

# PERBANDINGAN METODE KEMUNGKINAN MAKSIMUM DAN BAYES DALAM MENAKSIR KEMAMPUAN PESERTA TES PADA RANCANGAN TES ADAPTIF

Agus Santoso

Jurusan Statistik FMIPA Universitas Terbuka

email:aguss@ut.ac.id

## ABSTRAK

Penelitian ini bertujuan untuk membandingkan metode Kemungkinan Maksimum (*Maximum Likelihood*,) dan metode Bayes dalam menaksir kemampuan peserta pada rancangan tes adaptif. Penelitian ini dilakukan dengan prosedur simulasi, bank soal yang digunakan untuk simulasi sebanyak 500 butir soal yang dibangkitkan berdasarkan kriteria ideal. Dua algoritma tes adaptif menggunakan metode Kemungkinan Maksimum dan Bayes dikembangkan. Skala kemampuan peserta pada penelitian ini dibuat dari -3 sampai +3 dan di set sebanyak 21 titik kemampuan. Banyaknya butir yang diperlukan untuk mengestimasi kemampuan peserta, korelasi, bias, dan kesalahan pengukuran pada setiap titik kemampuan hasil simulasi dengan kedua metode tersebut dibandingkan.

Hasil penelitian menunjukkan bahwa: 1) banyaknya butir yang diperlukan untuk mengestimasi kemampuan peserta pada setiap titik kemampuan antara kedua metode tersebut adalah sama kecuali untuk peserta berkemampuan ekstrim, 2) korelasi kemampuan sesungguhnya dengan estimasinya untuk kedua metode tersebut adalah sangat kuat artinya kedua metode tersebut akurat dalam mengestimasi kemampuan peserta tes, 3) pola bias dengan metode Kemungkinan Maksimum cenderung lebih rendah dibandingkan metode Bayes, sedangkan pola kesalahan pengukuran menunjukkan hal yang sebaliknya yaitu metode Bayes lebih rendah dibandingkan metode Kemungkinan Maksimum. Keterbatasan dari metode Kemungkinan Maksimum adalah tidak dapat mengestimasi kemampuan peserta dengan pola jawaban benar (salah) semua, sehingga untuk mengestimasi jawaban siswa yang belum berpola digunakan metode *step-size*, sedangkan pada metode Bayes tetap dapat mengestimasi kemampuan peserta walaupun jawaban siswa belum berpola. Oleh karena itu disarankan untuk menggunakan kombinasi keduanya yaitu menggunakan metode Bayes untuk mengestimasi kemampuan pada awal tes dan menggunakan metode Kemungkinan Maksimum untuk mengestimasi akhir (final) kemampuan peserta pada tes adaptif.

**Kata Kunci:** Metode Kemungkinan Maksimum, Metode Bayes, tes adaptif.

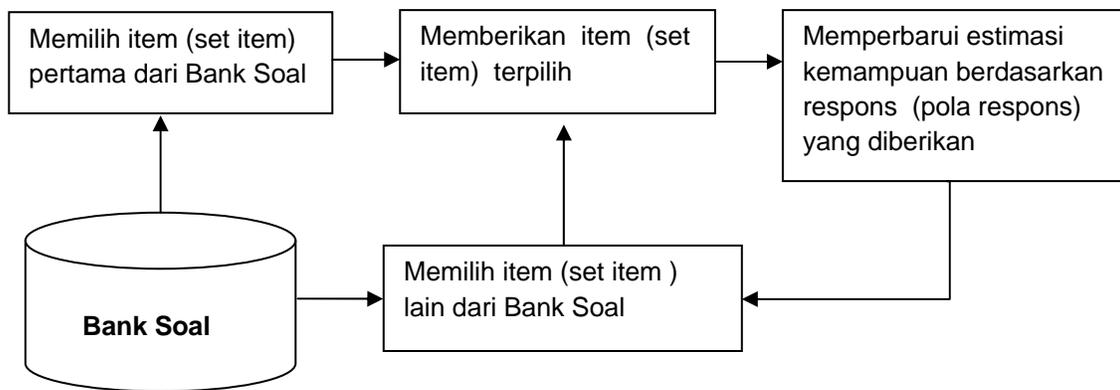
## PENDAHULUAN

Evaluasi dalam pendidikan dilaksanakan untuk memperoleh informasi tentang aspek yang berkaitan dengan pendidikan. Dalam evaluasi pendidikan diperlukan suatu instrument/alat. Alat yang digunakan untuk melakukan evaluasi salah satunya adalah tes. Penyelenggaraan tes kepada peserta tes dapat dilakukan dengan berbagai cara, mulai dari cara konvensional, yaitu dengan menggunakan kertas (*paper-based test*, PBT), hingga pemanfaatan teknologi, seperti tes menggunakan komputer (*computer-based test*, CBT).

*Computerized Adaptive Testing* (CAT) merupakan generasi kedua dari penggunaan komputer untuk pengujian (Bunderson, Inouye, & Olsen, 1989: 383). *Adaptive* berarti bahwa butir soal (tes) yang diberikan disesuaikan dengan tingkat kemampuan individu peserta tes. Pada CAT yang berbasiskan *item response theory* (IRT), komputer tidak hanya sekedar memindahkan butir soal ke dalam komputer, tetapi komputer diatur untuk menyeleksi dan menyajikan butir soal (tes) menurut perkiraan tingkat kemampuan peserta

tes. Individu peserta tes yang memiliki tingkat kemampuan tinggi akan mendapatkan butir soal yang lebih sulit dibandingkan dengan individu yang memiliki tingkat kemampuan rendah. Sebaliknya individu peserta tes yang memiliki tingkat kemampuan rendah akan mendapatkan butir soal yang lebih mudah dibandingkan dengan individu peserta tes yang memiliki tingkat kemampuan tinggi. Dengan demikian CAT lebih efisien karena dapat mengestimasi kemampuan peserta tes dengan jumlah butir soal yang lebih sedikit dibandingkan *PBT* maupun *CBT* tanpa mengurangi ketepatan pengukuran (Wainer, 1990: 10; Hambleton, Swaminathan, & Rogers, 1991: 146, Weiss & Schleisman, 1999: 130).

Proses *adaptive testing* secara skematik dapat dilihat pada Gambar 1.



Gambar 1. Proses *Adaptive Testing*

Berdasarkan Gambar 1, proses *adaptive testing* dimulai dengan memilih butir soal atau kelompok butir soal pertama dari bank soal. Biasanya butir soal pertama dipilih disesuaikan dengan tingkat kemampuan awal optimal dari populasi peserta tes atau dipilih dengan tingkat kesukaran sedang. Setelah butir soal atau kelompok butir soal dipilih, selanjutnya butir soal diberikan kepada peserta tes. Setelah peserta tes merespons (benar atau salah) butir soal atau kelompok butir soal pertama, kemudian tingkat kemampuan peserta diperbarui atau diestimasi kembali. Selanjutnya, berdasarkan estimasi tingkat kemampuan terbaru, butir soal atau kelompok butir soal yang lain dipilih kembali dari bank soal. Kemudian butir soal atau kelompok butir soal yang lain diberikan lagi kepada peserta tes, begitu seterusnya proses ini berlangsung dan diberhentikan setelah sebanyak butir soal yang ditentukan sudah diberikan atau setelah presisi estimasi tingkat kemampuan atau tingkat kesalahan baku pengukuran yang diinginkan telah dicapai.

Dalam mengaplikasikan sebuah tes ke dalam CAT perlu diperhatikan beberapa komponen. Menurut Wainer (1990) secara umum sistem CAT memiliki empat komponen, yaitu: bank soal (*item bank*), prosedur pemilihan butir soal (*item selection procedure*), pengestimasian kemampuan (*ability estimation*), dan aturan pemberhentian (*stopping rule*), sedangkan dua komponen CAT lain yang sering diperhatikan pada sistem CAT adalah keseimbangan konten (*content balance*) dan kontrol butir soal yang sering muncul (*item exposure control*).

Metode yang umum digunakan untuk mengestimasi kemampuan peserta tes adalah metode Kemungkinan Maksimum (*Maximum Likelihood Estimation*) (Baker, 1992), dan metode Bayes (Bock & Mislevy, 1982; Baker, 1992). Penelitian ini bertujuan untuk membandingkan kedua metode tersebut dalam menaksir kemampuan peserta tes pada rancangan tes adaptive. Berikut dipaparkan secara singkat metode Kemungkinan Maksimum dan metode Bayes.

### Metode Kemungkinan Maksimum

Misalkan seorang peserta tes dengan tingkat kemampuan  $\theta$  menjawab tes yang berisi  $n$  butir soal pilihan ganda dengan parameter butir soal diketahui (diestimasi sebelumnya). Peluang bersama dari peserta tes dapat dituliskan sebagai  $P(U_1, U_2, \dots, U_n | \theta)$ . Dalam praktek,  $U_1, U_2, \dots, U_n$  diganti dengan skor butir soal sesungguhnya untuk peserta yang dituliskan sebagai  $u_1, u_2, \dots, u_n$  ( $u_i = 0$  jika jawaban pada butir soal ke  $i$  salah, dan  $u_i = 1$  jika jawaban benar). Selanjutnya fungsi kemungkinannya (*likelihood function*);  $L(\theta)$ , dituliskan sebagai berikut

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

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

Tujuan metode Kemungkinan Maksimum (KM) adalah menemukan nilai yang memaksimumkan  $L(\theta)$ . Nilai parameter kemampuan yang memaksimumkan fungsi kemungkinan,  $L$  disebut dengan *the maximum likelihood estimate of ability* (Hambleton, 1993). Secara matematik, Nilai ini dapat diperoleh dengan membuat turunan pertama dari logaritma naural dari  $L(\theta)$  terhadap  $\theta$  sama dengan nol.

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{i=1}^n [u_i - P_i(\theta)] \frac{P_i'(\theta)}{P_i(\theta)Q_i(\theta)} = 0. \quad (2)$$

Pada praktiknya, untuk menyelesaikan sistem persamaan (2) dilakukan dengan menggunakan prosedur iterasi Newton-Raphson. Nilai  $\theta$  pada iterasi ke  $(m+1)$  dapat dinyatakan menggunakan relasi berulang sebagai berikut,

$\theta_{m+1} = \theta_m - h_m$  ;  $h_m$  merupakan faktor koreksi yang dirumuskan sebagai berikut.

$$h_m = D \left[ r - \sum_{i=1}^l P_i(\theta_m) \right] / \left[ -D^2 \left[ \sum_{i=1}^l P_i(\theta_m) \cdot Q(\theta_m) \right] \right] \quad (3)$$

dengan  $r$  adalah banyak jawaban benar dan  $D = 1,7$ . Proses iterasi berhenti ketika  $|h_m| = |\theta_{m+1} - \theta_m| < \varepsilon$ , dengan  $\varepsilon$  bilangan yang sangat kecil (Hambleton & Swaminathan, 1985: 83).

### Metode Bayes

Metode Bayes didasarkan pada penerapan teorema Bayes yang terkait dengan peluang bersyarat dan marginal dari dua atau lebih kejadian. Untuk peluang bersyarat dari dua kejadian A dan B, teorema Bayes menyatakan sebagai berikut

$$P(B|A) = P(A|B)P(B)/P(A) \quad (4)$$

Berdasarkan teorema Bayes maka:

$$P(\theta | u_1, u_2, \dots, u_n) = P(u_1, u_2, \dots, u_n | \theta)P(\theta) / P(u_1, u_2, \dots, u_n), \quad (5)$$

dimana parameter kemampuan (ability,  $\theta$ ) pengganti B, dan skor butir soal untuk individu peserta pengganti untuk A.  $P(\theta | u_1, u_2, \dots, u_n)$  adalah sebaran posterior,  $P(u_1, u_2, \dots, u_n | \theta)$  adalah fungsi kemungkinan, dan  $P(\theta)$  adalah sebaran awal (*prior distribution*) dari  $\theta$ .

Selanjutnya penaksir Bayes untuk kemampuan peserta diperoleh dari mean sebaran posteriornya yang berupa

$$E(\theta | \underline{u}) = \int_{-\infty}^{\infty} \theta p(\theta | \underline{u}) d\theta, \quad (6)$$

Untuk menyelesaikan sistem persamaan (6) dilakukan menggunakan prosedur pendekatan Gauss-Hermite quadrature (Bock & Aitken, 1981),

$$\hat{\theta} = \frac{\sum_{k=1}^q X_k L_i(X_k) W(X_k)}{\sum_{k=1}^q L_i(X_k) W(X_k)}, \quad (7)$$

dimana  $X_k$  adalah salah satu dari  $q$  quadrature points,  $W(X_k)$  adalah bobot yang berkaitan dengan quadrature point  $X_k$  (Stroud & Secrest, 1966) dan  $L_i(X_k)$  adalah

fungsi kemungkinan dari jawaban terhadap butir soal ke- $i$  jika diberikan tingkat kemampuan,  $\theta$  pada  $X_k$ .

## METODE

Bank soal untuk keperluan simulasi rancangan tes adaptif terdiri atas 500 butir soal, yang berasal dari data bangkitan menggunakan model *item response theory* (IRT) 3 parameter dengan tingkat daya beda butir antara 0,4 sampai 2, tingkat kesukaran butir soal antara -3 sampai +3, dan faktor *guessing* antara 0 sampai 0,25. Berdasarkan model IRT 3 parameter, peluang seorang peserta tes yang memiliki tingkat kemampuan tertentu akan menjawab butir soal yang diberikan dengan benar tergantung pada nilai dari tiga parameter butir soal yaitu daya beda, tingkat kesukaran, dan faktor *guessing* (Hambleton & Swaminathan, 1985 : 49; Hambleton, Swaminathan, & Rogers, 1991: 17).

Prosedur simulasi untuk rancangan tes adaptif pada penelitian ini didasarkan pada 2100 simuli peserta tes yang disimulasikan, yang mewakili 100 simuli peserta tes untuk setiap 21 titik skala tingkat kemampuan,  $\theta$  dari -3,0 sampai +3,0 dengan kenaikan 0,3.

Langkah simulasi untuk rancangan tes adaptif sebagai berikut.

1. Untuk tingkat kemampuan peserta tes,  $\theta$  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 peubah acak  $x$  yang diambil dari sebaran uniform  $[0,1]$ . Jika  $P_i(\theta)$  lebih besar dari  $x$  maka respons diskor 1, sebaliknya jika  $P_i(\theta)$  kurang dari atau sama dengan  $x$  maka respons diskor 0. Berdasarkan respons dan parameter butir soal selanjutnya kemampuan peserta tes,  $\theta$  diestimasi. Estimasi  $\theta$  dicatat untuk dianalisis lebih lanjut.
2. Berdasarkan metode pemilihan butir soal, diberikan butir soal berikutnya untuk peserta tes.
3. Langkah 1 dan 2 diulang untuk seluruh 2100 simuli peserta tes.
4. 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 metode Kemungkinan Maksimum dan Bayes. Metode pemilihan butir soal berikutnya menggunakan kriteria fungsi informasi maksimum (Hambleton, Swaminathan, & Rogers, 1991: 91) yaitu butir soal yang mempunyai nilai fungsi informasi terbesar pada kemampuan tertentu dipilih untuk diberikan pada peserta tes.

Pada penelitian ini, kriteria pemberhentian tes yang digunakan adalah tes dihentikan jika nilai estimasi kesalahan baku estimasi (Hambleton, Swaminathan, & Rogers 1991 : 94) atau sering disebut dengan 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-based test* (Thissen, 1990).

Perbandingan kedua metode estimasi kemampuan peserta dilakukan dengan cara membandingkan: 1) banyaknya butir yang diperlukan untuk mengestimasi kemampuan peserta tes pada setiap titik kemampuan, 2) korelasi kemampuan sesungguhnya (*true-theta*) dengan estimasinya, 3) biasanya yaitu selisih antara kemampuan sesungguhnya dengan estimasinya, dan 4) kesalahan pengukurannya.

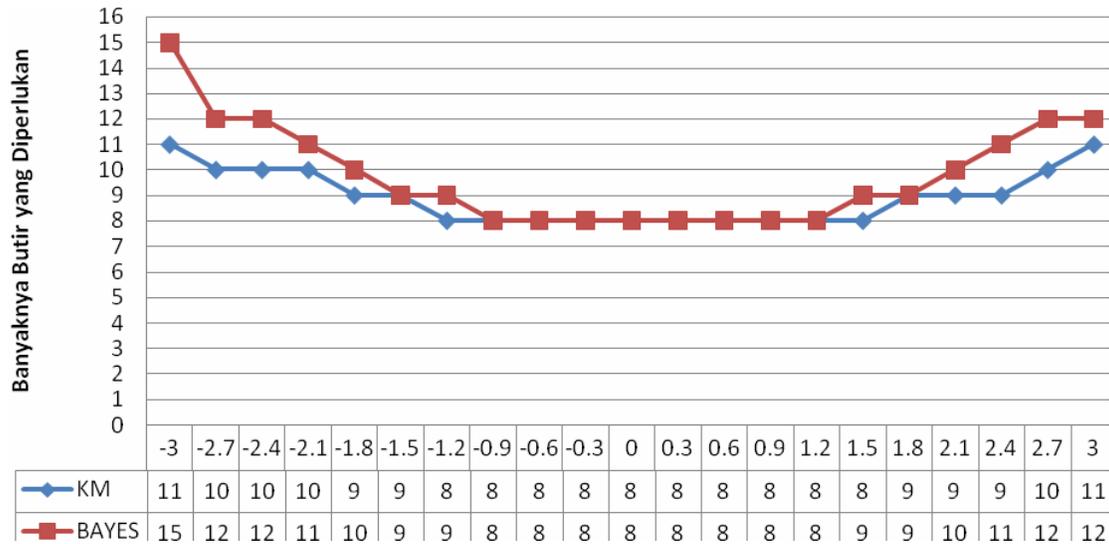
## HASIL DAN PEMBAHASAN

Ringkasan statistik parameter butir pada Bank Soal hasil bangkitan disajikan pada Tabel 1 berikut.

Tabel 1. Ringkasan Statistik Parameter Butir Soal pada Bank Soal

<b>Parameter</b>	<b>Mean</b>	<b>Std-deviasi</b>	<b>Min.</b>	<b>Maks.</b>
<b>Daya beda</b>	1,209734	0,454434	0,400	2,000
<b>Tingkat kesukaran</b>	-0,03357	1,749235	-3,000	2,999
<b>Guessing</b>	0,13375	0,072363	0,001	0,249

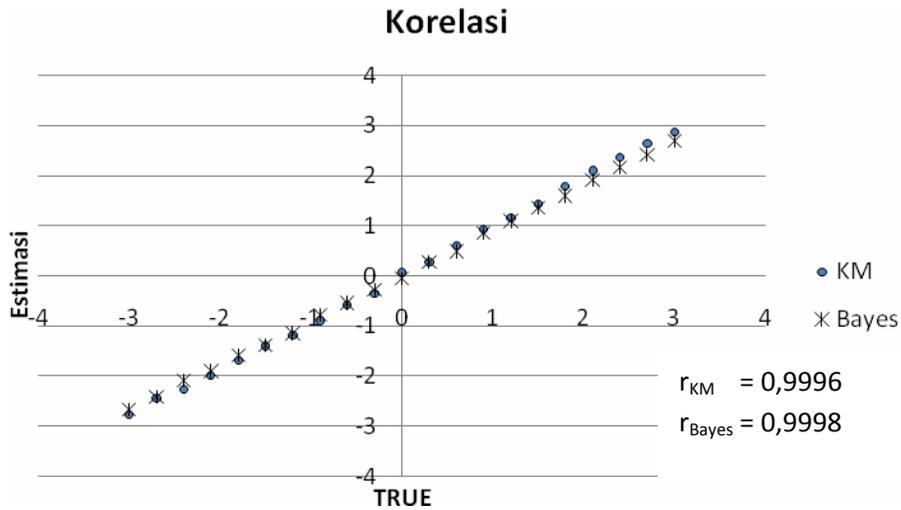
Banyak butir soal yang diperlukan untuk dapat mengestimasi kemampuan peserta pada setiap skala kemampuan pada rancangan tes adaptif dengan metode Kemungkinan Maksimum (KM) dan metode Bayes disajikan pada Gambar 2 berikut.



Gambar 2. Banyaknya Butir yang Diperlukan dengan Metode KM dan Bayes.

Dari Gambar 2 terlihat bahwa banyaknya butir yang diperlukan dengan kedua metode tersebut hampir sama yaitu berkisar antara 8 sampai 12 butir soal kecuali untuk tingkat kemampuan yang ekstrim rendah (skala kemampuan = -3) ada perbedaan sebesar 4 dan untuk tingkat kemampuan yang ekstrim tinggi sebanyak 2. Hal ini menunjukkan bahwa rancangan tes adaptif menggunakan metode KM maupun Bayes dapat mengestimasi kemampuan peserta dengan banyak butir soal lebih sedikit dibandingkan dengan tes konvensional menggunakan *paper-based test* maupun *computer-based test* yang umumnya memiliki panjang tes atau banyak butir dalam satu perangkat tes lebih dari 30 butir. Dengan demikian rancangan tes adaptif lebih efisien dibandingkan tes konvensional.

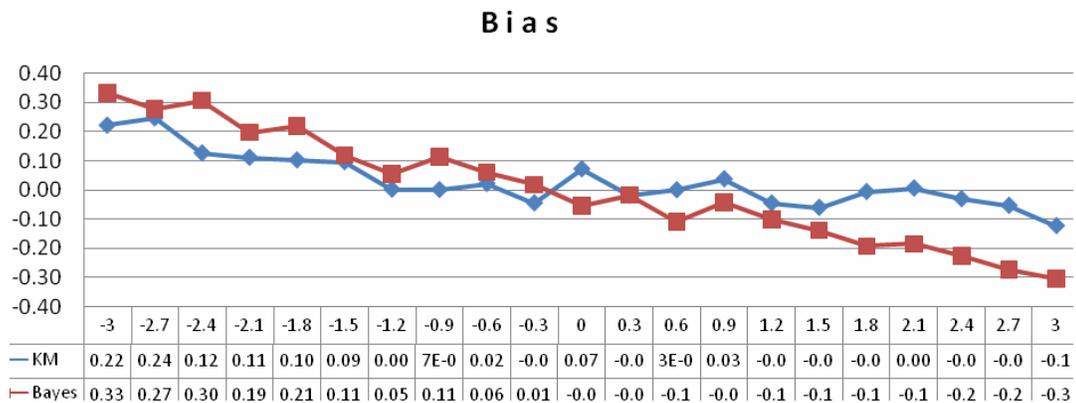
Gambar 3 berikut menyajikan korelasi kemampuan sesungguhnya dengan estimasinya dari kedua metode pengestimasian yang digunakan untuk mengestimasi kemampuan peserta tes.



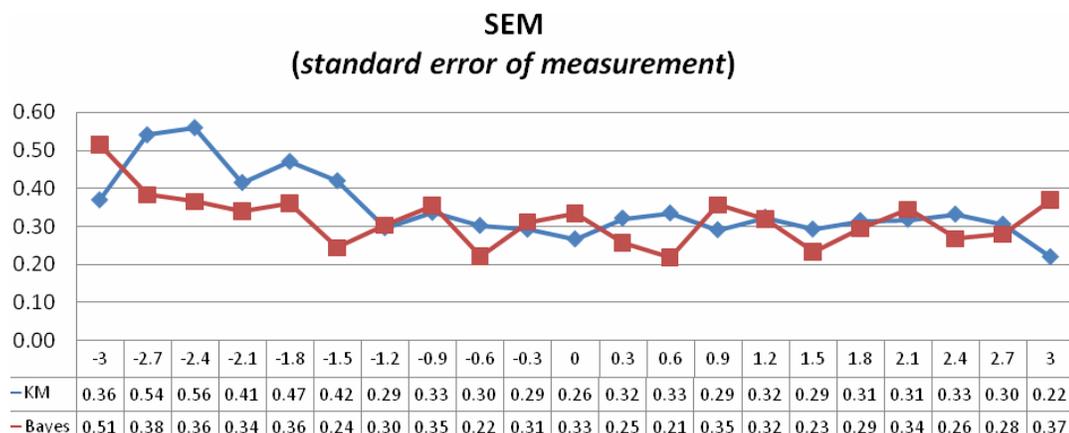
Gambar 3. Korelasi Kemampuan Sesungguhnya dengan Estimasinya.

Dari Gambar 3 terlihat bahwa korelasi antara kemampuan sesungguhnya (*true-theta*) dan estimasinya untuk metode KM dan Bayes membentuk garis lurus, hal ini menunjukkan bahwa kedua metode tersebut akurat untuk mengestimasi kemampuan peserta tes pada rancangan tes adaptif.

Pola bias dan kesalahan pengukuran untuk metode KM dan Bayes disajikan masing-masing pada Gambar 4 dan 5 berikut.



Gambar 4. Bias



Gambar 5. SEM

Berdasarkan Gambar 4 terlihat bahwa umumnya bias metode KM lebih rendah dibandingkan bias metode Bayes dengan kata lain pola bias dengan metode KM cenderung lebih rendah dibandingkan metode Bayes. Hal ini menunjukkan bahwa metode KM lebih akurat dibandingkan metode Bayes. Selanjutnya dari Gambar 5 terlihat bahwa umumnya pola kesalahan baku pengukuran (SEM) dengan metode Bayes lebih rendah dibandingkan metode KM. Hal ini menunjukkan bahwa metode Bayes lebih presisi dibandingkan metode KM.

Dari hasil awal simulasi terdapat keterbatasan dari penerapan metode KM pada rancangan tes adaptif, keterbatasan tersebut adalah ketidakmampuan fungsi kemungkinan untuk menemukan solusi maksimum ketika peserta tes menjawab semua butir soal dengan benar atau salah. Untuk mengatasi masalah ketidakmampuan metode KM dalam mengestimasi kemampuan peserta manakala respons peserta tes belum berpola Dodd (1990) dan Weiss (2004) menyarankan untuk menggunakan metode *step size*. Berdasarkan metode *step size* kemampuan peserta dinaikkan/diturunkan sebesar bilangan tertentu selama respons peserta tes belum berpola. Pada penelitian ini *step size* yang digunakan sebesar 0,5.

## KESIMPULAN DAN SARAN

Berdasarkan hasil simulasi dapat disimpulkan sebagai berikut.

1. Banyaknya butir yang diperlukan untuk mengestimasi kemampuan peserta pada setiap titik kemampuan dengan metode Kemungkinan Maksimum (KM) dan Bayes adalah antara 8 sampai 12 butir soal kecuali untuk peserta berkemampuan ekstrim (rendah maupun tinggi), metode Bayes memerlukan butir soal yang sedikit lebih banyak dibandingkan metode KM.

2. Korelasi kemampuan sesungguhnya dengan estimasinya untuk metode KM dan Bayes adalah sangat kuat artinya kedua metode tersebut akurat dalam mengestimasi kemampuan peserta tes.
3. Pola bias dengan metode KM cenderung lebih rendah dibandingkan metode Bayes, sedangkan pola kesalahan pengukuran menunjukkan hal yang sebaliknya yaitu metode Bayes lebih rendah dibandingkan metode KM.

Berdasarkan hasil penelitian simulasi ini, disarankan untuk menggunakan/menerapkan kombinasi dari kedua metode tersebut untuk mengestimasi kemampuan pada rancangan tes adaptif yaitu menggunakan metode Bayes untuk mengestimasi kemampuan pada awal tes dan menggunakan metode KM untuk mengestimasi akhir (final) kemampuan peserta pada tes adaptif.

## DAFTAR PUSTAKA

- Baker, F.B. (1992). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker, Inc.
- Bock, R.D., & Mislevy, R.J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 4, 431 –444.
- Bunderson, C.V., Inouye, D.K., & Olsen, J.B. (1989). The four generations of computerized educational measurement. Dalam R. L. Linn (Eds.), *Educational Measurement* (3<sup>rd</sup> ed., pp. 367–407). New York: American Council on Education & Macmillan Publishing Company.
- 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.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*, Boston, MA: Kluwer Academic Publishers.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.
- Hambleton, R.K. (1993). Principles and selected applications of item response theory. Dalam R. L. Linn (Eds.), *Educational Measurement* (3<sup>rd</sup> ed., pp. 147-200). Phoenix, AZ: American Council on Education and the Oryx Press.
- Stroud, A. H., & Secrest, D. (1966). *Gaussian quadrature formulas*. Englewood Cliffs, NJ: Prentice-Hall.
- Thissen, D. (1990). Reliability and measurement precision. Dalam H. Wainer (Eds.), *Computerized Adaptive Testing: A Primer* (2<sup>nd</sup> ed., pp. 161–186). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wainer, H. (1990). *Computerized adaptive testing: A primer* (2<sup>nd</sup> 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.
- Weiss, D.J. & Schleisman, J.L. (1999). Adaptive testing. Dalam G. N. Masters & J. P. Keeves (Eds.), *Advances in Measurement in Educational Research and Assessment* (pp. 129–137). Pergamon, NY: Elsevier Science Ltd.

**KEMBALI KE DAFTAR ISI**