PEMILIHAN BUTIR ALTERNATIF PADA TES ADAPTIF UNTUK PENINGKATAN KEAMANAN TES

Agus Santoso
FMIPA Universitas Terbuka
email: aguss@ut.ac.id

Abstrak

Penelitian ini bertujuan mendesksripsikan pengaruh pemilihan butir dengan pengacakan pada algoritma tes adaptif terhadap panjang tes dan estimasi kemampuan peserta tes. Penelitian ini dilakukan dengan prosedur simulasi. Bank soal untuk keperluan simulasi menggunakan 250 butir soal terkalibrasi menggunakan model *item response theory* 3 parameter. Dua desain algoritma tes adaptif, desain algoritma dengan dan tanpa pengacakan dikembangkan. Pada algoritma tes adaptif dengan pengacakan, pengacakan butir dilakukan pada urutan butir kedua dan ketiga, sedangkan pada algoritma tes adaptif tanpa pengacakan, tidak dilakukan pengacakan pada pemilihan butir soal berikutnya yang akan diberikan kepada peserta tes. Hasil simulasi menunjukkan bahwa secara statistik pengacakan tidak berpengaruh terhadap estimasi kemampuan peserta tes (nilai-*p*=0,306), dan panjang tes (nilai-*p*=0,328). Berdasarkan hal tersebut desain algoritma dengan pengacakan lebih tepat diterapkan pada algoritma tes adaptif.Tanpa mengurangi tingkat efisiensi dan presisi pengukuran, butir soal pada urutan awal yang diberikan kepada peserta tes lebih bervariasi sehingga dapat meningkatkan keamanan tes.

Kata kunci: tes adaptif, keamanan tes

SELECTION OF ALTERNATIVE ITEMS IN ADAPTIVE TESTS FOR THE INCREAMENT OF TEST SECURITY

Abstract

The study is aimed at describing the effects of the selection of test items using randomzation on the alogarithm of adaptive tests against test length and testees' ability estimation. The study is conducted using the simulation technique. The study uses an item bank of 250 items callibrated using the parameter 3 item response theory model. Two adaptive alogarittms are developed: one with and one without randomization. For the randomized model, item randomization is done on the second and third item, while for the non-randomization model, no randomization is applied. Results of the simulation show that, statistically, randomization dos not have an effect on the testees' ability estimation (*p value*=0,306), and test length (*p value*=0,328). Based on this, the randomized alogarithm model is more applicable to adaptive test alogarithm. With no decrease in the measurement efficiency and precision, first test items given to the testees are more varied such that it can increase test security.

Keywords: adaptive test, test security

PENDAHULUAN

Kemajuan teknologi komputer membawa dampak penting bagi kehidupan manusia, tidak terkecuali pada bidang pengukuran pendidikan dan pengujian. Perkembangan teori pengukuran modern, *item respons theory* (IRT) disertai ketersediaan *software* dan *hardware* sangat memungkinkan

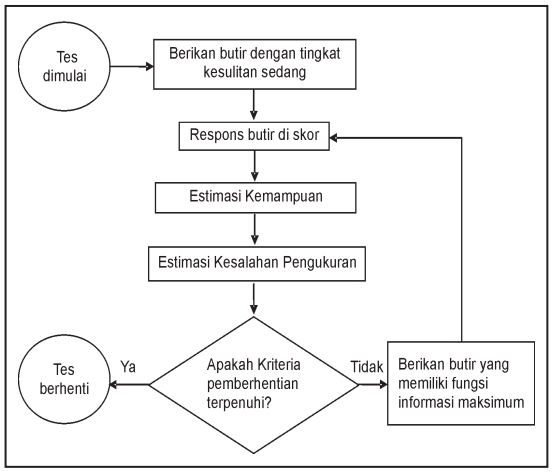
untuk menerapkan dan mengembangkan tes adaptif terkomputerisasi atau lebih populer dikenal dengan *Computerized Adaptive Testing* (CAT) (Lord, 1980; Wainer, 1990).

Pada CAT, komputer diatur untuk menyeleksi, memberikan butir soal, dan menyekor jawaban peserta. Selanjutnya komputer memilih butir soal baru untuk diberikan lagi kepada peserta. Butir soal yang diberikan adalah butir soal yang memberikan informasi tertinggi untuk peserta berdasarkan jawaban butir soal sebelumnya. Proses ini berlanjut terus sampai aturan pemberhentian telah tercapai. Melalui proses ini umumnya peserta tes akan menerima butir soal yang sesuai dengan kemampuan mereka dan menghindari butir soal yang terlalu sulit atau terlalu mudah. Ini berarti tes adaptif lebih efisien dibandingkan dengan tes konvensional.

Efisiensi CAT didukung oleh beberapa penelitian. McBride & Martin (1983) menyimpulkan bahwa untuk mencapai tingkat reliabilitas yang sama, pada tes konvensional masih memerlukan jumlah butir sebanyak 2,57 kali jumlah butir pada tes adaptif. Eignor, et al. (1993) juga menyimpulkan bahwa dengan rancangan tes adaptif hanya memerlukan panjang tes kurang lebih separoh dari panjang perangkat *paper and pencil test* pada tingkat presisi pengukuran yang sama. Weiss (2004) menyimpulkan bahwa tes adaptif juga efisien dan efektif untuk pengukuran di bidang konsultasi dan pendidikan. Santoso (2009) menyimpulkan hal yang senada bahwa rancangan tes adaptif efisien dan efektif untuk mengukur hasil belajar mahasiswa Universitas Terbuka.

Namun, penerapan metode *maximum likelihood* untuk mengestimasi kemampuan peserta tes dan kriteria fungsi informasi maksimum untuk menampilkan butir soal berikutnya pada tes adaptif mengakibatkan butir soal yang diberikan kepada peserta tes mudah dikenali, Hal itu terjadi khususnya pada urutan butir soal awal yang diberikan kepada peserta tes. Oleh karena itu diperlukan strategi atau modifikasi pada algoritma tes adaptif untuk mengurangi butir yang mudah dikenali karena butir soal itu sering dimunculkan atau dikenal sebagai *item exposure*. Jika muncul masalah *item exposure* maka walaupun rancangan tes adaptif ini lebih efisien dan reliabel tetapi keamanan tes tidak terjamin.

Beberapa peneliti telah mengembangkan metode pemilihan butir untuk mengontrol *item exposure* (Stocking, 1993: Stocking & Lewis, 1998: Chang & Ying, 1999). Kit & Chang (2001) juga mengajurkan bahwa untuk mencegah *item exposure* dengan mengaplikasikan kriteria pemilihan butir soal dengan tingkat daya beda rendah di awal tes adaptif, sedangkan Kingsbury & Zara (1989) menganjurkan dengan menerapkan metode randomisasi atau pengacakan pada pemilihan butir soal pada tes adaptif.



Gambar 1. Bagan Alur Pengujian Algoritma Tes Adaptif

Tujuan penelitian ini adalah untuk mengetahui pengaruh randomisasi atau pengacakan butir soal terhadap estimasi kemampuan peserta dan panjang tes pada pemilihan butir soal awal pada rancangan algoritma tes adaptif. Gambar 1 adalah bagan alur pengujian algoritma tes adaptif. Berdasarkan Gambar 1, tes dimulai dengan

memilih butir soal awal dari bank soal dengan tingkat kesukaran sedang. Berikutnya respons terhadap butir diskor. Kemudian diestimasi (sementara) tingkat kemampuan peserta. Selanjutnya dicari nilai fungsi informasi butir pada tingkat kemampuan peserta yang telah diperoleh dan dihitung pula estimasi kesalahan baku pengukurannya. Kemudian dipilih lagi butir yang memiliki nilai fungsi informasi tertinggi atau yang mengurangi kesalahan pengukuran terbesar. Begitu seterusnya sampai tes dihentikan jika kriteria pemberhentian terpenuhi.

METODE

Penelitian ini dilakukan dengan studi simulasi. Bank soal untuk keperluan simulasi sebanyak 250 butir soal ideal yang diperoleh dari pembangkitan data menggunakan *software* WinGen *versi* 2 (Han & Hambleton, 2007). Pembangkitan data didasarkan pada model IRT 3 parameter. Pada model ini, peluang seseorang yang berkemampuan (theta) tertentu menjawab butir soal dengan benar bergantung pada tiga parameter butir soal, yaitu tingkat kesukaran, daya beda, dan faktor *guessing* (Hambleton, Swaminathan, & Rogers, 1991). Butir soal ideal dalam Bank soal bangkitan mempunyai spesifikasi: daya beda butir antara 0,4 sampai 2, tingkat kesukaran butir soal antara -3 sampai +3, dan faktor *guessing* antara 0 sampai 0,30.

Selanjutnya, dua rancangan tes adaptif dikembangkan yaitu rancangan tes adaptif murni (tidak diacak) dan yang diacak. Prosedur simulasi untuk rancangan tes adaptif murni maupun rancangan yang diacak berdasarkan pada 2100 simuli peserta tes yang disimulasikan, yang mewakili 100 simuli peserta tes untuk setiap 21 titik skala tingkat kemampuan, theta (θ) dari -3,0 sampai +3,0 dengan kenaikan 0,3.

Langkah simulasi untuk rancangan tes adaptif murni sebagai berikut. *Pertama*, untuk tingkat kemampuan peserta tes, θ tertentu, tes adaptif diberikan. Berdasarkan metode pemilihan butir awal, satu butir soal dipilih dan diberikan. Peluang peserta tes menjawab benar pada butir soal ke-*i*, $P_i(\theta)$ dihitung. Untuk membangkitkan jawaban atau respons dari peserta tes, nilai $P_i(\theta)$ dibandingkan dengan perubah acak x yang diambil dari sebaran uniform [0,1]. Jika x kurang dari $P_i(\theta)$ maka respons diskor 1, sebaliknya jika x lebih dari atau sama dengan $P_i(\theta)$ maka respons diskor 0. Berdasarkan respons dan parameter butir soal selanjutnya kemampuan peserta tes, θ diestimasi. Estimasi theta dan butir soal yang diberikan dicatat untuk dianalisis lebih lanjut. *Kedua*, berdasarkan metode pemilihan butir soal, diberikan butir soal berikutnya untuk peserta tes, θ tersebut sampai mencapai tingkat kesalahan baku pengukuran (*standard error of measurement, SEM*) sebesar 0,30. *Ketiga*, langkah 1 dan 2 diulang untuk seluruh 2100 simuli peserta tes. *Keempat*, banyaknya butir soal dan estimasi tingkat kemampuan dicatat untuk dianalisis.

Metode pemilihan butir soal awal menggunakan tingkat kesukaran sedang yaitu dimulai dengan rentang antara -0,50 sampai 0,50 yang dipilih secara acak. Metode pendugaan tingkat kemampuan menggunakan *maximum likelihood estimation* (Baker, 1992), namun ketika pola respons belum berpola pendugaan tingkat kemampuan menggunakan metode *step size* berukuran 0,5 (Dodd, 1990). Metode pemilihan butir soal berikutnya menggunakan kriteria fungsi informasi maksimum yaitu butir soal yang mempunyai nilai fungsi informasi terbesar pada kemampuan tertentu dipilih untuk diberikan pada peserta tes.

Langkah simulasi untuk rancangan tes adaptif yang diacak prinsipnya sama dengan rancangan tes adaptif murni bedanya pada pemilihan butir kedua dipilih satu butir secara acak dari 10 butir yang memiliki fungsi informasi 10-terbesar, sedangkan untuk butir soal ketiga dipilih satu butir secara acak dari 5 butir soal yang memiliki fungsi informasi 5-terbesar. Selanjutnya untuk butir soal

keempat dan seterusnya kriteria pemilihan butir soal berikutnya kembali ke kriteria fungsi informasi maksimum yang tidak diacak.

Pada penelitian ini, kriteria pemberhentian tes yang digunakan adalah tes dihentikan jika nilai estimasi kesalahan baku pengukuran (*standard error of measurement,* SEM) sudah mencapai 0,30. Nilai SEM sebesar 0,30 ini setara dengan reliabilitas sebesar 0,91 pada tes konvensional dengan *paper and pencil test* (Thissen, 1990).

HASIL PENELITIAN DAN PEMBAHASAN

Ringkasan statistik 250 butir soal yang digunakan sebagai bank soal untuk keperluan simulasi disajikan pada Tabel 1.

Berikut dipaparkan contoh hasil simulasi rancangan tes adaptif murni atau rancangan yang tidak dirandom. Misalkan untuk theta = 0 yang diambil secara acak. Berdasarkan hasil simulasi, peserta ini sudah dapat diestimasi dengan butir soal sebanyak 9 butir. Nomor induk soal (NIS), parameter butir soal, pola respons untuk setiap urutan butir soal yang ditampilkan serta estimasi theta, kesalahan baku pengukuran dan nilai fungsi informasi disajikan pada Tabel 2.

Berdasarkan Tabel 2 terlihat bahwa butir pertama yang terpilih adalah butir soal dengan NIS 214, memiliki tingkat kesukaran, b = -0,248, artinya ini sesuai dengan kriteria yang diterapkan pada algoritma desain CAT murni bahwa butir soal awal yang dipilih adalah butir dengan tingkat kesukaran sedang, yang dipilih secara acak pada rentang tingkat kesukaran sedang (-0,5 sampai +0,5).

Berdasarkan Tabel 2 terlihat pula bahwa butir soal ini direspons 1, artinya dijawab benar, selanjutnya karena benar maka ditampilkan lagi butir soal dengan NIS 250. Butir soal dengan NIS 250 ini dipilih karena memiliki fungsi informasi terbesar pada theta sebesar 0,5, yaitu sebesar 0,7940. Hal ini juga telah sesuai dengan kriteria pemilihan butir soal berikutnya yang diterapkan pada algoritma CAT murni yang menggunakan kriteria *step size* sebesar 0,5. Berdasarkan kriteria *step size* ini maka ketika butir soal pertama dijawab benar maka butir soal kedua dipilih adalah butir soal yang mampu memberikan informasi maksimum bagi peserta dengan kemampuan pada tingkat 0,5, sebaliknya jika butir pertama dijawab salah maka butir soal kedua dipilih adalah butir soal yang memberikan informasi maksimum bagi peserta dengan kemampuan pada tingkatan -0,5. Pada butir soal pertama ini, kesalahan baku estimasi atau kesalahan pengukuran belum bisa ditentukan karena belum ada pola respons.

Selanjutnya ketika butir soal kedua direspons salah, maka pemilihan butir soal ketiga sudah didasarkan pada hasil pengestimasian theta. Hal ini karena metode MLE yang diterapkan pada algoritma desain tes adaptif murni akan berproses setelah respons sudah berpola (minimal ada satu benar atau satu salah). Berdasarkan metode MLE setelah menjawab butir soal nomor urut 1 benar dan nomor urut 2 salah, maka berdasarkan metode MLE kemampuan peserta ini diestimasi sebesar -0,1435, dan kesalahan pengukuran sudah dapat dihitung, yaitu sebesar 0,7099, dan karena kesalahan baku pengukuran belum mencapai 0,30 maka tes masih berlanjut.

Berdasarkan nilai fungsi informasi maksimum, maka butir soal ketiga yang dipilih adalah butir soal dengan NIS 83. Butir soal ini terpilih karena memiliki nilai fungsi informasi terbesar diantara butir-butir soal lainnya di bank soal untuk theta sebesar -0,1435. Seperti terlihat pada Tabel 2, nilai fungsi informasi butir soal ini sebesar 1,1902. Selanjutnya butir soal ketiga ini direspons, kemampuan dan kesalahan baku pengukuran diestimasi kembali, kemudian butir soal keempat dipilih, direspons, kemampuan diestimasi ulang, begitu seterusnya sampai tes dihentikan pada butir soal ke-9 karena pada butir ke-9 kesalahan baku pengukurannya

Tabel 1. Ringkasan Statistik Parameter Butir Soal pada Bank Soal

Parameter	Mean	Std-deviasi	Min.	Maks.
Daya beda	1,261	0,448	0,411	1,988
Tingkat kesukaran	0,060	1,696	-2,967	2,965
Guessing	0,153	0,088	0,004	0,297

Tabel 2. Nomor Induk Soal, Pola Respons, Estimasi Theta, SEM, dan Nilai Fungsi Informasi

No.Urut	1	2	3	4	5	6	7	8	9
N.I.S	214	250	83	175	51	139	180	105	96
a	1.38	1.924	1.93	1.719	1.976	1.695	1.857	1.419	1.535
b	-0.248	0.311	-0.064	0.589	-0.378	-0.119	0.551	-0.009	-0.402
c	0.284	0.061	0.082	0.022	0.126	0.194	0.087	0.118	0.108
Respons	1	0	1	0	1	1	0	0	0
Theta	0.5	-0.1435	0.1835	0.0778	0.1440	0.2207	0.1434	0.0221	-0.1404
SEM		0.7099	0.4991	0.4405	0.3963	0.3626	0.3375	0.3173	0.2986
Info	0.7940	1.1902	2.0302	1.1390	1.2144	1.2383	1.1737	1.1549	1.2843

Keterangan: Respons 1 = benar; 0 = salah

telah mencapai 0,30 dengan estimasi theta sebesar -0,1404.

Berdasarkan hasil simulasi rancangan tes adaptif murni atau yang tidak dirandom, menghasilkan variasi NIS diurutan butir soal pertama sebanyak 42 NIS, diurutan kedua ada 2 NIS yang sering dimunculkan, yaitu NIS 250 dan NIS 51. NIS 250 dimunculkan jika butir soal pertama direspons benar dan NIS 51 jika butir soal pertama diresponssalah. Diurutan ketiga ada 8 NIS yang dimunculkan yaitu NIS 165, 83, 215, 51, 139, 180, 250 dan NIS 194. NIS 215 dimunculkan jika pola responsnya: benar-benar. NIS 180 atau 83 atau 139 atau 51 dimunculkan jika pola responsnya: benar-salah. NIS 83 atau 194 atau 250 dimunculkan jika pola responsnya: salah-benar. NIS 165 dimunculkan jika pola responsnya: salah-salah. Diurutan keempat sampai urutan keduabelas, variasi NIS yang dimunculkan bervariasi antara 16 sampai 55.

Variasi butir soal yang dimunculkan di awal tes pada rancangan tes adaptif yang tidak diacak menyebabkan butir soal mudah dikenali atau pada tes tersebut muncul masalah *item exposure*, jika ini terjadi maka keamanan tes tidak terjamin.

Selanjutnya, berdasarkan hasil simulasi rancangan tes adaptif yang diacak, diurutan kedua terdapat variasi NIS sebanyak 20. Diurutan ketiga terdapat variasi NIS sebanyak 19. Diurutan keempat sebanyak 22, diurutan kelima 31, diurutan keenam 39, diurutan ketujuh 47, diurutan kedelapan 53 sedangkan diurutan kesembilan terdapat variasi NIS sebanyak 60.

Berdasarkan hasil simulasi, diperoleh hasil estimasi tingkat kemampuan untuk rancangan tes adaptif yang tidak diacak dan yang diacak untuk setiap tingkat kemampuan yang disimulasikan disajikan pada Tabel 3.

Selanjutnya, berdasarkan hasil analisis menggunakan uji-*t* menunjukkan bahwa secara statistik estimasi tingkat kemampuan peserta tes dengan rancangan yang diacak dan rancangan yang tidak diacak tidak berbeda nyata (nilai-*p* = 0,306) atau dengan kata lain rancangan tes adaptif yang diacak tidak berpengaruh terhadap estimasi kemampuan

peserta tes. Dengan demikian, estimasi kemampuan dari rancangan tes adaptif yang diacak tetap reliabel dengan keamanan tes lebih dijamin dibandingkan dengan rancangan yang tidak diacak.

Tabel 3. Estimasi Tingkat Kemampuan pada 21 Titik Theta yang Disimulasikan

Tingkat Kemampuan (Theta)	Tidak Diacak	Diacak
-3	-2.72409	-2.68222
-2.7	-2.49504	-2.48617
-2.4	-2.27374	-2.23848
-2.1	-2.04551	-1.97797
-1.8	-1.79421	-1.61959
-1.5	-1.39508	-1.42835
-1.2	-1.17839	-1.163
-0.9	-0.8662	-0.91762
-0.6	-0.64271	-0.56717
-0.3	-0.25213	-0.29552
0	0.05866	0.00591
0.3	0.26301	0.35268
0.6	0.64513	0.63706
0.9	0.91005	0.9397
1.2	1.1668	1.19032
1.5	1.49895	1.49554
1.8	1.79666	1.79793
2.1	2.10319	2.03381
2.4	2.43403	2.39569
2.7	2.63434	2.63036
3	2.83504	2.85056

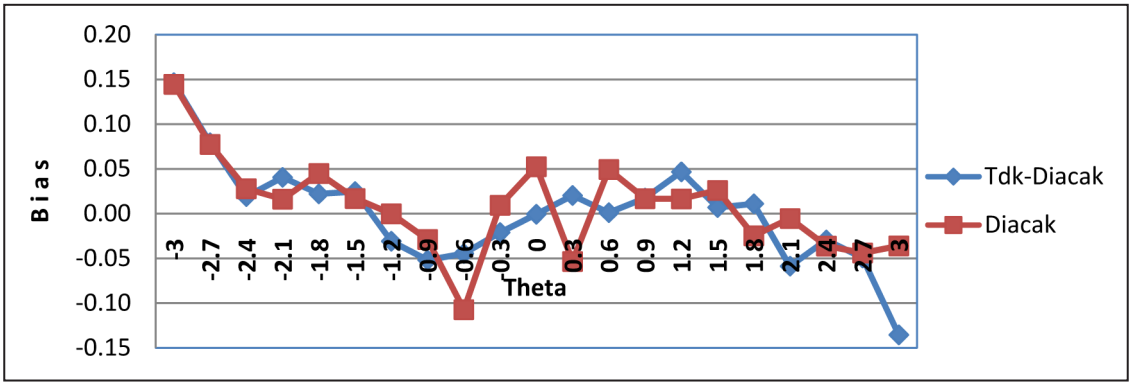
Berdasarkan hasil simulasi diperoleh bahwa panjang tes (banyaknya butir soal yang diperlukan) untuk rancangan tes adaptif yang tidak diacak dan yang diacak untuk setiap tingkat kemampuan yang disimulasikan disajikan pada Tabel 4. Dari Tabel 4 terlihat bahwa banyaknya butir yang diperlukan untuk mengestimasi tingkat kemampuan peserta tes untuk rancangan yang tidak diacak dan yang diacak berkisar antara 8 sampai 12 butir soal. Hal ini menunjukkan bahwa dengan rancangan tes adaptif tingkat peserta tes sudah dapat diestimasi kemampuannya hanya dengan 8 sampai 12 butir soal saja.

Tabel 4. Panjang Tes pada 21 Tingkatan Theta yang Disimulasikan

Theta	Tidak Diacak	Diacak
-3	12	11
-2.7	10	10
-2.4	10	10
-2.1	10	10
-1.8	10	10
-1.5	9	9
-1.2	9	9
-0.9	9	9
-0.6	9	9
-0.3	9	9
0	8	9
0.3	8	8
0.6	8	8
0.9	8	8
1.2	9	9
1.5	10	9
1.8	10	9
2.1	10	10
2.4	10	10
2.7	11	11
3	12	12

Selanjutnya, berdasarkan hasil analisis menunjukkan bahwa panjang tes dengan rancangan yang diacak tidak berbeda secara signifikan dibandingkan dengan rancangan yang tidak diacak (nilai-p = 0,328) atau dengan kata lain rancangan tes adaptif yang diacak tidak berpengaruh terhadap panjang tes. Dengan demikian, panjang tes dari rancangan tes adaptif yang diacak tetap efisien dengan keamanan tes lebih dijamin dibandingkan dengan rancangan yang tidak diacak.

Berdasarkan Gambar 2 terlihat pola bias kedua rancangan ini hampir mirip hanya



Gambar 2. Bias Rancangan Tes Adaptif Tidak Diacak dan Yang Diacak

di tingkat kemampuan, theta -0,6 dan 0,6 yang tidak begitu mirip yaitu dengan selisih bias sebesar 0,07 dan 0,08. Namun demikian, hasil analisis juga menyimpulkan bahwa bias kedua rancangan ini tidak berbeda secara signifikan.

SIMPULAN

Berdasarkan hasil simulasi maka dapat disimpulkan bahwa estimasi tingkat kemampuan pesertadan panjang tes dengan rancangan yang dirandom atau diacak dan rancangan yang tidak diacak tidak berbeda nyata atau dengan kata lain rancangan yang diacak tidak berpengaruh terhadap estimasi tingkat kemampuan dan panjang tes dengan nilai -p = 0,306 dan 0,328.

Dari penelitian ini, maka desain algoritma dengan randomisasi atau pengacakan lebih disarankan untuk diterapkan pada algoritma tes adaptif karena tanpa mengurangi tingkat efisiensi dan presisi pengukuran, butir soal pada urutan awal yang diberikan kepada peserta tes lebih bervariasi sehingga dapat meningkatkan keamanan tes.

DAFTAR PUSTAKA

Baker, F.B. 1992. *Item Response Theory: Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.

Chang, H., & Ying, Z. 1999. "A Stratified Multistage Computerized Adaptive Testing". *Applied Psychological Measurement*, 23, 211–222.

Dodd, B.G. 1990. "The Effect of Item Selection Procedure and Stepsize on Computerized Adaptive Attitude Measurement Using The Rating Scale Model". *Applied Psychological Measurement*, 4, 355 – 366.

Eignor, D.R., Stocking, M.L., Way, W.D., et al. 1993. "Case Studies in Computer Adaptive Test Design through Simulation" *Research Report*. Pg 93 – 56. Princeton, NJ: Educational Testing Service.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. 1991. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications, Inc.

Han, K.T., & Hambleton, R.K. 2007. *User's Manual for Wingen: Windows Software that Generates IRT Model Parameters and Item Responses*. Amherst, MA: University of Massachusetts.

Kingsbury, G.G., & Zara, A. R. 1989. "Procedures for Selecting Items for Computerized Adaptive Test". *Applied Measurement in Education*, 4, 359 – 375.

Kit, T.H., & Chang, H. H. 2001. "Item Selection in Computerized Adaptive Testing: Should More Discriminating Item be Used First?". *Journal of Educational Measurement*, 3, 249 – 266.

Lord, F. M. 1980. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale NJ: Lawrence Erlbaum Associates.

McBride, J.R., & Martin, J.T. 1983. "Reliability and Validity of Adaptive Ability Tests in A Military Setting",dalam D.J. Weiss, (Ed), *New Horizons in Testing*, (pp.223 – 236). New York, NY: Academic Press.

Santoso, A. 2009. "Pengembangan Computerized Adaptive Testing untuk Mengukur Hasil Belajar Mahasiswa Universitas Terbuka". *Disertasi*. Yogyakarta:Universitas Negeri Yogyakarta.

Stocking, M.L. 1993. "Controlling Item Exposure Rates in A Realistic Adaptive Testing Program". *Research Report* 93-2. Princenton, NJ:Educational Testing Service.

Stocking, M.L., & Lewis, C. 1998. "Controlling Item Exposure Conditional on Ability in Computerized Adaptive Testing". *Journal of Educational and Behavioral Statistics*, 1, 57 – 75.

Thissen, D. 1990. "Reliability and Measurement Precision", dalam H. Wainer (Eds.), *Computerized Adaptive Testing: A Primer* (2nd ed., pp. 161–186). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wainer, H. 1990. *Computerized Adaptive Testing: A Primer* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Weiss, D.J. 2004. "Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education". *Measurement and Evaluation in Counseling and Development*, 2, 70 - 84.