

BAB III

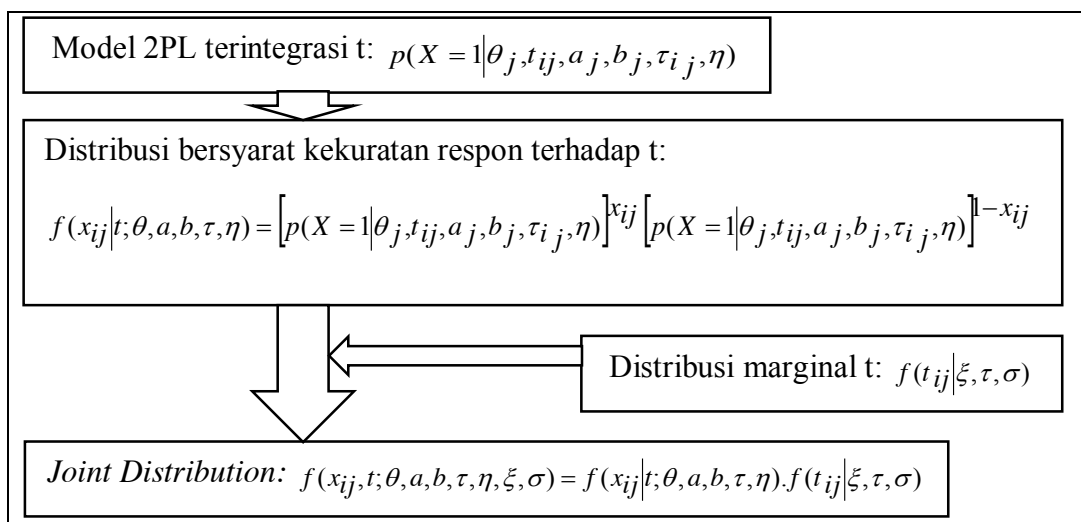
METODE PENELITIAN

A. Model Pengembangan

Jenis penelitian ini merupakan penelitian pengembangan. Penelitian pengembangan ini bertujuan untuk menghasilkan produk berupa model matematis yaitu model logistik dengan variabel random waktu respon yang dihasilkan secara simultan. Pengembangan model ini dilakukan dengan mempertimbangkan hasil kajian secara empiris dan teoritis.

Apabila dikaji secara empiris, suatu tes dalam pelaksanaannya selalu dibatasi waktu, sehingga ada faktor waktu yang seharusnya dilibatkan untuk mengungkapkan kondisi *real* dalam pelaksanaan suatu tes. Waktu respon dapat dipergunakan sebagai informasi tambahan bagi penyelenggara tes dalam memutuskan seorang peserta tes lulus atau gagal. Jika pada saat lampau mencatat waktu respon adalah hal yang sulit, maka pada era kemajuan Teknologi dan Informasi (TI) seperti saat ini bukanlah hal yang mustahil untuk melakukan pencatatan waktu respon dengan bantuan komputer. Oleh karena itu pemodelan yang melibatkan waktu respon ini sangat perlu dipertimbangkan untuk diterapkan, karena menurut para peneliti sebelumnya dapat memperbaiki hasil estimasi parameter kemampuan peserta tes khususnya pada model *Item Response Theory* (IRT) yang tidak memperhitungkan waktu respon sebagai indikator kemampuan peserta tes (Abdelfattah & Johanson, 2007; Oshima, 1994; Schnipke & Pashley, 1997; Schnipke & Scrams, 2002; Sinharay, 2018).

Selanjutnya apabila dikaji secara teoritis, setiap model waktu respon yang dikembangkan mempunyai spesifikasi tersendiri dalam penerapannya. Setiap model pasti mempunyai kekurangan dan kelebihan, sehingga tidak ada model standar yang dapat digunakan untuk memodelkan semua kasus waktu respon dalam berbagai tes. Selain itu, fakta lain juga menunjukkan bahwa distribusi waktu respon dari berbagai item soal bisa beragam, sehingga satu model tertentu tidak akan berlaku untuk segala macam tes (Ranger & Kuhn, 2012). Berdasarkan alasan tersebut maka pengembangan model yang memperhitungkan waktu respon sangat perlu terus dilakukan agar model semakin realistik dalam penerapannya. Pengembangan model ini dilakukan dengan cara memperbaiki kelemahan proses pengembangan model simultan dari Hidayah et al. (2016), yaitu melalui proses seleksi parameter yang masuk ke dalam model baru, pemilihan distribusi marginal waktu respon yang tepat dan pengujian kecocokan model yang dihasilkan dengan data empiris.



Gambar 7. Model 2PL dengan Variabel Random Waktu Respon

Model yang dikembangkan merupakan model simultan yang didapat dengan konsep *joint distribution*, yaitu perkalian antara model distribusi bersyarat keakuratan respon terhadap waktu respon dengan model distribusi marginal dari waktu respon. Model distribusi bersyarat keakuratan respon terhadap waktu respon adalah modifikasi dari model logistik terintegrasi waktu respon hasil pengembangan Ingrisone et al. (2008), Ingrisone II et al. (2008) dan Hidayah et al. (2016), sedangkan model distribusi marginal dari waktu respon adalah modifikasi dari model hasil pengembangan van der Linden (2009), van der Maas et al. (2011) dan Hidayah et al. (2016). Waktu respon di sini diasumsikan sebagai *random variable*, karena pada kenyataannya waktu respon tidak hanya dikontrol oleh petugas administrasi tes atau teknologi komputer saja tetapi ada faktor lain yang mempengaruhi seperti strategi dan *speed* yang digunakan oleh peserta tes untuk menyelesaikan suatu soal, sehingga apabila seseorang menjalankan tes yang sama dengan waktu pelaksanaan tes yang berbeda, maka waktu respon yang ditempuh bisa jadi akan berbeda dengan tes yang dilaksanakan sebelumnya (Veldkamp et al., 2017: 693).

B. Prosedur Pengembangan

Seperti yang telah dijelaskan sebelumnya, tujuan utama dari penelitian adalah mengembangkan model matematis secara simultan yang dapat meningkatkan keakuratan estimasi parameter kemampuan (θ/Θ) dengan memperhitungkan waktu respon sebagai indikator kemampuan, sehingga dapat

menjelaskan kondisi *real* tes dengan lebih baik. Pengembangan model simultan ini dilakukan dengan beberapa tahapan prosedur.

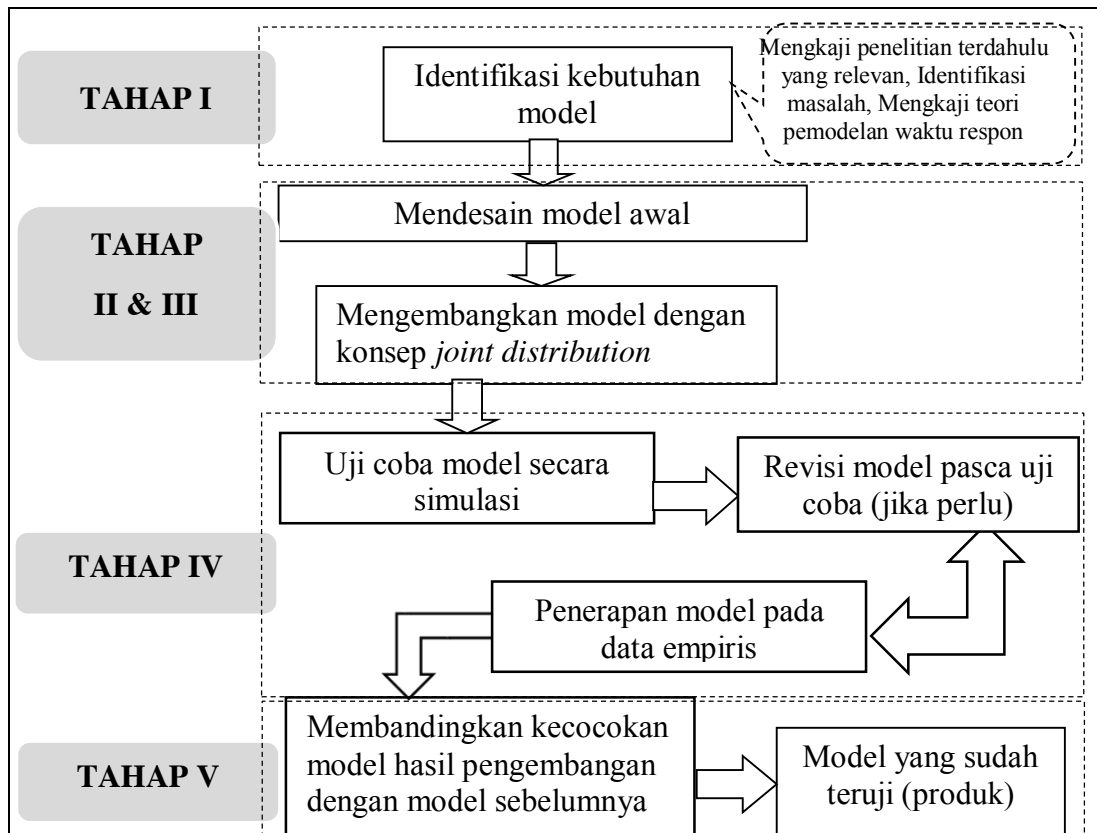
Tahap I, peneliti mengidentifikasi kebutuhan dalam penelitian. Pada tahapan ini peneliti mengkaji hasil penelitian terdahulu yang relevan, mengkaji literatur yang terkait dengan pengembangan model waktu respon, dan mengidentifikasi faktor-faktor apa yang menimbulkan permasalahan (kelemahan model waktu respon sebelumnya) sehingga perlu ada pengembangan model waktu respon yang baru.

Tahap II dan Tahap III, peneliti mendesain dan mengembangkan model. Pada Tahap II peneliti merumuskan model awal untuk memecahkan masalah yang telah ditemukan pada Tahap I. Model dirumuskan dengan konsep *joint distribution* antara model distribusi bersyarat keakuratan respon terhadap waktu respon dengan model distribusi marginal dari waktu respon. Pada Tahap III model dikembangkan dengan cara memformulasikan terlebih dahulu: (1) Model distribusi bersyarat keakuratan respon terhadap waktu respon berdasarkan kajian teori terkait; dan (2) Model waktu respon berdasarkan karakteristik distribusi marginal yang sesuai (pengujian distribusi dengan statistik uji *Anderson Darling*). Produk awal yang dihasilkan penelitian ini berupa Model Logistik 2 Parameter dengan variabel random waktu respon.

Tahap IV, peneliti melakukan penerapan model. Setelah model pada Tahap III siap untuk digunakan kegiatan selanjutnya adalah melakukan uji coba model tersebut. Uji coba dilakukan secara simulasi, dengan skenario yang sudah dibuat oleh peneliti. Parameter model yang dihasilkan, diestimasi dengan metode *Bayesian*

Markov Chain Monte Carlo (MCMC), dengan cara: (1) Menentukan distribusi *likelihood* yang sesuai dengan data; (2) Menentukan distribusi *prior* yang sesuai dengan setiap parameter yang ada dalam model; (3) Menghasilkan bentuk distribusi *posterior*; dan (4) Menggunakan algoritma *Gibbs Sampling* untuk membangkitkan nilai parameter berdasarkan distribusi *posterior* yang dihasilkan. Selanjutnya, melihat akurasi proses estimasi parameter model berdasarkan kriteria bias, SE (*Standard Error*), RMSE (*Root Mean Square Error*) dan korelasi. Model dapat disimpulkan layak untuk digunakan atau diterapkan jika mempunyai tingkat akurasi yang tinggi atau baik (bias, SE, RMSE mendekati 0 dan korelasi mendekati 1. Uji coba secara simulasi ini penting dilakukan untuk mengantisipasi kesalahan yang dapat terjadi selama penerapan model yang sesungguhnya berlangsung. Pada tahap ini juga model yang dihasilkan dibandingkan keakuratan parameter kemampuannya dengan model yang sudah dikembangkan sebelumnya (Hidayah et al., 2016) , sekaligus dengan model IRT. Dari hasil uji coba secara simulasi ini, juga dapat dilihat seberapa tepat Model Logistik 2 Parameter dengan variabel random waktu respon diterapkan pada tahap berikutnya.

Pada Tahap V dilakukan evaluasi model dengan cara membandingkan kecocokan model simultan yang dihasilkan dengan model *Item Response Theory* (IRT) menggunakan kriteria *Deviance Information Criterion* (DIC), dan memutuskan bahwa model dengan DIC terendah adalah model yang paling cocok untuk data empiris. Revisi model dapat dilakukan setelah model tersebut diterapkan atau diujikan di lapangan. Revisi ini dilakukan apabila ditemukan hasil yang kurang maksimal pada saat penerapan.

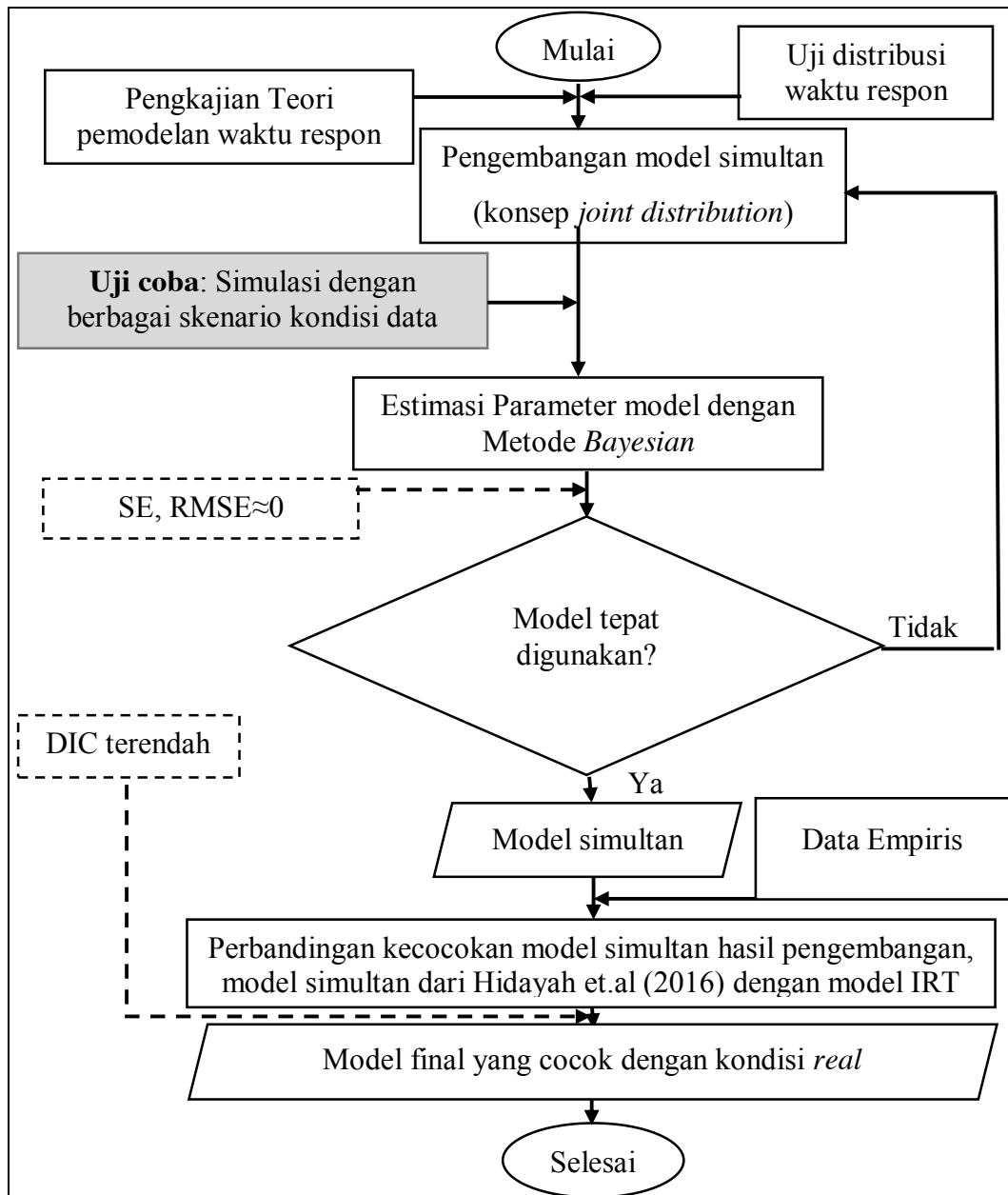


Gambar 8. Prosedur Pengembangan Model

C. Desain Uji Coba Produk

1. Desain Uji Coba

Uji coba Model Logistik 2 Parameter dengan variabel random waktu respon dilakukan untuk mengetahui apakah model yang dihasilkan tersebut memang cocok untuk tes terkomputerisasi.



Gambar 9. *Flowchart* Pengujian Produk

Uji coba dilakukan secara simulasi, dengan skenario Tabel 1 yang sudah dibuat oleh peneliti. Simulasi dilakukan untuk mendapatkan kesimpulan secara general dari berbagai macam kondisi data. Seperti yang sudah dijelaskan sebelumnya, data simulasi ini sifatnya ideal karena dibangkitkan berdasarkan suatu

algoritma. Uji coba dengan data simulasi ini penting dilakukan, karena dapat mengantisipasi kesalahan yang dapat terjadi selama penerapan model pada kondisi *real* di tahap selanjutnya. Untuk skenario pembangkitan data simulasi dapat dilihat pada Tabel 1.

Tabel 1. Skenario Simulasi

Jumlah Peserta Tes	Banyaknya Soal		
	10 Butir (Tes Pendek)	20 Butir (Tes Sedang)	40 Soal (Tes Panjang)
300 <i>Testee</i> (Sampel Kecil)	S ₁₁	S ₁₂	S ₁₃
600 <i>Testee</i> (Sampel Sedang)	S ₂₁	S ₂₂	S ₂₃
1200 <i>Testee</i> (Sampel Besar)	S ₃₁	S ₃₂	S ₃₃

Dengan mempertimbangkan banyaknya parameter yang harus dilibatkan dan keterbatasan kemampuan analisis data *software* yang digunakan, maka ada 9 skenario simulasi yang memungkinkan untuk diterapkan peneliti. Banyaknya butir soal dibagi menjadi 3 kategori, 10 soal untuk mewakili tes pendek, 20 soal untuk mewakili tes sedang dan 40 soal untuk mewakili tes panjang. Jumlah peserta tes terbatas pada 300 *testee* untuk mewakili kategori sampel kecil, 600 *testee* untuk mewakili kategori sampel sedang dan 1200 untuk mewakili kategori sampel besar.

2. Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data simulasi dan data empiris berupa pola respon peserta tes yang berbentuk dikotomis, dimana jawaban benar diberikan kode 1 dan jawaban salah diberikan kode 0, dan data waktu respon peserta tes dalam satuan detik. Data simulasi berasal dari suatu pembangkitan

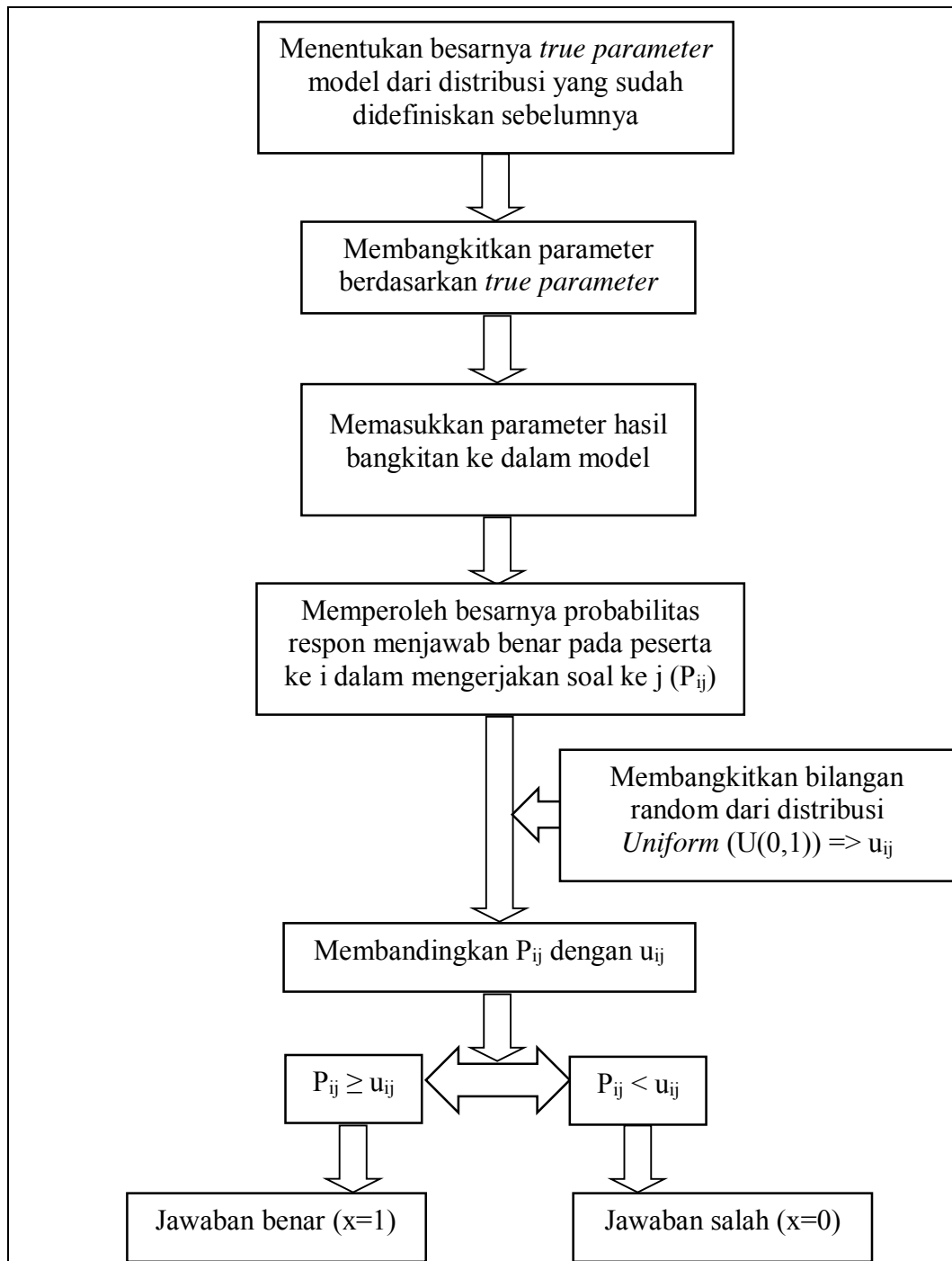
dengan algoritma yang dihasilkan dalam suatu model matematis. Untuk proses simulasi data, ditunjukkan pada Gambar 10.

Data empiris yang akan digunakan dalam penelitian ini merupakan data sekunder yang didapatkan dari Unit Pelaksana Teknis (UPT) Pusat Komputer melalui kantor Layanan Admisi Universitas Negeri Yogyakarta (UNY) berupa berupa data pola respon dari 30 soal tes dan data waktu respon dari tiap peserta tes seleksi penerimaan mahasiswa baru Pascasarjana UNY tahun 2017 dalam menyelesaikan tiap butir soal tes terkomputerisasi (CBT) yang berjumlah 717 orang sebagai sampel.

Soal seleksi masuk Program Pascasarjana UNY yang datanya diterima oleh peneliti berupa Tes Potensi Akademik (TPA) yang terdiri 3 macam soal, pertama soal untuk menguji kemampuan verbal (bahasa), kedua soal untuk menguji kemampuan analogi, ketiga soal untuk menguji kemampuan numerik (angka). Peneliti memilih untuk menggunakan data hasil tes analogi yang berjumlah 30 soal, karena dari ketiga macam tes tersebut data hasil tes analogi-lah yang paling lengkap datanya. Penerapan data empiris pada model ini akan sangat berguna untuk menguji apakah model tersebut mampu menjelaskan kondisi *real* dalam praktik pengukuran yang mempertimbangkan waktu tes.

Waktu respon (t_{ij}) yang dicatat dalam penelitian ini, dihitung dari waktu peserta tes i mulai mengklik butir soal ke j untuk dibaca, dijawab, sampai dengan klik ke soal selanjutnya. Apabila peserta tes i kesulitan atau ragu untuk menjawab butir soal ke j dan memilih untuk kembali lagi menjawab soal ke j tersebut setelah selesai mengerjakan soal lainnya, maka waktu responnya adalah total waktu saat

peserta tes i mulai mencoba mengerjakan butir soal ke j untuk pertama kalinya sampai dengan mencoba untuk kedua kalinya, dan seterusnya sampai dengan waktu tes habis.



Gambar 10. Proses Simulasi Data

3. Teknik Analisis Data

Seperti yang sudah dijelaskan sebelumnya, penelitian ini bermaksud untuk mengembangkan Model Logistik 2 Parameter dengan variabel random waktu respon secara simultan. Selain dilihat keakuratannya dengan memanfaatkan data bangkitan simulasi, model juga dilihat kecocokannya dengan menggunakan data empiris.

Model dengan data bangkitan simulasi dianalisis dengan cara melihat bias, ketepatan, dan keakuratan proses estimasi parameter modelnya. Ukuran keakuratan proses estimasi parameter model dapat dilihat dari nilai SE (*Standard Error*). Untuk ukuran keakuratan model dapat dilihat dari RMSE (*Root Mean Square Error*). Ukuran keakuratan proses dan hasil estimasi parameter model dapat dilihat dari korelasi antara *true parameter* dengan rata-rata parameter hasil estimasi, dapat digunakan kriteria korelasi. Apabila bias, SE, RMSE nilainya mendekati 0 dan nilai korelasinya mendekati 1, maka dapat dimaknai bahwa model, proses dan parameter yang dihasilkan sudah akurat. Analisis dengan data bangkitan simulasi ini dilakukan sesuai dengan skenario pada Tabel 1, dengan jumlah replikasi sebanyak 30 kali dan iterasi sebanyak 15000 untuk masing-masing skenario. Berikut ini adalah persamaan untuk mendapatkan bias, SE, RMSE dan korelasi (Hidayah et al., 2016: 84-85).

Besarnya bias dapat dinyatakan dalam persamaan 31 berikut ini:

$$Bias(\hat{\gamma}_m) = \frac{1}{R} \sum_{r=1}^R \hat{\gamma}_{mr} - \gamma_m . \quad (31)$$

Nilai bias dihitung di tiap soal ($m=1, 2, 3, \dots, M$). Bias adalah rata-rata hasil estimasi parameter model ($\hat{\gamma}_{mr}$) sebanyak R kali replikasi ($r=1, 2, 3, \dots, R$) dikurangi dengan besarnya nilai *true parameter* (γ_m).

Besarnya SE dapat dinyatakan dalam persamaan 32 berikut ini:

$$SE(\hat{\gamma}_m) = \sqrt{\frac{1}{R} \sum_{r=1}^R \left(\hat{\gamma}_{mr} - \frac{1}{R} \sum_{r=1}^R \hat{\gamma}_{mr} \right)^2}. \quad (32)$$

Nilai SE dihitung di tiap soal ($m=1, 2, 3, \dots, M$). *Standard Error* adalah akar dari rata-rata penjumlahan sebanyak R kali replikasi ($r=1, 2, 3, \dots, R$) dari selisih kuadrat antara hasil estimasi parameter model ($\hat{\gamma}_{mr}$) dengan rata-rata hasil estimasi parameter model ($\hat{\gamma}_{mr}$) sebanyak R kali replikasi ($r=1, 2, 3, \dots, R$).

Besarnya RMSE dapat dinyatakan dalam persamaan 33 berikut ini:

$$RMSE(\hat{\gamma}_m) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\gamma}_{mr} - \gamma_m)^2}. \quad (33)$$

Nilai RMSE dihitung di tiap soal ($m=1, 2, 3, \dots, M$). *Root Mean Square Error* adalah akar dari rata-rata penjumlahan sebanyak R kali replikasi ($r=1, 2, 3, \dots, R$) dari selisih kuadrat antara hasil estimasi parameter model ($\hat{\gamma}_{mr}$) dengan besarnya nilai *true parameter* (γ_m).

Besarnya korelasi dapat dinyatakan dalam persamaan 34 berikut ini:

$$\rho_{\hat{\gamma}\gamma} = \frac{\sum_{m=1}^M \gamma_m \bar{\hat{\gamma}}_m - \sum_{m=1}^M \gamma_m \sum_{m=1}^M \bar{\hat{\gamma}}_m}{\sqrt{\left[\sum_{m=1}^M \gamma_m^2 - (\sum_{m=1}^M \gamma_m)^2 \right] \left[\sum_{m=1}^M \bar{\hat{\gamma}}_m^2 - (\sum_{m=1}^M \bar{\hat{\gamma}}_m)^2 \right]}}. \quad (34)$$

Korelasi antara *true parameter* (γ_m) dengan rata-rata hasil estimasi parameter ($\hat{\gamma}_{mr}$) dihitung agar dapat dilihat sedekat apa hubungan antara *true parameter* (γ_m) dengan rata-rata hasil estimasi parameter ($\hat{\gamma}_{mr}$). Semakin besar nilai korelasinya maka proses dan hasil estimasi parameter modelnya semakin akurat.

Model dengan data empiris dianalisis dengan melakukan perbandingan kecocokan antara model simultan (*joint distribution model*) dengan model *Item Response Theory* (IRT) berdasarkan kriteria *Deviance Information Criterion* (DIC). Model yang paling cocok adalah model dengan nilai DIC terendah.

Salah satu cara untuk membandingkan model adalah dengan melihat kecocokan antara data dengan model yang dibandingkan dan kompleksitas dari masing-masing model. Kriteria perbandingan model *Bayesian* dapat dinyatakan dalam persamaan berikut ini (Spiegelhalter et al., 2002, 2014).

$$DIC = \text{goodness of fit} + \text{complexity} \quad (35)$$

Ukuran *goodness of fit* dapat dilihat dari besarnya *deviance*.

$$D(\theta) = -2 \log L(\text{data}|\theta) \quad (36)$$

Ukuran *complexity* dapat dilihat dari keefektifan jumlah parameter yang efektif.

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta}) \quad (37)$$

Sehingga DIC dapat didefinisikan kembali menjadi:

$$DIC = \bar{D} + p_D. \quad (38)$$