

## ANALISIS MODEL SIMULTAN MODEL LOGISTIK SATU PARAMETER DENGAN WAKTU RESPON BERDASARKAN DATA SIMULASI

Noer Hidayah <sup>1\*</sup>, Kumaidi <sup>2</sup>

<sup>1</sup>Sekolah Tinggi Agama Islam Negeri Kediri, <sup>2</sup>Universitas Muhamadiyah Surakarta

<sup>1</sup>Jln. Sunan Ampel No.7, Ngronggo, Rejomulyo, Kota Kediri, Jawa Timur 64127, Indonesia

<sup>2</sup>Jl. A. Yani, Mendungan, Pabelan, Kartasura, Sukoharjo, Jawa Tengah 57162, Indonesia

\* Corresponding Author. Email: noer\_hid@yahoo.com

### Abstrak

Penelitian ini bertujuan untuk menganalisis model simultan model logistik satu parameter (ML1P) dengan waktu respon. Analisis terhadap model menggunakan data simulasi, yang skenario pembangkitan data simulasi dilakukan berdasarkan banyaknya peserta tes (500, 1000) dan banyaknya soal tes (11, 20, 40). Setiap skenario direplikasi sebanyak 30 kali. Metode estimasi parameter model menggunakan metode Bayesian, Markov Chain Monte Carlo. Analisis terhadap model dilakukan dengan menghitung selisih antara besaran parameter bangkitan (*true value*) dengan besaran parameter estimasi. Metode analisis menggunakan *Root Mean Square Error* (RMSE), *Standart Error* (SE) dan bias. Hasil penelitian menunjukkan bahwa *performance* hasil estimasi parameter model yang terdapat dalam soal tes (tingkat kesulitan soal, kelambatan soal, dan besarnya usaha untuk soal ke-*j*), tidak dipengaruhi oleh banyaknya soal tes. *Performance* hasil estimasi parameter model dalam peserta tes (kecepatan dan kemampuan peserta tes) dipengaruhi oleh banyaknya soal tes, yang semakin banyak soal tes maka hasil estimasi parameternya akan semakin mendekati nilai parameter yang sebenarnya.

**Kata kunci:** *model simultan ML1P dengan waktu respon, data simulasi, metode analisis*

## ANALYSIS OF SIMULTANEOUS MODEL OF ONE PARAMETER LOGISTIC MODEL AND RESPONSE TIME BASED ON SIMULATION DATA

### Abstract

The aim of this research is to analyse simultaneous model One Parameter Logistic Model (1-PLM) and respon time. The analysis of model used the simulation data, where the data generation scenario was done based on the number of test takers (500, 1000) and the number of test items (11, 20, 40). Parameter estimation method used the Bayesian method, Markov Chain Monte Carlo. The analysis of model was done with the accounting of the distance of true value and estimated parameter. The Analysis methods use Root Mean Square Error (RMSE), Standart Error (SE) and bias. The result of research reveals the performance of parameter estimation result for the test item (the test item difficulty, test item slowness, and the effort to complete the item test) is not influenced by the number of the test items. However, the performance of parameter estimation result for the test takers (the speed and ability of the test takers) is influenced by the number of the test items. The more test items there are, the closer is the parameter estimation result to the true parameter.

**Keywords:** *simultaneous model one parameter logistic model (1-PLM) and respon time, simulation data, analysis methods*

Permalink/DOI: <http://dx.doi.org/10.21831/jpep.v20i2.8068>

## Pendahuluan

Perkembangan dan kemajuan Ilmu Pengetahuan Teknologi dan Seni (IPTEKS) khususnya bidang Teknologi Informasi (TI), dewasa ini menunjukkan peningkatan yang cukup signifikan bagi kepentingan umat manusia. Salah satu pemanfaatan teknologi informasi dalam dunia pendidikan adalah penggunaan komputer sebagai alat tes. *Computer Adaptive Test* (CAT) dan *Computer Based Test* (CBT) merupakan dua bentuk tes yang menggunakan komputer (Haryanto, 2013, p. 49). Pemanfaatan komputer dalam tes memiliki banyak keuntungan dibandingkan dengan *test* yang berbentuk *paper-pencil test*. CAT/CBT dapat menjamin pelaksanaan tes dilakukan secara jujur atau dengan kata lain menghindari perilaku menyontek (*cheating*). Tes yang berbentuk CAT/CBT akan menghemat biaya operasional tes karena tidak membutuhkan biaya penggandaan soal, biaya perawatan, biaya penyimpanan, dan biaya pengiriman, sehingga tes CAT/CBT dianggap lebih efektif dan efisien.

Penggunaan CAT dan CBT akan menghasilkan dua informasi yang berkaitan dengan hasil tes, yaitu pola respon peserta tes dan waktu respon. Pola respon peserta tes menunjukkan pola jawaban benar-salah dan waktu respon menunjukkan berapa lama peserta tes menyelesaikan sebuah soal. Waktu respon peserta tes dapat digunakan untuk melihat seberapa besar usaha peserta tes dalam menyelesaikan soal tes (Wise & DeMars, 2006), memberikan informasi tentang kecepatan peserta tes dan intensitas waktu yang dibutuhkan untuk menyelesaikan tes (van der Linden, 2007).

Pleskac & Busemeyer (2010, p. 864) menyebutkan tiga hal yang perlu diperhatikan dalam menilai kemampuan kognitif peserta yaitu jawaban (*response*), waktu respon dan kepercayaan diri (*confidence*) peserta tes. Beberapa penelitian lain menyebutkan batasan waktu dalam sebuah tes mempengaruhi hasil tes peserta tes *speededness* dalam tes mempengaruhi estimasi kemampuan dan parameter butir soal (Oshima, 1994; Bolt, Cohen, & Wollack, 2002). Hasil penelitian yang dilakukan Abdelfattah (2007) menyaa-

rankan waktu respon hendaknya dipertimbangkan dan diintegrasikan ke dalam proses penskoran, karena terdapat perbedaan estimasi parameter soal tes dan kemampuan peserta antara yang mempertimbangkan waktu respon dan yang tidak mempertimbangkan waktu respon. Schnipke & Pashley (1997), Schnipke & Scrams, (1999), Hornke (2000) mempertimbangkan waktu respon atau kecepatan peserta tes sebagai komponen untuk mengukur kemampuan peserta tes. Penelitian-penelitian tersebut memperkuat bahwa waktu respon seharusnya dipertimbangkan dalam pemodelan *latent trait*.

Waktu respon dalam tes dapat memberikan informasi berharga tentang proses kognitif jika dibandingkan dengan hanya menganalisis respon peserta tes saja. Analisis terhadap waktu respon akan menghasilkan pengetahuan baru, misalnya bagaimana hubungan antara tingkat kesulitan dengan banyaknya waktu yang dibutuhkan untuk menyelesaikan soal, apakah soal yang lebih sulit membutuhkan waktu yang lebih lama dalam menyelesaikannya, bagaimana hubungan antara kemampuan (integensi) dengan kecepatan, apakah peserta tes yang memiliki kemampuan tinggi cenderung lebih cepat menyelesaikan soal tes dibandingkan dengan siswa yang berkemampuan rendah, apakah format soal tes mempengaruhi hubungan-hubungan tersebut (Entink, 2009, p. 47).

Pertanyaan-pertanyaan tersebut menjadi pertimbangan untuk memanfaatkan waktu respon dan respon peserta tes dalam menganalisis parameter peserta tes dan parameter soal tes. Analisis terhadap waktu respon dapat dilakukan dengan memodelkan secara matematis waktu respon tersebut sebagaimana model matematis respons butir (dikenal dengan teori respons butir).

Teori tentang pemodelan waktu respon telah berkembang, tetapi perkembangan dan aplikasinya tidak sebanyak teori respon butir. Teori waktu respon cenderung diabaikan dalam psikometrik, meskipun telah lama digulirkan.

Seiring dengan perkembangan ilmu pengetahuan, topik tentang waktu respon

mengalami perkembangan pesat khususnya dalam bidang psikologi dan pendidikan. Salah satu topik tersebut berkaitan dengan pemodelan waktu respon untuk mengestimasi parameter soal tes dan kemampuan peserta tes. Pemodelan waktu respon dimaksudkan untuk memperbaiki konsep IRT ketika sebuah tes memperhitungkan waktu atau kecepatan peserta tes.

*Teori respons butir atau Item Response Theory (IRT)*, memodelkan *latent trait* (parameter peserta dan parameter soal tes seperti tingkat kesulitan, daya beda dan faktor tebakan) hanya berdasarkan pada pola respon peserta tes, tanpa memperhitungkan waktu responnya. Konsep IRT cocok diterapkan pada *pure power test*, yang tes tidak dibatasi oleh waktu. van der Linden & Hambleton (1997, p. 166) menyatakan bahwa *power test* murni hanya idealisme dan tidak ada dalam pelaksanaannya, karena dalam setiap tes selalu dibatasi oleh waktu. Batasan waktu dalam tes memungkinkan adanya aspek lain yang terlibat dalam penyelesaian tes, disamping kemampuan peserta tes. Aspek lain tersebut adalah kecepatan peserta tes.

Waktu respon seharusnya dapat menjadi informasi tambahan bagi lembaga penyelenggara tes, khususnya dalam menilai kemampuan peserta tes. Waktu respon merupakan variabel *dependent* yang dipertimbangkan dalam memahami perilaku peserta tes di dalam psikologi kognitif. Oleh karena itu, waktu respon seharusnya juga digunakan sebagai dasar untuk mengestimasi parameter peserta tes (proses *scoring*) dan parameter soal tes. Pelibatan waktu respon dalam proses pengukuran akan memperbaiki hasil estimasi parameter.

Pelibatan waktu respon dalam proses pengukuran dilakukan dengan memodelkan waktu respon dalam bentuk model matematis. Pengembangan model yang melibatkan waktu respon dilakukan berdasarkan jenis tes, yaitu apakah tes merupakan *power test* atau *speed test*. Karakteristik *speed test* adalah soal-soal tes disusun dengan tingkat kesulitan rendah, sehingga dalam waktu tertentu semua peserta tes baik yang kemampuan tinggi maupun yang berkemampuan rendah

dapat menjawab dengan benar soal tes. Dengan kata lain, probabilitas peserta tes menjawab benar soal tes sama dengan satu. *Power test* memiliki konsep yang berbeda dengan *speed test*. Soal-soal dalam *power test* disusun dengan tingkat kesulitan bervariasi, sehingga tidak semua peserta tes dapat menjawab dengan benar soal tes.

Pengembangan model selain berdasarkan jenis tes, juga dilakukan berdasarkan bentuk model matematis yang digunakan. Model matematis yang digunakan untuk memodelkan antara waktu respon dengan respon butir untuk tujuan *scoring* (mengestimasi parameter peserta tes dan parameter soal) ada 3 bentuk yaitu: (1) model respon waktu dan respon butir diolah dan dianalisis secara terpisah (van der Linden, 2006); (2) model waktu respon dan model respon butir diolah secara terpisah kemudian hasilnya dimodelkan secara simultan. (van der Linden, 2007; Entink, van der Linden, & Fox, 2009; Meyer, 2010; Meng, Tao, & Chang, 2015); (3) waktu respon dan respon butir dimodelkan secara simultan, (Roskam dalam van der Linden & Hambleton, 1997; Verhelst, Verstralen, & Jansen dalam van der Linden & Hambleton, 1997; Wang & Hanson, 2005; Wang, 2006; Ingrisone II, 2008; Hidayah, Kumaidi, & Kartowagiran, 2015).

Penelitian yang berkaitan dengan analisis waktu respon bertujuan untuk (1) mengestimasi kemampuan peserta tes (proses *scoring*) dan parameter soal (Thissen, 1983; Roskam dalam van der Linden & Hambleton, 1997; van der Linden, 2007; Verhelst, Verstralen, & Jansen dalam van der Linden & Hambleton, 1997; Wang & Hanson, 2005; Wang, 2006; Ingrisone, 2008; Ingrisone II, 2008); (2) menganalisis kecepatan peserta tes dalam mengerjakan soal dan estimasi waktu yang dibutuhkan untuk menyelesaikan soal (van der Linden, 2006); (3) menguji hubungan antara komponen kecepatan dan komponen keakuratan dalam menyelesaikan tes (van der Linden, 2007; Entink, van der Linden, & Fox., 2009); (4) memperbaiki estimasi parameter khususnya jika dalam test terjadi kondisi

*speededness* (Oshima, 1994; Yamamoto, 1995; Meyer, 2010; Wollack & Woo, 2009); (5) sebagai alat diagnostis terhadap kemampuan peserta tes dan proses pembelajaran (Meyer & Wise, 2005; Gvozdenko & Chambers, 2007); (6) mengetahui bagaimanakah usaha peserta tes dalam menyelesaikan tes, apakah mengerjakan sebuah soal dengan cara atau strategi yang sesuai atau sekedar menebak jawaban (Wise & DeMars, 2006); (7) mendeteksi pola jawaban yang tidak wajar dalam CAT (van der Linden & van Krimpen-Stoop, 2003).

Model yang akan dianalisis dalam penelitian ini adalah model yang dikembangkan oleh Hidayah, Kumaidi, & Kartowagiran (2015). Hidayah, Kumaidi, & Kartowagiran (2015) mengintegrasikan waktu respon ke dalam Model Logistik Satu Parameter (ML-1P), sehingga diperoleh model simultan ML-1P dengan waktu respon. Model digunakan untuk tujuan *power test*. Model simultan tersebut perlu dianalisis lebih lanjut untuk melihat ketepatan dan keakuratan model dalam beberapa situasi tes. Situasi tes selalu melibatkan banyaknya peserta tes dan banyaknya soal tes. Oleh karena itu, analisis dilakukan untuk melihat *performance* model simultan ML-1P dengan waktu respon berdasarkan banyaknya peserta tes dan banyaknya soal tes. Data yang digunakan untuk tujuan tersebut adalah data simulasi.

Berdasarkan uraian yang telah disampaikan tersebut, maka penelitian ini bertujuan untuk menganalisis model simultan model logistik satu parameter (ML1P) dengan waktu respon.

## Metode Penelitian

Penelitian ini pada dasarnya merupakan penelitian kuantitatif, yang bertujuan untuk menganalisis model simultan ML 1P dengan waktu respon. Model simultan ML 1P dengan waktu respon merupakan hasil pengembangan ML-1P. Model simultan ML 1P dengan waktu respon merupakan model matematis yang lebih realistis yang dapat digunakan dalam praktek pengukuran. Model IRT memodelkan respons peserta tes hanya berdasarkan jawaban peserta tes, yai-

tu jawaban benar-salah soal tes tanpa mempertimbangkan waktu responsnya. Padahal pada kenyataannya sebuah tes selalu dibatasi oleh waktu. Oleh karena itu, penggabungan waktu respons ke dalam model, diharapkan dapat memperbaiki konsep IRT. Estimasi kemampuan peserta tes dan parameter butir soal dikalibrasi sesuai dengan kondisi yang ada dalam sebuah tes.

Data yang digunakan untuk menganalisis model simultan ML 1P dengan waktu respon adalah data simulasi atau data bangkitan. Data simulasi merupakan data yang dibangkitkan berdasarkan algoritma tertentu yang sesuai dengan model matematis yang dianalisis. Data dibangkitkan menurut skenario tertentu, dimana skenario yang digunakan dalam penelitian ini berdasarkan banyaknya peserta tes dan banyaknya soal tes. Skenario pembangkitan data terdapat dalam Tabel 1.

Tabel 1. Skenario Pembangkitan Data Simulasi

Banyaknya peserta tes	Banyaknya soal tes		
	10 butir	20 butir	40 butir
500	A11	A12	A13
1000	A21	A22	A23

Setiap skenario di atas direplikasi sebanyak 30 kali.

Penelitian ini dilakukan dengan prosedur berikut. Prosedur pertama yang dilakukan adalah menentukan model simultan antara ML 1P dengan waktu respon yang akan dianalisis. Prosedur berikutnya adalah membangkitkan data. Langkah ini diperlukan untuk menganalisis model simultan antara ML 1P dengan waktu respon. Data dibangkitkan sesuai dengan algoritma yang disusun berdasarkan model simultan antara ML 1P dengan waktu respon yang akan dianalisis. Data dibangkitkan sesuai dengan skenario simulasi data yang telah ditentukan sebagaimana dalam Tabel 1.

Algoritma pembangkitan data dimulai dengan menentukan besaran-besaran parameter yang terdapat dalam model, baik parameter untuk model simultan ML-1P de-

ngan waktu respon maupun parameter untuk model waktu respon. Besaran-besaran parameter yang ditentukan ini disebut dengan besaran parameter bangkitan atau *true parameter*. Parameter bangkitan yang terdapat dalam model waktu respon kemudian dimasukkan ke dalam model waktu respon. Hasil yang diperoleh adalah besaran waktu respon,  $t_{ij}$ . Besaran  $t_{ij}$  dan besaran parameter lain yang terdapat dalam model simultan ML-1P dan waktu respon dimasukkan ke dalam model simultan ML-1P dan waktu respon. Hasil yang diperoleh adalah besarnya probabilitas respon benar bersyarat terhadap waktu ( $P_{ij}$ ). Proses berikutnya adalah membangkitkan data yang berdistribusi *uniform*,  $U(0,1)$ , sebanyak  $i$  peserta tes dan  $j$  soal tes. Hasil yang diperoleh adalah probabilitas berdistribusi *uniform*,  $p_{ij}$ .  $P_{ij}$  dibandingkan dengan  $p_{ij}$ . Jika  $P_{ij} > p_{ij}$  maka diberi kode 1 untuk menyatakan jawaban benar dan jika  $P_{ij} < p_{ij}$  maka diberi kode 0 untuk menyatakan jawaban salah. Implementasi pembangkitan data menggunakan program R.

Setelah data simulasi diperoleh, prosedur yang ketiga mengestimasi parameter model simultan ML-1P dengan waktu respon menggunakan pendekatan Bayesian Markov Chain Monte Carlo. Pendekatan Bayesian dipilih karena pendekatan ini bersifat fleksibel terutama jika model bersifat kompleks. Pendekatan Bayesian menggabungkan informasi prior (yaitu informasi tentang data yang diperoleh sebelum penelitian) dengan fungsi *likelihood* (yaitu informasi tentang data).

Langkah-langkah dalam estimasi dengan pendekatan Bayesian adalah (a) menentukan parameter-parameter yang terlibat dalam model simultan ML-P dan waktu respon, (b) menentukan distribusi *likelihood* berdasarkan data, (c) menentukan distribusi prior dari setiap parameter yang terdapat dalam model, (d) menentukan distribusi posterior, (e) menentukan distribusi bersyarat penuh dari masing-masing parameter, dan (f) mengestimasi parameter dengan Metode Markov Chain Monte Carlo menggunakan algoritma Gibbs Sampler. Imple-

mentasi pendekatan Bayesian menggunakan program WinBUGS 1.4.

Prosedur yang terakhir adalah menganalisis hasil estimasi parameter model berdasarkan banyaknya peserta tes dan banyaknya soal tes. Tujuan dari prosedur ini adalah mengetahui bagaimana *performance* hasil parameter estimasi model dalam berbagai situasi tes.

Analisis data pada dasarnya dilakukan dengan melihat besarnya selisih nilai antara besaran parameter bangkitan (*true value*) dengan besaran parameter estimasi. Semakin kecil selisih antara parameter bangkitan dengan besaran parameter estimasi menunjukkan hasil estimasi parameter semakin tepat dan semakin akurat. Ukuran yang digunakan untuk melihat besarnya selisih tersebut adalah RMSE, SE dan bias. Analisis juga dilakukan dengan membandingkan RMSE, SE dan bias tersebut berdasarkan banyaknya peserta tes dan banyaknya soal tes. Hal tersebut dilakukan untuk melihat *performance* model pada situasi tes tertentu.

Bias merupakan selisih antara rata-rata parameter estimasi dalam  $M$  replikasi dengan parameter bangkitan. Misalkan  $\psi_j$  menyatakan parameter bangkitan dari soal ke- $j$  (misalnya  $b_j$  beta),  $\hat{\psi}_{jm}$  menyatakan  $\psi_j$  estimasi (parameter estimasi) pada replikasi ke- $m$ ,  $j$  menyatakan urutan soal,  $j= 1,2,3,\dots,J$  dan  $m$  menyatakan urutan replikasi,  $m = 1,2,3,\dots, M$ .

Bias dinyatakan sebagai berikut.

$$Bias(\hat{\psi}) = \frac{1}{M} \sum_{m=1}^M \hat{\psi} - \psi_j \dots\dots\dots(1)$$

Standar *error* merupakan deviasi standar dari parameter estimasi dengan parameter bangkitan dalam  $M$  replikasi.

SE dinyatakan sebagai berikut.

$$SE(\hat{\psi}_j) = \sqrt{\frac{1}{M} \sum_{m=1}^M \left( \hat{\psi}_{jm} - \frac{1}{M} \sum_{m=1}^M \hat{\psi}_{jm} \right)^2} \dots\dots\dots(2)$$

*Root Mean Square Error* (RMSE) merupakan akar kuadrat dari rata-rata selisih kuadrat antara parameter estimasi dengan parameter bangkitan.

RMSE dinyatakan sebagai berikut.

$$RMSE(\hat{\psi}_j) = \sqrt{\frac{1}{M} \sum_{m=1}^M (\hat{\psi}_{jm} - \psi_j)^2} \quad (3)$$

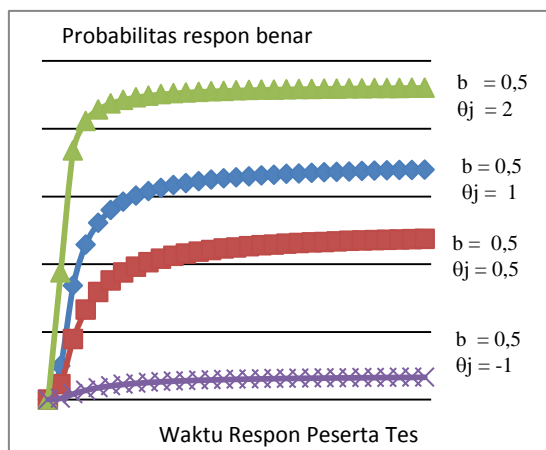
### Hasil Penelitian dan Pembahasan

Formulasi model simultan ML 1P dengan waktu respon yang akan dianalisis adalah sebagai berikut (Hidayah, Kumaidi, & Kartowagiran, 2015, p. 9).

$$P(x=1|\theta_j, t_{ij}, b_i) = \frac{\exp(\theta_j - \frac{d_j}{t_{ij}} - b_i)}{1 + \left( \exp(\theta_j - \frac{d_j}{t_{ij}} - b_i) \right)} \quad (4)$$

Model tersebut menunjukkan probabilitas respon benar peserta tes,  $P(x=1|\theta_j, t_{ij}, b_i)$ , dipengaruhi oleh kemampuan peserta tes ( $\theta_j$ ), waktu respon peserta tes ( $t_{ij}$ ) dan tingkat kesulitan soal tes ( $b_i$ ). Jika besarnya  $t_{ij}$  tak terhingga, maka persamaan 1 sama dengan dengan ML 1P. Besarnya  $t_{ij}$  yang tak berhingga dapat bermakna bahwa tes tidak dibatasi oleh waktu.

Hubungan antara probabilitas respon benar dengan parameter yang mempengaruhinya dalam ML 1P dengan integrasi waktu respon, sebagaimana persamaan 4, dijelaskan dalam CAF berikut.



Gambar 1. CAF model simultan ML 1P dengan Waktu Respon

CAF tersebut disusun pada kondisi tingkat kesulitan soal tes ( $b_i$ ) sama yaitu sebesar 0,5 dan tingkat kemampuan ( $\theta_j$ ) berbeda. Hal ini dimaksudkan untuk mem-

bandingkan beberapa kondisi yang menunjukkan selisih antara kemampuan peserta tes dan tingkat kesulitan soal, yang dalam Gambar 1 terdapat selisih kemampuan peserta tes dan tingkat kesulitan soal yang positif dan terdapat selisih kemampuan peserta tes dan tingkat kesulitan soal yang negatif. Grafik pada semua kondisi yang dibandingkan menunjukkan fungsi monoton naik. CAF tersebut bermakna semakin banyak waktu yang digunakan untuk menyelesaikan soal tes, semakin tinggi probabilitas menjawab benar soal tes. Semakin besar selisih antara tingkat kemampuan peserta tes dengan tingkat kesulitan soal akan mengakibatkan semakin tinggi probabilitas respon benar, artinya peserta tes dengan kemampuan lebih tinggi akan memiliki probabilitas respon benar lebih tinggi jika dibandingkan dengan peserta tes dengan kemampuan lebih rendah (Hidayah, Kumaidi, & Kartowagiran, 2015, p. 10).

Waktu respon,  $t_{ij}$ , pada persamaan 4 diatas merupakan variabel independen yang bersifat tetap (*fixed variable*), sehingga seolah-olah probabilitas waktu respon hanya ditentukan oleh waktu respon yang dicatat oleh administrasi tes pada saat tes berlangsung. Kenyataannya waktu respon peserta tes dapat berubah-ubah sesuai dengan kondisi peserta tes. Peserta tes yang sama, jika mengikuti tes berkali-kali dengan soal tes dan *setting* tes yang sama belum tentu membutuhkan waktu respon yang sama. Kondisi peserta tes dan strategi yang digunakan dalam menyelesaikan soal tes dapat berubah-ubah dari waktu ke waktu. Hal tersebut mempengaruhi kecepatan dan waktu respon peserta tes. Oleh karena itu, waktu respon seharusnya *random variable*.

Waktu respon sebagai *random variable* dalam statistika dinyatakan dalam bentuk distribusi. Waktu respon secara empiris memiliki bentuk *skewed* dan bernilai positif. Beberapa bentuk distribusi yang memenuhi karakteristik waktu respon tersebut adalah distribusi Lognormal, distribusi Weibull dan distribusi Gamma.

Distribusi lognormal yang dipilih sebagai distribusi marginal waktu respon da-

lam penelitian ini adalah distribusi log-normal yang telah dikembangkan oleh van der Linden (2006). van der Linden, 2006, pp. 257-259) menyatakan waktu respon dan kecepatan adalah dua konsep yang berbeda. Hal ini dapat dijelaskan berdasarkan dua soal yang terdapat dalam soal pada Tabel 2.

Tabel 2. Dua soal tes dengan Panjang Tes yang Berbeda

soal 1	soal 2
425	370
$\frac{325}{750}^+$	230
	064
	$\frac{086}{750}^+$

Siswa A mampu mengerjakan soal tes ke-1 dalam waktu 60 detik, sedangkan siswa B mampu menyelesaikan dalam waktu 90 detik. Apakah dapat disimpulkan bahwa siswa A lebih cepat mengerjakan soal tes dibandingkan dengan siswa B? Dua soal dalam Tabel 2 memiliki tingkat kompleksitas yang berbeda, dimana soal 1 hanya bertingkat 2 sedangkan soal 2 bertingkat 4, sehingga wajar kalau soal 2 membutuhkan waktu yang lebih lama dibandingkan soal 1. Hal tersebut menunjukkan perbedaan lamanya waktu yang dibutuhkan untuk menyelesaikan soal tidak disebabkan oleh perbedaan kecepatan dalam menyelesaikan soal tetapi ditentukan oleh kompleksitas soal. Oleh karena itu waktu respon tidak dapat dianggap sama dengan kecepatan. Waktu respon merupakan fungsi kecepatan peserta tes dan faktor lain yang berkaitan dengan soal tes. van der Linden (2006) menyebut faktor lain soal tes dengan intensitas waktu yang dibutuhkan untuk menyelesaikan soal tes (*time intensity*).

Hubungan antara waktu respon dan kecepatan dianalogikan sebagaimana dalam ilmu fisika. Kecepatan dalam ilmu fisika didefinisikan dengan besarnya jarak yang ditempuh dalam waktu tertentu. Definisi tersebut dinyatakan dalam persamaan berikut.

$$\text{Kecepatan} = \frac{d(t_2) - d(t_1)}{t_2 - t_1} \tag{5}$$

$d(t_2) - d(t_1)$  adalah jarak yang ditempuh dalam waktu antara  $t_1$  dan  $t_2$

Berkaitan dengan definisi kecepatan dalam menyelesaikan soal tes, maka jarak yang ditempuh sebagaimana dalam persamaan 5 dianalogikan dengan besarnya usaha yang dibutuhkan untuk menyelesaikan soal tes. Oleh karena itu, kecepatan dalam menyelesaikan soal tes didefinisikan sebagai besarnya usaha yang dibutuhkan untuk menyelesaikan soal tes dalam waktu tertentu. Misalkan  $\tau_i$  menyatakan kecepatan yang digunakan oleh peserta tes ke-i,  $\beta_j$  menyatakan besarnya usaha yang dibutuhkan untuk menyelesaikan soal ke-j,  $t_{ij}$  menyatakan waktu respon peserta tes ke-i untuk soal tes ke-j. Definisi waktu respon dapat dinyatakan dalam persamaan berikut.

$$\tau_i^* = \frac{\beta_j}{t_{ij}} \tag{6}$$

Persamaan 6 dapat dinyatakan dalam bentuk logaritma.

$$\ln(\tau_i^*) = \ln(\beta_j) - \ln(t_{ij}) \tag{7}$$

misalkan  $\beta_j = \ln \beta_j^*$  dan  $\tau_i = \ln \tau_i^*$  maka persamaan 6 dapat ditulis kembali dalam persamaan berikut.

$$\ln(t_{ij}) = \beta_j - \tau_i \tag{8}$$

Operasi matematika berupa ekspektasi dapat digunakan dalam persamaan 8 untuk menyatakan rata-rata waktu respon, sehingga diperoleh hasil berikut.

$$E(\ln(t_{ij})) = \beta_j - \tau_i \tag{9}$$

Dimana  $\beta_j \in (-\infty, \infty)$  dan  $\tau_i \in (-\infty, \infty)$ . Asumsi tambahan dalam persamaan di atas adalah kecepatan peserta tes dalam menyelesaikan setiap soal adalah konstant.

Perbedaan model waktu respon dalam penelitian ini dengan model waktu respon

van der Linden (2006) terletak pada pengertian varian pada distribusi lognormal. van der Linden (2006) mendefinisikan varian sebagai daya beda antara peserta tes, dimana  $\sigma^2 = \frac{1}{\sigma_i^2}$  Penelitian ini tidak mendefinisikan

varians sebagai fungsi tertentu. Model matematis waktu normal berdasarkan distribusi lognormal adalah sebagai berikut.

$$f(t_i; \tau_i, \beta_j) = \frac{1}{t_{ij} \sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[ \frac{\ln t_{ij} - (\beta_j - \tau_i)}{\sigma} \right]^2\right) \quad (10)$$

$\beta_j - \tau_i$  merupakan rata-rata waktu respon.  $\beta_j$  adalah besarnya usaha yang dibutuhkan untuk menyelesaikan soal tes ke-j,  $\beta_j \in (-\infty, \infty)$ . Semakin besar  $\beta_j$ , maka banyaknya usaha yang dibutuhkan oleh peserta tes ke-i juga semakin besar.  $\tau_i$  adalah kecepatan peserta tes ke-i dalam menyelesaikan soal,  $\tau_i \in (-\infty, \infty)$ . Semakin besar  $\tau_i$ , maka semakin tinggi kecepatan peserta tes ke-i, maka semakin sedikit waktu yang dibutuhkan untuk menyelesaikan soal tes ke-j.  $\sigma$  merupakan varian  $t_{ij}$ .

Model simultan antara respon butir dan waktu respon dalam persamaan 1 merupakan distribusi bersyarat. Distribusi bersyarat dalam statistika dapat dijabarkan kembali ke dalam bentuk *joint distribution*, sehingga konsep yang digunakan untuk pemodelan adalah *joint distribution*. *Joint distribution* merupakan perkalian antara distribusi bersyarat (sebagaimana dalam persamaan 1) dengan distribusi *marginal* waktu respon (sebagaimana dalam persamaan 2). *Joint distribution* tersebut dinyatakan sebagai berikut.

$$f(x_i, t_{ij} | \theta, b, d_j, \beta_j, \tau_i) = \left[ \frac{\exp\left(\theta - \frac{d_j}{t_{ij}} - b\right)}{1 + \exp\left(\theta - \frac{d_j}{t_{ij}} - b\right)} \right]^{y_i} \left[ \frac{\exp\left(\theta - \frac{d_j}{t_{ij}} - b\right)}{1 + \exp\left(\theta - \frac{d_j}{t_{ij}} - b\right)} \right]^{-y_i} \left[ \frac{1}{\sqrt{2\pi} \sigma} \exp\left\{-\frac{1}{2} \left[ \frac{\ln(t_{ij} - \beta_j - \tau_i)}{\sigma} \right]^2\right\} \right] \quad (11)$$

Parameter-parameter dalam model tersebut diestimasi dengan menggunakan metode Bayesian Markov Chain Monte Carlo. Semua jenis informasi dalam pendekatan Bayesian dinyatakan dalam bentuk

distribusi probabilitas. Estimasi parameter dilakukan dengan menentukan distribusi prior (informasi awal berkaitan dengan parameter) dan distribusi *likelihood* (distribusi yang merepresentasikan data).

Distribusi prior setiap parameter dipilih berdasarkan domain parameter tersebut yang ditentukan menurut teori. Tingkat kesulitan soal secara teoritis berkisar antara negatif tak hingga sampai dengan positif tak hingga, sehingga distribusi prior yang digunakan adalah distribusi normal,  $b \sim N(0,1)$ . Tingkat kemampuan peserta tes berkisar antara negatif tak hingga sampai dengan positif tak hingga sehingga  $\theta_i \sim N(0,1)$ . Besarnya usaha yang dibutuhkan untuk menyelesaikan soal tes ( $\beta_j$ ) adalah negatif tak hingga sampai dengan positif tak hingga sehingga  $\beta_i \sim N(0,1)$  Kecepatan peserta tes ke-j dalam menyelesaikan soal ( $\tau_j$ ) berada antara negatif tak hingga sampai dengan positif tak hingga, sehingga  $\tau_j \sim N(0,1)$ . Besarnya  $d_j$  adalah positif, sehingga distribusi prior yang dipilih adalah distribusi normal yang nilainya lebih dari nol atau yang bernilai positif. Sementara itu  $\sigma^2$  ditetapkan bernilai 1, sehingga distribusi prior yang dipilih adalah distribusi uniform  $\sigma^2 \sim U(0,1)$ .

Distribusi *likelihood* ditentukan berdasarkan data. Data dalam penelitian ini adalah respon atau jawaban peserta tes yang berbentuk dikotomus, yaitu jika jawaban benar diberi skor satu, dan jika jawaban salah diberi skor nol. Oleh karena itu distribusi *likelihood* yang dipilih adalah distribusi Bernoulli.

Distribusi posterior merupakan proporsional terhadap perkalian antara distribusi *likelihood* dengan distribusi prior setiap parameter. Distribusi posterior untuk parameter  $b_j, \theta_i, \beta_j, \tau_i, d_j, \sigma$  tidak dapat dinyatakan dalam bentuk pendekatan distribusi yang sudah dikenal dan bentuk distribusinya tidak sederhana. Oleh karena itu, estimasi parameternya dilakukan dengan iterasi numerik dengan implementasi MCMC, dengan algoritma Gibbs Sampler. Algoritma tersebut membangkitkan nilai parameter berda-



sarkan distribusi posterior tersebut. Implementasi terhadap algoritma tersebut dilakukan dengan bantuan program WinBUGS 1.4. Proses estimasi parameter dilakukan sebanyak skenario data bangkitan yaitu 180 kali (banyaknya skenario dikalikan dengan banyaknya replikasi,  $6 \times 30 = 180$ ).

Besaran parameter estimasi yang diperoleh dari *running* program Winbugs1.4 dibandingkan dengan parameter bangkitan (*true value*). Kesimpulan umum tentang parameter-parameter dalam model dibuat dengan menghitung rata-rata RMSE, bias dan SE terhadap banyaknya peserta tes dan banyaknya soal tes. Rata-rata RMSE, bias dan SE tersebut terdapat dalam Tabel 3

sampai dengan Tabel 5 berikut. Perbandingan rata-rata RMSE, bias dan SE berdasarkan banyaknya peserta tes dan banyaknya soal tes digambarkan dalam Gambar 2 dan Gambar 3.

Rata-rata RMSE, SE dan bias pada semua situasi tes (baik pada banyak peserta tes 500 dan 1000 serta pada banyak soal tes 10, 20, dan 40) adalah kecil, dimana nilainya mendekati nol. Hal tersebut menunjukkan bahwa semua besaran parameter estimasi mendekati besaran parameter bangkitan. Kondisi ini merepresentasikan bahwa proses estimasi parameter berjalan dengan baik dan hasil estimasi parameter model cukup akurat untuk digunakan sebagai hasil pengukuran.

Tabel 3. Rata-Rata RMSE terhadap Banyaknya Peserta Tes dan Banyaknya Soal Tes

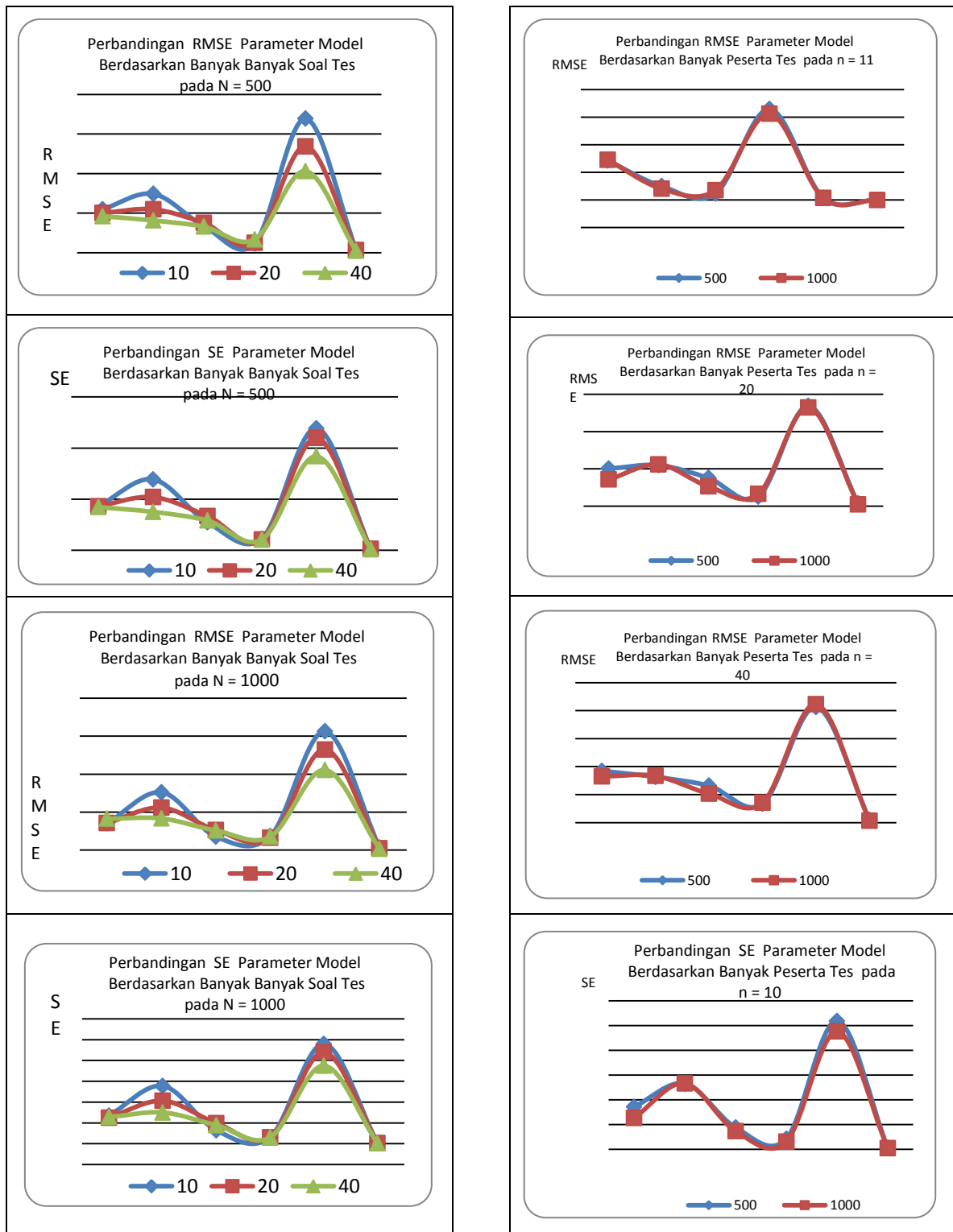
Banyak Peserta Tes	Banyak Soal Tes	RMSE					
		b	tau	d	Beta	theta	sigma
500	10	0,22	0,30	0,13	0,05	0,68	0,02
	20	0,20	0,22	0,15	0,05	0,54	0,01
	40	0,18	0,16	0,13	0,07	0,41	0,01
1000	10	0,14	0,30	0,07	0,07	0,63	0,01
	20	0,14	0,22	0,11	0,06	0,53	0,01
	40	0,16	0,17	0,10	0,07	0,42	0,01

Tabel 4. Rata-Rata Bias terhadap Banyaknya Peserta Tes dan Banyaknya Soal Tes

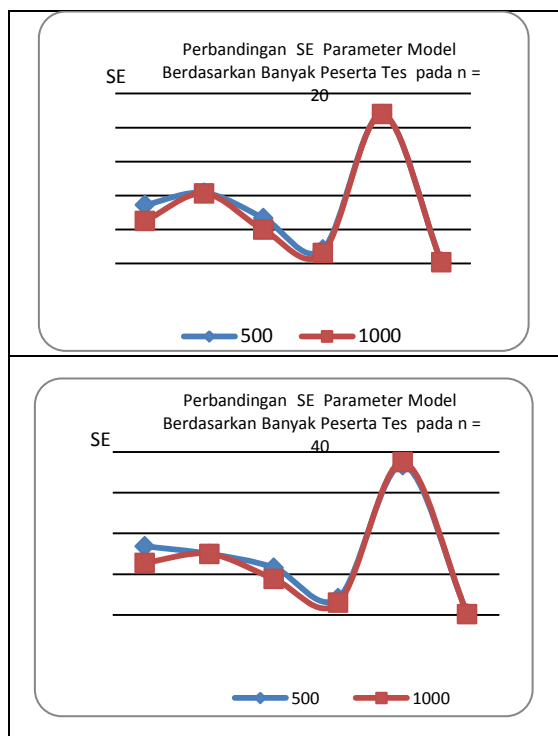
Banyak Peserta Tes	Banyak Soal Tes	Bias					
		b	tau	d	Beta	theta	sigma
500	10	0,03	-0,01	0,06	-0,01	0,11	0,02
	20	-0,02	-0,03	0,04	-0,03	0,00	-0,01
	40	-0,01	0,05	0,05	0,05	0,02	-0,01
1000	10	-0,04	-0,07	0,02	0,07	-0,01	-0,01
	20	-0,03	-0,06	0,03	-0,06	-0,01	-0,01
	40	-0,06	-0,06	0,04	-0,06	-0,01	-0,01

Tabel 5. Rata-Rata SE terhadap Banyaknya Peserta Tes dan Banyaknya Soal Tes

Banyak Peserta Tes	Banyak Soal Tes	SE					
		b	tau	d	Beta	theta	sigma
500	10	0,17	0,28	0,11	0,05	0,48	0,01
	20	0,17	0,21	0,13	0,04	0,44	0,01
	40	0,17	0,15	0,12	0,04	0,37	0,00
1000	10	0,13	0,28	0,06	0,03	0,48	0,00
	20	0,13	0,21	0,10	0,03	0,44	0,00
	40	0,13	0,15	0,09	0,03	0,37	0,00



Gambar 2. Perbandingan RMSE dan SE parameter model Berdasarkan Banyak soal tes pada  $N = 500$  dan  $N = 1000$



Gambar 3. Perbandingan RMSE dan SE Parameter Model Berdasarkan Banyak Peserta Tes pada  $n = 10$ ,  $n=20$  dan  $n=40$

Gambar 2 menunjukkan bahwa baik pada peserta tes sebanyak 500 atau 1000, *performance* RMSE dan SE semua parameter dalam model simultan antara ML 1P dengan waktu respon adalah sama yang terdapat pola yang sangat mirip pada semua level banyaknya soal tes. Parameter tingkat kesulitan soal tes (b), parameter besarnya usaha untuk soal tes (beta), parameter kelambatan soal tes (d) memiliki RMSE dan SE yang sangat dekat nilainya. Parameter yang terdapat dalam peserta tes yaitu kecepatan peserta tes dan kemampuan peserta tes juga memiliki RMSE dan SE dengan kecenderungan pola yang sama untuk ketiga level banyaknya soal, dimana semakin banyak soal tes semakin kecil RMSE dan SE.

Berdasarkan Gambar 3 dapat disimpulkan bahwa banyaknya peserta tes tidak berpengaruh secara signifikan terhadap perbedaan hasil estimasi parameter-parameter yang terdapat dalam model simultan antara ML 1P dengan waktu respon. Berapapun banyaknya peserta tes yang dilibatkan dalam proses estimasi parameter memberikan RM-

SE yang sama atau sangat mendekati hasilnya. Kondisi tersebut berlaku untuk semua parameter yang terdapat dalam model, yaitu parameter tingkat kesulitan soal (b), besarnya usaha untuk soal ke-j ( $\beta$ ), kelambatan soal tes (d), kecepatan peserta tes ( $\tau$ ) dan kemampuan peserta tes ( $\theta$ ) hal tersebut menunjukkan bahwa metode estimasi parameter yang digunakan dalam penelitian ini cukup *robust* terhadap perbedaan banyaknya peserta tes.

Hal tersebut menunjukkan *performance* hasil estimasi parameter yang terdapat dalam soal tes (tingkat kesulitan soal, kelambatan soal, dan besarnya usaha untuk soal ke-j, tidak dipengaruhi oleh banyaknya soal tes. *Performance* hasil estimasi parameter dalam peserta tes (kecepatan dan kemampuan peserta tes) dipengaruhi oleh banyaknya soal tes, yang semakin banyak soal tes maka hasil estimasi parameternya akan semakin mendekati nilai parameter yang sebenarnya.

## Simpulan dan Saran

### Simpulan

Berdasarkan bias, SE, dan RMSE dapat disimpulkan bahwa proses estimasi dengan pendekatan Bayesian berjalan dengan baik. Besaran parameter estimasi akurat untuk digunakan sebagai ukuran parameter soal tes dan parameter peserta tes dalam pelaksanaan tes.

Hasil penelitian juga menunjukkan bahwa perbedaan banyaknya peserta tes (500 atau 1000) tidak berpengaruh secara signifikan terhadap hasil estimasi parameter-parameter yang terdapat dalam model simultan ML-1P dengan waktu respon. Kondisi tersebut berlaku untuk semua parameter yang terdapat dalam model. Banyaknya soal tes tidak mempengaruhi *performance* hasil estimasi parameter soal tes (tingkat kesulitan soal, kelambatan soal, dan besarnya usaha untuk soal ke-j). Berapapun banyak soal tes *performance* hasil estimasi pada parameter tersebut hampir sama. Banyaknya soal tes mempengaruhi *performance* hasil estimasi parameter peserta tes (kecepatan dan kemampuan peserta tes) yang semakin banyak soal

tes maka hasil estimasi parameternya akan semakin mendekati nilai parameter yang sebenarnya.

#### Saran

Hasil estimasi parameter dengan metode estimasi Bayesian sangat dipengaruhi oleh pemilihan distribusi prior. Distribusi prior yang berbeda akan menghasilkan besaran parameter estimasi yang berbeda. Kesimpulan yang berkaitan dengan besaran parameter estimasi dalam penelitian ini hanya berlaku untuk distribusi prior yang digunakan dalam penelitian. Penelitian berikutnya dapat menggunakan distribusi prior yang berbeda dengan yang digunakan dalam penelitian ini. Hasil yang diharapkan adalah besaran parameter estimasi semakin baik. Kriteria yang digunakan adalah besaran parameter estimasi lebih mendekati besaran parameter bangkitan, sehingga ukuran kriteria RMSE, bias dan SE menjadi semakin kecil atau semakin mendekati nol. Pendekatan atau metode estimasi yang lain dapat digunakan untuk mengestimasi parameter-parameter dalam model, misalnya Metode Maksimum Likelihood.

#### Daftar Pustaka

- Abdelfattah, F. A. (2007). *Response latency effects on classical and item response theory parameters using different scoring procedures*. Disertasi tidak dipublikasikan. Ohio University.
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item Parameter Estimation Under Conditions of Test Speededness: Application of a Mixture Rasch Model With Ordinal Constraints. *Journal of Educational Measurement*, 39(4), 331–348. <https://doi.org/10.1111/j.1745-3984.2002.tb01146.x>
- Entink, R. H. K. (2009). *Statistical Models for Response and Response Time*. Thesis, tidak dipublikasikan. University of Twente, Enchede.
- Entink, R. H. K., van der Linden, W. J., & Fox, J.-P. (2009). A Box-Cox normal model for response times. *The British Journal of Mathematical and Statistical Psychology*, 62(Pt 3), 621–640. <https://doi.org/10.1348/000711008X374126>
- Gvozdenko, E., & Chambers, D. (2007). Beyond test accuracy: Benefits of measuring response time in computerised testing. *Australasian Journal Of Educational Technology*, 23(4), 542–558. Retrieved from <http://dx.doi.org/10.14742/ajet.1251>
- Haryanto. (2013). Pengembangan computerized adaptive testing (CAT) Dengan algoritma logika fuzzy. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 15(1), 47–70. Retrieved from <http://journal.uny.ac.id/index.php/jpep/article/view/1087/1139>
- Hidayah, N., Kumaidi, & Kartowagiran, B. (2015). Integrated model of one parameter logistic model and response time model. In *Prosiding, International Conference and Workshop on Basic and Applied Sciences*. Surabaya: Universitas Airlangga.
- Hornke, L. F. (2000). Response times in CAT as an additional means to differential assessment. *Psicologica*, 21(1), 175–189. Retrieved from <https://www.uv.es/psicologica/articulos1y2.00/hornke.pdf>
- Ingrison, S. J. (2008). An extended item response theory model incorporating item response time. *Disertasi Doktor, tidak dipublikasikan*. Florida State University.
- Ingrison II, J. N. (2008). Modeling the joint distribution of response accuracy and response time. *Disertasi, tidak dipublikasikan*. Florida State University.
- Meng, X.-B., Tao, J., & Chang, H.-H. (2015). A Conditional Joint Modeling Approach for Locally Dependent Item Responses and Response Times. *Journal of Educational Measurement*, 52(1), 1–27. <https://doi.org/10.1111/jedm.12060>
- Meyer, J. P. (2010). A Mixture Rasch Model With Item Response Time

- Components. *Applied Psychological Measurement*, 34(7), 521–538.  
<https://doi.org/10.1177/0146621609355451>
- Meyer, J. P., & Wise, S. L. (2005). Item response time and distractor analysis including item response time in distractor analysis via multivariate kernel smooting. *Makalah* disajikan dalam The meeting of the National Council on Measurement in Education.
- Oshima, T. C. (1994). The Effect of Speededness on Parameter Estimation in Item Response Theory. *Journal of Educational Measurement*, 31(3), 200–219. <https://doi.org/10.1111/j.1745-3984.1994.tb00443.x>
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117(3), 864–901. Retrieved from <http://dx.doi.org/10.1037/a0019737>
- Schnipke, D. L., & Scrams, D. J. (1999). *Exploring issues of examinee behavior: Insights gained from response-time analyses*. Law School Admission Council Computerized Testing Report. LSAC Research Report Series.
- Schnipke, D. L., & Pashley, P. J. (1997). Assessing Subgroup Differences in Item Response Times. In *American Educational Research Association Annual Meeting 1997*. Retrieved from <https://www.learntechlib.org/p/81203>
- Thissen, D. (1983). Latent trait scoring of timed ability tests, university of kansas. Retrieved November 25, 2012, from <http://www.psych.umn.edu/psylab>.
- van der Linden, W. J. (2006). A Lognormal Model for Response Times on Test Items. *Journal of Educational and Behavioral Statistics*, 31(2), 181–204. <https://doi.org/10.3102/10769986031002181>
- van der Linden, W. J. (2007). A Hierarchical Framework for Modeling Speed and Accuracy on Test Items. *Psychometrika*, 72(3), 287. <https://doi.org/10.1007/s11336-006-1478-z>
- van der Linden, W. J., & Hambleton, R. K. (Eds). (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.
- van der Linden, W. J., & van Krimpen-Stoop, E. M. L. A. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika*, 68(2), 251–265. <https://doi.org/10.1007/BF02294800>
- Wang, T. (2006). *A model for the joint distribution of item response and response time using one-parameter weibull distribution*. CASMA Research Report 20. Iowa City, IA : Center for Advance Studies in Measurement and Assessment.
- Wang, T., & Hanson, B. A. (2005). Development and Calibration of an Item Response Model That Incorporates Response Time. *Applied Psychological Measurement*, 29(5), 323–339. <https://doi.org/10.1177/0146621605275984>
- Wise, S. L., & DeMars, C. E. (2006). An Application of Item Response Time: The Effort-Moderated IRT Model. *Journal of Educational Measurement*, 43(1), 19–38. <https://doi.org/10.1111/j.1745-3984.2006.00002.x>
- Wollack, J. A., & Woo, V. (2009). Using response time to improve parameter estimation for speeded test item. *Paper* in The annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Yamamoto, K. (1995). *Estimating the effects of test length and test time on parameter estimation using the hybrid model*. ETS Technical Report TR-10, Princeton, NJ: Educational Testing Service.