

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Pada era kemajuan TI (Teknologi dan Informasi) saat ini, sangat memungkinkan untuk menyelenggarakan suatu tes dengan menggunakan bantuan komputer (tes terkomputerisasi), seperti CBT (*Computerized-Based Testing*) dan CAT (*Computerized-Adaptive Testing*). Desain CBT sebenarnya sama dengan *Paper Based Test* (PBT), namun pada CBT pengerjaannya menggunakan komputer dan tidak perlu penggandaan soal tes seperti PBT (Istiyono et al., 2018: 145). Pada CAT setiap peserta tes memperoleh butir-butir soal yang sesuai dengan tingkat kemampuannya (Hadi, 2013: 12). Kesesuaian butir-butir soal tersebut didasarkan pada respon jawaban yang diberikan oleh peserta tes (Haryanto, 2011: 50). Jadi sangat memungkinkan kalau antar peserta tes menyelesaikan tes dengan butir soal yang tidak sama karena kemampuan peserta tes yang berbeda-beda. Pada CAT, apabila peserta tes dapat menjawab soal sebelumnya dengan benar, maka akan mendapatkan soal yang lebih sulit lagi dan sebaliknya jika peserta tes tidak dapat menjawab soal sebelumnya dengan benar maka dia akan mendapatkan soal yang lebih mudah.

Pada CBT dan PBT, waktu pengerjaan tes dan jumlah butir soal sudah ditentukan sebelumnya, oleh karena itu tidak diperlukan bank soal yang banyak, hanya saja pada CBT hasil tesnya sudah dapat langsung diketahui karena proses *scoring*-nya dilakukan oleh komputer melalui perintah program sedangkan pada PBT memerlukan waktu dalam penilaiannya. Pada CAT, waktu pengerjaannya

akan berhenti apabila estimasi kemampuan peserta tes sudah bisa diketahui (dengan estimasi *error* 1%). Pada CAT peserta tes akan mendapatkan penilaian secara langsung berdasarkan tingkat kesulitan yang diterima. Penerapan di lapangan membuktikan bahwa tes terkomputerisasi seperti CBT dan CAT terbukti lebih unggul dibandingkan PBT. Keuntungan pelaksanaan tes terkomputerisasi adalah fleksibilitas manajemen tes, peningkatan keamanan tes, peningkatan motivasi dalam literasi TI, dan efisiensi waktu (Georgiadou et al., 2006: 3). Tes terkomputerisasi ini dapat memberikan informasi yang sifatnya kompleks, karena selain informasi tentang pola respon peserta tes, tes dengan menggunakan bantuan komputer juga dapat memberikan informasi tentang waktu respon yang digunakan peserta tes dalam menjawab suatu butir soal (Lee & Chen, 2011: 360).

Waktu respon (*response time*) atau yang kerap kali disebut sebagai *response latency* (Abdelfattah & Johanson, 2007; Halkitis, 1996), adalah waktu yang dibutuhkan oleh peserta tes untuk membaca sekaligus menjawab suatu butir soal tes (Verbic & Tomic, 2009: 1). Suatu tes dalam pelaksanaannya pasti dibatasi waktu, sehingga ada faktor waktu yang seharusnya dilibatkan sebagai parameter (dicatat atau tidak) untuk mengungkapkan kondisi *real* dalam pelaksanaan suatu tes. Waktu respon merupakan variabel penting yang patut dipertimbangkan dalam memahami perilaku (*performance*) peserta tes, karena waktu respon dapat mendeteksi bagaimana peserta tes memproses soal atau strategi apa yang digunakan, apakah hanya sekedar menebak atau dengan hati-hati membaca dan menjawab setiap butir soal. Perilaku menebak kemungkinan dilakukan sebagai strategi mengerjakan soal ketika peserta tes menyadari bahwa mereka kehabisan

waktu (Bridgeman & Cline, 2004: 137; Hadadi & Luecht, 1998: 47). Selain itu waktu respon juga bisa menjadi bukti yang sangat informatif untuk validitas nilai tes, karena waktu respon dapat memberikan informasi pada nomor berapa peserta tes mengerjakan soal berlama-lama dan pada nomor berapa peserta tes dapat mengerjakan soal dengan cepat (Li et al., 2017: 159).

Di Indonesia, sistem tes dengan bantuan komputer sudah mulai diterapkan, seperti adanya ujian nasional yang lebih dikenal sebagai UNBK (Ujian Nasional Berbasis Komputer) yang mulai dilaksanakan pada tahun 2014 secara online, walaupun pada saat itu masih terbatas pada SMP Indonesia Singapura dan SMP Indonesia Kuala Lumpur. Penggunaan komputer dalam ujian nasional ini menunjukkan bahwa Indonesia sudah mulai masuk dalam babak baru, karena keberanian Kemdikbud (Kementerian Pendidikan dan Kebudayaan) untuk menerapkan tes terkomputerisasi dengan CBT dalam ujian nasional, yang merupakan terobosan untuk kemajuan pendidikan di Indonesia, selain itu adanya UNBK juga dapat mendorong meningkatnya literasi siswa terhadap TI. Pemerintah berusaha untuk menyeleraskan pendidikan dengan tuntutan jaman dan diberlakukan secara proporsional, kontekstual sesuai dengan kemajuan daerah.

Sampai dengan saat ini sistem ujian nasional memang masih dilaksanakan dengan 2 macam yaitu sistem terkomputerisasi (CBT) dan PBT, tetapi dengan melihat keunggulan tes dengan menggunakan komputer, maka bukanlah hal yang mustahil jika pemerintah melaksanakan UNBK secara terus menerus (walaupun dalam perjalanannya sangat mungkin berganti nama) sampai akhirnya akan menjadi suatu kebutuhan, dan akan meninggalkan ujian nasional dengan sistem

PBT. Salah satu yang menjadi alasan pemerintah menerapkan ujian nasional dengan sistem terkomputerisasi adalah untuk meminimalkan tingkat kecurangan, hal ini sejalan dengan motto Ujian Nasional 2015/2016 yaitu, “Ujian Nasional Jujur dan Berprestasi (Prestasi Ya, Jujur Harus)”. Ujian nasional dengan sistem terkomputerisasi dapat membuat dimungkinkannya pencatatan informasi waktu respon, selain informasi tentang pola respon peserta tes dalam menjawab suatu butir soal.

Selain ujian nasional sekolah, beberapa tahun terakhir ini ujian masuk perguruan tinggi juga memulai babak barunya, dengan ikut menyelenggarakan tes seleksi masuk berbasis komputer. Universitas Negeri Yogyakarta (UNY) adalah salah satu perguruan tinggi di Indonesia yang mulai tahun 2017 melakukan inovasi dalam seleksi calon mahasiswa baru. Pada awalnya seleksi memang dilakukan dengan PBT, namun seiring perkembangan teknologi, tes seleksi masuk diubah menjadi CBT. Melalui inovasi ini dimungkinkan pelaksanaan tes bisa menjadi lebih efektif dan efisien.

Dalam suatu tes, akan selalu ada dua parameter utama yang dilibatkan untuk menggambarkan bagaimana pengerjaannya. Pertama adalah akurasi respon, yang menunjukkan seberapa benar peserta tes mengerjakan soal, dan yang kedua adalah waktu respon, yang dapat menunjukkan ukuran *speed* dari peserta tes dalam mengerjakan soal. *Speed* tidaklah mengukur konstruk yang sama dengan keakuratan (Goldhammer & Entink, 2011: 108). Dari kedua parameter inilah akan bisa didapatkan keputusan peserta tes tersebut lulus atau gagal maupun masuk kategori cepat atau lambat (Dodonova & Dodonov, 2013: 1). Catatan waktu respon

yang dihasilkan dalam suatu tes akan menjadi penting untuk dipertimbangkan dalam pengambilan keputusan (Rubinstein, 2013: 540). Pada suatu tes, catatan waktu respon tersebut dapat memberikan informasi tambahan bagi lembaga penyelenggara tes, terutama pada saat ada dua peserta tes yang memiliki respon sama terhadap suatu butir soal, tetapi waktu yang dibutuhkan untuk menyelesaikan butir soal tersebut berbeda, sehingga dengan adanya catatan waktu respon dapat mempermudah pelaksanaan seleksi. Tetapi sayangnya masih banyak lembaga penyelenggara tes (khususnya di Indonesia) yang belum memanfaatkan informasi catatan waktu respon tersebut dalam proses seleksi.

Kekurangan pemodelan IRT (*Item Response Theory*) sebagai pendekatan teori tes modern adalah, informasi yang diberikan hanya terbatas pada keakuratan yaitu seberapa benar atau salah jawaban yang diberikan oleh peserta tes saja tanpa mempertimbangkan aspek *speed* (Hambleton & Swaminathan, 2013: 30), sedangkan pemodelan waktu respon dapat memberikan informasi *speed* atau seberapa cepat peserta tes menjawab soal. Waktu respon yang lambat dari seorang peserta tes akan dapat mencerminkan pemrosesan yang memang lambat atau kehati-hatian dari peserta tes tersebut, kemudian seorang peserta tes dapat merespon dengan cepat dan benar karena bisa saja dia menebak dan beruntung, selanjutnya jika peserta tes tidak menjawab benar, hal itu bisa saja terjadi karena dia memang tidak tahu jawabannya, atau kurang cukup waktu dalam memproses jawabannya (Kyllonen & Zu, 2016: 2). Penelitian terkait perilaku menebak ini, pernah dilakukan sebelumnya oleh Meyer (2010) yang mengusulkan sebuah *Mixed Model* dengan waktu respon yang diaplikasikan melalui tes *real* (data empiris) dan studi simulasi

untuk menghadapi perilaku menebak siswa pada sebuah ujian, kemudian ada van der Maas et al. (2011) yang meneliti tentang kesesuaian model IRT dalam mengukur kemampuan, keputusan lulus atau tidak lulus, dimana model juga mempertimbangkan tentang perilaku menebak dan *speed* dari peserta tes.

Dalam tes terkomputerisasi, parameter tambahan yang muncul seperti *pseudo guessing* dan *careless* dapat menyebabkan perbedaan hasil estimasi parameter kemampuan peserta tes yang signifikan antara Model Logistik 3 Parameter dan Model Logistik 4 Parameter (Liao et al., 2012: 1679). Hal ini senada dengan alasan kenapa waktu respon sebagai parameter tambahan (yang harusnya diakomodasi dalam suatu model) sebaiknya dilibatkan dalam proses pengukuran, karena dapat memperbaiki hasil estimasi parameter kemampuan peserta tes. Pernyataan ini didukung oleh Schnipke & Pashley (1997) dan Schnipke & Scrams (2002) yang dalam penelitiannya mempertimbangkan waktu respon untuk mengukur kemampuan peserta tes. Hornke (2000: 175) juga berpendapat bahwa waktu respon penting untuk dipertimbangkan dalam suatu tes, karena waktu respon adalah indikator untuk ciri-ciri kepribadian peserta tes yang harus dibedakan dari skor. Selain itu dalam suatu tes, informasi waktu respon juga dapat digunakan untuk mendeteksi perilaku menyontek (Man et al., 2018: 155). Abdelfattah & Johanson (2007: 85) dalam penelitiannya menghasilkan perbedaan estimasi parameter kemampuan peserta tes dan *item parameter* antara model yang mempertimbangkan waktu respon dan tidak mempertimbangkan waktu respon. Hasil penelitian Abdelfattah & Johanson (2007) ini juga dikuatkan oleh penelitian Oshima (1994: 200) yang menyatakan bahwa *speed* dari peserta tes dalam mengerjakan suatu butir

soal dapat mempengaruhi estimasi parameter kemampuan peserta tes dan estimasi *item parameter*-nya.

Pemodelan waktu respon secara matematis dapat dilakukan dengan tiga cara (Entink & van der Linden, 2009: 47-48). Cara pertama adalah pemodelan waktu respon secara eksklusif, pemodelan seperti ini dilakukan jika soal yang diujikan mudah dan batasan waktu yang diberikan sangatlah ketat, dari hasil ujian ini didapatkan informasi yang sangat sedikit. Cara kedua adalah melakukan analisis terpisah pada model waktu respon dan model keakuratan skor, sehingga didapatkan informasi waktu respon dan keakuratan skor secara terpisah atau independen. Cara ketiga adalah menggabungkan waktu respon dan keakuratan skor ke dalam satu model secara simultan. Model simultan bisa didapatkan dengan konsep *joint distribution*, yang mana dari konsep tersebut akan dihasilkan *joint model*, seperti yang dilakukan oleh Fox & Marianti (2017: 244) yang menggabungkan parameter akurasi dengan parameter waktu respon ke dalam model logistik dengan variabel random waktu respon secara simultan. Pada penelitiannya, Fox & Marianti (2017) melakukan uji kecocokan model terhadap *joint model*-nya dan untuk mendeteksi perilaku menyimpang pada pola respon. Statistik dari uji kecocokan model tersebut dihitung menggunakan metode *Bayesian Markov Chain Monte Carlo* (MCMC). Studi simulasi dan penerapan data empiris diterapkan di sini, dan dari sini dihasilkan kesimpulan bahwa uji kecocokan model dengan metode *Bayesian* dapat mendeteksi perilaku menyimpang pada pola respon.

Berdasarkan kajian penelitian terdahulu, ada beberapa metode yang dapat digunakan untuk mengestimasi parameter pada model, seperti *Maximum Likelihood*

Estimation (MLE) dan *Bayesian* (Retnawati, 2015: 146). Ide dibalik metode MLE ini adalah menentukan parameter-parameter dengan memaksimumkan probabilitas (*likelihood*) dari data sampel. Pada metode MLE parameter populasi dipandang bernilai tetap (*fixed variable*), sedangkan metode *Bayesian* memandang parameter populasi sebagai *random variable* yang memiliki distribusi, disebut sebagai distribusi *prior*. Distribusi *prior* merupakan distribusi subjektif yang didasarkan oleh keyakinan peneliti terhadap suatu parameter. Penentuan distribusi *prior* secara subjektif inilah yang dapat menjadi kelemahan dari metode *Bayesian*, karena pemilihannya terbatas pada representasi kepercayaan peneliti, sehingga perlu adanya pengujian lebih lanjut. Dari distribusi *prior*, selanjutnya dapat ditentukan distribusi *posterior* sehingga diperoleh estimator *Bayesian* yang merupakan *mean* atau modus dari distribusi *posterior*. Metode MLE hanya mendasarkan inferensinya pada informasi yang dikandung oleh sampel (informasi *prior* tidak dimasukkan), sehingga apabila distribusi populasinya tidak diketahui, maka metode MLE tidak dapat digunakan. Berbeda dengan metode *Bayesian* yang menggabungkan informasi yang dikandung sampel dengan informasi lain yang tersedia sebelumnya (*prior*).

Metode *Bayesian* dapat mengestimasi parameter model yang sifatnya kompleks, melanggar asumsi dasar IRT, dan ukuran sampel yang kecil (Natesan, 2011: 550). Dalam penelitiannya tentang pengembangan Model Logistik 4 Parameter, Magis (2013: 306) menerapkan metode *Bayesian* untuk mengestimasi modelnya yang mempunyai parameter yang cukup kompleks. Model matematis yang digunakan untuk mengestimasi *person parameter* dan *item parameter* dalam

pemodelan waktu respon dengan keakuratan skor merupakan model yang kompleks karena memiliki banyak parameter, sehingga metode *Bayesian* dianggap tepat digunakan untuk mengestimasi metode simultan pada penelitian ini. Penyelesaian secara numerik dalam model yang kompleks tersebut dapat dilakukan dengan teknik *Markov Chain Monte Carlo* (Entink, Fox, et al., 2009; Entink, Kuhn, et al., 2009; Levy, 2009; Meng et al., 2015; Tuerlinckx & de Boeck, 2005; van der Linden, 2006, 2007).

Seiring berjalannya waktu, topik penelitian yang membahas tentang waktu respon ini mengalami perkembangan, khususnya yang terkait dengan pemodelan waktu respon. Topik ini memang sudah lama dibahas dalam psikometrik tetapi relatif diabaikan, sehingga pengembangannya tidak sebanyak pemodelan IRT (Thissen, 1979: 257). Seperti yang telah dijelaskan sebelumnya, penelitian tentang pemodelan waktu respon ini dimaksudkan untuk memperbaiki konsep IRT pada saat sebuah tes memperhitungkan waktu respon.

Pengembangan model waktu respon secara teoritis bisa dilakukan berdasarkan jenis tes, apakah merupakan jenis *speed test* atau *power test*. *Power test* murni dan *speed test* murni mempunyai konsep yang berbeda (Gulliksen, 1950). *Speed test* murni adalah tes yang dengan waktu terbatas dan butir soalnya mudah untuk dijawab dengan benar (tingkat kesulitan sama), tes seperti ini dapat diskor berdasarkan total waktu yang digunakan peserta tes untuk mengisi semua butir soal. *Power test* murni adalah tes dengan waktu yang tidak dibatasi, bedanya dengan *speed test* murni adalah tingkat kesulitan butir soalnya bervariasi, tes seperti

ini cara penskorannya adalah dengan menghitung jumlah butir soal yang dapat dijawab dengan benar.

Rouder et al. (2015), mengembangkan model respon butir dengan waktu respon pada *speed test* untuk memahami proses kognitif peserta tes yang disebut sebagai *Lognormal Race Model*, di sini *finishing time* (waktu respon) diasumsikan berdistribusi *Lognormal* karena distribusi tersebut dianggap sesuai dengan karakteristik waktu respon (nilainya positif dan bentuknya *positive skewed*) dan mudah diinterpretasikan.

Ada beberapa karakteristik yang harus dipertimbangkan dalam pemilihan distribusi waktu respon: (1) haruslah bernilai positif; (2) waktu respon yang singkat lebih sering terjadi dibandingkan waktu respon yang lama, atau dengan kata lain besarnya probabilitas waktu respon yang singkat sangat besar jika dibandingkan besarnya probabilitas waktu respon yang lama atau *positive skewed* (Lindsey, 2004: 197). Distribusi-distribusi yang cocok dengan kedua karakteristik tersebut adalah distribusi *Lognormal*, *Weibull* dan *Gamma* (Lindsey, 2004: 203-206). Menurut Attali (2010: 458-459), Teori *Generalizability* (Teori G) dan *Analysis of Variance* (ANOVA) adalah salah satu metode yang bisa digunakan untuk mengestimasi parameter dari fungsi distribusi dari waktu respon dalam suatu tes. Dalam Teori G dan ANOVA, variabilitas waktu respon dari seluruh peserta ujian dalam mengerjakan tiap item harus dipertimbangkan untuk mengembangkan batasan waktu dalam suatu tes.

Pemodelan dalam IRT memodelkan *person parameter* dan *item parameter* (tingkat kesulitan, daya beda, dan *pseudo guessing*) hanya berdasarkan pola respon

peserta tes yaitu jawaban salah atau benar dari peserta tes. Konsep IRT hanya cocok diterapkan pada *power test* murni, yang mana tes tidak dibatasi waktu (van Breukelen, 2005: 359). Tetapi pada pelaksanaannya *power test* murni itu tidak ada, karena tiap tes selalu dibatasi waktu (van der Linden & Hambleton, 2013: 166). Adanya batasan waktu tes, membuat adanya aspek lain yang terlibat dalam pengerjaan suatu butir soal yaitu aspek *speed* dari peserta tes dalam menjawab suatu butir soal. *Speed* peserta tes dapat diketahui dari waktu respon peserta tes dalam menjawab suatu butir soal. Lee & Ying (2015) mengembangkan model simultan antara model respon butir dengan waktu respon yang disebut sebagai *Mixture Cure Rate Model*, model ini analog dengan model pada analisis survival yaitu terjadinya sensor waktu saat peserta tes belum selesai mengerjakan tetapi waktu sudah habis, model ini dikembangkan karena adanya batasan waktu pada *power test*.

Setiap model waktu respon yang dikembangkan mempunyai spesifikasi tersendiri dalam penerapannya, ada kekurangan juga kelebihan. Tidak ada model standar yang dapat digunakan untuk memodelkan semua kasus waktu respon dalam berbagai tes. Karena faktanya distribusi waktu respon dari berbagai item soal bisa beragam, sehingga model tertentu tidak akan berlaku untuk segala macam tes (Ranger & Kuhn, 2012: 31). Pengembangan model yang memperhitungkan waktu respon perlu terus dilakukan untuk memperbaiki model yang sudah ada agar lebih realistis.

Penelitian ini bermaksud untuk untuk mengembangkan Model Logistik 2 Parameter dengan variabel random waktu respon secara simultan untuk tes terkomputerisasi. Pengembangan model dilakukan dengan cara mengkaji kembali

model dari Hidayah et al. (2016) agar didapatkan pengembangan model yang lebih realistis penerapannya di lapangan. Hidayah et al. (2016) mengembangkan model simultan antara Model Logistik 1 Parameter dengan waktu respon. Model simultan ini didapatkan dengan konsep *joint distribution*, yaitu perkalian antara distribusi bersyarat keakuratan respon (Model 1PL) terhadap waktu respon dengan distribusi marginal dari waktu respon (*Lognormal*). Penelitian Hidayah et al. (2016) merupakan penelitian simulasi yang datanya dibangkitkan berdasarkan banyaknya peserta tes dan banyaknya soal tes.

Hidayah et al. (2016) mengembangkan modelnya dengan cara memperbaiki kelemahan dari model Ingrisone II et al. (2008: 3) yaitu penggunaan model untuk *power test*, padahal menurut Hidayah et al. (2016: 86) model Ingrisone II et al. (2008) terbukti lebih cocok untuk *speed test* bukan *power test* karena lebih tepat dan cocok dalam memprediksi butir-butir soal mudah untuk dijawab dengan benar dengan waktu yang terbatas (cepat). Kelebihan sekaligus kelemahan dari model yang dikembangkan Hidayah et al. (2016) adalah, model hanya diuji dengan menggunakan data simulasi saja dalam penelitiannya. Studi simulasi berbasis IRT ini dapat mendorong para peneliti untuk menempatkan “gangguan” realistis dan faktor-faktor penting ke dalam model (Luecht & Ackerman, 2018: 75). Selain itu, *output* dari data simulasi ini memang bersifat ideal, namun di sisi lainnya teknik simulasi ini hanya menghasilkan sekumpulan *output* dari sistem pada berbagai kondisi yang berbeda, bukan berupa jawaban. Kelemahan lainnya pada model (Hidayah et al., 2016) adalah dalam pemilihan distribusi waktu respon juga hanya berdasarkan pertimbangan asumsi saja, bukan pembuktian secara empiris. Oleh

karena itu selain data simulasi, penggunaan data empiris juga akan diterapkan pada penelitian ini, yaitu berupa data hasil tes seleksi masuk Program Pascasarjana UNY berbasis komputer (CBT) tahun 2017. Selain dengan penerapan data empiris pada penelitian, pengembangan model pada penelitian ini juga dilakukan dengan cara mengkaji ulang parameter, yaitu mempertimbangkan parameter akurasi dengan parameter waktu respon yang masuk ke dalam model logistik dengan variabel random waktu respon secara simultan berdasarkan kajian teori dari penelitian relevan sebelumnya. Model matematis pada penelitian ini dikembangkan berdasarkan kondisi tes dengan model terkomputerisasi (agar informasi waktu respon dapat terekam) dan dengan data pola respon peserta tes yang berbentuk dikotomus.

B. Identifikasi Masalah

Berdasarkan latar belakang di atas, ada beberapa masalah yang dapat diidentifikasi oleh peneliti.

1. Suatu tes dalam pelaksanaannya selalu dibatasi waktu, sehingga ada faktor waktu yang seharusnya dilibatkan untuk mengungkapkan kondisi *real* dalam pelaksanaan suatu tes. Selama ini model IRT lebih sering diterapkan untuk mengungkap kondisi pelaksanaan tes, baik itu pada tes terkomputerisasi maupun tidak, namun sayangnya IRT tidak melibatkan variabel waktu respon dalam modelnya, padahal apabila waktu respon dilibatkan dalam suatu pemodelan maka akan dapat memperbaiki hasil estimasi parameter kemampuan peserta tes.

2. 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.
3. Kelebihan sekaligus kekurangan model simultan dari Hidayah, Kumaidi & Kartowagiran pada tahun 2016 yang akan dikembangkan dalam penelitian ini adalah pengujian terbatas data simulasi. Di satu sisi, *output* dari data simulasi ini bersifat ideal karena dibangkitkan dengan algoritma tertentu sesuai dengan model. Tetapi di sisi lain, hasil dari simulasi ini bisa jadi tidak akurat karena teknik ini bukan proses optimasi dan tidak menghasilkan jawaban tetapi hanya menghasilkan sekumpulan *output* dari sistem pada berbagai kondisi yang berbeda sehingga keakuratannya masih harus dipertanyakan.
4. Ada beberapa parameter model yang dikembangkan Hidayah, Kumaidi & Kartowagiran pada tahun 2016 yang pada praktiknya (secara empiris) perlu dikaji lagi, misalnya parameter kelambatan (d_j), dan parameter besarnya usaha yang dibutuhkan untuk menyelesaikan soal tes ke- j (β_j) pada pendefinisian *mean log natural* waktu respon ($\mu_{\ln t_{ij}}$) distribusi *Lognormal*.
5. Dalam penentuan distribusi marginal dari waktu respon, Hidayah, Kumaidi & Kartowagiran pada tahun 2016 menentukan secara langsung distribusi

marginal dari waktu responnya adalah *Lognormal* dengan $\ln t_{ij} \approx N((\beta_j - \tau_i), \sigma^2)$. Keputusan dibuat hanya berdasarkan pertimbangan bahwa distribusi tersebut nilainya positif, bentuknya *positive skewed* dan lebih mudah diinterpretasikan, padahal bisa saja ada distribusi lain yang secara uji empiris lebih cocok dengan karakter waktu respon, misalnya *Weibull* atau *Gamma*. Apabila menggunakan data simulasi seperti pada penelitian Hidayah, Kumaidi & Kartowagiran tahun 2016, peneliti hanya dapat menentukan distribusi marginal waktu respon yang dianggap cocok berdasarkan asumsi peneliti saja.

C. Pembatasan Masalah

Agar masalah yang dikaji dalam penelitian ini menjadi lebih fokus, maka diperlukan pembatasan masalah sebagai berikut:

1. Penelitian ini bermaksud untuk mengembangkan Model Logistik 2 Parameter dengan variabel random waktu respon secara simultan untuk tes terkomputerisasi, dengan parameter yang berkaitan dengan soal tes adalah daya beda soal tes (a_j), dan tingkat kesulitan soal tes (b_j), dan *time intensity* untuk soal tes ke- j (ξ_j) kemudian parameter yang berkaitan dengan peserta tes adalah kemampuan peserta tes (Θ_i) dan waktu respon peserta tes (t_{ij}) dan *speed* peserta tes dalam menyelesaikan soal (τ_i).
2. Pengembangan model dilakukan dengan cara mengkaji kembali model yang dikembangkan Hidayah, Kumaidi & Kartowagiran pada tahun

2016 agar didapatkan pengembangan model yang lebih realistis penerapannya di lapangan, yaitu: (1) mengkaji ulang parameter pada model yang dikembangkan Hidayah, Kumaidi & Kartowagiran pada tahun 2016, dengan cara menyeleksi parameter-parameter apa saja yang dapat dimasukkan ke dalam model, yang tentu saja parameter-parameter tersebut pada praktiknya logis; serta (2) mengkaji ulang distribusi marginal dari waktu respon, apakah berdistribusi *Lognormal*, *Weibull* atau *Gamma*, karena penelitian ini menggunakan data empiris, maka dapat melakukan uji kecocokan distribusi marginal terlebih dahulu dalam penerapannya.

3. Selain data simulasi, data empiris juga akan digunakan dalam penelitian ini. Data empiris diperoleh dari tes berbasis komputer (CBT) pada seleksi masuk Program Pascasarjana UNY tahun 2017, dimana pada tes ini dicatat waktu respon dari setiap peserta tes dalam menyelesaikan tiap butir soal yang dikerjakan.
4. Model simultan yang dihasilkan merupakan model yang memiliki banyak parameter, sehingga membutuhkan perhitungan yang kompleks, oleh karena itu estimasi parameter model dilakukan dengan menggunakan metode *Bayesian*. Metode *Bayesian* mampu mengestimasi parameter pada model yang sifatnya kompleks, data yang melanggar beberapa asumsi dasar IRT, dan ukuran sampel yang kecil.

D. Rumusan Masalah

Berdasarkan latar belakang dan identifikasi masalah di atas, maka rumusan masalah yang dapat diformulasikan adalah:

1. Bagaimana model matematis yang dihasilkan untuk menjelaskan kondisi realistis dalam praktik pengukuran tes dengan sistem terkomputerisasi, apabila waktu respon juga diperhitungkan sebagai indikator kemampuan?
2. Bagaimana keakuratan estimasi parameter model matematis yang dihasilkan apabila waktu respon diperhitungkan sebagai indikator kemampuan?
3. Bagaimana perbandingan keakuratan hasil estimasi parameter kemampuan (θ/Θ) antara model matematis yang memperhitungkan waktu respon sebagai indikator kemampuan, dengan model *Item Response Theory* (IRT)?
4. Bagaimana perbandingan kecocokan antara model matematis yang memperhitungkan waktu respon sebagai indikator kemampuan dan model IRT, dengan menggunakan data empiris?

E. Tujuan Pengembangan

Tujuan utama dari penelitian ini adalah untuk menghasilkan produk berupa model matematis yang dikembangkan secara simultan, yang dapat meningkatkan keakuratan estimasi parameter θ (Θ) dengan memperhitungkan waktu respon sebagai indikator kemampuan, sehingga dapat menjelaskan kondisi

realistis tes dengan lebih baik. Dari tujuan utama tersebut dapat dipecah lagi menjadi beberapa tujuan untuk mencapainya.

1. Untuk menghasilkan model matematis yang dapat menjelaskan kondisi realistis dalam praktik pengukuran tes dengan sistem terkomputerisasi, apabila waktu respon juga diperhitungkan sebagai indikator kemampuan.
2. Untuk mengetahui keakuratan estimasi parameter model matematis yang dihasilkan apabila waktu respon diperhitungkan sebagai indikator kemampuan.
3. Untuk mengetahui perbandingan keakuratan hasil estimasi parameter kemampuan (θ/Θ) antara model matematis yang memperhitungkan waktu respon sebagai indikator kemampuan, dengan model *Item Response Theory* (IRT).
4. Untuk mengetahui perbandingan kecocokan antara model matematis yang memperhitungkan waktu respon sebagai indikator kemampuan dan model IRT, dengan menggunakan data empiris.

F. Spesifikasi Produk yang Dikembangkan

Adapun spesifikasi produk yang dikembangkan adalah model matematis yang dapat menjelaskan kondisi realistis dalam praktik pengukuran khususnya tes dengan sistem terkomputerisasi (CBT) 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 *speed* peserta tes dalam menyelesaikan soal (τ_i). Model simultan ini didapatkan dengan konsep *joint distribution*, yaitu perkalian antara distribusi bersyarat keakuratan respon (Model 2PL) terhadap waktu respon dengan distribusi marginal yang cocok dengan karakteristik dari waktu responnya.

G. Manfaat Pengembangan

Berikut ini adalah manfaat penelitian apabila ditinjau dari segi :

1. Teoritis

Penelitian tentang pemodelan waktu respon ini dimaksudkan untuk memperbaiki konsep IRT pada tes dengan sistem terkomputerisasi, karena pada saat waktu respon diperhitungkan sebagai bagian dari kemampuan peserta tes (θ/Θ), maka keakuratan estimasi parameter kemampuan peserta tes (θ/Θ) akan meningkat, sehingga didapatkan model waktu respon menjadi lebih realistis (informatif) penerapannya dibandingkan dengan model IRT. Pada kenyatannya akan selalu ada dua parameter utama yang terlibat untuk menggambarkan bagaimana pengerjaan suatu tes, yaitu akurasi respon dan waktu respon, namun parameter waktu respon relatif diabaikan dalam pemodelan IRT. Pengembangan Model Logistik 2 Parameter dengan variabel random waktu respon ini diharapkan dapat memberikan sumbangan dalam perkembangan teori pemodelan, yang realitanya dalam bidang psikometrik penelitian semacam ini memang sudah lama dibahas, namun pengembangannya tidak sebanyak pemodelan IRT, karena pada

masa lampau mencatat atau merekam waktu respon bukanlah hal yang mudah untuk dilakukan (keterbatasan teknologi).

2. Praktis

Selain itu dari penelitian ini dapat diketahui bahwa, catatan waktu respon yang dihasilkan dari tes dengan sistem terkomputerisasi akan menjadi penting untuk dipertimbangkan dalam suatu tes, apalagi tujuannya adalah untuk seleksi. Ilustrasi secara sederhananya, jika ada dua peserta tes yang memiliki respon sama terhadap suatu butir soal (jawaban keduanya benar), tetapi waktu yang dibutuhkan untuk menyelesaikan butir soal tersebut berbeda, maka catatan waktu respon ini dapat memberikan informasi tambahan bagi lembaga penyelenggara tes dalam pengambilan keputusan (lulus atau gagal), sehingga dapat mempermudah pelaksanaan seleksi. Dengan dikembangkannya model yang dapat meningkatkan keakuratan estimasi parameter theta (Θ) ini, maka lembaga penyelenggara tes akan lebih mudah lagi untuk memetakan kemampuan sebenarnya dari peserta tes, sehingga rekomendasi keputusan seleksi yang diberikan akan dapat lebih tepat lagi

H. Asumsi Pengembangan

Ada berapa asumsi yang harus dipenuhi agar model dapat dikembangkan:

1. Waktu respon diasumsikan sebagai *random variable*, karena apabila seseorang menjalankan tes yang sama (soalnya) dengan waktu pelaksanaan tes yang berbeda, maka waktu respon yang ditempuh bisa jadi akan berbeda dengan tes yang dilaksanakan sebelumnya, karena diasumsikan bahwa waktu respon tidak hanya dikontrol oleh petugas administrasi tes atau teknologi komputer saja tetapi ada faktor lain yang

mempengaruhi waktu respon peserta tes dalam mengerjakan soal, seperti perbedaan strategi dan *speed* yang digunakan oleh peserta tes dalam mengerjakan soal yang sama di waktu yang berbeda.

2. Probabilitas menjawab benar peserta tes dipengaruhi oleh beberapa parameter. Parameter yang berkaitan dengan soal tes adalah daya beda soal tes (a_j), dan tingkat kesulitan soal tes (b_j), dan *time intensity* untuk soal tes ke- j (ξ_j), sedangkan parameter yang berkaitan dengan peserta tes adalah kemampuan peserta tes (Θ_i) dan waktu respon peserta tes (t_{ij}) dan *speed* peserta tes dalam menyelesaikan soal (τ_i).
3. Model matematis yang dikembangkan tidak mempertimbangkan perbedaan kemampuan peserta tes dalam mengoperasikan komputer, karena diasumsikan semua peserta tes dapat memanfaatkan komputer dengan sama baiknya.