

BAB V

SIMPULAN DAN SARAN

A. Simpulan

Berdasarkan hasil analisis data dan pembahasan pada Bab IV, maka didapatkan beberapa simpulan seperti berikut ini.

1. Penelitian ini menghasilkan produk berupa model matematis yang dikembangkan untuk menjelaskan kondisi tes dengan sistem terkomputerisasi, sebagai berikut:

$$f(x_{ij}|t, \theta, a, b, \tau, \eta) = \left[P(x=1|\theta_j, t_{ij}, a_j, b_j, \tau_{ij}, \eta) \right]^{x_{ij}} \left[P(x=0|\theta_j, t_{ij}, a_j, b_j, \tau_{ij}, \eta) \right]^{1-x_{ij}} \cdot f(t_{ij}|\xi, \tau, \sigma)$$

dengan,

$$P(x=1|\theta_j, t_{ij}, a_j, b_j, \tau_{ij}, \eta) = \frac{\exp(1.7(a_j(\theta_i - \tau_i - \eta \left(\frac{1}{t_{ij}} \right) - b_j)))}{1 + \exp(1.7(a_j(\theta_i - \tau_i - \eta \left(\frac{1}{t_{ij}} \right) - b_j)))}$$

dan

$$f(t_{ij}|\xi, \tau, \sigma) = \left\{ \frac{1}{t_{ij}\sigma\sqrt{2\pi}} \exp \left(-\frac{1}{2} \left[\frac{(\ln t_{ij} - (\xi_j - \tau_i))^2}{\sigma^2} \right] \right) \right\}$$

Model matematis tersebut berupa Model Logistik 2 Parameter dengan variabel random waktu respon yang dihasilkan secara simultan yang totalnya disusun oleh 6 parameter yaitu kemampuan peserta tes (θ_i), waktu respon peserta tes (t_{ij}), tingkat kesulitan soal tes (b_j), daya

beda soal tes (a_j), *time intensity* untuk soal tes ke- j (ξ_j), dan

kecepatan peserta tes dalam menyelesaikan soal (τ_i).

2. Berdasarkan pengujian ketepatan estimasi parameter dengan kriteria bias, SE, RMSE dan korelasi, dapat disimpulkan bahwa hasil estimasi parameter akurat digunakan sebagai ukuran *item parameter* dan *person parameter*. Banyaknya peserta tes berpengaruh terhadap hasil estimasi *item parameter* (semakin banyak peserta tes, maka hasil estimasi *item parameter* semakin akurat), sedangkan banyaknya butir soal tes berpengaruh terhadap hasil estimasi *person parameter* (semakin banyak butir soal tes, maka hasil estimasi *person parameter* semakin mendekati nilai parameter sebenarnya).
3. Untuk *person parameter* (theta), hasil perbandingan nilai kriteria SE dan RMSE menunjukkan bahwa Model 2PL dengan variabel random waktu respon (hasil pengembangan) lebih tepat dalam mengungkap kondisi tes yang memperhitungkan waktu respon sebagai indikator kemampuan peserta tes jika dibandingkan model IRT dan model simultan dari Hidayah et al. (2016), sehingga tujuan utama dari penelitian ini yaitu untuk menghasilkan model yang dapat meningkatkan keakuratan estimasi parameter kemampuan peserta tes (θ/Θ) tercapai.
4. Berdasarkan pengujian kecocokan dengan kriteria nilai DIC terendah, model yang paling cocok untuk mengungkapkan kondisi *real* tes terkomputerisasi adalah Model Logistik 2 Parameter dengan variabel

random waktu respon (hasil pengembangan), sehingga dari penelitian ini didapatkan model yang lebih cocok untuk menggambarkan keadaan tes sebenarnya, dibandingkan model IRT dan model simultan dari Hidayah et al. (2016). Selain itu berdasarkan kriteria parameter daya beda (a) dan tingkat kesulitan (b) dapat diketahui bahwa butir soal yang tergolong baik atau ideal adalah butir soal nomor 12, 14, 17, 18, 19 dan 21, sedangkan kurun waktu ideal yang dibutuhkan untuk mengerjakan masing-masing soal Tes Potensi Akademik (TPA) dengan subtes analogi adalah 72,78 detik. Tes ini juga paling cocok untuk peserta tes dengan estimasi kemampuan (θ) sebesar -2,2.

B. Saran

Berikut ini adalah saran-saran yang bisa peneliti berikan untuk pengembangan model logistik dengan variabel random selanjutnya.

1. Parameter *time intensity* pada penelitian ini hanya diasumsikan sebagai *fixed variable*, padahal pada kenyataannya parameter *time intensity* bisa dipengaruhi oleh beberapa faktor, seperti tingkat kesulitan soal, kompleksitas soal dan banyaknya kata dalam soal. Oleh karena itu pada penelitian selanjutnya bisa dipertimbangkan untuk pendefinisian kembali parameter *time intensity* sebagai *random variable* sehingga didapatkan model yang lebih realistis lagi dalam mengungkapkan kondisi tes sebenarnya
2. Pengujian kecocokan distribusi waktu respon terbatas dengan cara membandingkan nilai *Anderson Darling* (AD) pada beberapa distribusi

yang karakteristiknya dianggap sesuai dengan waktu respon (nilainya positif dan bentuknya *positive skewed*), seperti distribusi *Weibull*, *Gamma* dan *Lognormal*. Pada penelitian selanjutnya akan lebih baik lagi jika ukuran kecocokan distribusinya tidak hanya sekedar dengan menampilkan nilai AD saja, tetapi bisa juga menampilkan bentuk histogram dari masing-masing data waktu respon dari sejumlah butir yang akan dianalisis sebagai informasi tambahan, sehingga bisa juga dibuktikan secara visual bahwa pola distribusi data memang *positive skewed*.

3. Pemilihan distribusi *prior* pada proses estimasi dengan menggunakan metode *Bayesian* adalah subyektivitas dari peneliti, jadi antar peneliti bisa mendefinisikan suatu parameter dengan distribusi *prior* yang berbeda. Ada baiknya jika penelitian selanjutnya menggunakan distribusi *prior* yang berbeda dengan penelitian ini, sehingga hasilnya bisa dibandingkan dan dengan cara ini akan didapatkan distribusi *prior* yang lebih tepat lagi. Selanjutnya untuk metode estimasi parameternya juga bisa dicoba untuk menerapkan metode selain *Bayesian Markov Chain Monte Carlo* dengan algoritma *Gibbs Sampling* sehingga bisa dibandingkan metode mana yang paling *robust* dalam mengestimasi parameter.
4. Untuk penelitian selanjutnya model bisa dicobakan pada data *power test* dan *speed test* sehingga bisa diketahui kecenderungan model lebih cocok untuk diterapkan pada tes jenis apa. Kriteria kecocokan model

pada penelitian ini terbatas menggunakan ukuran DIC (*Deviance Information Criterion*), padahal ada ukuran kecocokan model lainnya yang bisa disertakan untuk memperkuat hasil keputusan terkait kecocokan model, seperti grafik PPMC (*Posterior Predictive Model Checking*). Walaupun metode ini tidak menghasilkan nilai statistik seperti DIC, namun dari grafik PPMC ini akan dapat dilihat sejauh apa selisih antara data empiris dengan hasil prediksi model berdasarkan skor jawaban benar soal dari kedua garis pada grafik PPMC.