

PENGEMBANGAN MODEL *MULTILEVEL ITEM RESPONSE THEORY* UNTUK DATA BERSTRUKTUR HIERARKI

ALONA DWINATA



**PROGRAM STUDI DOKTOR STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University

PERNYATAAN MENGENAI DISERTASI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa disertasi dengan judul “Pengembangan Model *Multilevel Item Response Theory* Untuk Data Berstruktur Hierarki” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir disertasi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2026

Alona Dwinata
G161190031



@Hak cipta milik IPB University

IPB University

RINGKASAN

ALONA DWINATA. Pengembangan Model *Multilevel Item Response Theory* Untuk Data Berstruktur Hierarki. Dibimbing oleh ANANG KURNIA, AJI HAMIM WIGENA, dan MUHAMMAD NUR AIDI.

Data respons butir merupakan data hasil pengukuran yang diperoleh dari jawaban peserta tes terhadap sejumlah butir soal. Data ini digunakan untuk menggambarkan kemampuan laten peserta tes berdasarkan pola jawabannya. Salah satu model yang banyak digunakan adalah *Item Response Theory* dua parameter logistik (IRT 2PL), karena model ini mempertimbangkan dua karakteristik penting butir, yaitu daya beda dan tingkat kesukaran. Jadi, peluang peserta menjawab benar suatu butir tidak hanya ditentukan oleh kemampuan laten, tetapi juga ditentukan oleh karakteristik setiap butir soal. Pada tes bidang pendidikan, data respons butir sering tidak berasal dari populasi yang homogen dan antarpeserta tes tidak bebas. Peserta tes biasanya berada dalam struktur berkelompok, misalnya siswa tersarang dalam sekolah. Siswa dari sekolah yang sama cenderung memiliki kesamaan lingkungan belajar, fasilitas, sumber daya, dan sistem pembelajaran. Kondisi ini menyebabkan asumsi independensi pada model IRT klasik tidak terpenuhi. Untuk mengakomodasi struktur tersebut, diperlukan model Multilevel IRT 2PL. Model ini mengintegrasikan model respons butir dengan pemodelan multilevel, sehingga variasi kemampuan laten dapat dipisahkan ke dalam variasi pada level siswa dan variasi pada level sekolah. Selain pendekatan pengukuran, analisis respons butir juga dapat dikembangkan melalui pendekatan penjelas, yaitu dengan memasukkan prediktor yang relevan untuk menerangkan variasi kemampuan laten. Dalam konteks ini dikembangkan model Multilevel *Latent Regression* IRT 2PL (MLR-IRT 2PL) sebagai perluasan Multilevel IRT 2PL dengan melibatkan prediktor pada level siswa dan level sekolah.

Model Multilevel IRT 2PL dikembangkan dengan memodelkan kemampuan laten peserta tes sebagai gabungan efek acak siswa dan efek acak sekolah, yaitu kemampuan siswa dalam sekolah dan variasi antarsekolah. Secara konseptual, model ini menuliskan kemampuan laten sebagai $\theta_{jk} = \delta_{jk} + \delta_k$, dengan δ_{jk} sebagai efek acak siswa dalam sekolah dan δ_k sebagai efek acak sekolah. Karena kemampuan laten siswa dan efek sekolah tidak diamati secara langsung, pendugaan parameter memerlukan pendekatan yang mampu memarginalkan efek acak tersebut. Salah satu pendekatan yang tersedia untuk model Multilevel IRT adalah paket PLmixed pada R. Pendugaan parameter pada paket ini menggunakan pendekatan GLMM dengan struktur faktor, termasuk model IRT 2PL dengan menggunakan metode *profile likelihood* dan *nested maximizations*. Oleh karena itu, PLmixed digunakan sebagai metode pembandingan dalam penelitian ini. Penelitian ini mengembangkan model Multilevel IRT 2PL dengan metode pendugaan Marginal Maximum Likelihood dan Gauss-Hermite Quadrature (MML-GHQ). Pendugaan parameter dilakukan menggunakan metode *Marginal Maximum Likelihood* (MML), integral efek acak dihampiri dengan Gauss-Hermite Quadrature (GHQ), dan optimisasi dilakukan melalui algoritma *Expectation-Maximization* (EM). Kemampuan laten kemudian diduga dengan pendekatan *Expected A Posteriori* (EAP). Hasil kajian simulasi model Multilevel IRT 2PL menunjukkan bahwa model memiliki kinerja baik. Simulasi menggunakan faktor



Intraclass Correlation Coefficient (ICC), jumlah sekolah, jumlah siswa per sekolah, dan jumlah butir. Nilai korelasi antara kemampuan laten sebenarnya dan kemampuan laten hasil dugaan relatif tinggi pada semua skenario. Faktor yang paling kuat meningkatkan korelasi dan menurunkan RMSE parameter kemampuan laten adalah jumlah butir. Faktor jumlah sekolah, jumlah siswa per sekolah, dan jumlah butir berpengaruh signifikan terhadap parameter daya beda butir, sedangkan parameter kesukaran butir dipengaruhi oleh ICC, jumlah sekolah, dan jumlah siswa per sekolah.

Model MLR-IRT 2PL dikembangkan sebagai perluasan Multilevel IRT 2PL dengan memasukkan prediktor pada struktur kemampuan laten. Kemampuan laten peserta tidak hanya dipengaruhi oleh efek acak siswa dan sekolah, tetapi juga oleh prediktor level siswa dan prediktor level sekolah. Dengan demikian, model ini tidak hanya mengukur kemampuan laten, tetapi juga menjelaskan variasi kemampuan laten berdasarkan prediktor. Faktor yang digunakan dalam simulasi meliputi ICC, jumlah sekolah, jumlah siswa per sekolah, jumlah butir, efek prediktor siswa, dan efek prediktor sekolah. Hasil simulasi menunjukkan bahwa model MLR-IRT 2PL memiliki kinerja yang baik dalam menduga kemampuan laten pada data respons butir berstruktur hierarki. Korelasi antara kemampuan laten sebenarnya dan kemampuan laten hasil dugaan relatif tinggi pada seluruh skenario. Jumlah butir merupakan faktor utama yang meningkatkan kualitas dugaan kemampuan laten, semakin banyak butir, korelasi meningkat dan RMSE menurun. Hasil ANOVA menunjukkan bahwa ICC, jumlah butir, efek prediktor siswa, dan efek prediktor sekolah berpengaruh signifikan terhadap korelasi kemampuan laten, sedangkan jumlah sekolah dan jumlah siswa per sekolah tidak signifikan. Untuk parameter butir, jumlah sekolah paling berperan dalam menurunkan RMSE daya beda dan kesukaran, sehingga nilai dugaan parameter semakin akurat.

Pada penelitian ini, hasil analisis data PISA menggunakan model Multilevel IRT 2PL dengan pendekatan MML-GHQ dibandingkan dengan hasil estimasi menggunakan PLmixed. Kedua pendekatan menghasilkan pola estimasi yang relatif sejalan, baik pada kemampuan laten peserta tes maupun pada parameter daya beda dan kesukaran butir. Pendekatan MML-GHQ menghasilkan sebaran estimasi kemampuan laten yang lebih lebar daripada PLmixed. Hal ini menunjukkan bahwa MML-GHQ memberikan diferensiasi yang lebih besar dalam mengidentifikasi peserta tes dengan kemampuan laten rendah, sedang, dan tinggi. Selanjutnya, penerapan model MLR-IRT 2PL pada data PISA menunjukkan bahwa dugaan kemampuan laten bergeser ke arah positif dan memiliki rentang lebih lebar dibandingkan Multilevel IRT 2PL. Meskipun demikian, korelasi antara estimasi kemampuan laten dari kedua model adalah 0,97. Nilai tersebut menunjukkan bahwa urutan kemampuan laten peserta tes tetap konsisten, meskipun kedua model menghasilkan skala dan rentang dugaan yang berbeda. Model MLR-IRT 2PL juga menunjukkan bahwa prediktor kelas dan lingkungan sekolah memiliki kontribusi dalam menjelaskan keragaman kemampuan laten. Dengan demikian, model Multilevel IRT 2PL memiliki kinerja yang baik untuk memisahkan keragaman level siswa dan sekolah, sedangkan model MLR-IRT 2PL memperluas interpretasi dengan memperhitungkan pengaruh prediktor pada level siswa dan sekolah.

Kata kunci: data respons butir, Multilevel IRT 2PL, MLR-IRT 2PL, kemampuan laten



SUMMARY

ALONA DWINATA. Development of a Multilevel Item Response Theory Model for Hierarchically Structured Data. Supervised by ANANG KURNIA, AJI HAMIM WIGENA, and MUHAMMAD NUR AIDI.

Item response data is measurement data obtained from participants' answers to a number of test items. This data is used to describe participants' latent abilities based on their response patterns. One widely used model is the two-parameter logistic Item Response Theory (IRT 2PL), because this model considers two important item characteristics: discrimination and difficulty level. Therefore, the probability of a participant answering an item correctly is determined not only by latent ability but also by the characteristics of each item. In educational tests, item response data often do not come from a homogeneous population and are not independent among participants. Participants are usually grouped, for example, students nested within schools. Students from the same school tend to have similar learning environments, facilities, resources, and learning systems. This condition violates the independence assumption in the classical IRT model. To accommodate this structure, a multilevel IRT 2PL model is needed. This model integrates the item response model with multilevel modeling, allowing latent ability variation to be separated into student-level and school-level components. In addition to the measurement approach, item response analysis can also be developed through an explanatory approach, namely by incorporating relevant predictors to explain variations in latent abilities. In this context, the Multilevel Latent Regression IRT 2PL (MLR-IRT 2PL) model was developed as an extension of Multilevel IRT 2PL by involving predictors at the student and school levels.

The Multilevel IRT 2PL model was developed by modeling participants' latent ability as a combination of student random effects and school random effects, namely student ability within schools and variation between schools. Conceptually, this model writes the latent ability as $\theta_{jk} = \delta_{jk} + \delta_k$, where δ_{jk} is the student random effect within schools and δ_k is the school random effect. Because student latent ability and school effects are not directly observed, parameter estimation requires an approach that can marginalize these random effects. One approach for Multilevel IRT models is the PLmixed package in R. This package estimates GLMMs with factor structures, including IRT 2PL models via profile likelihood and nested maximization. Therefore, PLmixed is used as a comparison method in this study. This study develops a Multilevel IRT 2PL model and estimates it using the Marginal Maximum Likelihood with Gauss-Hermite Quadrature (MML-GHQ) method. Parameter estimation was performed using Marginal Maximum Likelihood (MML), the random effect integral was approximated with Gauss-Hermite Quadrature (GHQ), and optimization was performed using the Expectation-Maximization (EM) algorithm. Latent ability was then estimated using the Expected A Posteriori (EAP) approach. The results of the Multilevel IRT 2PL model simulation study showed that the model performed well. The simulation used the Intraclass Correlation Coefficient (ICC), the number of schools, the number of students per school, and the number of items. The correlation between the actual and estimated latent abilities was relatively high across all scenarios. The strongest factor that increases the correlation and decreases the RMSE of the latent

@Hak cipta milik IPB University

IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ability parameter is the number of items. The number of schools, the number of students per school, and the number of items significantly influence the item discrimination parameter, while the item difficulty parameter is influenced by the ICC, the number of schools, and the number of students per school.

The MLR-IRT 2PL model was developed as an extension of Multilevel IRT 2PL by incorporating predictors into the latent ability structure. Participants' latent ability is influenced not only by students' and schools' random effects, but also by student- and school-level predictors. Thus, this model not only measures latent ability but also explains variations in latent ability based on predictors. In the simulation, the following factors are used: ICC, number of schools, number of students per school, number of items, student predictor effects, and school predictor effects. The simulation results show that the MLR-IRT 2PL model performs well in estimating latent ability in hierarchically structured item response data. The correlation between actual and estimated latent ability is relatively high across all scenarios. The number of items is the main factor that improves the quality of latent ability estimation; as more items are added, the correlation increases and the RMSE decreases. ANOVA results show that ICC, number of items, student predictor effects, and school predictor effects significantly influence the correlation of latent ability, while the number of schools and the number of students per school are not significant. For item parameters, the number of schools played the most important role in reducing RMSE for discrimination and difficulty, leading to more accurate parameter estimates.

In this study, the results of PISA data analysis using the Multilevel IRT 2PL model with the MML-GHQ approach were compared with those obtained using PLmixed. Both approaches produced relatively consistent estimation patterns across examinees' latent abilities and item discrimination and difficulty parameters. The MML-GHQ approach produced a wider distribution of latent ability estimates than PLmixed. This indicates that the MML-GHQ provides greater differentiation in identifying examinees with low, medium, and high latent abilities. Furthermore, applying the MLR-IRT 2PL model to the PISA data showed that the latent ability estimates shifted in a positive direction and had a wider range than those from the Multilevel IRT 2PL model. However, the correlation between the latent ability estimates from the two models was 0.97. This value indicates that the order of examinees' latent abilities remained consistent, even though the two models produced different scales and ranges of estimates. The MLR-IRT 2PL model also showed that classroom and school environment predictors contributed to explaining the diversity of latent abilities. Thus, the Multilevel IRT 2PL model performs well at separating student- and school-level variability, while the MLR-IRT 2PL model broadens interpretation by accounting for the influence of predictors at both student- and school-levels.

Keywords: item response data, Multilevel IRT 2PL, MLR-IRT 2PL, latent ability



@Hak cipta milik IPB University

IPB University

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PENGEMBANGAN MODEL *MULTILEVEL ITEM RESPONSE THEORY* UNTUK DATA BERSTRUKTUR HIERARKI

ALONA DWINATA

Disertasi
sebagai salah satu syarat untuk memperoleh gelar
Doktor pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI DOKTOR STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

IPB University

Penguji Luar Komisi Pembimbing pada Ujian Tertutup Disertasi:

- HS
/R
1. Prof. Dr. Ir. Hari Wijayanto, M.Si.
 2. Sachnaz Desta Oktarina S.Stat., M.Agr., Ph.D.

Promotor Luar Komisi Pembimbing pada Sidang Promosi Terbuka Disertasi:

1. Dr. Suryadi, S.P., S.II., M.II.
2. Sachnaz Desta Oktarina S.Stat., M.Agr., Ph.D.





Judul Disertasi : Pengembangan Model Multilevel Item Response Theory
Untuk Data Berstruktur Hierarki
Nama : Alona Dwinata
NIM : G161190031

Disetujui oleh

Pembimbing 1:
Prof. Dr. Anang Kumia S.Si., M.Si.

Pembimbing 2:
Prof. Dr. Ir. Aji Hamim Wigena, M.Sc.

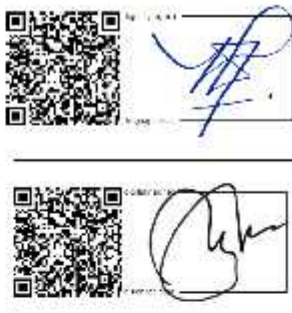
Pembimbing 3:
Prof.Dr. Ir. Muhammad Nur Aidi, M.S.



Diketahui oleh

Ketua Program Studi:
Dr. Kusman Sadik, S.Si., M.Si.
NIP. 196909121997021001

Dekan Sekolah Sains Data, Matematika, dan Informatika:
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.
NIP. 197807232007011001



Tanggal Ujian:

Tanggal Lulus:

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PRAKATA

Alhamdulillah Rabbil ‘Alamin. Puji dan syukur penulis panjatkan kepada Allah subhanahu wa taala atas segala karunia-Nya sehingga disertasi berjudul “Pengembangan Model Multilevel Item Response Theory untuk Data Berstruktur Hierarki” dapat diselesaikan. Keberhasilan penyusunan disertasi ini tidak terlepas dari bimbingan, bantuan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis menyampaikan terima kasih kepada:

1. Komisi Pembimbing, Prof. Dr. Anang Kurnia, S.Si., M.Si., Prof. Dr. Ir. Aji Hamim Wigena, M.Sc., dan Prof. Dr. Ir. Muhammad Nur Aidi, M.S., yang telah membimbing, memberikan arahan, dan memberikan dorongan kepada penulis dengan penuh kesabaran.
2. Penguji pada ujian kualifikasi lisan, yaitu Dr. Kusman Sadik, S.Si., M.Si., dan Dr. Budi Susetyo, M.S.; penguji pada ujian tertutup dan sidang promosi, yaitu Prof. Dr. Ir. Hari Wijayanto, M.Si., Sachnaz Desta Oktarina, S.Stat., M.Agr., Ph.D., dan Dr. Suryadi, S.P., S.H., M.H. yang telah mengulas, memberikan masukan, dan mengomentari disertasi ini.
3. Ketua Program Studi Statistika dan Sains Data serta Bapak dan Ibu Dosen yang telah memberikan bimbingan dan ilmu pengetahuan kepada penulis
4. Sekretariat Program Studi Statistika dan Sains Data yang senantiasa membantu penulis dalam berbagai proses administrasi.
5. Rektor Universitas Maritim Raja Ali Haji (UMRAH) dan Dekan Fakultas Keguruan dan Ilmu Pendidikan (FKIP UMRAH) yang telah memberikan kesempatan kepada penulis untuk melanjutkan studi program doctoral di IPB.
6. Kemdikbud atas Beasiswa Pendidikan Pascasarjana Dalam Negeri (BPPDN) yang telah mendukung studi dan penelitian ini secara finansial.

Ungkapan terima kasih secara khusus penulis sampaikan kepada suami tercinta, Febrian Firza; kedua orang tua penulis, Mama Mulyani T. dan Papa Usman; mertua penulis, Mak Siti Zaleka dan alm. Bapak Ali Azwar, yang senantiasa mengiringi langkah penulis dengan doa dan kasih sayangnya; kakak dan adik (Uda Ari, Uni Erni, Rifki, dan Tilla); keponakan (Azalia, Ayesha, Zeea, dan Nadhif); serta seluruh keluarga besar yang telah memberikan dukungan dan motivasi kepada penulis. Terima kasih atas kebersamaan teman-teman S-3 Statistika dan Sains Data yang telah berkenan menjadi tempat diskusi dan berbagi pengetahuan kepada penulis selama menjalani proses perkuliahan hingga penyelesaian disertasi ini.

Penulis menyadari bahwa disertasi ini masih memiliki keterbatasan. Oleh karena itu, penulis mohon maaf atas segala kekurangan dan kesalahan dalam penyusunan disertasi ini. Penulis berharap disertasi ini dapat menjadi rujukan bagi pengembangan kajian selanjutnya, serta memberikan kontribusi bagi kemajuan ilmu pengetahuan, khususnya dalam bidang statistika pendidikan.

Bogor, Juli 2026

Alona Dwinata



DAFTAR ISI

DAFTAR TABEL	iv
DAFTAR GAMBAR.....	v
DAFTAR LAMPIRAN	vi
GLOSARIUM	vii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	5
1.4 Manfaat	5
1.5 Ruang Lingkup	5
1.6 Kebaruan (<i>novelty</i>)	5
II TINJAUAN PUSTAKA	7
2.1 <i>Item Response Theory</i> (IRT)	7
2.2 <i>Item Response Theory</i> dalam Perspektif GLMM	9
2.3 Data Respons Butir Berstruktur Hierarki	9
2.4 Model Multilevel	10
2.5 Pendekatan <i>Marginal Maximum Likelihood</i> pada model IRT 2PL	10
2.6 <i>Programme for International Student Assessment</i> (PISA)	11
III MULTILEVEL ITEM RESPONSE THEORY DUA PARAMETER LOGISTIK (Multilevel IRT 2PL)	12
3.1 Pendahuluan	12
3.2 Pengembangan Model Multilevel IRT 2PL	13
3.3 Simulasi Model Multilevel IRT 2PL	27
3.4 Kesimpulan	36
IV MULTILEVEL LATENT REGRESSION ITEM RESPONSE THEORY DUA PARAMETER LOGISTIK (MLR-IRT 2PL)	37
4.1 Pendahuluan	37
4.2 Pengembangan Model MLR-IRT 2PL	38
4.3 Simulasi Model MLR-IRT 2PL	55
4.4 Kesimpulan	67
V PENERAPAN MODEL MULTILEVEL IRT 2PL DAN MLR-IRT 2PL PADA DATA PISA INDONESIA TAHUN 2018	68
5.1 Pendahuluan	68
5.2 Metode Penelitian	69
5.3 Penerapan Multilevel IRT 2PL pada Data PISA	71
5.4 Penerapan Model MLR-IRT 2PL pada Data PISA	73
5.5 Kesimpulan	76
VI PEMBAHASAN UMUM	78
VII SIMPULAN DAN SARAN	82
7.1 Simpulan	82

7.2 Saran	83
DAFTAR PUSTAKA	84
LAMPIRAN	88
RIWAYAT HIDUP	92



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



DAFTAR TABEL

Tabel 1 Struktur Data Respons Butir	7
Tabel 2 Skenario Simulasi Model Multilevel IRT 2PL	28
Tabel 3 Rata-rata Korelasi Parameter Kemampuan Laten	30
Tabel 4 ANOVA Efek Utama Terhadap Korelasi Parameter Kemampuan Laten	31
Tabel 5 Uji Lanjut Efek Utama Terhadap Korelasi Parameter Kemampuan Laten	32
Tabel 6 Rata-rata RMSE Parameter Kemampuan Laten	32
Tabel 7 ANOVA Efek Utama Terhadap RMSE Parameter Kemampuan Laten	32
Tabel 8 Uji Lanjut Efek Utama terhadap RMSE Parameter Kemampuan Laten	33
Tabel 9 Rata-rata RMSE Parameter Daya Beda Butir	34
Tabel 10 ANOVA Efek Utama Terhadap RMSE Parameter Daya Beda Butir	34
Tabel 11 Uji Lanjut Efek Utama terhadap RMSE Daya Beda Butir	34
Tabel 12 Rata-Rata RMSE Parameter Tingkat Kesukaran Butir	35
Tabel 13 ANOVA Efek Utama RMSE Parameter Tingkat Kesukaran Butir	35
Tabel 14 Uji Lanjut Efek Utama RMSE Parameter Tingkat Kesukaran Butir	35
Tabel 15 Faktor dan level yang digunakan pada model MLR-IRT 2PL	56
Tabel 16 Skenario Simulasi Model MLR-IRT 2 PL	56
Tabel 17 Rata-rata Korelasi dan RMSE Parameter Kemampuan Laten	61
Tabel 18 Uji ANOVA Korelasi Parameter Kemampuan Laten	61
Tabel 19 Uji lanjut terhadap Nilai Korelasi Kemampuan Laten	63
Tabel 20 ANOVA RMSE Parameter Kemampuan Laten	63
Tabel 21 Uji lanjut RMSE Kemampuan Laten	63
Tabel 22 Rata-rata RMSE Parameter Butir	65
Tabel 23 ANOVA Parameter Daya Beda Butir	65
Tabel 24 Uji Lanjut Parameter Daya Beda Butir	66
Tabel 25 ANOVA Parameter Kesukaran Butir	66
Tabel 26 Uji lanjut Parameter Kesukaran butir	67

DAFTAR GAMBAR

Gambar 1 Kebaruan penelitian	6
Gambar 2 Rata-rata Korelasi Parameter Kemampuan Laten	31
Gambar 3 Rata-rata Korelasi Parameter Kemampuan Laten	62
Gambar 4 Sebaran Pendugaan Kemampuan Literasi Matematis	72
Gambar 5 Pendugaan Parameter Daya Beda Butir	72
Gambar 6 Pendugaan Parameter Tingkat Kesukaran Butir	73
Gambar 7 Sebaran Pendugaan Kemampuan Literasi Matematis	74
Gambar 8 Pendugaan Parameter Daya Beda Soal Domain Matematika PISA	74
Gambar 9 Pendugaan Parameter Kesukaran Soal Domain Matematika PISA	75



DAFTAR LAMPIRAN

Lampiran 1 Kode program R untuk menentukan kombinasi faktor simulasi kinerja Model Multilevel IRT 2PL	89
Lampiran 2 Kode program R untuk menentukan kombinasi faktor simulasi kinerja Model MLR-IRT 2PL	90

GLOSARIUM

Butir soal

Satuan pertanyaan dalam tes yang digunakan untuk mengukur kemampuan laten peserta tes.

Data respons butir

Data jawaban peserta tes terhadap setiap butir soal. Pada penelitian ini, respons butir bersifat dikotomis, yaitu bernilai 1 jika jawaban benar dan 0 jika jawaban salah.

Data respons butir berstruktur hierarki

Data respons butir yang memiliki struktur bersarang, misalnya respons butir berada pada peserta tes, dan peserta tes tersarang dalam sekolah.

Data respons butir berstruktur hierarki murni

Data hierarki ketika setiap unit pada level lebih rendah hanya menjadi anggota satu unit pada level yang lebih tinggi, misalnya setiap siswa hanya berasal dari satu sekolah.

Efek acak sekolah

Komponen acak yang merepresentasikan variasi kemampuan laten antarsekolah.

Efek acak siswa

Komponen acak yang merepresentasikan variasi kemampuan laten antarsiswa di dalam sekolah.

Gauss–Hermite Quadrature (GHQ)

Metode integrasi numerik untuk menghampiri integral terhadap sebaran Normal. Pada penelitian ini, GHQ digunakan untuk menghitung likelihood marginal yang melibatkan efek acak siswa dan sekolah.

IRT (Item Response Theory)

Teori pengukuran yang menjelaskan hubungan antara kemampuan laten peserta tes, karakteristik butir, dan peluang peserta menjawab butir dengan benar.

IRT 2PL (Two-Parameter Logistic Item Response Theory)

Model IRT dua parameter logistik yang memodelkan peluang menjawab benar berdasarkan kemampuan laten, daya beda butir, dan tingkat kesukaran butir.

Kemampuan laten

Kemampuan yang tidak dapat diamati secara langsung, tetapi diduga berdasarkan respons peserta terhadap butir soal.

Kesukaran butir

Parameter yang menunjukkan tingkat kemampuan laten yang diperlukan peserta agar memiliki peluang menjawab butir dengan benar.

MLR-IRT 2PL (Multilevel Latent Regression Item Response Theory Two-Parameter Logistic)

Perluasan model Multilevel IRT 2PL yang memasukkan prediktor level siswa dan level sekolah ke dalam struktur kemampuan laten.

MML (Marginal Maximum Likelihood)

Metode pendugaan parameter dengan memaksimumkan likelihood marginal.

Model multilevel

Model statistik yang digunakan untuk menganalisis data berstruktur hierarki dengan memisahkan sumber keragaman pada beberapa level.

Multilevel IRT 2PL

Model IRT 2PL yang dikembangkan untuk data respons butir berstruktur hierarki dengan memasukkan efek acak siswa dan efek acak sekolah dalam kemampuan laten.

PISA (Programme for International Student Assessment)

Program asesmen internasional yang mengukur kemampuan siswa usia 15 tahun pada domain membaca, matematika, sains, dan kompetensi global. Dalam penelitian ini digunakan data respons butir domain matematika PISA Indonesia tahun 2018.