

PEMILIHAN BUTIR SOAL PADA RANCANGAN TES ADAPTIF BERDASARKAN *EFFICIENCY BALANCED INFORMATION*

Agus Santoso (aguss@ut.ac.id)
Jurusan Statistika FMIPA Universitas Terbuka

ABSTRAK

Salah satu metode pemilihan butir soal yang populer digunakan dalam rancangan tes adaptif adalah metode Informasi Maksimum. Melalui metode ini, butir soal yang memiliki informasi maksimum pada tingkat kemampuan tertentu akan dipilih dan diberikan kepada peserta tes. Namun kelemahan dari metode ini adalah kurang akurat dalam mengestimasi tingkat kemampuan peserta pada awal tes dan memiliki kecenderungan untuk memilih butir dengan nilai daya pembeda parameter butir yang tinggi dibandingkan butir dengan nilai parameter daya pembeda yang rendah, sehingga menimbulkan masalah pemeliharaan butir soal dalam bank soal. Karena itu dicari cara lain untuk mengatasi masalah tersebut. Penelitian ini bertujuan untuk mengetahui performa hasil estimasi dari penerapan metode *Efficiency Balanced Information (EBI)* pada rancangan tes adaptif. Penelitian ini dilakukan melalui studi simulasi dalam setting penyelenggaraan ujian akhir semester Universitas Terbuka. Bank soal untuk keperluan simulasi dibangkitkan berdasarkan model *Item Response Theory* 3 parameter. Sebanyak 900 butir soal dalam bank soal bangkitan dengan spesifikasi parameter butir yang ideal. Dua kriteria pemilihan butir soal yang disimulasikan yaitu Informasi Maksimum dan *EBI* Maksimum yang juga dirancang agar memenuhi keseimbangan isi. Hal ini agar menjamin bahwa algoritma yang dihasilkan sesuai dengan pembelajaran modular yang diterapkan UT, artinya butir soal setiap modul secara proporsional terwakili dan sesuai kisi-kisi. Aturan pemberhentian tes menggunakan kesalahan baku estimasi (*standard error of estimation=SEE*) sebesar 0,3. Hasil penelitian menyimpulkan bahwa algoritma rancangan tes adaptif dengan kriteria *EBI* menghasilkan performa hasil estimasi kemampuan peserta yang lebih akurat dibandingkan kriteria Informasi Maksimum. Hal ini ditunjukkan oleh nilai bias dan simpangan baku pengukuran yang lebih kecil dibandingkan kriteria Informasi Maksimum. Kelebihan lain dari penerapan kriteria *EBI* Maksimum adalah kebermanfaatan bank soal lebih optimal karena butir-butir soal dengan tingkat daya beda rendah juga dimunculkan khususnya pada awal tes. Sedangkan kriteria Informasi Maksimum walaupun lebih efisien dari sisi panjang tes tetapi kurang optimal dalam memanfaatkan bank soal.

Kata kunci: *EBI*, informasi maksimum, tes adaptif

ABSTRACT

One of the most popular item selection methods in the design of adaptive testing is Maximum Information Method. This method provides the items with the maximum information at a certain level of selecting ability. The lack of this method is rarely accurate in estimating the level of ability of the examinees at the beginning of the test and tends to select items with higher discrimination-parameter value than items with lower discrimination-parameter. It creates problem in maintaining item bank. Therefore, another method should be found. The objective of the study is to determine the performance of the application of the estimation method Balanced Efficiency Information (EBI) on the design of an adaptive test.

This research was carried out through a simulation study in the setting of organizing the Open University final exams. Item bank for the purposes of simulation models based on 3 parameters Item Response Theory was constructed a total of 900 items were generated base on the ideal parameter of the item specifications. Two selection criteria items were simulated, namely Information Maximum and EBI Maximum. Those two criteria were also designed to meet the content balancing. This is to ensure that the algorithm is appropriate with the applicable UT modular learning, meaning items of each module were proportionally represented and suited the blueprint. The tests will be stopped at when the standard error of estimate (standard error of estimation = SEE) is 0.3. The study summarized that the algorithm of EBI was more accurate than the Maximum Information criteria in estimating performance capabilities of participants. This is indicated by the value of the bias and the standard deviation of EBI is smaller than Maximum Information criteria's. Another advantage of the application of the EBI Maximum is optimally utilizing of the item bank. The items with low level of the discrimination-parameter will also be chosen at the beginning of the test. The maximum information criterion is more efficient in terms of test length but less optimally of the item bank utilization.

Keywords: adaptive test, maximum information, EBI

Penerapan berbagai metode pemilihan butir soal dalam rancangan tes adaptif merupakan hal yang menarik untuk dikaji dan diteliti. Penelitian mengenai efektivitas dan efisiensi berbagai metode pemilihan butir soal dalam rancangan tes adaptif telah dilakukan sejak sistem tes adaptif ini diperkenalkan sekitar tahun 1980 an. Metode pemilihan butir yang paling populer dan sering digunakan adalah metode Informasi Maksimum. Berdasarkan metode ini butir soal yang memberikan informasi maksimum atau yang mengurangi kesalahan pengukuran pada tingkat kemampuan tertentu akan terpilih dan diberikan kepada peserta tes. Pada penyelenggaraan tes pilihan ganda yang butir soalnya dibangun melalui model *item response theory (IRT)* 3 parameter, pemilihan butir soal berdasarkan metode Informasi Maksimum akan memilih butir soal ke-*i* yang memaksimalkan fungsi berikut.

$$I_i(\hat{\theta}_j) = \frac{2,89a_i^2(1-c_i)}{\left[(c_i + \exp(1,702.a_i(\hat{\theta}_j - b_i))) \left[1 + \exp(-1,702.a_i(\hat{\theta}_j - b_i)) \right]^2 \right]} \quad (1)$$

dengan a_i , b_i , dan c_i masing-masing adalah daya pembeda, tingkat kesukaran, dan faktor tebakan butir ke-*i*.

Penerapan metode Informasi Maksimum akan menghasilkan penyelenggaraan tes yang efektif karena menghasilkan informasi tes paling maksimum untuk setiap individu peserta tes. Namun demikian, penerapan metode ini sering kurang akurat dalam mengestimasi tingkat kemampuan peserta pada awal tes serta kurang optimal dalam memanfaatkan bank soal. Untuk mengatasi kelemahan metode Informasi Maksimum, beberapa metode pemilihan butir soal telah dikembangkan diantaranya adalah metode *Efficiency Balanced Information = EBI* (Han, 2012). Berdasarkan metode *EBI* Maksimum, butir soal yang memiliki nilai *EBI* terbesar akan dipilih dan diberikan kepada peserta tes.

$$EBI_i(\hat{\theta}_j) = \left(1 + \frac{1}{I_i(\theta_i^*)} \right) \int_{\theta_j - 2\varepsilon_j}^{\theta_j + 2\varepsilon_j} I_i(\theta) d\theta \quad (2)$$

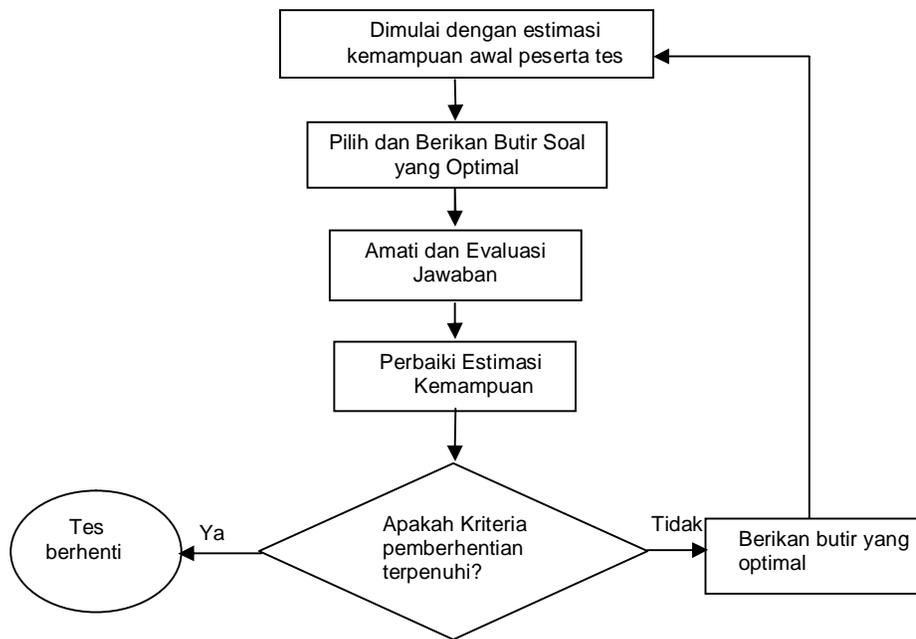
$$\text{dengan } \theta_i^* = b_i + \frac{1}{Da_i} \ln \left(\frac{1 + \sqrt{1 + 8c_i}}{2} \right)$$

dan $I_i(\theta_i^*)$ = fungsi informasi butir pada saat tingkat kemampuan (theta) yang disesuaikan dengan tingkat kesukaran butir.

Dalam prosedur *EBI*, estimasi tingkat kemampuan peserta akan disesuaikan dengan tingkat kesukaran butir dan di tetapkan pada $\pm 2\varepsilon$ atau 2 *standard error of estimation (SEE)* dari estimasi kemampuan setelah butir ke-*j* diberikan kepada peserta tes.

Artikel ini membahas performa hasil estimasi tingkat kemampuan peserta tes pada algoritma rancangan tes adaptif dengan kriteria Informasi Maksimum dan *EBI* Maksimum. Hasil dari kedua rancangan ini lalu dibandingkan.

Menurut Wainer (1990) pengembangan algoritma tes adaptif memerlukan perhatian pada empat komponen, yaitu: bank soal, prosedur pemilihan butir soal, pendugaan kemampuan, dan aturan pemberhentian. Berikut adalah alur pengujian algoritma tes adaptif.



Gambar 1. Bagan alur pengujian algoritma tes adaptif

Ketika tes adaptif dimulai, belum ada butir soal yang diberikan pada peserta tes, belum ada respons atau jawaban yang diberikan oleh peserta tes sehingga tes adaptif dimulai dengan mengestimasi tingkat kemampuan awal peserta tes (Green, et al. 1984; Vispoel, 1999; Mills, 1999). Jika tidak ada informasi tingkat kemampuan peserta, maka tingkat kemampuan awal peserta tes umumnya di estimasi pada rentang tingkat kemampuan sedang. Selanjutnya komputer memilih dan

menyajikan atau memberikan butir soal yang optimal dengan estimasi tingkat kemampuan awal peserta tes tersebut.

Setelah memperoleh jawaban peserta tes terhadap butir soal yang diberikan, selanjutnya komputer menskor jawaban dengan benar atau salah, kemudian memutuskan apakah tes dilanjutkan ataukah tidak. Ada dua langkah proses untuk melanjutkan tes adaptif. Langkah pertama adalah memperbaiki atau mengestimasi tingkat kemampuan peserta tes; langkah kedua adalah bagaimana memilih butir soal berikutnya.

Metode Estimasi Tingkat Kemampuan: *Maximum Likelihood Estimation (MLE)*

Misalkan seorang peserta tes dengan tingkat kemampuan θ menjawab tes yang berisi n butir soal pilihan ganda dengan parameter butir soal diketahui. Peluang bersama dari peserta tes dapat dituliskan sebagai $P(U_1, U_2, \dots, U_n | \theta)$. Selanjutnya jika asumsi independensi lokal diterapkan maka fungsi kemungkinannya dituliskan sebagai berikut

$$L(\theta) = P(U_1 = u_1, U_2 = u_2, \dots, U_n = u_n | \theta) = \prod_{i=1}^n P_i(\theta)^{u_i} (1 - P_i(\theta))^{1-u_i} \quad (3)$$

$$i = 1, 2, \dots, n, \quad -\infty < \theta < \infty$$

$P_i(\theta)$ adalah peluang menjawab benar untuk butir soal ke- i oleh peserta tes yang berkemampuan θ .

Tujuan *MLE* adalah menemukan nilai θ yang memaksimalkan fungsi kemungkinan $L(\theta)$, yang pada prakteknya, sering dilakukan menggunakan prosedur Newton-Raphson (Hambleton dan Swaminathan, 1985; Embretson dan Raise, 2000). Satu masalah dengan penerapan metode *MLE* pada tes adaptif adalah ketidakmampuan fungsi kemungkinan untuk menemukan solusi maksimum ketika peserta tes menjawab semua butir soal dengan benar atau salah. Jalan keluar untuk mengatasi ketidakmampuan ini adalah menggunakan metode *step-size* (Dodd, 1990).

Setelah kemampuan peserta diperkirakan kembali, kemudian komputer mesti memilih butir soal berikutnya. Dua metode yang digunakan untuk memilih butir soal berikutnya, yaitu metode pemilihan menggunakan Informasi butir soal Maksimum (Lord, 1980) dan metode *Efficiency and Balanced Information* (Han, 2012). Kedua metode pemilihan butir soal selanjutnya dikombinasikan dengan keseimbangan konten (*content balance*).

Penelitian ini dilakukan melalui studi simulasi. Bank soal untuk keperluan simulasi dibangun melalui data bangkitan dengan model IRT 3 parameter dengan karakteristik butir-butir soal yang ideal, yaitu tingkat daya beda pada rentang 0,7 sampai 2,0, tingkat kesukaran berdistribusi normal baku, dan faktor tebakan pada rentang 0,0 sampai 0,25. Sebanyak 900 butir soal dijadikan sebagai bank soal untuk keperluan simulasi. Kemampuan peserta tes yang disimulasikan pada 31 titik tingkat kemampuan (*theta*) dari rentang tingkat kemampuan -3,0 sampai +3,0. Setiap titik tingkat kemampuan diulang sebanyak 100 kali.

Rancangan algoritma tes adaptif yang disimulasikan adalah sebagai berikut.

1. Estimasi awal kemampuan peserta dari -0,5 sampai +0,5
2. Pemilihan butir soal menggunakan metode Informasi Maksimum dan *EBI* Maksimum
3. Metode estimasi peserta setelah peserta menjawab butir soal menggunakan metode *Maximum Likelihood*
4. Aturan pemberhentian tes menggunakan batas $SEE = 0,30$.

Nilai *SEE* sebesar ini setara dengan tingkat reliabilitas sebesar 91 persen pada penyelenggaraan tes konvensional menggunakan *paper and pencil test* (Weiner, 1990).

Rancangan algoritma tes adaptif di set sesuai dengan penyelenggaraan Sistem Ujian Online Universitas Terbuka (SUO-UT). Rancangan ini tidak hanya menerapkan metode *EBI* untuk memilih butir soal yang akan diberikan kepada peserta, tetapi juga mengikuti setting pembelajaran modular yang diterapkan UT. Dengan kata lain butir-butir soal yang dipilih dan diberikan harus urut modul sehingga keterwakilan butir soal setiap modul secara proporsional dan sesuai kisi-kisi terpenuhi. Butir soal pertama dipilih dari modul pertama. Setelah sejumlah tertentu dari modul pertama secara proporsional dipilih berdasarkan kisi-kisi, maka soal berikutnya dipilih dari modul kedua, begitu seterusnya sampai butir soal dari modul terakhir dipilih dan dimunculkan. Jika kriteria pemberhentian telah dipenuhi maka tes akan dihentikan, namun jika belum terpenuhi maka tes akan dilanjutkan dengan memunculkan butir soal dari modul pertama lagi dan seterusnya secara berurutan sampai kriteria pemberhentian terpenuhi.

Simulasi dilakukan menggunakan program SAS versi 9. Program ini dipilih karena kelebihanannya dalam mengolah data berstruktur matriks menggunakan prosedur *IML (Interactive Matrix Language)*. Selanjutnya evaluasi penerapan metode pemilihan butir soal dalam rancangan tes adaptif dilakukan terhadap banyaknya butir soal yang diperlukan atau panjang tes, performa hasil estimasi melalui tiga kriteria, yaitu: korelasi, bias, dan kesalahan baku pengukuran (*standard error of measurement = SEM*). Korelasi adalah hubungan tingkat kemampuan sesungguhnya (*true theta, θ*) dengan rata-rata estimasi tingkat kemampuan (*theta estimasi, $\hat{\theta}$*). Semakin besar korelasi semakin dekat rata-rata estimasi tingkat kemampuan terhadap tingkat kemampuan sesungguhnya. Bias adalah selisih antara estimasi tingkat kemampuan dengan kemampuan sesungguhnya, sedangkan kesalahan baku pengukuran (*Standard Error of Measurement = SEM*) mengindikasikan kesalahan acak pada pengestimasian tingkat kemampuan tertentu. Kesalahan ini juga menggambarkan presisi dari pendugaan kemampuan. Evaluasi juga dilakukan untuk mengetahui proporsi penggunaan butir soal berdasarkan dua metode pemilihan butir soal yang disimulasikan.

PEMBAHASAN

Sebanyak 900 butir soal dalam Bank Soal untuk keperluan simulasi dibangkitkan menggunakan *software WinGen* (Han dan Hambleton, 2007). Ringkasan statistika parameter bank soal bangkitan disajikan pada Tabel 1.

Tabel 1. Ringkasan Statistik Parameter Butir Soal pada Bank Soal Bankitan

Parameter	Mean	Std-deviasi	Min.	Maks.
Daya beda (a)	1,325	0,366	0,70	2,00
Tingkat kesukaran (b)	0,032	0,980	-3,27	3,62
Faktor Tebakan atau <i>Guessing</i> (c)	0,127	0,722	0,00	0,25

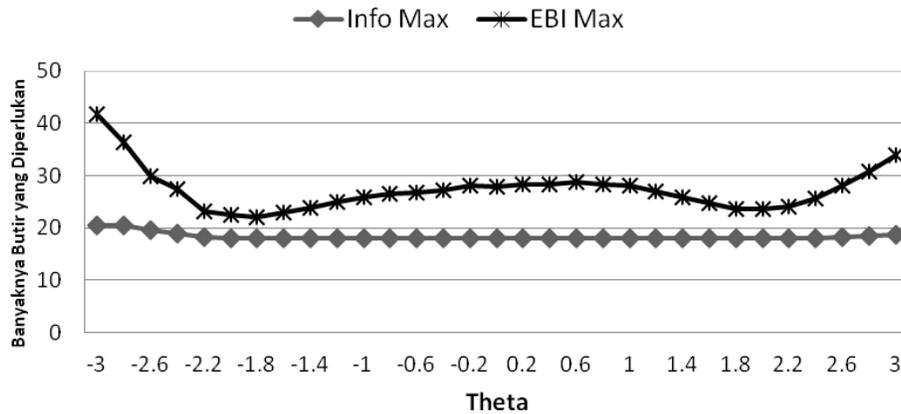
Berdasarkan Tabel 1 terlihat bahwa dari 900 butir soal dari data bangkitan yang dibangun sudah sesuai dengan kriteria parameter butir soal yang ideal. Daya beda butir merentang dari 0,7 sampai 2,0; tingkat kesukaran butir soal menyebar normal baku. Faktor tebakan merentang dari 0,0 sampai 0,25 yang sesuai dengan banyaknya option/pilihan pada perangkat tes SUO-UT yaitu sebanyak 4, sehingga peluang untuk menebak jawaban yang benar adalah 0,25.

Panjang Tes

Panjang tes atau banyaknya butir soal yang diperlukan untuk rancangan tes adaptif dengan metode Informasi Maksimum dan EBI Maksimum pada setiap tingkat kemampuan peserta tes yang disimulasikan disajikan pada Tabel 2 dan Gambar 2.

Tabel 2. Panjang Rancangan Tes Adaptif Kriteria Info Maks dan EBI Maks.

Theta	Info Max	EBI Max
-3	21	42
-2.8	21	36
-2.6	20	30
-2.4	19	27
-2.2	18	23
-2	18	22
-1.8	18	22
-1.6	18	23
-1.4	18	24
-1.2	18	25
-1	18	26
-0.8	18	27
-0.6	18	27
-0.4	18	27
-0.2	18	28
0	18	28
0.2	18	28
0.4	18	28
0.6	18	29
0.8	18	28
1	18	28
1.2	18	27
1.4	18	26
1.6	18	25
1.8	18	24
2	18	24
2.2	18	24
2.4	18	26
2.6	18	28
2.8	18	31
3	19	34

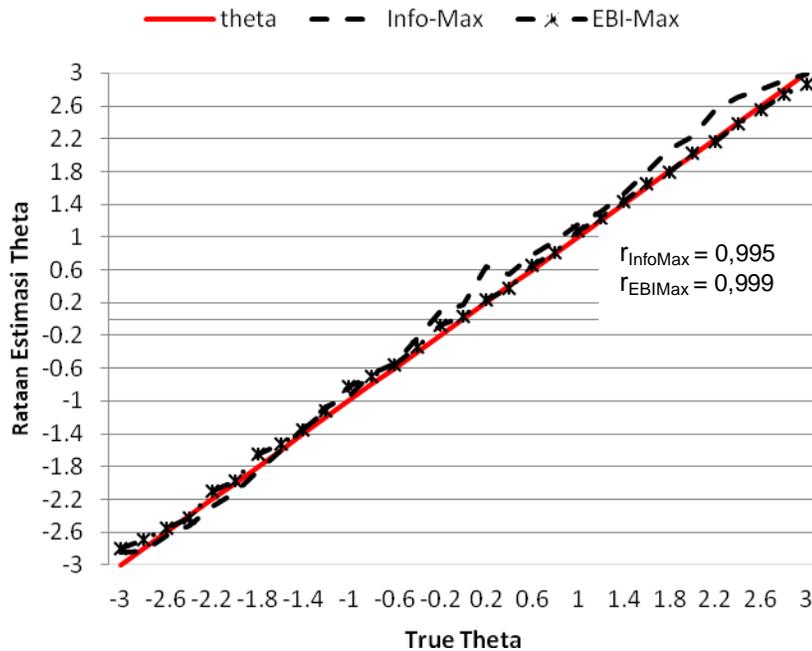


Gambar 2. Panjang tes rancangan tes adaptif metode info Maks dan EBI Maks.

Berdasarkan Tabel 2 dan Gambar 2 terlihat bahwa penerapan metode Informasi Maksimum memerlukan rata-rata sebanyak 18 butir soal, sedangkan untuk kriteria EBI Maksimum memerlukan rata-rata sebanyak 25 butir soal sampai peserta tes diberhentikan tesnya. Dengan demikian dari sisi panjang tes metode Informasi Maksimum lebih efisien dibandingkan metode EBI Maksimum.

Performa Estimasi Tingkat Kemampuan

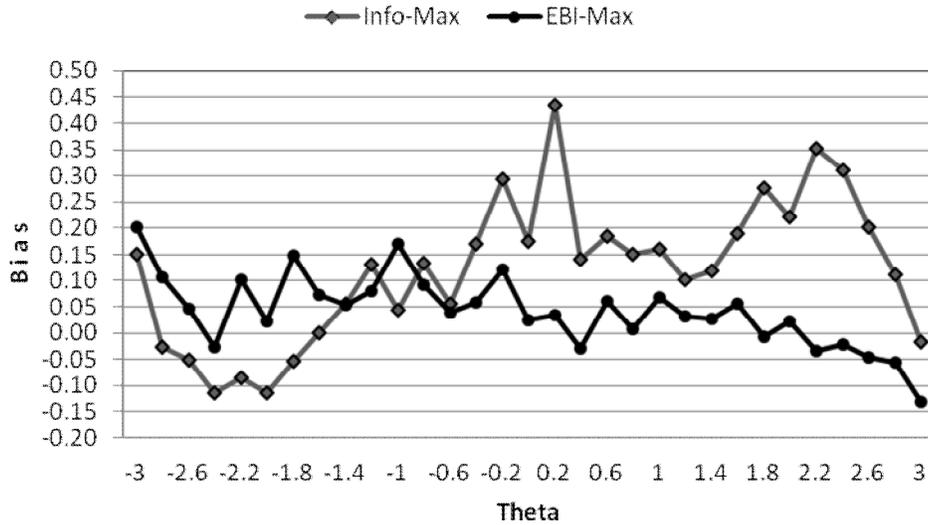
Performa hasil estimasi tingkat kemampuan dapat dilihat melalui nilai korelasi, bias dan kesalahan baku pengukuran. Hubungan antara *true theta* dengan rata-rata estimasi theta dapat dilihat pada Gambar 3.



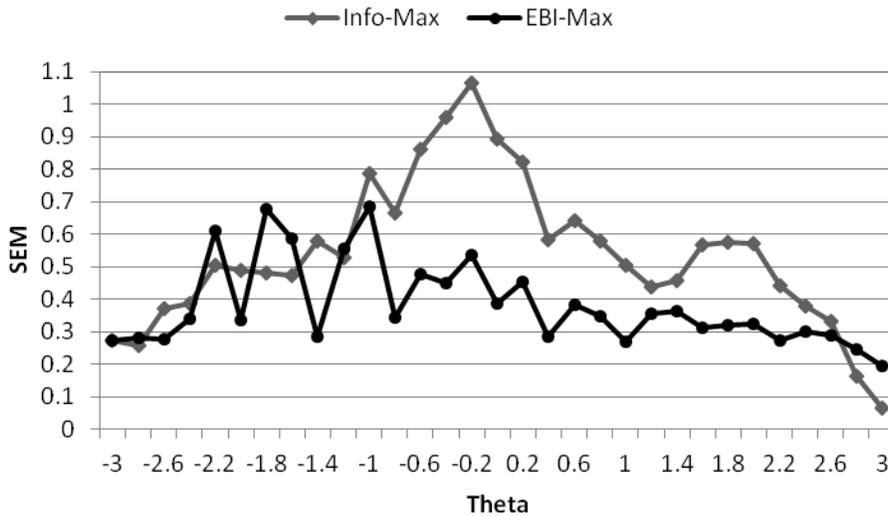
Gambar 3. Korelasi *True Theta* dan Rataan Estimasi Theta

Berdasarkan Gambar 3 terlihat bahwa hubungan antara *true theta* dengan rata-rata estimasi theta dari penerapan metode *EBI* Maksimum lebih dekat dibandingkan metode Informasi Maksimum. Hal ini juga ditunjukkan dari nilai korelasi untuk metode *EBI* Maksimum sebesar 0,999 yang lebih besar dibandingkan metode Informasi Maksimum sebesar 0,995.

Bias dan kesalahan baku pengukuran dari penerapan metode Informasi Maksimum dan *EBI* Maksimum dapat dilihat pada Gambar 4 dan 5.



Gambar 4. Bias info Maks dan *EBI* Maks.

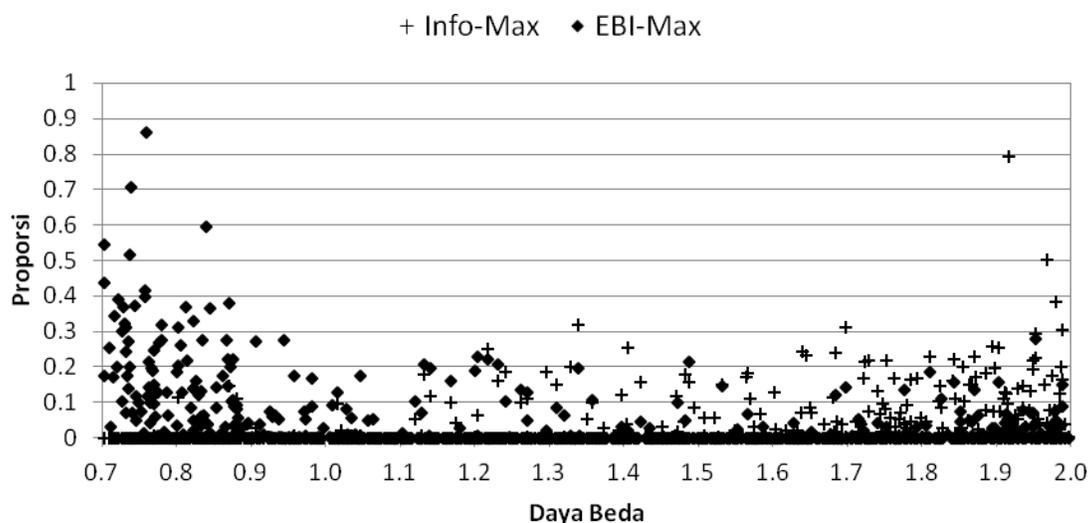


Gambar 5. SEM Info Maks dan *EBI* Maks.

Berdasarkan Gambar 4 dan 5 dapat dilihat bahwa bias dan kesalahan baku pengukuran (*standard error of measurement = SEM*) rancangan tes adaptif menggunakan metode *EBI* Maksimum

lebih rendah dibandingkan metode Informasi Maksimum. Ini berarti metode *EBI* Maksimum lebih akurat dibandingkan metode Informasi Maksimum. Berdasarkan kriteria korelasi, bias, dan *SEM* maka penerapan metode pemilihan butir soal dalam rancangan tes adaptif menggunakan *EBI* Maksimum lebih baik dibandingkan metode Informasi Maksimum.

Frekuensi atau proporsi butir soal yang digunakan disajikan pada gambar 6. Dari Gambar 6 terlihat bahwa berdasarkan daya pembeda butir, proporsi penggunaan butir soal tidak sama untuk dua metode pemilihan butir soal. Butir-butir soal dengan tingkat daya pembeda butir tinggi lebih sering terpilih untuk metode pemilihan butir menggunakan Informasi Maksimum. Sebaliknya untuk metode *EBI*, butir-butir soal dengan daya pembeda rendah lebih sering terpilih dibandingkan butir-butir soal dengan daya pembeda tinggi. Hasil simulasi juga menunjukkan bahwa pada pemilihan butir dengan metode Informasi Maksimum, butir-butir soal dengan daya pembeda tinggi lebih sering dimunculkan pada awal tes sedangkan butir-butir soal dengan daya pembeda rendah dimunculkan pada akhir-akhir tes.



Gambar 6. Proporsi Penggunaan Butir Soal

Selanjutnya metode *EBI* Maksimum menunjukkan hal yang sebaliknya yaitu butir-butir soal dengan daya beda rendah lebih sering dimunculkan pada awal tes sedangkan butir-butir soal dengan daya beda tinggi lebih sering dimunculkan di akhir tes. Dengan demikian metode *EBI* Maksimum lebih memanfaatkan bank soal dibandingkan metode Informasi Maksimum yang jarang memanfaatkan butir-butir soal dengan daya beda rendah. Pertanyaan tentang apakah butir soal pertama dalam rancangan tes adaptif yang dipilih harus yang memiliki daya beda butir tinggi (Veerkamp dan Berger, 1997; Kit dan Chang, 2001) dapat terjawab melalui metode *EBI* Maksimum.

Metode Informasi Maksimum menghasilkan panjang tes lebih sedikit dibandingkan metode *EBI* Maksimum. Hal ini merupakan konsekuensi logis dari metode Informasi Maksimum yang memilih butir soal dengan nilai fungsi informasi yang tertinggi. Nilai Informasi butir soal tertinggi akan diperoleh dari butir soal dengan tingkat daya beda yang tinggi. Semakin sering butir-butir soal dengan daya beda butir yang tinggi terpilih dan diberikan pada awal tes semakin sedikit jumlah butir soal yang diperlukan untuk menghentikan tes. Karena metode Informasi Maksimum sangat dipengaruhi oleh

nilai daya beda, maka butir soal yang dipilih sangat dipengaruhi oleh daya beda butir tanpa menghiraukan kesesuaian dengan tingkat estimasi kemampuan setiap individu peserta tes khususnya di awal tes. Akibatnya penerapan metode Informasi Maksimum kurang akurat dalam mengestimasi kemampuan peserta khususnya di awal tes. Frekuensi atau proporsi butir soal yang dipilih dipengaruhi oleh tingkat daya beda butir dan tidak dipengaruhi oleh tingkat kesukaran butir dan faktor tebakan. Berdasarkan metode Informasi Maksimum, butir-butir soal dengan tingkat daya beda tinggi lebih sering dipilih pada awal tes dan butir soal dengan daya beda rendah jarang dipilih, sedangkan berdasarkan metode *EBI* Maksimum butir-butir soal dengan daya beda rendah sering dipilih di awal tes dan butir-butir soal dengan daya beda tinggi dipilih di akhir. Hal ini menunjukkan bahwa penerapan metode *EBI* lebih mengoptimalkan kebermanfaatan parameter-parameter di bank soal dibandingkan metode Informasi Maksimum. Hal ini juga merupakan konsekuensi logis dari metode Informasi Maksimum yang sangat dipengaruhi oleh tingkat daya beda butir. Kurang optimalnya kebermanfaatan bank soal karena butir-butir soal dengan daya beda rendah jarang terpilih untuk diberikan kepada peserta tes. Hal sebaliknya terjadi pada penerapan metode *EBI* Maksimum. Berdasarkan metode ini butir soal dengan daya beda rendah akan terpilih di awal tes dengan tingkat kemampuan setiap individu peserta tes disesuaikan dengan tingkat kesukaran butir soal. Butir-butir soal dengan daya beda tinggi akan terpilih di akhir tes. Hasil simulasi menunjukkan bahwa penerapan metode *EBI* Maksimum akan menghasilkan estimasi tingkat kemampuan yang akurat dan menjamin kebermanfaatan dari bank soal. Penelitian simulasi ini hasilnya sejalan dengan penelitian Han (2012). Perbedaan penelitian ini dengan penelitian Han adalah pada algoritma rancangan tes adaptif, bank soal, dan software simulasi.

KESIMPULAN

Performa estimasi tingkat kemampuan berdasarkan metode *EBI* Maksimum lebih baik dibandingkan metode Informasi Maksimum. Hal ini ditunjukkan oleh nilai korelasi antara tingkat kemampuan sesungguhnya dengan estimasi tingkat kemampuan yang lebih tinggi pada metode *EBI* Maksimum lebih baik daripada Informasi Maksimum. Hal senada juga ditunjukkan dari nilai bias dan kesalahan baku pengukuran dari metode *EBI* Maksimum yang lebih kecil dibandingkan Informasi Maksimum. Kebermanfaatan butir soal pada bank soal juga lebih optimal untuk metode *EBI* Maksimum. Berdasarkan hasil simulasi maka jika UT akan mengembangkan tes adaptif kedalam SUO-UT sebaiknya menerapkan metode *EBI* untuk pemilihan butir soal, karena di samping menghasilkan performa hasil estimasi yang akurat juga dapat mengoptimalkan penggunaan bank soal.

REFERENSI

- 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.
- Embretson, S.E. & Rouse, S.P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates, Inc.
- Green, B.F., Bock, R.D., Humphreys, L.G., et al. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 4, 347-360.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*, Boston, MA: Kluwer Academic Publishers.

- Han, K.T. (2012). An efficiency balanced information criterion for item selection in computerized adaptive testing. *Journal of Educational Measurement*, 3, 225-246.
- 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.
- 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.
- Mills, C.N. (1999). Development and introduction of a computerized adaptive graduate record examinations general test. Dalam F. Drasgow & J. B. Olson-Buchanan (Eds), *Innovations in Computerized Assessment* (pp. 117–136). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Veerkamp, W.J.J. & Berger, M.P.F. (1997). Some new item selection criterion for adaptive testing. *Journal of Educational and Behavioral Statistics*, 2, 203-226.
- Vispoel, W.P. (1999). Creating computerized adaptive test of music aptitude: Problem, solutions, and future directions. Dalam F. Drasgow, & J. B. Olson-Buchanan (Eds.), *Innovations in Computerized Assessment* (pp. 151-176). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Wainer, H., et al. (1990). *Computerized adaptive testing: A primer* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.