penalaran) were flagged for having r_{bis} of less than 0.25 and 6 other items for having abnormally high pseudo-guessing parameters. Dimensionality analyses found out that Penalaran subtest was local-dependent while Verbal and Kuantitatif subtests both were local-independent.

In addition, MIRT analyses failed to fully describe item characteristics of the test due to effect of interaction among probabilities of correct response of the three subtests.

Keywords: PAPs UGM; dimensionality; UIRT; MIRT

Berbagai pola respons para peserta dalam suatu tes, dituangkan oleh *item response theory* (IRT) dalam suatu model pengukuran. Salah-satu asumsi utama yang mendasari IRT adalah unidimensionalitas, yang berarti bahwa hanya terdapat satu atribut laten yang mendasari kemampuan atau abilitas para peserta tes dalam menjawab aitem (Hambleton, Swaminathan, & Rogers, 1991; Lord, 1980). Sekumpulan aitem-aitem dalam tes dapat disebut unidimensional bila kinerja pada peserta tes dapat dijelaskan oleh hanya sebuah atribut laten (Hambleton & Rovinelli, 1986). Lebih jauh, probabilitas untuk menjawab benar pada sebuah aitem tertentu hanya dipengaruhi oleh parameter aitem dan sebuah atribut laten θ , dan bukan yang lain. Inilah yang disebut dengan prinsip independensi lokal (*local independence*, LI) (Lord, 1980). Bila sebuah atribut laten belum cukup mampu untuk menjelaskan, dengan sendirinya independensi lokal tidak terpenuhi (Stout, 1984, 1989, 2002). Akibatnya asumsi unidimensionalitas tidak dapat dipertahankan dan implikasinya lebih lanjut adalah bahwa sekumpulan aitem disebut bersifat multidimensional.

Asumsi unidimensional kadang juga bersifat problematik, yaitu ketika aitem-aitem tes didesain untuk mengukur satu atribut laten tertentu namun ternyata para peserta memerlukan lebih dari satu atribut laten saja untuk dapat menjawab benar sebuah aitem. Apabila data respons yang bersifat multidimensi seperti itu kemudian diperlakukan sebagai data unidimensi maka berarti telah menyimpang dari asumsi unidimensionalitas dalam UIRT dan juga tidak sesuai dengan aspek struktural dari konstruk yang diukur (Messick, 1995). Solusi terhadap problem tersebut kemudian memunculkan model *multidimensional item response theory* (MIRT) (Reckase, 1985; Reckase & Ackerman, 1986). MIRT adalah semacam pengembangan *unidimensional item response theory* (UIRT) yang memungkinkan analisis terhadap aitem-aitem yang direspons benar oleh para peserta tes berdasarkan pada atribut laten lebih dari satu. IRT yang pada mulanya didasarkan pada asumsi unidimensi tentu mengalami kendala dalam melakukan penskoran pada tes-tes yang bersifat multidimensi (misalnya Ackerman, 1989; Cheng, Wang, & Ho, 2009; DeMars, 2006; Dirir & Sinclair, 1996; Oshima & Miller, 1990; Reise, Moore, & Haviland, 2010; Yao, 2011). Dengan kata lain, tes yang bersifat multidimensional akan mengalami ketidaktepatan bila diskor berdasarkan paradigma unidimensional.

Guna menyaring mahasiswa yang hendak mengikuti pendidikan pascasarjana, Universitas Gadjah Mada (UGM) Yogyakarta menggunakan skor tes potensi sebagai salah-satu kriteria penerimaan. Tes ini dikembangkan oleh Tim Fakultas Psikologi UGM dan diberi nama Tes Potensi Akademik Pascasarjana (PAPs). Skor peserta pada mata uji ini ikut menentukan diterima atau tidaknya dia pada program studi yang menjadi pilihannya. Mengacu pada terminologi yang dikemukakan oleh Thomas (2005), Liu, Harris, dan Schmidt (2007), dan Togut (2011), PAPs dapat disebut sebagai *high-stakes testing* karena konsekuensi yang akan diterima peserta tes berimplikasi pada masa depan mereka. Oleh sebab itu, sebagai sebuah tes yang berisiko tinggi, sudah semestinya mengandung kekeliruan sekecil mungkin dari sudut pandang pengukuran.

Tes yang mengukur potensi akademik dirancang untuk mengungkap kemampuan individu dalam menghadapi problem kognitif yang perlu diselesaikan dengan strategis dan cepat. PAPs, sebagaimana umumnya tes potensi, terdiri dari 3 subtes yang masing-masing mengukur abilitas Verbal, Kuantitatif, dan Penalaran (Azwar, 2008). Ketiga subtes tersebut diasumsikan bersifat unidimensional dan membentuk struktur potensi yang diukur. Persoalannya adalah pada dimensi atribut laten yang mendasari kemampuan peserta yang menjawab, apakah juga bersifat unidimensional atau multidimensional. Isu dimensi dalam tes ini penting untuk diteliti karena hal tersebut mempengaruhi penskoran, analisis data dan laporan hasilnya (Abedi, 1997; Kahraman & Thompson, 2011). Isu psikometrik lain yang perlu

diperhatikan adalah cara interpretasi terhadap kombinasi skor dari beberapa subtes (Ackerman, 1994; Reckase & McKinley, 1991).

Hasil penelitian ini akan berusaha menjawab pertanyaan-pertanyaan mengenai a) karakteristik aitem-aitem Subtes Verbal, Kuantitatif, dan Penalaran dalam PAPs berdasarkan *unidimensional item response theory*; b) dimensionalitas aitem-aitem dalam Tes PAPs; dan c) karakteristik aitem-aitem dalam Tes PAPs berdasarkan *multidimensional item response theory*.

Dimensionalitas Tes

Nunnally (1981), seorang pionir psikometrika, menegaskan bahwa sebuah tes idealnya berisikan aitem-aitem yang bersifat homogen; atau paling tidak tiap kluster berisikan aitem-aitem homogen. Hal ini sejalan dengan asumsi penting yang mendasari UIRT yaitu independensi lokal dapat terpenuhi terkait dengan sebuah atribut laten atau unidimensional.

Pentingnya memastikan unidimensionalitas dimana hanya sebuah atribut laten dapat menjelaskan keseluruhan matriks respons peserta tes sudah lama disarankan oleh Lord (1980). Informasi mengenai dimensionalitas tes ini juga akan memberikan bukti struktural terkait konsistensi antara struktur internal tes dan struktur konstruk (Fiske, 2002). Lebih jauh, informasi mengenai struktur dimensi ini dapat dijadikan fondasi dalam melaporkan skor atau subskor.

Multidimensionalitas akan terjadi manakala tes didesain mengukur atribut laten yang kompleks (Camilli, Wang, & Fesq, 1995). Bila sebuah tes didesain untuk mengukur atribut laten yang kompleks, sulit kiranya mengklaim konstruk yang diukur bersifat unidimensional murni. Apalagi, bila memang sejak awal sebuah tes didesain dengan domain isi yang bersifat multidimensi.

Pengondisian agar skor bersifat komparabel antar kelompok atau antar waktu seharusnya menjadi perhatian serius karena menyangkut validitas, utamanya aspek generalisasi (Messick, 1995). Perbedaan struktur antar kelompok atau waktu dapat ditelusuri berdasarkan dimensionalitasnya (Tate, 2002, 2003). Sementara dalam kenyataan analisis data, banyak peneliti yang menemukan bahwa dalam data respons tes riil tidak dapat dimodelkan secara baik menggunakan UIRT (Ackerman, 1989; Way, Ansley, & Forsyth, 1988). Dengan demikian diperlukan suatu model yang lebih mampu menjelaskan data matriks respons peserta tes.

Secara formal, dimensionalitas tes dapat didefinisikan sebagai jumlah dimensi minimum yang dapat menjelaskan data dan model sehingga bersifat independen secara lokal dan monoton (*monotone locally independent*, MLI) (Stout, 1989, 2002). Dimensionalitas dalam pengukuran dapat pula dimaknai sebagai banyaknya atribut laten yang mendasari kemampuan peserta dalam merespons aitem-aitem tes (Chou & Wang, 2010). Dalam konteks tes kemampuan, dimensionalitas disebut sebagai banyaknya kemampuan yang diukur oleh tes atau oleh kumpulan aitem.

Bila dikaitkan dengan konten materi tes, dimensionalitas dapat dipandang sebagai aspek-aspek pengukuran yang didesain untuk diukur oleh tes (Mislevy, Almond, & Lukas, 2003) atau bisa pula dipandang sebagai analisis terhadap data respons pada sekumpulan aitem (Hattie, Krakowski, Rogers, & Swaminathan, 1996; Nandakumar, Yu, & Zhang, 2011; Reckase, 2009; B. Zhang, 2008). Penelitian ini akan mengacu pada kedua sudut pandang ini. Di satu sisi, sebuah tes didesain dengan tujuan ukur pada domain atau dimensi-dimensi tertentu. Namun demikian, pada kenyataannya, perlu diselidiki interaksi para peserta dengan aitem-aitem tes yang tercermin dalam data respons yang diperoleh.

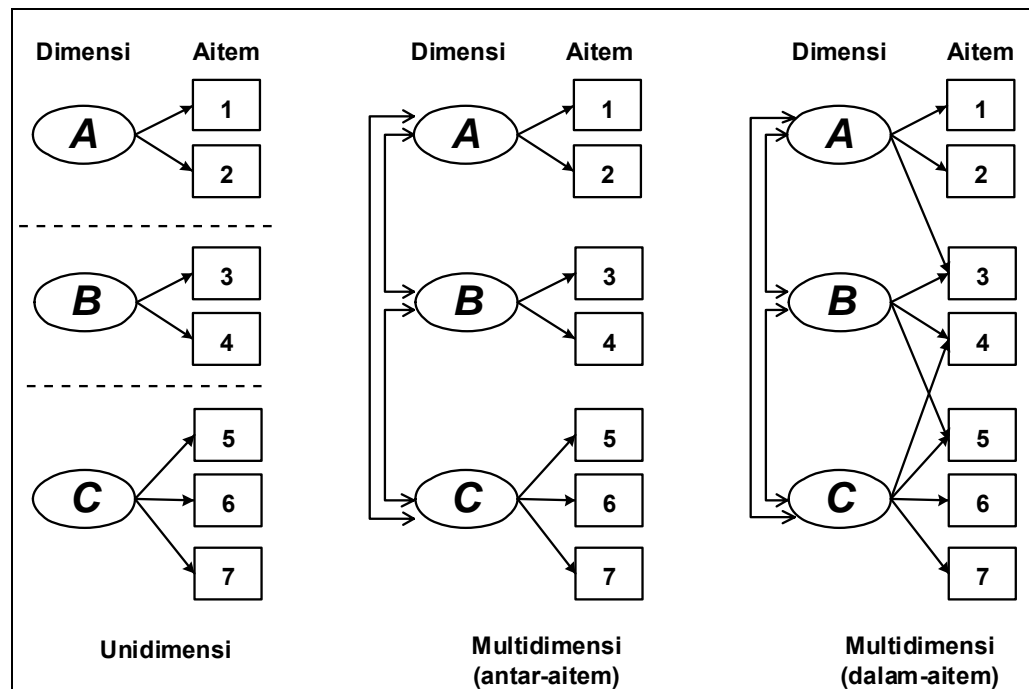
Meskipun pengertian dimensionalitas dapat dilihat dari sudut pandang yang berbeda, eksplorasi ataupun konfirmasi struktur dimensi merupakan bagian dari proses validasi yang bersifat komprehensif (Jang & Roussos, 2007). Oleh sebab itu,

dalam konteks sebuah tes yang terdiri dari beberapa subtes, aitem-aitem dalam tiap pasang subtes perlu diuji unidimensionalitasnya. Bila terbukti unidimensional maka dua subtes tersebut sebaiknya diperlakukan sebagai satu kesatuan sehingga, sebagai konsekuensinya, tidak perlu melaporkan subskor secara terpisah (Tate, 2000, h. 205).

Multidimensional Item Response Theory (MIRT)

Untuk menjawab benar sebuah aitem, sering kali peserta tes memerlukan lebih dari satu atribut laten (Ackerman, 1994) sehingga disebut multidimensi. Dalam kondisi seperti inilah, berdasarkan pendapat Reckase (1997), MIRT sangat berguna untuk memahami struktur atribut laten yang diperlukan untuk merespons benar aitem-aitem.

Dalam beberapa tahun terakhir, telah banyak penelitian berlandaskan sudut pandang teori MIRT. Reckase (1997) menulis tentang ringkasan anteseden MIRT dengan analisis faktor dan UIRT sebagai asal muasalnya. Dia meneruskan upaya beberapa ahli sebelumnya seperti Spearman, Thurstone, Lord dan Novick, serta Samejima. Lebih jauh, Ackerman, Gierl, dan Walker (2003) melihat dimungkinkannya aplikasi MIRT sekaligus mendiskusikannya dalam konteks mengevaluasi pengukuran dalam pendidikan. Asumsinya adalah bahwa setiap tes secara alamiah bersifat multidimensional, lebih sering mengukur lebih dari satu konstruk. Konstruk yang valid ialah sesuai dengan tujuan ukur yang telah dideskripsikan oleh pengembang tes.



Gambar 1. Representasi Grafis Model Unidimensi, Multidimensi antar aitem, dan Multidimensi dalam aitem [diadaptasi dari Cheng dkk. (2009)]

Aitem-aitem dalam tes seringkali mengukur abilitas komposit, di luar kehendak pengembang tes sebagaimana tergambar dalam *blueprint*. Bila sebuah aitem tidak cukup sensitif untuk mengukur lebih dari satu atribut laten atau peserta tes bervariasi dalam atribut laten yang sama, maka interaksi antara aitem dan peserta akan bersifat unidimensional (Ackerman, 1992, 1994).

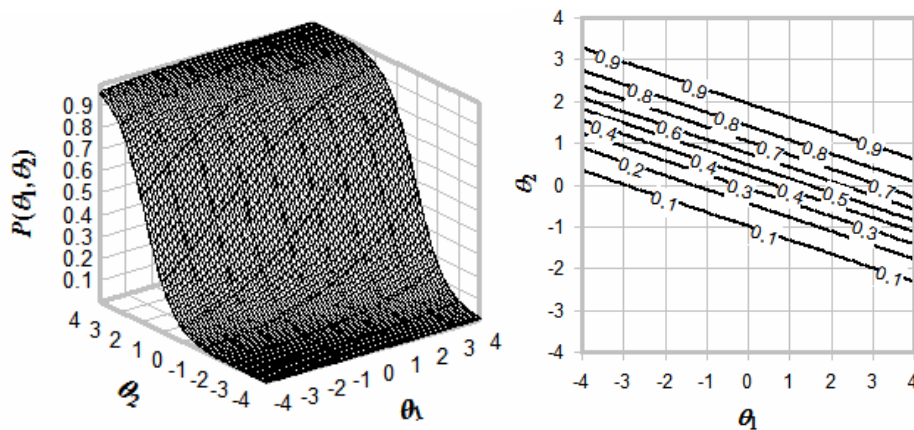
Konsepsi MIRT dapat dipandang sebagai kasus khusus dari analisis faktor atau model persamaan struktural, atau pengembangan dari UIRT (Reckase, 1997). Beberapa model yang mungkin terjadi sehingga mampu menjelaskan interaksi antara peserta dan aitem dapat direpresentasikan dalam Gambar 1 (Cheng, dkk., 2009).

MIRT terbagi menjadi dua jenis: model kompensasi (*compensatory*) dan model nonkompensasi (*noncompensatory*) (Reckase, 2009). Model kompensasi didasarkan pada hubungan kombinasi linier koordinat vektor atribut laten, θ . Sedangkan model nonkompensasi memisahkan atribut-atribut laten dalam merespons aitem serta menggunakan model UIRT untuk setiap atribut laten. Dalam penelitian ini hanya dibahas model kompensasi.

Satu di antara fakta dalam aitem dengan bentuk pilihan ganda adalah bahwa peserta dimungkinkan untuk menjawab aitem dengan benar melalui tebakan sehingga probabilitas menjawab benar melibatkan komponen tambahan, yaitu parameter tebakan. Model yang cocok dengan komponen ini adalah MIRT 3PL (Reckase, 1985, 2009):

$$P(U_{is} = 1 | \theta_s, \mathbf{a}_i, c_i, d_i) = c_i + (1 - c_i) \frac{e^{\mathbf{a}_i \theta_s + d_i}}{1 + e^{\mathbf{a}_i \theta_s + d_i}}. \quad (1)$$

dimana Vektor \mathbf{a}_i menunjukkan vektor $1 \times m$ parameter daya beda. Parameter d adalah intercept yang bersesuaian dengan garis sehingga menghasilkan $P(\theta_1, \theta_2) = 0,5$. Embretson dan Reise (2000) menyebut d sebagai *easiness intercept*. Makin tinggi harga d maka akan makin rendah tingkat kesukarannya. Sebuah aitem dengan parameter $a_1 = 0,5$, $a_2 = 1,5$, $d = -0,7$, dan $c = 0$ dapat diketahui karakteristiknya secara lebih jelas sebagaimana ditampakkan oleh Gambar2.



Gambar 2. Plot Permukaan dan Kontur Aitem, $a_1 = 0,5$, $a_2 = 1,5$, $d = -0,7$, dan $c = 0$

METODE

Data penelitian ini adalah skor Tes PAPs Seri A1 dari sebanyak 2035 orang calon mahasiswa pascasarjana UGM. Dengan demikian, variabel dalam penelitian berupa aitem-aitem dalam Subtes Verbal, Kuantitatif, dan Penalaran PAPs. Masing-masing Subtes terdiri dari 40 aitem sehingga keseluruhan aitem berjumlah 120.

Data respons peserta Tes PAPs pada tiga Subtes (Verbal, Kuantitatif, dan Penalaran) diperlakukan sebagai berikut.

Untuk mengetahui karakteristik aitem tiap Subtes berdasarkan *unidimensional item response theory* (UIRT), dilakukan kalibrasi parameter aitem pada tiap subtes dengan metode *marginal maximum likelihood* (MML) hingga diperoleh kecocokan data baik pada level aitem maupun level tes. Prosedur ini dilakukan dengan bantuan BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 2003).

Analisis dimensi dalam penelitian ini mengacu pada saran Jang dan Roussos (2007) yaitu dengan menerapkan teknik eksploratori dan konfirmatori pada struktur dimensi PAPs, dibantu software HCA/ CCPROX (Roussos, Stout, & Marden, 1998), DETECT (J. Zhang & Stout, 1999), dan DIMTEST (Stout & Nandakumar, 2006). Untuk mengetahui karakteristik aitem berdasarkan *multidimensional item response theory* (MIRT), dilakukan kalibrasi parameter aitem secara keseluruhan dengan memperhatikan struktur subtes dengan metode *bayesian* MIRT melalui bantuan BMIRT (Yao & Boughton, 2007).

HASIL

Analisis UIRT

Analisis UIRT dilakukan hanya pada aitem-aitem tes PAPs yang memenuhi persyaratan $r_{bis} \geq 0,25$ yaitu sebanyak 93 aitem dengan rincian: 30 aitem verbal; 32 aitem kuantitatif; dan 31 aitem penalaran. Pada tiap subtes, data respons dikalibrasi berdasarkan model logistik 3-parameter (3PL) dengan metode *marginal maximum likelihood* (MML). Hasilnya sebagai berikut,

Tabel 1. Parameter Aitem berdasarkan UIRT pada 30 Aitem Subtes Verbal

NO.	I	% B	a	b	c	NO.	I	% B	a	b	c
1	V001	67,8	0,497	-0,786	0,096	16	V024	50,3	0,533	0,168	0,060
2	V002	71,2	0,548	-0,999	0,065	17	V025	90,4	1,623	-1,515	0,064
3	V003	86,6	0,476	-2,210	0,210	18	V026	83,9	1,142	-1,247	0,040
4	V012	36,6	0,421	1,228	0,079	19	V027	78,4	0,592	-1,406	0,069
5	V013	64,9	0,669	-0,379	0,136	20	V028	80,7	1,157	-1,043	0,047
6	V014	88,9	1,571	-1,417	0,048	21	V029	75,4	0,698	-0,991	0,108
7	V015	51,5	0,586	0,070	0,048	22	V030	68,5	0,805	-0,558	0,099
8	V016	87,7	1,065	-1,571	0,046	23	V031	70,9	0,481	-0,818	0,176
9	V017	80,9	0,921	-1,199	0,050	24	V032	52,3	0,486	0,162	0,094
10	V018	62,1	0,626	-0,362	0,086	25	V033	83,9	1,275	-0,481	0,500
11	V019	53,2	0,607	0,094	0,089	26	V034	54,9	0,772	0,136	0,141
12	V020	44,4	0,541	0,445	0,047	27	V035	71,6	0,784	-0,808	0,046
13	V021	52,0	0,438	0,232	0,105	28	V036	71,7	0,717	-0,876	0,043
14	V022	42,5	0,646	0,500	0,050	29	V038	27,5	0,476	1,771	0,073
15	V023	77,0	0,536	-1,398	0,078	30	V039	90,7	1,479	-1,596	0,064

Keterangan: I = nama aitem; %B = persentase peserta menjawab benar, a = daya beda; b = kesukaran; dan c = tebakan

Hasil kalibrasi aitem-aitem Subtes Verbal dalam Tabel 1 menampakkan bahwa parameter daya beda (a), harga rata-ratanya adalah 0,772. Parameter daya beda tertinggi dimiliki aitem nomor 25 dengan $a_{25} = 1,623$, sementara terendah dimiliki oleh aitem nomor 12 dengan $a_{12} = 0,421$. Tingkat kesukaran rata-rata pada subtes ini sebesar -0,562; tertinggi pada aitem nomor 38 dengan $b_{38} = 1,771$ dan terendah aitem nomor 3 dengan $b_3 = -2,210$. Pada parameter tebakan semu, ditemukan rata-rata sebesar 0,095; tertinggi pada aitem nomor 33 dengan $c_{33} = 0,5$ dan terendah pada nomor 26 dengan $c_{26} = 0,04$.

Tabel 2. Parameter Aitem berdasarkan UIRT pada 32 Aitem Subtes Kuantitatif

NO.	I	% B	a	b	c	NO.	I	% B	a	b	c
1	K041	67,9	0,769	-0,658	0,062	17	K058	43,5	1,296	0,445	0,108
2	K042	77,0	1,070	-0,947	0,065	18	K059	53,3	0,464	0,179	0,120
3	K043	54,4	0,892	0,867	0,361	19	K060	41,9	0,782	0,956	0,194
4	K044	78,3	0,403	-1,515	0,225	20	K061	50,2	1,066	0,313	0,151
5	K045	55,1	1,667	0,777	0,393	21	K062	52,0	1,008	0,180	0,124
6	K047	76,9	1,107	-0,528	0,327	22	K063	26,3	0,990	1,721	0,167
7	K048	68,1	0,751	-0,066	0,334	23	K066	45,5	1,195	1,202	0,332
8	K049	81,3	1,196	-1,087	0,090	24	K067	61,7	0,737	0,032	0,240
9	K050	53,0	0,973	-0,012	0,049	25	K069	34,2	1,737	0,888	0,150
10	K051	62,1	0,993	-0,327	0,068	26	K070	50,2	1,116	0,425	0,195
11	K052	11,1	1,129	1,936	0,039	27	K071	49,7	0,513	0,687	0,195
12	K053	43,8	1,128	0,361	0,068	28	K072	50,1	2,242	0,506	0,255
13	K054	38,7	1,150	0,808	0,154	29	K073	24,7	1,526	1,664	0,178
14	K055	65,8	1,189	0,001	0,308	30	K077	50,4	0,980	0,375	0,174
15	K056	40,3	1,090	0,668	0,129	31	K078	53,3	0,776	0,597	0,271
16	K057	64,3	1,757	-0,148	0,192	32	K079	53,1	0,612	0,627	0,253

Keterangan: I = nama aitem; %B = persentase peserta menjawab benar, a = daya beda; b = kesukaran; dan c = tebakan

Untuk Subtes Kuantitatif yang disajikan pada Tabel 2, dapat dideskripsikan rata-rata daya beda Subtes Kuantitatif adalah 1,072, minimum $a_{44} = 0,403$ dan maksimum $a_{72} = 2,242$. Tingkat kesukaran rata-rata berharga 0,341 dengan harga minimum adalah $b_{44} = -1,515$ dan maksimum $b_{52} = 1,936$. Sementara itu, parameter tebakan memiliki harga rata-rata sebesar 0,187 dengan minimum $c_{52} = 0,039$ dan maksimum $c_{45} = 0,393$.

Tabel 3. Parameter Aitem berdasarkan UIRT pada 31 Aitem Subtes Penalaran

NO.	I	% B	a	b	c	NO.	I	% B	a	b	c
1	P081	86,2	1,005	-1,494	0,064	17	P100	32,7	0,923	1,054	0,118
2	P082	74,0	0,488	-1,274	0,081	18	P101	71,3	0,969	-0,650	0,093
3	P083	65,0	0,322	-0,863	0,105	19	P102	63,3	0,949	0,519	0,420
4	P084	51,4	0,775	0,129	0,079	20	P103	60,9	0,877	-0,152	0,135
5	P085	51,6	1,336	0,561	0,275	21	P104	76,8	0,765	-1,009	0,123
6	P086	59,5	0,838	-0,222	0,072	22	P105	75,8	1,133	-0,842	0,055
7	P087	70,4	0,560	-0,728	0,167	23	P108	23,7	0,480	2,358	0,101
8	P088	79,8	1,012	-1,025	0,112	24	P109	51,6	0,608	0,279	0,129
9	P090	12,6	1,497	1,986	0,084	25	P113	90,2	0,621	-2,139	0,254
10	P091	33,7	0,501	1,129	0,050	26	P114	69,2	0,577	-0,869	0,058
11	P093	26,8	0,674	1,480	0,084	27	P115	90,7	0,736	-1,791	0,353
12	P094	53,6	0,860	1,080	0,386	28	P116	38,7	0,367	1,255	0,089
13	P095	50,0	0,827	0,223	0,094	29	P117	84,8	1,441	-1,155	0,094
14	P096	49,4	0,833	0,563	0,206	30	P119	68,7	0,614	-0,793	0,064
15	P097	53,9	0,940	0,589	0,292	31	P120	64,7	0,840	-0,459	0,062
16	P099	61,4	0,929	-0,291	0,066						

Keterangan: %B = persentase peserta menjawab benar, a = daya beda; b = kesukaran; dan c = tebakan

Deskripsi karakteristik aitem-aitem pada Subtes Penalaran disajikan dalam Tabel 3. Tampak bahwa rata-rata daya beda adalah 0,816 dengan harga minimum $a_{83} = 0,322$ dan maksimum $a_{90} = 1,497$. Tingkat kesukaran memiliki rata-rata -0,082 dengan minimum $b_{113} = -2,139$ dan maksimum $b_{108} = 2,358$. Parameter tebakan memiliki harga rata-rata = 0,141 dengan minimum pada $c_{91} = 0,05$ dan maksimum pada $c_{102} = 0,420$.

Analisis Dimensionalitas

Analisis dimensionalitas untuk melihat apakah aitem-aitem tes PAPs bersifat unidimensional dilakukan dengan proses eksploratori dan konfirmatori. Proses eksploratori dilakukan dengan prosedur DIMTEST, HCA/CCPROX, dan DETECT. Sementara itu proses konfirmatori dilakukan dengan prosedur DIMTEST dan DETECT.

Eksplorasi melalui prosedur DIMTEST menghasilkan dua kluster dimana aitem-aitem Subtes Verbal dan Penalaran menjadi satu sebagai *partitioned test* (PT), sedangkan kluster ke dua berisikan aitem-aitem Subtes Kuantitatif yang berfungsi sebagai *assessment test* (AT). Hasil ini disajikan pada Tabel 4.

Tabel 4. Hasil Analisis Eksploratori dengan Prosedur DIMTEST

K	Nomor Aitem	Jumlah	T	p
PT	1, 2, 3, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 38, 39, 60, 81, 82, 83, 84, 85, 86, 87, 88, 90, 91, 93, 94, 95, 96, 97, 99, 100, 101, 102, 103, 104, 105, 108, 109, 113, 114, 115, 116, 117, 119, 120	60	6,281	<.001
AT	14, 36, 41, 42, 43, 44, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 62, 63, 66, 67, 69, 70, 71, 72, 73, 77, 78, 79	33		
TOTAL		93		

Keterangan: AT = assessment test; PT = partitioned test; K = kluster;

Statistik $T = 6,281$ dengan $p < 0,001$ menunjukkan bahwa aitem-aitem pada kedua kluster tidak bersifat lokal independen sehingga dapat dikatakan bahwa sebuah atribut laten saja tidak memadai dalam menjelaskan interaksi peserta tes dengan aitem-aitem PAPs. Jadi, aitem-aitem dalam dua kluster ini bersifat multidimensional. Implikasinya, data respons akan dapat dijelaskan secara lebih baik bila dimodelkan dengan MIRT, bukan dengan UIRT.

Walaupun analisis DIMTEST eksploratori di atas menghasilkan dua kluster dominan, mungkin saja dua kluster tidak cukup memadai untuk menjelaskan data respons peserta tes. Karena itu prosedur DETECT akan menentukan jumlah dan struktur dimensinya serta memartisi aitem-aitem kedalam kluster-kluster sehingga kovarians kondisional antaritem dalam satu kluster bernilai positif sedangkan dengan lain kluster bernilai negatif, yaitu aitem-aitem dalam kluster yang sama bersifat koheren satu sama lain. Hasilnya disajikan pada Tabel 5.

Tabel 5. Hasil Eksploratori dengan Prosedur DETECT

K	Nomor Aitem	Jumlah	Indeks
1	1, 2, 3, 12, 13, 14, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 35, 36, 38, 39, 81, 82, 83, 84, 85, 86, 87, 88, 90, 91, 93, 94, 95, 96, 108	41	Det = 0,195; IDN = 0,616; $r_{\max} = 0,388$
2	41, 42, 43, 44, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 62, 63, 66, 69, 70, 71, 72, 73, 77, 78	29	
3	15, 20, 33, 34, 60, 67, 79 , 97, 99, 100, 101, 102, 103, 104, 105, 109, 113, 114, 115, 116, 117, 119, 120	23	
TOTAL		93	

Keterangan: Aitem-aitem yang tercetak miring dan tebal tidak sesuai dengan yang diharapkan

Tampak bahwa sebagian aitem-aitem Subtes Penalaran dan hampir seluruh aitem Subtes Verbal mengerucut pada kluster 1. Hal ini mengindikasikan bahwa

dalam menjawab benar aitem-aitem penalaran, diperlukan pula kemampuan verbal. Pada kluster 2 seluruh isinya adalah aitem-aitem Subtes Kuantitatif. Hal yang menarik adalah terdapat beberapa aitem Subtes Verbal (nomor 15, 20, 33, dan nomor 34) serta aitem Subtes Kuantitatif (nomor 60, 67, dan nomor 79) yang bersama-sama dengan sebagian aitem-aitem Subtes Penalaran mengelompok menjadi kluster 3.

Partisi aitem sebagaimana disajikan pada Tabel 5. Tabel 5 merupakan partisi terbaik yang dapat dilakukan sehingga data respons pada tiap kluster bersifat homogen. Pengelompokan aitem-aitem menjadi kluster-kluster tersebut menunjukkan adanya multidimensionalitas di antara aitem-aitem namun indeks DETECT sebesar $Det = 0,193$ (mendekati 0) menunjukkan bahwa multidimensional tersebut sangat kecil (Monahan, Stump, Finch, & Hambleton, 2007; Roussos & Ozbek, 2006).

Bila dilakukan analisis eksploratori menggunakan prosedur HCA/CCPROX, yang ringkasan hasilnya disajikan dalam Tabel 6, tampak bahwa secara umum, aitem-aitem Subtes Verbal dan Kuantitatif mengerucut pada kluster yang diharapkan sekalipun masih terdapat beberapa aitem yang perlu mendapatkan perhatian pada kluster 1, yaitu aitem nomor 59 (Kuantitatif). Selain itu, aitem-aitem nomor 81, 82, 83, 84, 85, 86, 87, 88, 90, 91, 93, 94, 95, 96, 97, 99, 100, 101, 102, 103, 104, 108, 116 (Penalaran) ikut mengerucut bersama-sama dengan aitem-aitem Subtes Verbal pada kluster 1. Hal ini menguatkan indikasi sebelumnya bahwa untuk menjawab benar aitem-aitem penalaran, diperlukan pula kemampuan verbal.

Tabel 6. Hasil Eksploratori dengan Prosedur HCA/CCPROX

K	Nomor Aitem	Jumlah
1	1, 2, 3, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 59, 81, 82, 83, 84, 85, 86, 87, 88, 90, 91, 93, 94, 95, 96, 97, 99, 100, 101, 102, 103, 104, 108, 116	54
2	41, 42, 43, 44, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 60, 61, 62, 63, 66, 67, 69, 70, 71, 72, 73, 77, 78, 79	31
3	105, 109, 113, 114, 115, 117, 119, 120	8
TOTAL		93

Keterangan: Aitem-aitem yang tercetak miring dan tebal tidak sesuai dengan yang diharapkan

Dari hasil analisis eksploratori yang dilakukan melalui prosedur DIMTEST, DETECT, dan HCA/CCPROX, disimpulkan bahwa respons peserta pada aitem-aitem Subtes Penalaran menghasilkan data dengan kovarians kondisional yang terbagi dengan Subtes Verbal. Hal itu tidak terjadi pada aitem-aitem Subtes Kuantitatif. Kenyataan tersebut diperkuat oleh hasil analisis konfirmatori berupa uji independensi lokal pada tiap subtes yang disajikan pada Tabel 7, yang menunjukkan bahwa aitem-aitem Subtes Penalaran bersifat independen secara lokal dengan aitem-aitem Subtes Verbal. Dengan kata lain, secara umum, aitem-aitem dalam Subtes Verbal dan Penalaran berbagi varians satu sama lain.

Tabel 7. Hasil Konfirmatori Aitem-aitem antar Subtes dengan Prosedur DIMTEST

Pasangan	TL	TGbar	T	p	Independensi lokal aitem
Verbal – Kuantitatif	14,451	10,516	3,915	< 0,01	Tidak
Verbal – Penalaran	8,437	7,645	0,788	0,215	Ya
Kuantitatif – Penalaran	14,109	9,271	4,814	< 0,01	Tidak

Keterangan: TL = statistik T yang diperoleh dari data respons; TGbar = statistik T yang diperoleh dari rata-rata data simulasi;

Aitem-aitem Subtes Penalaran dan Subtes Verbal yang bersifat independen lokal mengisyaratkan adanya dua kemungkinan. Pertama, kedua kelompok aitem mengukur sebuah dimensi secara bersama. Ke dua, dengan adanya dua dimensi (dalam hal ini verbal dan penalaran), kedua kelompok aitem mengandung bobot yang hampir sama dalam mengungkap kedua dimensi tersebut. Dikatakan bahwa untuk dapat menjawab benar aitem-aitem Subtes Penalaran, diperlukan dua kemampuan laten, yaitu kemampuan penalaran dan kemampuan verbal.

Konfirmatori ke dua dilakukan dengan prosedur DETECT untuk mengetahui sejauh mana tingkat dan kompleksitas dimensionalitasnya bila aitem dikelompokkan sesuai dengan dimensi (subtes) masing-masing, yaitu Verbal (θ_1), Kuantitatif (θ_2), dan Penalaran (θ_3).

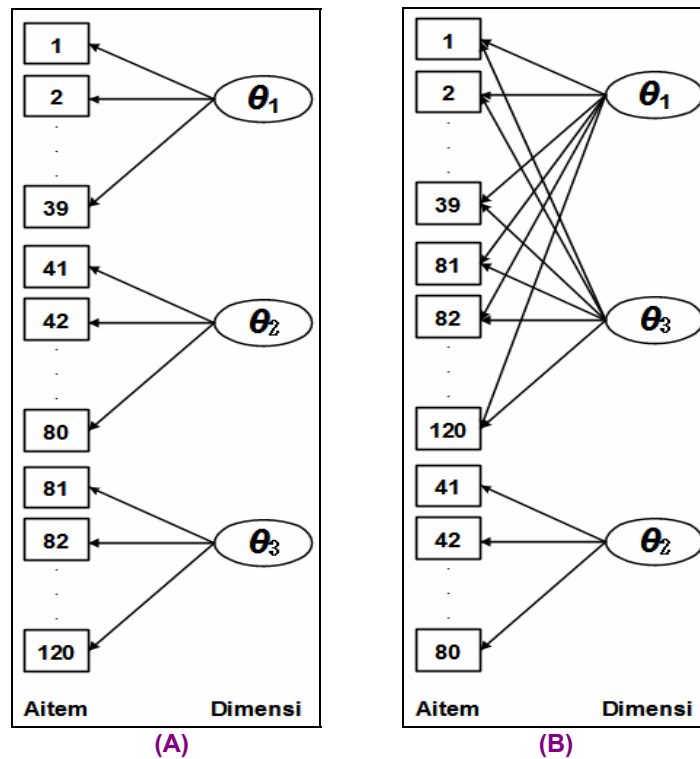
Tabel 8. Hasil Konfirmatori 3 Dimensi dengan Prosedur DETECT

K	Nomor Aitem	Jumlah	Indeks
1	1, 2, 3, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39	30	Det = 0,179; IDN = 0,631; $r_{\max} = 0,439$
2	41, 42, 43, 44, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 66, 67, 69, 70, 71, 72, 73, 77, 78, 79	32	
3	81, 82, 83, 84, 85, 86, 87, 88, 90, 91, 93, 94, 95, 96, 97, 99, 100, 101, 102, 103, 104, 105, 108, 109, 113, 114, 115, 116, 117, 119, 120	31	
TOTAL		93	

Sebagaimana dimuat pada Tabel 8, klaster 1, 2, dan 3 secara berturut-turut mengacu pada dimensi Verbal (θ_1), Kuantitatif (θ_2), dan Penalaran (θ_3) menunjukkan indeks DETECT Det = 0,179 yang mengindikasikan multidimensionalitas dengan tingkat yang rendah. Indeks IDN = 0,631 dan $r_{\max} = 0,439$ mengindikasikan bentuk multidimensi yang bersifat kompleks. Penjelasan lebih detil tentang indeks Det, IDN, dan r_{\max} dapat dilihat dalam Zhang dan Stout (1999) dan Monahan, dkk. (2007).

Hasil dari rangkaian proses analisis eksplorasi dan konfirmasi sebagaimana dideskripsikan di atas membawa pada kesimpulan bahwa dimensi yang mengukur aitem-aitem Tes PAPs adalah dimensi Verbal (θ_1), Kuantitatif (θ_2), dan Penalaran (θ_3). Hal ini bermakna bahwa terdapat tiga sumbu ortogonal dimana masing-masing sumbu mencerminkan masing-masing dimensi. Berdasarkan sudut pandang model MIRT, interaksi antara peserta tes dan aitem-aitem akan menghasilkan karakteristik aitem berupa vektor aitem yang dapat diproyeksikan pada ketiga sumbu tersebut. Aitem-aitem Subtes Verbal dan Penalaran dapat diproyeksikan pada sumbu θ_1 dan θ_3 , sedangkan aitem-aitem Subtes Kuantitatif dapat diproyeksikan pada satu sumbu yaitu sumbu θ_2 .

Dalam bentuk ilustrasi grafis, diagram jalur struktur dimensi dan hubungannya dengan aitem-aitem Tes PAPs disajikan dalam Gambar 3. Bagian kiri (A) merupakan struktur teoritik sebagaimana semula didesain oleh perancang tes, sementara bagian kanan (B) adalah struktur yang menunjukkan interaksi peserta tes dengan aitem-aitem secara empirik.



Gambar 3. Struktur Dimensi PAPs: (A) Teoritik; (B) Empirik
 Keterangan: θ_1 = Verbal; θ_2 = Kuantitatif; θ_3 = Penalaran.
 Angka-angka dalam kotak adalah nomor aitem.

Analisis MIRT

Hasil analisis dimensionalitas yang telah dijelaskan di atas menjadi alasan perlunya dilakukan analisis MIRT terkait dengan struktur kontrak PAPs. Secara teoretik struktur kontrak PAPs dirancang mengikut Gambar 3A, namun data empirik jawaban dari 2035 peserta tes menunjukkan bahwa Gambar 3B lebih dapat mencerminkan struktur data respons yang diperoleh. Dengan struktur seperti Gambar 3B, dapat dikatakan bahwa aitem-aitem PAPs bersifat multidimensi dalam-aitem. Justifikasi ini didasarkan pada kenyataan terdapatnya aitem-aitem yang mengukur dua dimensi (Verbal dan Penalaran).

Berbeda dari kerangka analisis UIRT yang dilakukan secara terpisah untuk tiap subtes PAPs, dalam analisis MIRT kalibrasi dilakukan secara utuh mengikuti struktur kontrak sebagaimana disajikan dalam Gambar 3B serta besarnya korelasi di antara dimensi penyusunnya. Dalam konteks ini dimensi yang dimaksud disesuaikan dengan kontrak awal PAPs, yaitu dimensi Verbal (θ_1), dimensi Kuantitatif (θ_2), dan dimensi Penalaran (θ_3).

Berdasarkan ketiga dimensi yang telah teridentifikasi sebelumnya dengan struktur sebagaimana pada Gambar 3B, dalam Tabel 9 disajikan ringkasan hasil estimasi parameter daya beda bagi setiap dimensi (masing-masing a_1 , a_2 , dan a_3), parameter tingkat kemudahan (d), dan parameter peluang tebakan (c). Dapat disimpulkan beberapa karakteristik PAPs seri A1 sebagai berikut: a) Dimensi Verbal memiliki daya beda yang sedang, b) Dimensi Kuantitatif memiliki daya beda yang tinggi, c) Dimensi Penalaran memiliki daya beda yang sedang, d) Tingkat kemudahan tes berada pada taraf sedang, dan e) Peluang tebakan berada pada taraf sedang.

Tabel 9. Ringkasan Parameter PAPs menurut Analisis MIRT

Parameter	Minimum	Maksimum	Rata-rata
Daya Beda Dimensi Verbal ($a1$)	$a1_{90} = 0,284$	$a1_{25} = 2,043$	1,046
Daya Beda Dimensi Kuantitatif ($a2$)	$a2_{44} = 0,589$	$a2_{72} = 3,024$	1,452
Daya Beda Dimensi Penalaran ($a3$)	$a3_{38} = 0,390$	$a3_{33} = 1,520$	0,892
Kemudahan (d)	$b_{25} = -2,543$	$b_{52} = 2,368$	0,030
Peluang Tebakan (c)	$c_{90} = 0,122$	$c_{102} = 0,227$	0,168

Kenyataan bahwa secara empirik struktur konstruk tes PAPs seri A1 tersusun sebagaimana nampak pada Gambar 3B, ada beberapa implikasi. Pertama, abilitas potensi akademik peserta tes perlu dilaporkan dalam bentuk atribut laten yang bersifat komposit. Ke dua, abilitas potensi akademik peserta tes dicerminkan oleh: (a) skor laten gabungan antara dimensi Verbal–Penalaran; dan (b) skor laten dimensi Kuantitatif. Mengacu pada Yen dan Walker (2007), skor abilitas komposit PAPs dapat dilakukan dengan cara merata-rata skor Verbal–Penalaran dan Kuantitatif melalui UIRT atau secara langsung dengan mengestimasi kombinasi linier antara Verbal–Penalaran dan Kuantitatif melalui MIRT.

KESIMPULAN

Kesimpulan yang dapat diambil dari hasil penelitian ini adalah:

1. Sebanyak 27 aitem tidak diikuti dalam kalibrasi UIRT karena r_{bis} terlalu rendah (<0,25), yaitu aitem-aitem nomor 4 – 11, 37 dan 40 (Verbal); 46, 64, 65, 68, 74, 75, 76, dan 80 (Kuantitatif); 89, 92, 98, 106, 107, 110 – 112, dan 118 (Penalaran). Hasil kalibrasi UIRT memperoleh hasil parameter tebakan terlalu tinggi (>0,35) terdapat pada 6 aitem, yaitu aitem-aitem nomor 33 (Verbal); 43 dan 45 (Kuantitatif); 94, 102, dan 115 (Penalaran).
2. Berdasarkan rata-rata parameter aitem dalam kerangka UIRT,
 - a. Subtes Verbal memiliki daya beda baik, tingkat kesukaran yang agak mudah, dan peluang tebakan yang rendah.
 - b. Subtes Kuantitatif memiliki daya beda baik, tingkat kesukaran yang agak sulit, dan peluang tebakan yang rendah.
 - c. Subtes Penalaran memiliki daya beda baik, tingkat kesukaran yang sedang, dan peluang tebakan yang rendah.
3. Hasil analisis dimensionalitas menunjukkan bahwa aitem-aitem dalam Tes PAPs bersifat multidimensional. Secara lebih rinci, aitem-aitem Verbal – Kuantitatif tidak bersifat independen lokal, Verbal – Penalaran bersifat independen lokal, dan Kuantitatif – Penalaran bersifat tidak independen lokal. Dengan kata lain, aitem-aitem PAPs bersifat multidimensi dalam-aitem.
4. Karakteristik aitem-aitem PAPs berdasarkan MIRT adalah sebagai berikut.
 - a. Dimensi Verbal memiliki daya beda yang sedang.
 - b. Dimensi Kuantitatif memiliki daya beda yang tinggi.
 - c. Dimensi Penalaran memiliki daya beda yang sedang.
 - d. Tingkat kemudahan berada pada taraf sedang.
 - e. Tebakan berada pada taraf sedang.

Rekomendasi yang diberikan berdasarkan hasil studi ini adalah:

1. Bahwasanya terdapat aitem-aitem yang mengukur lebih dari satu dimensi, maka direkomendasikan bagi pengembang Tes PAPs untuk mempelajari karakter dan penyebab terjadinya hal tersebut dengan analisis isi (*content analysis*).

2. Berdasarkan kenyataan bahwa aitem-aitem Subtes Verbal dan Penalaran bersifat independen secara lokal, pengembang Tes PAPs dapat memperlakukan kedua subtes sebagai satu kesatuan subtes yang mengukur penalaran verbal.

oo0oo

DAFTAR PUSTAKA

- Abedi, J. (1997). Dimensionality of NAEP Subscale Scores in Mathematics *CSE Technical Report 428*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing (CRESST) University of California.
- Ackerman, T. A. (1989). Unidimensional IRT Calibration of Compensatory and Noncompensatory Multidimensional Items. *Applied Psychological Measurement, 13*(2), 113-127.
- Ackerman, T. A. (1992). *Assessing Construct Validity Using Multidimensional Item Response Theory*. Paper dipresentasikan pada the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Ackerman, T. A. (1994). Using Multidimensional Item Response Theory to Understand What Items and Tests Are Measuring. *Applied Measurement in Education, 7*(4), 255-278.
- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using Multidimensional Item Response Theory to Evaluate Educational and Psychological Tests. *Educational and Psychological Measurement, 22*(3), 37–51.
- Azwar, S. (2008). Kualitas Tes Potensi Akademik Versi 07A. *Jurnal Penelitian dan Evaluasi Pendidikan, 12*(2), 232-250.
- Camilli, G., Wang, M.-m., & Fesq, J. (1995). The Effects of Dimensionality on Equating the Law School Admission Test. *Journal of Educational Measurement, 32*(1), 79-96.
- Cheng, Y.-Y., Wang, W.-C., & Ho, Y.-H. (2009). Multidimensional Rasch Analysis of a Psychological Test With Multiple Subtests: A Statistical Solution for the Bandwidth--Fidelity Dilemma. *Educational and Psychological Measurement, 69*(3), 369-388.
- Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models With Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement, 70*(5), 717-731.
- DeMars, C. E. (2006). *Scoring Subscales Using Multidimensional Item Response Theory Models*. Paper dipresentasikan pada the Annual Meeting of the American Psychological Association, Washington, DC.
- Dirir, M. A., & Sinclair, N. (1996). *On Reporting IRT Ability Scores When the Test Is Not Unidimensional*. Paper dipresentasikan pada the Annual Meeting of the National Council on Measurement in Education, New York.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologist*. NJ: Lawrence Erlbaum Associates Inc.
- Fiske, D. W. (2002). Validity for what? Dalam H. I. Braun, D. N. Jackson & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (Edisi 1, hh. 169-178). Mahwah, NJ: Lawrence Erlbaum.
- Hambleton, R. K., & Rovinelli, R. J. (1986). Assessing the Dimensionality of a Set of Test Items. *Applied Psychological Measurement, 10*(3), 287-302.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Application*. Boston, MA: Kluwer Inc.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. CA: Sage Publication Inc.
- Hattie, J., Krakowski, K., Rogers, H. J., & Swaminathan, H. (1996). An Assessment of Stout's Index of Essential Unidimensionality. *Applied Psychological Measurement, 20*(1), 1-14.
- Jang, E. E., & Roussos, L. A. (2007). An Investigation into the Dimensionality of TOEFL Using Conditional Covariance-Based Nonparametric Approach. *Journal of Educational Measurement, 44*(1), 1–21.

- Kahraman, N., & Thompson, T. (2011). Relating Unidimensional IRT Parameters to a Multidimensional Response Space: A Review of Two Alternative Projection IRT Models for Scoring Subscales. *Journal of Educational Measurement*, 48(2), 146-164.
- Liu, J., Harris, D. J., & Schmidt, A. (2007). Statistical Procedures Used in College Admissions Testing. Dalam C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics 26: Psychometrics* (Edisi 1, hh. 1057-1091). Amsterdam: Elsevier.
- Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.
- Messick, S. J. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A Brief Introduction to Evidence Centered Design. *Rsearch Report RR-03-16*. Princeton: Educational Testing Services.
- Monahan, P. O., Stump, T. E., Finch, H., & Hambleton, R. K. (2007). Bias of Exploratory and Cross-Validated DETECT Index Under Unidimensionality. *Applied Psychological Measurement*, 31(6), 483-503. doi: 10.1177/0146621606292216
- Nandakumar, R., Yu, F., & Zhang, Y. (2011). A Comparison of Bias Correction Adjustments for the DETECT Procedure. *Applied Psychological Measurement*, 35(2), 127-144.
- Nunnally, J. C. (1981). *Psychometric Theory*. New Delhi: McGraw-Hill Company Limited.
- Oshima, T. C., & Miller, M. D. (1990). Multidimensionality and IRT-Based Item Invariance Indexes: The Effect of Between-Group Variation in Trait Correlation. *Journal of Educational Measurement*, 27(3), 273-283.
- Reckase, M. D. (1985). The Difficulty of Test Items That Measure More Than One Ability. *Applied Psychological Measurement*, 9(4), 401-412.
- Reckase, M. D. (1997). The Past and Future of Multidimensional Item Response Theory. *Applied Psychological Measurement*, 21(1), 25-36.
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. New York: Springer.
- Reckase, M. D., & Ackerman, T. A. (1986). *Building a Test Using Items That Require More than One Skill to Determine a Correct Answer*. Paper dipresentasikan pada the The Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Reckase, M. D., & McKinley, R. L. (1991). The Discriminating Power of Items That Measure More Than One Dimension. *Applied Psychological Measurement*, 15(4), 361-373.
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor Models and Rotations: Exploring the Extent to Which Multidimensional Data Yield Univocal Scale Scores. *Journal of Personality Assessment*, 92(6), 544-559.
- Roussos, L. A., & Ozbek, O. Y. (2006). Formulation of the DETECT Population Parameter and Evaluation of DETECT Estimator Bias. *Journal of Educational Measurement*, 43(3), 215-243.
- Roussos, L. A., Stout, W. F., & Marden, J. I. (1998). Using New Proximity Measures with Hierarchical Cluster Analysis to Detect Multidimensionality. *Journal of Educational Measurement*, 35(1), 1-30.
- Stout, W. F. (1984). A Statistical Procedure for Assessing Test Dimensionality. *Measurement Series 84-2*. Washington, D.C.: ERIC Clearinghouse.
- Stout, W. F. (1989). A New Item Response Theory Modeling Approach with Applications to Unidimensionality Assessment and Ability Estimation.

- Cognitive Science Program*. Champaign, IL: Department of Statistics - Univ. of Illinois.
- Stout, W. F. (2002). Psychometrics: From Practice to Theory and Back (15 Years of Nonparametric Multidimensional IRT, DIF/Test Equity, and Skills Diagnostic Assessment). *Psychometrika*, *67*(4), 485-518.
- Stout, W. F., & Nandakumar, R. (2006). DIMTEST 2.1 [Computer Software]. Missoula: Assessment System Corporation.
- Tate, R. (2000). Performance of a Proposed Method for the Linking of Mixed Format Tests with Constructed Response and Multiple Choice Items. *Journal of Educational Measurement*, *37*(4), 329-346.
- Tate, R. (2002). Test Dimensionality. Dalam G. Tindal & T. M. Haladyna (Eds.), *Large-Scale Assessment Program for All Students: Validity, Technical Adequacy, and Implementation* (Edisi 1, hh. 181-211). Mahwah, NJ: Lawrence Erlbaum.
- Tate, R. (2003). A Comparison of Selected Empirical Methods for Assessing the Structure of Responses to Test Items. *Applied Psychological Measurement*, *27*(3), 159-203.
- Thomas, R. M. (2005). *High-Stakes Testing: Coping with Collateral Damage*. New Jersey: Lawrence Erlbaum.
- Togut, T. D. (2011). High-Stakes Testing: Educational Barometer for Success, or False Prognosticator for Failure. Diterima 15 Agustus 2011, dari <http://www.harborhouselaw.com/articles/highstakes.togut.htm#1>
- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The Comparative Effects of Compensatory and Noncompensatory Two-Dimensional Data on Unidimensional IRT Estimates. *Applied Psychological Measurement*, *12*(3), 239-252.
- Yao, L. (2011). Multidimensional Linking for Domain Scores and Overall Scores for Nonequivalent Groups. *Applied Psychological Measurement*, *35*(1), 48-66.
- Yao, L., & Boughton, K. A. (2007). A Multidimensional Item Response Modeling Approach for Improving Subscale Proficiency Estimation and Classification. *Applied Psychological Measurement*, *31*(2), 83-105.
- Yen, S. J., & Walker, L. (2007). *Multidimensional IRT Models for Composite Scores*. Paper dipresentasikan pada the Annual Meeting of the National Council of Measurement in Education, Chicago, IL.
- Zhang, B. (2008). Application of Unidimensional Item Response Models to Tests With Items Sensitive to Secondary Dimensions. *Journal of Experimental Education*, *77*(2), 147.
- Zhang, J., & Stout, W. F. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, *64*(2), 231-249.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). BILOG-MG (Version 3). Lincolnwood, IL: Scientific Software International.

LAMPIRAN

Lampiran 1. Ringkasan Parameter Aitem Subtes Verbal dalam Kerangka UIRT

Parameter	Minimum	Maksimum	Rata-rata
Daya Beda (a)	$a_{12} = 0,421$	$a_{25} = 1,623$	0,772
Kesukaran (b)	$b_3 = -2,210$	$b_{38} = 1,771$	-0,562
Tebakan (c)	$c_{26} = 0,04$	$c_{33} = 0,5$	0,095

Lampiran 2. Ringkasan Parameter Aitem Subtes Kuantitatif dalam Kerangka UIRT

Parameter	Minimum	Maksimum	Rata-rata
Daya Beda (a)	$a_{44} = 0,403$	$a_{72} = 2,242$	1,072
Kesukaran (b)	$b_{44} = -1,515$	$b_{52} = 1,936$	0,341
Tebakan (c)	$c_{52} = 0,039$	$c_{45} = 0,393$	0,187

Lampiran 3. Ringkasan Parameter Aitem Subtes Penalaran dalam Kerangka UIRT

Parameter	Minimum	Maksimum	Rata-rata
Daya Beda (a)	$a_{83} = 0,322$	$a_{90} = 1,497$	0,816
Kesukaran (b)	$b_{113} = -2,139$	$b_{108} = 2,358$	-0,082
Tebakan (c)	$c_{91} = 0,05$	$c_{102} = 0,420$	0,141

Lampiran 4. Parameter Aitem berdasarkan kerangka MIRT

AITEM	a1	a2	a3	d	c
VER001	0,647	0	0,521	-0,316	0,175
VER002	1,020	0	0,466	-0,517	0,163
VER003	0,766	0	0,491	-1,553	0,195
VER012	0,574	0	0,469	1,432	0,174
VER013	1,102	0	0,949	-0,021	0,175
VER014	1,400	0	1,153	-1,882	0,201
VER015	1,009	0	0,841	0,714	0,166
VER016	1,711	0	0,857	-1,884	0,138
VER017	1,149	0	0,993	-1,254	0,156
VER018	0,997	0	0,913	0,087	0,158
VER019	0,910	0	0,968	0,661	0,176
VER020	0,697	0	0,907	0,945	0,153
VER021	0,687	0	0,520	0,496	0,141
VER022	0,838	0	0,930	1,144	0,143
VER023	0,689	0	0,536	-0,873	0,185
VER024	0,812	0	0,736	0,674	0,165
VER025	2,043	0	1,000	-2,543	0,160
VER026	1,636	0	0,997	-1,559	0,174
VER027	0,944	0	0,488	-1,037	0,182
VER028	1,519	0	1,124	-1,270	0,170
VER029	0,890	0	0,887	-0,755	0,176
VER030	1,309	0	1,174	-0,192	0,175
VER031	0,823	0	0,515	-0,444	0,172
VER032	0,539	0	0,863	0,581	0,163
VER033	1,715	0	1,520	-1,245	0,215
VER034	1,036	0	0,890	0,540	0,165
VER035	1,069	0	0,964	-0,543	0,158
VER036	0,984	0	0,651	-0,570	0,160
VER038	0,830	0	0,390	1,825	0,139
VER039	1,289	0	1,243	-2,314	0,170
KUA041	0	1,315	0	-0,771	0,175
KUA042	0	1,306	0	-1,370	0,143
KUA043	0	0,815	0	0,233	0,196
KUA044	0	0,589	0	-1,145	0,160

AITEM	a1	a2	a3	d	c
KUA045	0	1,103	0	0,264	0,201
KUA047	0	1,267	0	-1,320	0,153
KUA048	0	0,965	0	-0,629	0,193
KUA049	0	2,011	0	-2,229	0,198
KUA050	0	1,864	0	0,141	0,154
KUA051	0	1,799	0	-0,416	0,186
KUA052	0	0,750	0	2,368	0,124
KUA053	0	2,099	0	0,842	0,154
KUA054	0	1,769	0	1,317	0,151
KUA055	0	1,386	0	-0,640	0,166
KUA056	0	1,565	0	0,920	0,134
KUA057	0	2,806	0	-0,879	0,178
KUA058	0	2,378	0	1,095	0,174
KUA059	0	0,826	0	0,246	0,177
KUA060	0	1,226	0	1,107	0,189
KUA061	0	1,813	0	0,480	0,185
KUA062	0	1,844	0	0,305	0,171
KUA063	0	1,076	0	1,830	0,135
KUA066	0	0,965	0	0,869	0,210
KUA067	0	1,076	0	-0,242	0,188
KUA069	0	2,176	0	1,652	0,123
KUA070	0	1,660	0	0,479	0,181
KUA071	0	0,792	0	0,531	0,205
KUA072	0	3,024	0	0,935	0,198
KUA073	0	0,875	0	1,993	0,164
KUA077	0	1,466	0	0,386	0,170
KUA078	0	0,993	0	0,237	0,178
KUA079	0	0,868	0	0,305	0,187
PEN081	1,332	0	1,146	-1,865	0,165
PEN082	0,738	0	0,677	-0,756	0,165
PEN083	0,628	0	0,457	-0,202	0,163
PEN084	1,294	0	1,209	0,830	0,140
PEN085	1,419	0	1,328	1,079	0,165
PEN086	1,488	0	1,165	0,321	0,145
PEN087	0,938	0	0,493	-0,453	0,167
PEN088	1,287	0	1,227	-1,118	0,155
PEN090	0,284	0	0,808	2,210	0,122
PEN091	0,763	0	0,977	1,755	0,158
PEN093	0,920	0	0,684	1,527	0,134
PEN094	0,608	0	0,695	0,514	0,181
PEN095	1,363	0	0,985	0,925	0,161
PEN096	1,184	0	0,955	1,076	0,183
PEN097	0,852	0	1,072	0,776	0,200
PEN099	1,303	0	1,284	0,210	0,149
PEN100	1,069	0	1,168	1,902	0,144
PEN101	1,293	0	1,321	-0,483	0,151
PEN102	0,725	0	0,714	0,254	0,227
PEN103	1,220	0	1,279	0,354	0,187
PEN104	0,857	0	0,890	-0,852	0,206
PEN105	1,714	0	1,477	-0,684	0,192
PEN108	0,763	0	0,489	2,090	0,125
PEN109	1,024	0	0,840	0,739	0,162
PEN113	1,043	0	0,811	-1,650	0,172
PEN114	0,834	0	0,792	-0,368	0,168
PEN115	0,903	0	0,914	-1,838	0,164
PEN116	0,661	0	0,601	1,285	0,153
PEN117	1,639	0	0,800	-1,708	0,132

AITEM	a1	a2	a3	d	c
PEN119	1,043	0	0,737	-0,277	0,178
PEN120	0,998	0	1,471	-0,037	0,160

Keterangan: a1 = parameter daya beda pada dimensi verbal (θ_1)

a1 = parameter daya beda pada dimensi kuantitatif (θ_2)

a1 = parameter daya beda pada dimensi penalaran (θ_3)

d = parameter tingkat kemudahan

c = parameter tebakan