

EVALUASI KEANDALAN DAN KARAKTERISTIK SOAL EKSAKTA  
PADA *COMPUTER-BASED ADAPTIVE TEST* DENGAN METODE  
*ITEM RESPONSE-THEORY THREE-PARAMETER LOGISTICS*

**SKRIPSI**

**FIKIH FIRMANSYAH**

**181401010**



**PROGRAM STUDI S-1 ILMU KOMPUTER**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**  
**UNIVERSITAS SUMATERA UTARA**  
**MEDAN**  
**2024**

**EVALUASI KEANDALAN DAN KARAKTERISTIK SOAL EKSAKTA  
PADA *COMPUTER-BASED ADAPTIVE TEST* DENGAN METODE  
*ITEM RESPONSE-THEORY THREE-PARAMETER LOGISTICS***

**SKRIPSI**

**Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh  
ijazah Sarjana Ilmu Komputer**

**FIKIH FIRMANSYAH**

**181401010**



**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2024**

## PERSETUJUAN

Judul : EVALUASI KEANDALAN DAN KARAKTERISTIK SOAL EKSAKTA PADA COMPUTER-BASED ADAPTIVE TEST DENGAN METODE ITEM RESPONSE-THEORY THREE-PARAMETER LOGISTICS

Kategori : SKRIPSI

Nama : FIKRI FIRMANSYAH

Nomor Induk Mahasiswa: 181401010

Program Studi : SARJANA (S-1) ILMU KOMPUTER

Departemen : ILMU KOMPUTER

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI INFORMASI UNIVERSITAS SUMATERA UTARA

Medan, Juni 2024

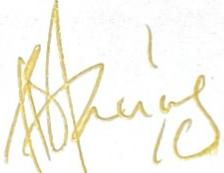
Komisi Pembimbing :

Pembimbing 1



Dr. Amalia, S.T., M.T.  
NIP: 197812212014042001

Pembimbing 2



Sri Melvani Hardi S.Kom., M.Kom.  
NIP. 198805012015042006

Diketahui/disetujui oleh  
Program Studi S-1 Ilmu Komputer

Ketua



Dr. Amalia, S.T., M.T.  
NIP: 197812212014042001

**PERNYATAAN**

EVALUASI KEANDALAN DAN KARAKTERISTIK SOAL EKSAKTA PADA  
*COMPUTER-BASED ADAPTIVE TEST DENGAN METODE ITEM*  
*RESPONSE-THEORY THREE-PARAMETER LOGISTICS*

**SKRIPSI**

Saya menyatakan bahwa skripsi ini adalah hasil karya saya sendiri, kecuali untuk beberapa kutipan dan ringkasan yang telah disebutkan sumbernya.

Medan, 24 Juni 2024



Fikih Firmansyah

181401010

## PENGHARGAAN

Alhamdulillahirrabbil'alamin dengan menyebut nama Allah, penulis ingin menyampaikan rasa syukur dan terima kasih atas berkah dan petunjuk-Nya yang telah membimbing penulis dalam menyelesaikan skripsi ini, yang merupakan syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi S-1 Ilmu Komputer, Universitas Sumatera Utara.

Penulis ingin menyampaikan penghargaan dan terima kasih kepada:

1. Orang tua penulis, bapak Muliadi dan Ibu Syafrita Hanum, adik Vauzi Firmansyah dan Dimas Aditia, yang selalu memberikan doa dan dukungan selama penulis menempuh perkuliahan hingga menyelesaikan skripsi ini.
2. Dr. Muriyanto Amin, S.Sos., M.Si., Rektor Universitas Sumatera Utara.
3. Dr. Maya Silvi Lydia, B.Sc., M.Sc., Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
4. Dr. Amalia, S.T., M.T., Ketua Program Studi S-1 Ilmu Komputer Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara dan Dosen Pembimbing I, yang memberikan motivasi, saran, dan masukan yang berharga sehingga penulis dapat menyelesaikan skripsi ini.
5. Sri Melvani Hardi, S.Kom., M.Kom., Sekretaris Program Studi S-1 Ilmu Komputer Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara dan Dosen Pembimbing II, yang memberikan semangat, saran, dan masukan yang berharga sehingga penulis dapat menyelesaikan skripsi ini.
6. Sahabat penulis selama kuliah dan menjalani tugas akhir yang paling banyak berkontribusi dalam memberikan semangat, ide, motivasi, dan kenangan yang paling berharga kepada penulis, Alvin Adam Nst, Angga Maulana P, Dwi Martin S, M Iqbal Maulana, Rasyid Hafiz, Khairul A, Tsabitah Syafanaura, teman-teman dari komunitas HIMALA USU, SUARA USU, dan GDG Medan.
7. Serta kepada semua pihak yang dengan sangat baik telah membantu penulis yang belum dapat penulis sertakan satu-persatu. Semoga Allah SWT memberkahi kebaikan dan bantuan yang telah diberikan.

Medan, Juni 2024  
Penulis

**EVALUASI KEANDALAN DAN KARAKTERISTIK SOAL EKSAKTA  
PADA *COMPUTER-BASED ADAPTIVE TEST* DENGAN METODE  
*ITEM RESPONSE-THEORY THREE-PARAMETER LOGISTICS***

**ABSTRAK**

*Computer-Based Adaptive Test* (CAT) telah menjadi metode penilaian yang semakin populer karena kemampuannya untuk menyesuaikan tingkat kesulitan soal dengan kemampuan peserta secara real-time. Penelitian ini bertujuan untuk mengevaluasi keandalan dan karakteristik soal eksakta pada CAT dengan menggunakan model *Item Response Theory* (IRT) *Three-Parameter Logistics* (3PL). Data respons 20 siswa pada 20 soal ujian eksakta dianalisis untuk mengestimasi parameter model 3PL, yaitu diskriminasi (a), kesulitan (b), dan tebakan (c). Hasil analisis menunjukkan bahwa sebagian besar soal memiliki karakteristik yang baik, dengan nilai diskriminasi yang cukup tinggi dan nilai tebakan yang rendah. Parameter kesulitan bervariasi, menunjukkan adanya soal dengan tingkat kesulitan yang mudah, sedang, dan sulit. Skor siswa pada tes klasik dibandingkan dengan skor yang diperoleh melalui penyetaraan IRT. Hasilnya menunjukkan bahwa model 3PL mampu membedakan karakteristik soal dengan lebih baik daripada tes klasik. Uji fit dengan *Chi-Square* digunakan untuk menguji kesesuaian model 3PL dengan data respons. Penelitian ini menunjukkan bahwa model IRT 3PL dapat digunakan untuk mengevaluasi dan meningkatkan kualitas soal pada CAT, sehingga memiliki validitas dan reliabilitas dengan hasil uji chi-square kurang menunjukkan nilai  $\chi^2 = 2$  dengan df (Derajat Kebebasan)= 3 dan P-value = 0.57. Dengan  $\alpha = 0.05$ , karena P-value >  $\alpha$ , tidak ada cukup bukti untuk menyatakan perbedaan signifikan antara distribusi kemampuan siswa yang diamati, ini menunjukkan bahwa model 3PL cukup sesuai dengan data respons sehingga diharapkan dapat digunakan sebagai alternatif efektif untuk CBT tradisional.

**Kata kunci:** *Computer-Based Adaptive Test* (CAT), *Item Response Theory* (IRT), *Three-Parameter Logistics* (3PL), Evaluasi Pendidikan, Karakteristik Butir Soal.

**EVALUATION OF RELIABILITY AND CHARACTERISTICS OF EXACT  
QUESTIONS ON COMPUTER-BASED ADAPTIVE TESTS WITH ITEM  
RESPONSE-THEORY THREE-PARAMETER LOGISTICS METHOD**

**ABSTRACT**

*Computer-Based Adaptive Test (CAT) has become an increasingly popular assessment method due to its ability to adjust the difficulty level of questions to the ability of participants in real-time. This study aims to evaluate the reliability and characteristics of exact sciences questions on CAT using the Item Response Theory (IRT) Three-Parameter Logistics (3PL) model. The response data of 20 students on 20 exact exam questions were analyzed to estimate the 3PL model parameters, namely discrimination ( $a$ ), difficulty ( $b$ ), and guessing ( $c$ ). The results of the analysis showed that most questions had good characteristics, with fairly high discrimination values and low guessing values. The difficulty parameters varied, indicating that there were questions with easy, moderate, and difficult difficulty levels. Students' scores on the classical test were compared with scores obtained through IRT equating. The results showed that the 3PL model was able to distinguish the characteristics of the questions better than the classical test. The Chi-Square fit test was used to test the fit of the 3PL model to the response data. This study shows that the IRT 3PL model can be used to evaluate and improve the quality of questions in CAT, so that it has validity and reliability with the results of the chi-square test showing a value of  $\chi^2 = 2$  with  $df$  (Degrees of Freedom) = 3 and  $P\text{-value} = 0.57$ . With  $\alpha = 0.05$ , since the  $P\text{-value} > \alpha$ , there is not enough evidence to declare a significant difference between the observed student ability distributions, this indicates that the 3PL model fits the response data reasonably well so it is expected to be used as an effective alternative to traditional CBT.*

**Keyword:** Computer-Based Adaptive Test (CAT), Item Response Theory (IRT), Three-Parameter Logistics (3PL), Learning Evaluation, Item Characteristics.

## DAFTAR ISI

<b>PERNYATAAN .....</b>	i
<b>PENGHARGAAN .....</b>	ii
<b>ABSTRAK .....</b>	iii
<b><i>ABSTRACT</i> .....</b>	iv
<b>DAFTAR ISI.....</b>	viii
<b>DAFTAR GAMBAR.....</b>	xi
<b>DAFTAR TABEL .....</b>	xii
<b>BAB 1 PENDAHULUAN .....</b>	1
<b>1.1 Latar Belakang.....</b>	1
<b>1.2 Rumusan Masalah .....</b>	2
<b>1.3 Batasan Masalah .....</b>	3
<b>1.4 Tujuan Penelitian.....</b>	4
<b>1.5 Manfaat Penelitian.....</b>	4
<b>1.6 Metodologi Penelitian .....</b>	4
<b>1.7 Sistematika Penulisan .....</b>	5
<b>BAB 2 LANDASAN TEORI .....</b>	7
<b>2.1 Teori Respons Item (IRT) .....</b>	7
<b>2.1.1 Asumsi Dasar IRT.....</b>	7
<b>2.1.1.1 Sifat Unidimensi .....</b>	8
<b>2.1.1.2 Independensi Lokal dari Item.....</b>	8
<b>2.1.1.3 Invariansi Parameter.....</b>	8
<b>2.1.2 Model Respon Item (Item Response Function - IRF).....</b>	8
<b>2.1.3 Parameter Model IRT .....</b>	9
<b>2.1.3.2 Parameter Kesulitan (<i>b</i>) .....</b>	10
<b>2.1.3.3 Parameter Tebakan (<i>c</i>).....</b>	10
<b>2.1.4 Model IRT.....</b>	11
<b>2.1.4.1 Model Logistik Satu Parameter (1PL).....</b>	11
<b>2.1.4.2 Model <i>Two-Parameter Logistic</i> (2PL) .....</b>	12
<b>2.1.4.3 Model <i>Three-Parameter Logistic</i> (3PL) .....</b>	13
<b>2.1.5 Skor Laten (<i>θ</i>).....</b>	14
<b>2.2 Sistem Ujian Adaptif Berbasis Komputer (CAT) .....</b>	14
<b>2.3 Laravel .....</b>	15

<b>2.4 PostgreSQL.....</b>	<b>16</b>
<b>2.5 Penelitian yang Relevan .....</b>	<b>16</b>
<b>BAB 3 ANALISIS DAN PERANCANGAN SISTEM .....</b>	<b>19</b>
<b>3.1 Analisis Masalah .....</b>	<b>19</b>
<b>3.2 Analisis Sistem.....</b>	<b>20</b>
<b>3.3 Perancangan Sistem.....</b>	<b>20</b>
<b>3.3.1 Pengumpulan Dataset .....</b>	<b>20</b>
<b>3.3.1.1 Pengumpulan Soal Ujian .....</b>	<b>21</b>
<b>3.3.1.2 Pengumpulan Respon Peserta .....</b>	<b>21</b>
<b>3.2 Perancangan Database .....</b>	<b>26</b>
<b>3.3 Perancangan Interface .....</b>	<b>28</b>
<b>3.3.1 Peserta Ujian - Tampilan Perancangan Home .....</b>	<b>28</b>
<b>3.3.2 Peserta Ujian - Tampilan Perancangan Ujian .....</b>	<b>29</b>
<b>3.3.3 Admin - Halaman Daftar Ujian.....</b>	<b>30</b>
<b>3.3.4 Admin - Halaman Bank Soal .....</b>	<b>31</b>
<b>BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM .....</b>	<b>32</b>
<b>4.1 Implementasi Sistem .....</b>	<b>32</b>
<b>4.1.1 Sumber Data.....</b>	<b>32</b>
<b>4.1.2 Karakteristik Responden .....</b>	<b>32</b>
<b>4.1.3 Distribusi Data .....</b>	<b>33</b>
<b>4.1.4.1 Estimasi Parameter Soal .....</b>	<b>36</b>
<b>4.1.4.1.1 Parameter Diskriminasi (a).....</b>	<b>36</b>
<b>4.1.4.1.2 Parameter Kesulitan (b).....</b>	<b>39</b>
<b>4.1.4.1.3 Parameter Tebakan (c).....</b>	<b>41</b>
<b>4.1.4.2 Estimasi Skor Laten (Theta).....</b>	<b>41</b>
<b>4.2.4.2 Perbandingan Skor Klasik dengan Penyetaraan IRT .....</b>	<b>43</b>
<b>4.2.4.2 Analisis Item Characteristic Curve (ICC) .....</b>	<b>45</b>
<b>4.2.4.3 Evaluasi Karakteristik Butir .....</b>	<b>48</b>
<b>4.2 Implementasi Interface.....</b>	<b>49</b>
<b>4.2.1 Tampilan Halaman Home .....</b>	<b>50</b>
<b>4.2.2 Tampilan Halaman Ujian.....</b>	<b>51</b>
<b>4.2.3 Tampilan Halaman Login Admin.....</b>	<b>51</b>
<b>4.2.4 Tampilan Halaman Daftar Peserta .....</b>	<b>52</b>
<b>4.2.5 Tampilan Halaman Daftar Bank Soal .....</b>	<b>53</b>
<b>4.2.6 Tampilan Halaman Analisis Karakteristik Soal.....</b>	<b>54</b>
<b>4.3 Pengujian .....</b>	<b>55</b>

<b>4.3.1 Metode Uji Kesesuaian Model .....</b>	<b>55</b>
<b>4.3.2 Hasil Uji Kesesuaian Model .....</b>	<b>58</b>
<b>BAB 5 PENUTUP .....</b>	<b>59</b>
<b>5.1 Kesimpulan .....</b>	<b>59</b>
<b>5.2 Saran .....</b>	<b>59</b>
<b>DAFTAR PUSTAKA .....</b>	<b>61</b>

**DAFTAR GAMBAR**

Gambar 3.1 Diagram Ishikawa .....	19
Gambar 3.2: Arsitektur Umum .....	20
Gambar 3.3 Rancangan <i>Database</i> .....	27
Gambar 3.4: Peserta Ujian - Tampilan Perancangan Home .....	28
Gambar 3.5: Peserta Ujian - Tampilan Perancangan Ujian .....	29
Gambar 3.6: Admin - Halaman Daftar Ujian.....	30
Gambar 3.7: Admin - Halaman Bank Soal .....	31
Gambar 4.1: Visualisasi Data Skor Siswa .....	36
Gambar 4.2: Kurva Karakteristik Item (ICC) .....	47
Gambar 4.3: Peserta Ujian - Halaman Home.....	50
Gambar 4.4: Tampilan Halaman Ujian .....	51
Gambar 4.5 Tampilan Login Admin.....	51
Gambar 4.6: Tampilan Halaman Daftar Peserta .....	52
Gambar 4.0.7: Tampilan Halaman Daftar Bank Soal .....	53
Gambar 4.8: Tampilan Halaman Analisis Karakteristik Soal.....	54

**DAFTAR TABEL**

Tabel 3.1: Contoh Soal dengan Parameter Dasar .....	21
Tabel 3.2: Contoh Respon Hasil Ujian Peserta.....	21
Tabel 3.3: Contoh Hasil Kalibrasi Parameter .....	23
Tabel 3.4: Contoh Ledger Jawaban Siswa.....	24
Tabel 4.1: Distribusi Data Jawaban .....	34
Tabel 4.2: Distribusi Skor .....	35
Tabel 4.3: Nilai (a) yang terkalibrasi .....	38
Tabel 4.4: Nilai (b) yang terkalibrasi .....	40
Tabel 4.5 Hasil perulangan theta menggunakan Log-Likelihood.....	42
Tabel 4.6: Perbandingan Skor Klasik dengan Penyetaraan .....	43
Tabel 4.7: Data plot ICC .....	46
Tabel 4.8: Hasil Evaluasi Karakteristik Butir .....	48
Tabel 4.9: Theta Akhir Peserta .....	56
Tabel 4.10: Frekuensi yang Diamati ( $O$ ) .....	57
Tabel 4.11 Hitung Statistik Chi-Square .....	57

## BAB 1

### PENDAHULUAN

#### 1.1 Latar Belakang

Dalam era digital saat ini, pendidikan dan penilaian kompetensi telah mengalami transformasi signifikan melalui adopsi teknologi berbasis komputer. Pengujian adaptif berbasis komputer atau *Computer-based Adaptive Test* (CAT), yang menggunakan algoritma untuk menyesuaikan tingkat kesulitan soal berdasarkan jawaban peserta, menjadi topik penelitian penting dalam bidang psikometri. Fokus untuk meningkatkan akurasi dan efisiensi dalam sistem CAT selama lima tahun terakhir ini memang telah menekankan pada penggunaan Teori Respon Butir atau *Item Response Theory* (IRT), khususnya model *Three-Parameter Logistics* (3PL). Pergeseran ini mencerminkan evolusi yang sedang berlangsung dalam metodologi asesmen pendidikan dan psikologis. Sistem CAT mengadaptasi tingkat kesulitan dan jumlah pertanyaan berdasarkan kemampuan peserta ujian. Sistem ini dimulai dengan pertanyaan dengan tingkat kesulitan sedang, kemudian menyesuaikan pertanyaan selanjutnya berdasarkan respon peserta ujian. Metode ini tidak hanya meningkatkan pengalaman ujian dengan memberikan tingkat tantangan yang sesuai, tetapi juga meningkatkan keamanan ujian karena setiap peserta ujian akan menghadapi serangkaian pertanyaan yang unik (Nathan Thompson, 2023).

IRT, berbeda dengan Teori Tes Klasik atau *Classic Test Theory* (CTT), menawarkan pendekatan yang lebih kompleks dan komprehensif untuk analisis tes. IRT berperan penting dalam mengembangkan bank soal, mendesain bentuk tes, dan penilaian. IRT sangat penting dalam pengujian adaptif karena IRT menempatkan butir soal dan peserta tes pada skala laten yang sama, sehingga memfasilitasi proses penilaian yang lebih akurat dan efisien (Reid et al., 2020).

Meskipun IRT dan CTT digunakan untuk memastikan reliabilitas dan validitas tes, namun keduanya memiliki perbedaan yang signifikan dalam pendekatannya. Sebagai contoh, IRT mempertimbangkan tingkat kesulitan dan diskriminasi melalui parameter seperti  $b$  dan  $a$ , serta menyertakan parameter  $c$  untuk menebak jawaban. Sebagai perbandingan, CTT menggunakan metrik yang lebih sederhana seperti proporsi-koreksi dan korelasi point-biserial seperti 1 dan 0.

Selain itu, IRT memungkinkan pengaitan dan penyetaraan nilai tes yang lebih kuat di berbagai bentuk dan tingkatan, yang sangat berharga dalam penilaian ujian (Anak Agung Purwa Antara, 2020).

Secara keseluruhan, pergeseran ke arah IRT dan khususnya model 3PL dalam penelitian CAT menggarisbawahi perlunya metode penilaian yang lebih baik, efisien, dan akurat dalam tes pendidikan dan psikologi. IRT menyediakan kerangka kerja matematis untuk menganalisis sifat respons individu terhadap soal tes. Model 3PL, yang memperhitungkan kesulitan soal (*difficulty*), daya beda soal (*discrimination*), dan tebakan (*guessing*), telah menunjukkan potensi besar dalam mempersonalisasi penilaian (Hamidah & Istiyono, 2022). Daya beda butir soal menunjukkan seberapa besar kemampuan individu dapat memprediksi apakah peserta tes akan menjawab butir soal tersebut dengan benar. Tingkat kesukaran butir soal menunjukkan seberapa besar kemungkinan peserta tes akan menjawab butir soal tersebut dengan benar. Parameter pengecoh menunjukkan seberapa besar kemungkinan peserta tes akan menjawab butir soal tersebut dengan benar meskipun memiliki kemampuan yang rendah (Syamsuddin, 2023).

Studi-studi terbaru juga mengeksplorasi aplikasi model 3PL dalam berbagai disiplin dan konteks, seperti pendidikan medis, teknik, dan bahasa asing. Penelitian ini menyoroti bagaimana karakteristik butir soal dapat diidentifikasi dan dioptimalkan menggunakan model 3PL, membantu pendidik dan peneliti untuk merancang tes yang lebih efektif dan adil. Keandalan pengukuran kompetensi juga menjadi topik penting, dengan studi yang menunjukkan bagaimana model 3PL dapat membantu mengurangi bias dan meningkatkan validitas hasil tes.

Berdasarkan latar belakang diatas, maka pada kesempatan kali ini penulis ingin melakukan penelitian yang berfokus pada evaluasi keandalan dan karakteristik soal eksakta melalui Model *Three-Parameter Logistic* dari *Item Response Theory* (IRT) dalam sistem CAT.

## 1.2 Rumusan Masalah

Penerapan model *Three-Parameter Logistic* (3PL) dari *Item Response Theory* (IRT) melalui penilaian *Computer-based Adaptive Test* (CAT) untuk meningkatkan kualitas soal dibandingkan dengan metode pengujian klasik (CTT),

dengan mengidentifikasi parameter diskriminasi (a), kesulitan (b), dan tebakan (c) dari soal, dan menguji kesesuaian model dengan data respons siswa menggunakan uji Chi-Square untuk menilai validitas dan reliabilitas sistem CAT.

### 1.3 Batasan Masalah

Berdasarkan latar belakang dan rumusan masalah, maka penulis membatasi ruang masalah agar tidak menyimpang dari tujuan yang diharapkan. Batasan-batasan masalah yang digunakan adalah sebagai berikut:

- 1) Fokus pada penerapan model *Three-Parameter Logistic* (3PL) dalam *Item Response Theory* (IRT) untuk sistem *Computer-based Adaptive Test* (CAT).
- 2) Analisis menggunakan data ujian yang telah diujangkan, tanpa mengumpulkan data primer baru. Soal berjumlah 20 pilihan ganda mata pelajaran Fisika kelas 12 SMA yang tidak menggunakan sub-topik.
- 3) Ruang lingkup terbatas pada konteks pendidikan, khususnya pengujian kompetensi akademik eksakta.
- 4) Fokus pada peningkatan kualitas soal ujian, bukan pada aspek administratif atau logistik ujian.
- 5) Tidak mengkaji dampak penggunaan sistem *Computer-based Adaptive Test* (CAT) terhadap motivasi dan kepuasan peserta ujian, dan tidak menguji kompatibilitas atau integrasi software, interoperabilitas sistem CAT dengan berbagai platform teknologi pendidikan lainnya.
- 6) Uji Fungsionalitas: Memastikan semua fitur sistem bekerja sesuai desain, termasuk input soal, melaksanakan tes adaptif, pencatatan jawaban, dan hasil analisis *Item Response Theory* (IRT).
- 7) Validasi Model IRT 3PL: Membandingkan hasil analisis sistem dengan hasil analisis manual dan uji fit dengan *Chi-Square*.
- 8) Sistem yang dibuat menggunakan *framework backend* Laravel (PHP), *frontend* menggunakan VueJS, *database management system* PostgreSQL, dan menggunakan server cloud.

## 1.4 Tujuan Penelitian

Sebagaimana yang menjadi permasalahan pada penelitian ini, tujuannya adalah mengembangkan dan menerapkan sistem ujian adaptif berbasis komputer yang menggunakan model *Three-Parameter Logistic* dari *Item Response Theory*. Sistem ini dirancang untuk menganalisis secara akurat karakteristik butir soal dan meningkatkan keandalan pengukuran kompetensi.

## 1.5 Manfaat Penelitian

Manfaat yang diharapkan dapat diraih dari penelitian ini adalah:

- 1) Mengembangkan sistem penilaian berbasis komputer yang lebih efisien dan adaptif dengan menggunakan metode Item Response Theory (IRT) Three-Parameter Logistics.
- 2) Meningkatkan keandalan dan validitas evaluasi soal eksakta, memastikan bahwa penilaian lebih akurat dan adil bagi siswa.
- 3) Memberikan dasar yang kuat bagi institusi pendidikan untuk mengoptimalkan proses evaluasi dan meningkatkan kualitas pendidikan secara keseluruhan.

## 1.6 Metodologi Penelitian

Metodologi penelitian yang dilakukan pada penelitian ini adalah:

### 1) Studi Pustaka

Melakukan penelitian literatur dengan mencari dan menganalisis referensi dari buku, jurnal, artikel ilmiah, dan sumber online yang relevan dengan IRT, model 3PL, dan sistem CAT.

### 2) Analisa dan Perancangan

Membangun dan mengintegrasikan model 3PL ke dalam sistem CAT yang dikembangkan. Ini termasuk *coding*, pengaturan database, dan integrasi algoritma.

### **3) Implementasi**

Menerapkan sistem CAT yang dikembangkan dalam lingkungan yang terkontrol untuk menguji fungsionalitas dan kinerjanya.

### **4) Pengujian dan Evaluasi**

Melakukan pengujian sistem untuk menilai keakuratan, efisiensi, dan keandalan sistem dalam menganalisis karakteristik butir soal dan mengukur kompetensi.

### **5) Analisis Data dan Interpretasi**

Menganalisis data yang dihasilkan dari pengujian, menggunakan teknik statistik untuk mengevaluasi hasil model 3PL.

### **6) Dokumentasi**

Mendokumentasikan seluruh proses penelitian dan hasilnya untuk disajikan dalam skripsi.

## **1.7 Sistematika Penulisan**

Berikut merupakan sistematika yang digunakan pada penulisan ini:

### **BAB 1 PENDAHULUAN**

Bab ini membahas mengenai latar belakang pemilihan judul skripsi "Evaluasi Keandalan Dan Karakteristik Soal Eksakta Pada Computer-Based Adaptive Test Dengan Metode *Item Response Theory Three-Parameter Logistics*", termasuk penjelasan mengenai konteks penelitian, rumusan masalah yang dihadapi, batasan-batasan penelitian, tujuan yang ingin dicapai, manfaat dari penelitian tersebut, metode yang digunakan dalam penelitian, dan juga sistematika penulisan yang akan diikuti dalam skripsi.

### **BAB 2 LANDASAN TEORI**

Bab ini membahas teori-teori yang relevan dengan penelitian, termasuk *Computer-Based Adaptive Test* (CAT), *Item Response Theory* (IRT), model *Three-Parameter Logistics* (3PL), serta penelitian terdahulu yang relevan dengan topik tersebut. Selain itu,

bab ini juga membahas konsep keandalan dan validitas dalam evaluasi pendidikan.

**BAB 3****ANALISIS DAN PERANCANGAN**

Bab ini membahas analisis masalah penelitian dan perancangan sistem CAT menggunakan metode IRT 3PL. Pembahasan mencakup analisis kebutuhan, perancangan basis data, perancangan modul soal adaptif, perancangan antarmuka pengguna, serta diagram-diagram seperti diagram *use case*, diagram aktivitas, *sequence diagram*, dan flowchart yang diperlukan..

**BAB 4****IMPLEMENTASI DAN PENGUJIAN SISTEM**

Bab ini membahas implementasi sistem CAT yang telah dibuat, meliputi pembuatan basis data, pengembangan modul soal adaptif, pengembangan antarmuka pengguna, dan integrasi sistem secara keseluruhan. Selanjutnya, bab ini mencakup tahapan pengujian sistem untuk memastikan bahwa sistem berjalan sesuai dengan perancangan. Pengujian juga melibatkan perbandingan dengan hitungan manual untuk memverifikasi akurasi dan keandalan sistem.

**BAB 5****KESIMPULAN DAN SARAN**

Bab terakhir ini mencakup rangkuman hasil penelitian beserta saran untuk pengembangan penelitian selanjutnya. Kesimpulan meliputi hasil dari implementasi dan pengujian sistem CAT, serta implikasi dari temuan penelitian tersebut. Saran diberikan untuk pengembangan lebih lanjut dari sistem CAT dan kemungkinan penelitian lanjutan yang dapat dilakukan..

## **BAB 2**

### **LANDASAN TEORI**

#### **2.1 Teori Respons Item (IRT)**

Konsep fungsi respon item sudah ada sejak sebelum tahun 1950, tetapi perintisan penelitian mengenai Teori Respons Item (IRT) terjadi pada tahun 1950-an dan 1960-an. Tiga tokoh pionir dalam pengembangan IRT adalah Frederick M. Lord, Georg Lasch, dan Paul Lazarsfeld. Mereka melakukan penelitian independen yang mendahului pengembangan teori ini. Tokoh penting lainnya dalam kemajuan IRT adalah Benjamin Drake Wright dan David Andrich. IRT mulai populer pada akhir tahun 1970-an dan 1980-an, ketika praktisi mendapat informasi tentang manfaatnya dan komputasi menjadi lebih mudah dengan hadirnya komputer pribadi (Livingston, 2020).

Tujuan utama IRT adalah memberikan kerangka kerja untuk mengevaluasi kualitas suatu penilaian dan seberapa baik setiap elemen penilaian berfungsi. IRT banyak digunakan dalam pendidikan untuk mengembangkan tes, mengelola bank soal, dan menyeimbangkan tingkat kesulitan soal. Model IRT sering disebut sebagai model sifat laten karena mengacu pada respons individu sebagai manifestasi atribut atau karakteristik yang tidak dapat diamati secara langsung.

IRT dianggap sebagai perbaikan dari Teori Tes Klasik (CTT) karena lebih fleksibel dan memberikan informasi yang lebih kompleks. IRT memungkinkan aplikasi seperti tes adaptif berbasis komputer, yang tidak dapat dilakukan dengan CTT. Keuntungan utama IRT dibandingkan CTT adalah kemampuannya menyediakan informasi yang lebih kompleks, yang dapat membantu peneliti meningkatkan keandalan penilaian mereka (Ridho et al., 2007).

##### **2.1.1 Asumsi Dasar IRT**

Penggunaan IRT memerlukan asumsi-asumsi pendukung yang secara tidak langsung dapat diukur dan diverifikasi. Asumsi-asumsi yang diperlukan dalam teori respons butir meliputi unidimensionalitas, independensi lokal, dan invariansi parameter. Hal ini penting untuk

memperhatikan pemilihan model yang tepat dalam penggunaan IRT (Sijtsma & van der Ark, 2022).

#### **2.1.1.1 Sifat Unidimensi**

Asumsi ini menyatakan bahwa sebuah tes atau kuesioner mengukur satu konstruk atau dimensi psikologis saja. Artinya, semua item atau pertanyaan dalam sebuah tes seharusnya berkaitan dengan satu konsep atau kemampuan tertentu. Misalnya, jika sebuah tes bertujuan mengukur matematika, maka semua item di dalamnya harus berkaitan dengan kemampuan matematika dan tidak mencampur dengan kemampuan lain seperti kemampuan verbal.

#### **2.1.1.2 Independensi Lokal dari Item**

Asumsi ini menekankan bahwa respon seseorang terhadap suatu item soal tidak dipengaruhi oleh respon terhadap item soal lainnya dalam tes yang sama. Dengan kata lain, bagaimana seseorang menjawab satu pertanyaan tidak tergantung atau tidak mempengaruhi cara mereka menjawab pertanyaan lain. Ini penting untuk memastikan bahwa setiap item berkontribusi secara independen terhadap pengukuran keseluruhan konstruk.

#### **2.1.1.3 Invariansi Parameter**

Syarat utama ini mengindikasikan bahwa fungsi atau kurva respons atau karakteristik butir tetap konsisten, bahkan jika kelompok peserta yang menjawab butir tersebut berubah-ubah. Selain itu, untuk kelompok yang sama, karakteristik mereka tetap konstan meskipun butir yang mereka jawab berbeda-beda.

### **2.1.2 Model Respon Item (Item Response Function - IRF)**

Ini berkaitan dengan bagaimana kemungkinan respon seseorang terhadap suatu item dapat diprediksi atau dimodelkan. IRF menggambarkan

hubungan antara kemampuan latent (kemampuan yang tidak langsung terlihat namun diukur melalui tes) seorang individu dan kemungkinan mereka menjawab suatu item dengan benar. Misalnya, dalam model 3PL yang sering digunakan dalam IRT, IRF dipengaruhi oleh tiga parameter: kesulitan item ( $b$ ), daya beda item ( $a$ ), dan tebakan ( $c$ ). Parameter ini membantu dalam memprediksi kemungkinan seseorang dengan tingkat kemampuan tertentu menjawab item tersebut dengan benar. Sifat ini selanjutnya diasumsikan dapat diukur dalam sebuah skala (keberadaan tes mengasumsikan hal ini), biasanya diatur ke skala standar dengan rata-rata 0,0 dan standar deviasi 1,0 dan 0,0 hingga 3,0.

Dalam Teori Respons Item (IRT), rumus umum yang digunakan untuk menggambarkan hubungan antara kemampuan responden dan probabilitas memberikan jawaban yang benar pada suatu item dikenal sebagai Fungsi Respon Item (*Item Response Function - IRF*). Rumus ini bervariasi tergantung pada model IRT yang digunakan.

### 2.1.3 Parameter Model IRT

Salah satu model yang paling sering digunakan adalah model IRT tiga parameter (3PL), atau model logistik tiga parameter, karena hampir selalu dinyatakan dalam bentuk logistik. Seperti semua model IRT, model ini berusaha memprediksi probabilitas respon tertentu berdasarkan tingkat kemampuan atau sifat peserta ujian dan beberapa parameter yang menggambarkan kinerja butir soal. Pada model 3PL, parameter tersebut adalah  $a$  (diskriminasi),  $b$  (tingkat kesulitan atau lokasi), dan  $c$  (tebakan semu) (Gyamfi & Acquaye, 2023).

#### 2.1.3.1 Parameter Diskriminasi ( $a$ )

Parameter diskriminasi, dilambangkan dengan  $a$ , mengukur seberapa baik suatu item dapat membedakan antara responden dengan kemampuan yang berbeda. Ini menunjukkan kemiringan fungsi respons item pada titik kesulitannya (Gyamfi & Acquaye, 2023).

- 1) Item dengan nilai  $a$  yang tinggi sangat sensitif terhadap perbedaan dalam kemampuan responden, sehingga sangat efektif dalam membedakan antara responden dengan kemampuan di atas dan di bawah tingkat kesulitan item.
- 2) Item dengan nilai  $a$  yang rendah kurang sensitif terhadap perbedaan kemampuan, sehingga tidak efektif dalam membedakan antara responden dengan kemampuan yang berbeda.

Parameter ini sangat berguna untuk menentukan item mana yang paling efektif dalam mengukur kemampuan atau sifat yang diinginkan pada tingkat kemampuan yang berbeda.

#### **2.1.3.2 Parameter Kesulitan ( $b$ )**

Parameter kesulitan, sering dilambangkan dengan  $b$ , mengukur tingkat kesulitan suatu item tes. Nilai  $b$  menunjukkan titik pada skala kemampuan di mana responden memiliki 50% peluang untuk menjawab item tersebut dengan benar (Gyamfi & Acquaye, 2023).

- 1) Item dengan nilai  $b$  yang rendah dianggap lebih mudah, karena responden dengan kemampuan yang lebih rendah masih memiliki peluang baik untuk menjawabnya dengan benar.
- 2) Item dengan nilai  $b$  yang tinggi dianggap lebih sulit, karena memerlukan tingkat kemampuan yang lebih tinggi untuk meningkatkan peluang menjawab dengan benar.

Parameter ini sangat penting untuk menentukan pada titik mana sepanjang kontinum kemampuan suatu item paling informatif.

#### **2.1.3.3 Parameter Tebakan ( $c$ )**

Parameter tebakan, yang sering dilambangkan dengan  $c$ , mengukur kemungkinan responden menebak jawaban yang benar, terutama dalam konteks tes pilihan ganda. Ini mencerminkan probabilitas minimal menjawab item dengan benar karena menebak, bukan karena kemampuan (Gyamfi & Acquaye, 2023).

- 1) Item dengan nilai  $c$  yang tinggi menunjukkan peluang yang lebih besar bagi responden untuk menebak jawaban dengan benar, meskipun mereka tidak memiliki kemampuan yang cukup.
- 2) Item dengan nilai  $c$  yang rendah menunjukkan bahwa peluang menebak jawaban dengan benar sangat kecil, sehingga lebih bergantung pada kemampuan nyata responden.

Parameter ini penting dalam model IRT untuk mengakomodasi efek menebak dalam penilaian, terutama dalam tes objektif di mana menebak jawaban bisa menjadi faktor signifikan.

#### **2.1.4 Model IRT**

Model IRT memprediksi jawaban responden terhadap item tes berdasarkan posisi mereka pada kontinum sifat laten dan karakteristik item, yang dikenal sebagai parameter. Asumsi dasarnya adalah setiap respons terhadap item memberikan indikasi tentang tingkat sifat laten atau kemampuan individu. Kemampuan seseorang ( $\theta$ ) diartikan sebagai probabilitas memberikan jawaban yang benar untuk item tersebut. Semakin tinggi kemampuan seseorang, semakin besar peluangnya menjawab benar. Probabilitas menjawab benar meningkat seiring dengan meningkatnya kemampuan responden. Secara teoritis, kemampuan ( $\theta$ ) berkisar antara  $-\infty$  hingga  $+\infty$ , tetapi dalam praktiknya biasanya berkisar antara -3 hingga +3 (Adebule, 2013).

##### **2.1.4.1 Model Logistik Satu Parameter (1PL)**

Model Logistik Satu Parameter (1PL), juga dikenal sebagai Model Rasch, adalah model dasar dalam Teori Respons Item (IRT). Model ini khususnya berfokus pada kesulitan item sebagai satu-satunya parameter untuk mengukur respon tes.

Rumus Model Rasch adalah sebagai berikut:

$$P(X_i = 1 | \theta) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)}$$

Di mana:

- 1)  $P(X_i = 1 | \theta)$  adalah probabilitas seorang responden dengan kemampuan menjawab item  $i$  dengan benar.  $\theta$  adalah kemampuan responden.
- 2)  $b_i$  adalah parameter kesulitan item.

Model ini mengasumsikan bahwa semua item memiliki diskriminasi yang sama dan tidak memperhitungkan efek tebakan. Kelebihan utamanya adalah kesederhanaan dan kemudahan interpretasinya. Model ini sangat berguna dalam membandingkan responden atau item tes secara langsung, seperti dalam penelitian pendidikan dan psikometrik. Model 1PL/Rasch sering digunakan untuk memvalidasi skala pengukuran dan memastikan bahwa item-item dalam tes memenuhi prinsip pengukuran yang konsisten. (Rahim & Haryanto, 2021).

#### **2.1.4.2 Model *Two-Parameter Logistic* (2PL)**

Model Logistik Dua Parameter (2PL) dalam Teori Respons Item (IRT) adalah pengembangan dari Model Rasch (1PL) yang menambahkan parameter kedua untuk diskriminasi item. Dengan demikian, model ini memberikan analisis lebih mendalam pada item tes dengan mempertimbangkan kedua aspek utama: tingkat kesulitan item dan kemampuannya untuk membedakan responden dengan tingkat kemampuan yang berbeda.

Rumus dasar Model 2PL adalah sebagai berikut:

$$P(X_i = 1 | \theta) = \frac{\exp(a_i(\theta - b_i))}{1 + \exp(a_i(\theta - b_i))}$$

Di mana:

- 1)  $P(X_i = 1 | \theta)$  adalah probabilitas seorang responden dengan kemampuan menjawab item  $i$  dengan benar.  $\theta$  adalah kemampuan responden.
- 2)  $a_i$  adalah parameter diskriminasi item yang menunjukkan seberapa baik item tersebut membedakan antara responden dengan kemampuan yang berbeda.

3)  $b_i$  adalah parameter kesulitan item.

Model Logistik Dua Parameter (2PL) dalam Teori Respons Item (IRT) memperhitungkan nilai diskriminasi ( $a$ ) yang berbeda untuk setiap item, yang memungkinkan item menjadi lebih efektif dalam membedakan responden dengan tingkat kemampuan yang beragam. Parameter kesulitan ( $b$ ) menunjukkan tingkat kesulitan item, di mana probabilitas responden menjawab dengan benar adalah 50%. Parameter diskriminasi ( $a$ ) menggambarkan seberapa tajam perbedaan antara responden dengan kemampuan di atas dan di bawah tingkat kesulitan item.

Model ini berguna ketika item dalam tes memiliki tingkat diskriminasi yang berbeda, memungkinkan analisis yang lebih tepat terhadap bagaimana setiap item mengukur kemampuan yang diinginkan. Berbeda dengan Model Rasch, di mana semua item diasumsikan memiliki diskriminasi yang sama, Model 2PL memberikan fleksibilitas lebih dalam menganalisis data tes, di mana item mungkin membedakan secara berbeda di berbagai tingkat kemampuan (Gyamfi & Acquaye, 2023).

#### **2.1.4.3 Model *Three-Parameter Logistic* (3PL)**

Model Logistik Tiga Parameter (3PL) dalam Teori Respons Item (IRT) lebih kompleks daripada model 1PL dan 2PL karena memperkenalkan parameter ketiga, yaitu parameter tebakan. Parameter ini memungkinkan model untuk mempertimbangkan kemungkinan menebak jawaban dengan benar, terutama dalam tes pilihan ganda.

Rumus dasar Model 3PL adalah sebagai berikut:

$$P(X_i = 1 | \theta) = c_i + (1 - c_i) \frac{\exp(a_i(\theta - b_i))}{1 + \exp(a_i(\theta - b_i))}$$

Di mana:

- 1)  $P(X_i = 1 | \theta)$  adalah probabilitas seorang responden dengan kemampuan menjawab item  $i$  dengan benar.  $\theta$  adalah kemampuan responden.

- 2)  $a_i$  adalah parameter diskriminasi item, yang menunjukkan seberapa baik item tersebut membedakan antara responden dengan kemampuan yang berbeda.
- 3)  $b_i$  adalah parameter kesulitan item.
- 4)  $c_i$  adalah parameter tebakan, yang menggambarkan probabilitas minimal menjawab item dengan benar karena menebak.

Model Logistik Tiga Parameter (3PL) dalam Teori Respons Item (IRT) memungkinkan setiap item memiliki nilai diskriminasi dan tebakan yang berbeda. Parameter tebakan penting dalam tes pilihan ganda, di mana responden dapat menebak jawaban dengan benar tanpa memahami konten. Model ini berguna dalam situasi di mana efek menebak tidak bisa diabaikan dan item tes memiliki tingkat diskriminasi yang beragam (Gyamfi & Acquaye, 2023).

### **2.1.5 Skor Laten ( $\theta$ )**

Skor laten dalam Teori Respons Item (IRT) adalah nilai yang tidak diamati secara langsung, tetapi mencerminkan kemampuan atau atribut laten peserta tes yang ingin diukur. Skor ini diperoleh dari respons peserta terhadap butir tes, dan dapat menggambarkan konstruk seperti kemampuan matematika atau tingkat motivasi. Dalam IRT, skor laten dihitung menggunakan fungsi respons butir, yang menghubungkan kemampuan laten peserta dengan peluang mereka untuk menjawab benar suatu butir tes (Herwin et al., 2019).

## **2.2 Sistem Ujian Adaptif Berbasis Komputer (CAT)**

*Computer-based Adaptive Test* (CAT) adalah metode penilaian modern di mana tingkat kesulitan soal-soal tes disesuaikan dengan tingkat kemampuan peserta tes secara real-time. Adaptif artinya pertanyaan yang diajukan disesuaikan dengan tingkat kinerja masing-masing peserta ujian, atau tes yang disesuaikan. Berdasarkan *Item Response Theory* (IRT), CAT diatur agar komputer tidak hanya mengirimkan item soal ke komputer, tetapi juga memilih dan menyajikan soal berdasarkan

perkiraan tingkat keterampilan peserta tes. Sehingga, peserta tes individu dengan tingkat kemampuan tinggi akan menjawab pertanyaan yang lebih sulit dibandingkan peserta tes dengan tingkat kemampuan lebih rendah. Sebaliknya, peserta tes individu yang kurang terampil diberikan soal-soal yang lebih mudah dibandingkan peserta tes individu yang lebih terampil. Pendekatan dinamis ini menawarkan beberapa keuntungan, termasuk efisiensi waktu tes, mengurangi kecemasan tes karena pertanyaan yang lebih tepat, dan meningkatkan ketepatan dalam mengukur kemampuan. Sifat CAT yang adaptif memastikan bahwa tes ini tidak terlalu mudah atau terlalu sulit bagi individu, sehingga memberikan pengalaman penilaian yang personal. Selama lima tahun terakhir, penggunaan CAT telah menjadi terkenal di berbagai bidang, termasuk pendidikan, sertifikasi, dan psikotes, karena keefektifannya dalam memberikan penilaian yang handal dan valid (Yağcı, 2023).

### 2.3 Laravel

Laravel merupakan kerangka kerja PHP yang sangat terkenal dalam pengembangan aplikasi web. Dibuat oleh Taylor Otwell, Laravel dirancang dengan berbagai fitur canggih untuk memudahkan proses pembuatan aplikasi web modern. Salah satu keunggulan utama Laravel adalah sistem templating Blade-nya yang kuat, yang memungkinkan pengembang membuat tampilan web dengan mudah dan efisien. Selain itu, Laravel juga menyediakan ORM (Object-Relational Mapping) yang kuat yang disebut Eloquent, mempermudah interaksi dengan basis data. Keamanan menjadi fokus utama dalam Laravel dengan adanya berbagai lapisan perlindungan seperti proteksi terhadap serangan SQL injection dan XSS (Cross-Site Scripting). Laravel juga menyediakan mekanisme otentikasi yang siap pakai, memungkinkan pengembang mengimplementasikan sistem otentikasi dengan mudah. Dengan menggunakan pola desain MVC (Model-View-Controller), Laravel memisahkan logika aplikasi menjadi tiga komponen utama, meningkatkan skalabilitas dan pemeliharaan aplikasi (Artturi Jalli, 2022).

## 2.4 PostgreSQL

PostgreSQL adalah sistem basis data relasional objek sumber terbuka yang kuat. Dikembangkan dari proyek POSTGRES di University of California, Berkeley pada tahun 1986, PostgreSQL telah mengalami lebih dari 35 tahun pengembangan aktif. Sistem ini menggunakan dan memperluas bahasa SQL dengan banyak fitur untuk menyimpan dan menskalakan beban kerja data yang kompleks dengan aman. PostgreSQL terkenal karena arsitekturnya yang terbukti, keandalannya, integritas datanya, dan fitur-fitur tangguh. Kemampuannya dapat diperluas dan didukung oleh komunitas open source yang aktif, membuatnya menjadi solusi yang konsisten dalam kinerja dan inovatif. PostgreSQL dapat dijalankan pada semua sistem operasi utama, memenuhi standar ACID sejak tahun 2001, dan memiliki pengaya seperti PostGIS untuk basis data geospasial. Karena kualitasnya, PostgreSQL telah menjadi pilihan utama bagi banyak orang dan organisasi dalam basis data relasional open source (Hagander, 2024).

## 2.5 Penelitian yang Relevan

Berikut adalah beberapa penelitian terdahulu yang relevan dengan penelitian yang akan dilakukan oleh penulis:

- 1) Penelitian yang dilakukan oleh Muhammad Lukman Hakim, Muslim, dan Taufik Ramalan Ramalis dengan judul "Karakteristik Tes Hasil Belajar Ranah Kognitif Materi Elastisitas Menggunakan Analisis *Item Response Theory*" membahas pada evaluasi tes hasil belajar ranah kognitif pada materi elastisitas dengan menggunakan *Item Response Theory* (IRT). Penelitian yang dilakukan terhadap 116 siswa dari dua sekolah menengah atas di Bandung ini menggunakan metode deskriptif dan pengambilan sampel secara purposif. Penelitian ini menegaskan validitas tes yang tinggi dalam hal materi, konstruksi, dan bahasa, dengan menggunakan indeks Aiken V. Analisis reliabilitas melalui IRT menunjukkan keefektifan tes untuk siswa dengan tingkat kemampuan sedang hingga tinggi. Secara keseluruhan, tes ini dianggap baik berdasarkan kriteria seperti daya pembeda, tingkat kesulitan, dan faktor tebakan semu, yang memberikan wawasan yang berguna untuk menilai pendidikan fisika (Hakim & Ramalis, 2019).

- 2) Penelitian "Analisis Butir Soal dengan Pendekatan *Item Response Theory* (IRT) pada Penilaian Akhir Mata Pelajaran Al-Quran Hadits" yang dilakukan oleh Ahmad Syafii dan kawan-kawan dari Universitas Negeri Yogyakarta berfokus pada penerapan Teori Respon Butir untuk menilai butir-butir soal pada mata pelajaran Pendidikan Agama Islam. Penelitian ini bertujuan untuk mengevaluasi kelayakan butir soal ujian mata pelajaran Al-Quran Hadits secara komprehensif dan canggih. Penelitian ini melibatkan analisis terhadap 25 soal pilihan ganda yang dijawab oleh 265 siswa. Dengan menggunakan IRT, para peneliti mengevaluasi tingkat kesulitan butir soal, daya pembeda, dan faktor tebakan semu. Temuan menunjukkan distribusi tingkat kesulitan soal yang seimbang dan daya pembeda yang baik di antara soal-soal tes. Penelitian ini berkontribusi pada peningkatan kualitas tes dalam Pendidikan Agama Islam, memastikan penilaian yang lebih akurat dan dapat diandalkan terhadap hasil belajar siswa (Sudaryono, 2011).
- 3) Penelitian dengan judul "ALICAT: Sebuah Pendekatan Khusus untuk Proses Pemilihan Item dalam *Computerized Adaptive Testing*" berfokus pada peningkatan *Computerized Adaptive Testing* (CAT) dengan menganalisis kinerja *Item Selection Rules* (ISR) yang berbeda. Penelitian ini bertujuan untuk mengurangi ukuran tes dikotomis (yang hanya mempertimbangkan jawaban benar atau salah) dalam CAT tanpa kehilangan akurasi yang signifikan dalam memperkirakan kemampuan peserta ujian. Penelitian ini memperkenalkan ALICAT (*personalized CAT*), yang memilih ISR secara dinamis berdasarkan kinerja pengguna dan tahap tes. Hasil penelitian menunjukkan bahwa ALICAT dapat secara signifikan mengurangi ukuran tes (hingga 53,3%) dengan tetap mempertahankan akurasi, yang merupakan kemajuan yang berharga dalam konstruksi CAT (G. Jatobá et al., 2020).
- 4) Penelitian "*The Science and Practice of Item Response Theory in Organizations*" oleh Jonas W.B. Lang dan Louis Tay menawarkan pandangan yang komprehensif mengenai penerapan Teori Respons Butir (IRT) dalam pengaturan organisasi. Makalah ini membahas berbagai model IRT dan aplikasinya di berbagai bidang seperti pengujian, respon kuesioner, validasi konstruk, dan kesetaraan pengukuran. Makalah ini juga mengeksplorasi

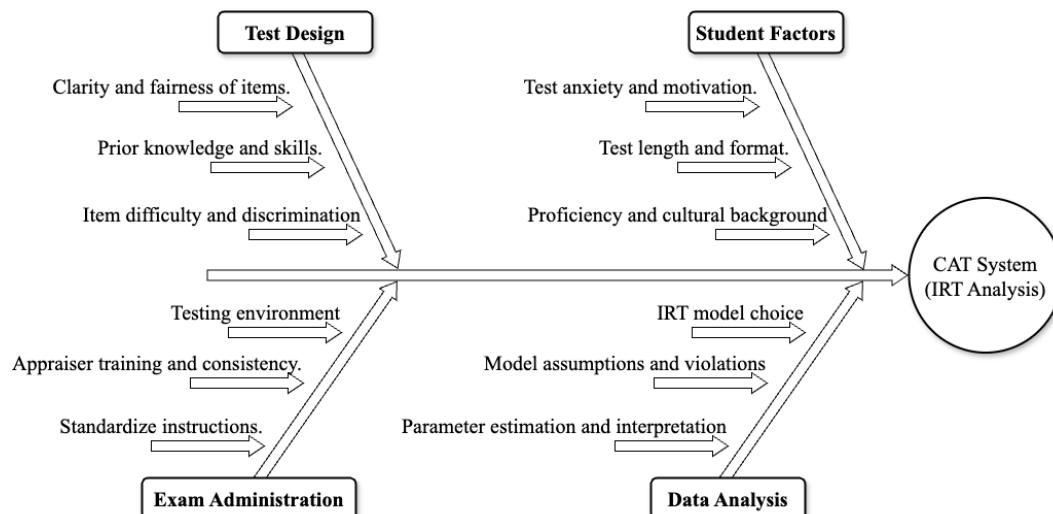
kemajuan terbaru dalam IRT, seperti IRT eksplanatori dan multidimensi, serta potensinya untuk memajukan ilmu pengetahuan dan praktik organisasi. Penelitian ini bertujuan untuk memberikan pemahaman yang rinci mengenai IRT dan manfaatnya bagi penelitian organisasi (Lang & Tay, 2021).

## BAB 3

### ANALISIS DAN PERANCANGAN SISTEM

#### 3.1 Analisis Masalah

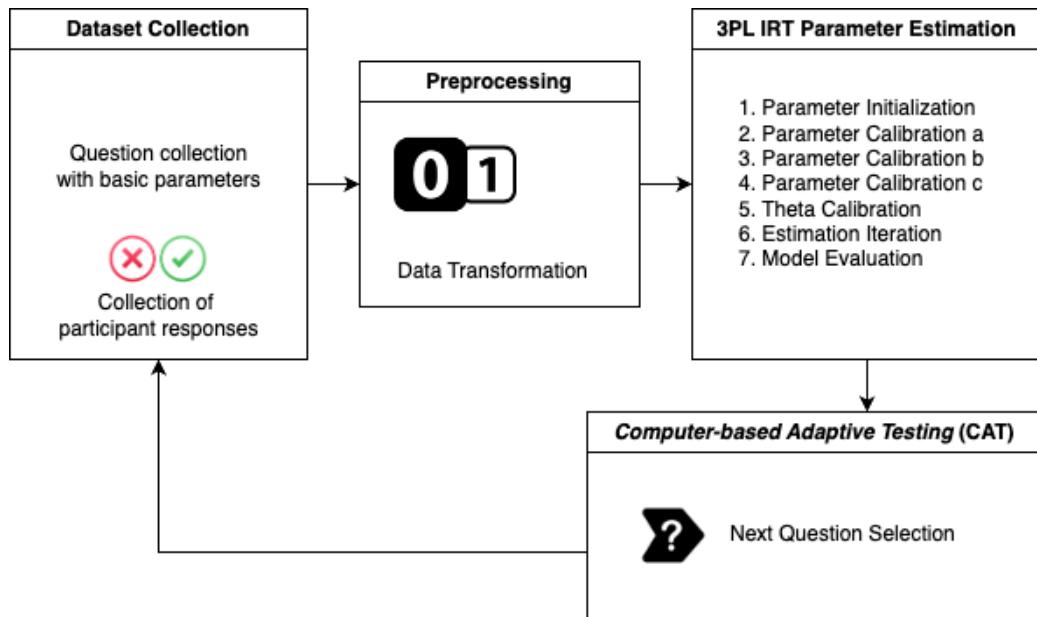
Dalam menganalisis masalah penelitian tersebut, maka dilakukan perancangan menggunakan diagram Ishikawa yang bertujuan sebagai alat bantu dalam identifikasi dan menguraikan penyebab dari permasalahan yang ada. Diagram di bawah ini mengidentifikasi faktor-faktor utama yang mempengaruhi efektivitas sistem *Computer-Based Adaptive Testing* (CAT) dengan analisis *Item Response Theory* (IRT). Faktor-faktor tersebut mencakup desain ujian, faktor siswa, administrasi ujian, dan analisis data. Dalam desain ujian, pemilihan soal, jumlah soal, format ujian, waktu ujian, dan petunjuk ujian sangat menentukan kualitas tes. Faktor siswa meliputi kemampuan awal, tingkat kecemasan, motivasi, kesiapan teknologi, dan latar belakang pendidikan yang mempengaruhi kinerja ujian mereka. Analisis data mencakup pengolahan data, validasi soal, estimasi parameter, interpretasi hasil, dan penyusunan laporan hasil. Semua faktor ini bersatu dalam sistem CAT yang adaptif, memungkinkan penilaian yang lebih akurat dan efisien berdasarkan kemampuan siswa yang diukur secara real-time.



**Gambar 3.1** Diagram Ishikawa

### 3.2 Analisis Sistem

Analisis sistem dari penelitian yang dilakukan dapat dilihat melalui arsitektur umum berikut.



**Gambar 3.2:** Arsitektur Umum

### 3.3 Perancangan Sistem

Perancangan sistem CAT dilakukan berdasarkan analisis sistem yang diilustrasikan dalam arsitektur umum, arsitektur ini menggambarkan langkah-langkah yang terlibat dalam membangun dan mengevaluasi sistem CAT dengan analisis IRT yang adaptif dan efisien. Berikut adalah tahapan dan proses yang diimplementasikan dalam perancangan sistem identifikasi kemampuan peserta dengan model 3PL IRT.

#### 3.3.1 Pengumpulan Dataset

Pengumpulan data merupakan tahap awal yang krusial dalam membangun sistem CAT. Tahap ini melibatkan dua proses utama yakni pengumpulan soal ujian dan pengumpulan respon peserta.

### 3.3.1.1 Pengumpulan Soal Ujian

Pengumpulan soal ujian berfokus pada identifikasi dan pengumpulan soal-soal yang relevan dengan domain pengujian. Setiap soal harus dilengkapi dengan parameter dasar seperti tingkat kesulitan, daya pembeda, dan tebakan. Soal-soal yang dikumpulkan dapat berasal dari berbagai sumber, seperti buku teks, bank soal, atau dibuat sendiri oleh pembuat ujian. Tingkat kesulitan soal menunjukkan seberapa sulit suatu soal, sementara daya pembeda soal menunjukkan seberapa baik suatu soal mampu membedakan antara peserta yang memiliki kemampuan tinggi dan peserta yang memiliki kemampuan rendah.

**Tabel 3.1:** Contoh Soal dengan Parameter Dasar

Nomor	Soal	Parameter (a) diskriminasi	Parameter (b) tingkat kesulitan	Parameter (c) tebakan
1	Soal 1	1.00	0.50	0.20
2	Soal 2	1.00	1.00	0.20
Dst..				

Sebagai contoh, dalam sebuah tes fisika, sebuah soal tentang hukum newton mungkin dikategorikan sebagai soal dengan tingkat kesulitan sedang dan daya pembeda sedang, sedangkan sebuah soal tentang gaya listrik mungkin dikategorikan sebagai soal dengan tingkat kesulitan tinggi dan daya pembeda tinggi.

### 3.3.1.2 Pengumpulan Respon Peserta

**Tabel 3.2:** Contoh Respon Hasil Ujian Peserta

Nomor	Peserta	Soal 1	Soal 2	Soal 3	Soal 4	Soal 5	Dst..
1	Peserta 1	Benar	Benar	Benar	Benar	Salah	
2	Peserta 2	Salah	Benar	Benar	Benar	Salah	
Dst..							

Data respon peserta dikumpulkan melalui pelaksanaan tes CAT. Peserta akan menjawab soal-soal yang disajikan oleh sistem CAT, dan

sistem akan merekam jawaban mereka. Data respon peserta kemudian akan digunakan untuk memperkirakan kemampuan peserta dan untuk memperbarui estimasi parameter model 3PL IRT. Misalnya, dalam sebuah tes CAT, seorang peserta menjawab 10 soal matematika. Dari 10 soal tersebut, 7 soal dijawab benar dan 3 soal dijawab salah. Data respon ini akan direkam oleh sistem CAT dan digunakan untuk memperkirakan kemampuan peserta dalam matematika. Pengumpulan data yang akurat dan lengkap merupakan kunci keberhasilan sistem CAT. Data soal yang berkualitas tinggi dan data respon peserta yang lengkap akan menghasilkan estimasi parameter yang akurat dan penilaian kemampuan peserta yang lebih reliabel.

### 3.3.2 *Preprocessing*

*Preprocessing* data bertujuan untuk menyiapkan data agar siap dianalisis dengan model 3PL IRT. Langkah kunci dalam praproses data adalah transformasi data. Transformasi data melibatkan perubahan format data yang dikumpulkan agar sesuai dengan persyaratan model 3PL IRT. Transformasi data mengubah data kualitatif menjadi data kuantitatif, seperti mengubah jawaban "Benar" dan "Salah" menjadi kode numerik 1 dan 0.

### 3.3.3 Estimasi Parameter 3PL IRT

Tahap estimasi parameter merupakan jantung dari sistem CAT berbasis 3PL IRT. Tahap ini melibatkan proses menghitung dan memperkirakan nilai parameter model 3PL IRT yang menggambarkan karakteristik soal dan kemampuan peserta.

#### 3.3.3.1 Inisialisasi Parameter

Tahap inisialisasi parameter merupakan langkah awal dalam proses estimasi parameter model 3PL IRT. Pada tahap ini, nilai awal untuk parameter model, yaitu parameter pembedaan (a), parameter kesulitan (b), dan parameter tebakan (c) untuk setiap soal, akan dipersiapkan dari database awal soal.

### 3.3.3.2 Kalibrasi Parameter

Setelah parameter model 3PL IRT diinisialisasi, tahap selanjutnya adalah kalibrasi parameter. Kalibrasi parameter merupakan proses mencari nilai parameter yang paling optimal, yang sesuai dengan data yang dikumpulkan dari soal dan respon peserta. Proses kalibrasi biasanya melibatkan algoritma estimasi parameter, seperti metode Maximum Likelihood Estimation (MLE).

Proses kalibrasi melibatkan perhitungan dan penyesuaian nilai parameter model 3PL IRT secara berulang, berdasarkan data yang dikumpulkan. Setiap iterasi dalam algoritma estimasi parameter menghasilkan penyesuaian nilai parameter, mendekati nilai optimal yang sesuai dengan data. Algoritma ini bekerja dengan meminimalkan perbedaan antara data yang diamati (respon peserta) dan data yang diprediksi oleh model 3PL IRT dengan parameter dasar yang diberikan.

Misalkan kita memiliki soal tentang gaya newton, dan parameter kesulitannya diinisialisasi dengan nilai 0,5. Setelah data respon peserta dikumpulkan, algoritma estimasi parameter mungkin menemukan bahwa banyak peserta menjawab soal tersebut dengan benar, menunjukkan bahwa soal tersebut mungkin lebih mudah dari yang diperkirakan. Maka, parameter kesulitan akan dikalibrasi ke nilai yang lebih rendah, seperti 0,3, untuk lebih mencerminkan tingkat kesulitan soal yang sebenarnya.

**Tabel 3.3:** Contoh Hasil Kalibrasi Parameter

No	Peserta Benar	Soal	(a) awal	(b) awal	(c) awal	(a) kalibrasi	(b) kalibrasi	(c) kalibrasi
1	9	Soal 1	1.00	0.50	0.20	0.70	0.20	0.20
2	4	Soal 2	1.00	1.00	0.20	1.70	2.00	0.20
Dst..		Soal 10						

### 3.3.3.3 Kalibrasi Theta

Selain parameter soal, parameter theta ( $\theta$ ) yang mewakili kemampuan peserta juga perlu dikalibrasi. Proses kalibrasi theta melibatkan penggunaan data respon peserta terhadap soal-soal yang telah dijawab untuk

memperkirakan kemampuan peserta ( $\theta$ ). Proses ini biasanya dilakukan secara bersamaan dengan kalibrasi parameter soal, menggunakan algoritma estimasi parameter yang sama.

Misalkan dalam sebuah tes CAT dengan model 3PL IRT, seorang peserta menjawab 12 soal dengan benar dari 20 soal yang diberikan. Data respon ini akan digunakan untuk memperkirakan kemampuan peserta, yang direpresentasikan oleh parameter theta ( $\theta$ ). Umumnya semua kemampuan siswa awal adalah 1, dengan anggapan semua siswa mampu menjawab semua soal. Nilai cukup standar mengingat nilai theta biasanya berkisar antara -3 hingga 3, dengan nilai positif menunjukkan kemampuan yang lebih tinggi.

**Tabel 3.4:** Contoh Ledger Jawaban Siswa

No	Peserta	Soal	Status	Theta Awal	Theta Akhir
1	Peserta 1	Soal 1	Benar	1.00	1.20
2	Peserta 2	Soal 1	Salah	1.00	0.90
Dst..					

Namun, estimasi theta ini bukanlah nilai akhir. Proses kalibrasi theta dilakukan secara bersamaan dengan kalibrasi parameter soal. Algoritma estimasi akan terus memperbarui nilai theta berdasarkan data respon peserta terhadap soal-soal yang telah dijawab. Jika peserta tersebut kemudian menjawab soal-soal berikutnya dengan benar, nilai theta mereka mungkin akan meningkat, menandakan peningkatan kemampuan. Sebaliknya, jika peserta tersebut menjawab soal-soal berikutnya dengan salah, nilai theta mereka mungkin akan menurun, menandakan penurunan kemampuan. Proses iteratif ini akan terus dilakukan hingga model mencapai konvergensi, di mana nilai theta tidak lagi berubah secara signifikan setelah beberapa iterasi. Hasil akhir dari kalibrasi theta adalah estimasi kemampuan peserta yang lebih akurat, berdasarkan data respon yang telah dikumpulkan.

### 3.3.3.4 Iterasi Estimasi

Iterasi Estimasi merupakan proses mencari parameter model IRT 3 parameter yang optimal secara iteratif. Proses ini dimulai dengan menginisialisasi parameter awal. Selanjutnya, parameter soal ( $a$ ,  $b$ , dan  $c$ ) dan kemampuan peserta ( $\theta$ ) dihitung dengan metode *Maximum Likelihood Estimation* (MLE). Langkah ini diulang beberapa kali (iterasi), dengan hasil iterasi sebelumnya digunakan sebagai parameter awal untuk iterasi berikutnya. Melalui pengulangan ini, parameter model IRT 3 parameter secara bertahap disesuaikan sehingga mencapai estimasi yang lebih akurat.

### 3.3.3.4 Evaluasi Model

Setelah model IRT 3 parameter selesai diestimasi, langkah selanjutnya adalah model evaluation, yaitu memeriksa kualitas model yang telah dihasilkan. Proses ini bisa dilakukan dengan menggunakan uji fit dengan chi-square untuk melihat apakah ada perbedaan yang signifikan antara data yang diamati dan data yang diprediksi oleh model. Hasil dari model evaluation memberikan informasi tentang kualitas model IRT 3 parameter yang dihasilkan. Informasi ini dapat digunakan untuk mengevaluasi apakah model tersebut dapat diandalkan untuk mengukur kemampuan peserta, dan untuk mengidentifikasi potensi masalah yang mungkin perlu ditangani, seperti soal yang tidak berfungsi dengan baik atau estimasi parameter yang tidak akurat.

## 3.3.4 Computer-based Adaptive Test(CAT)

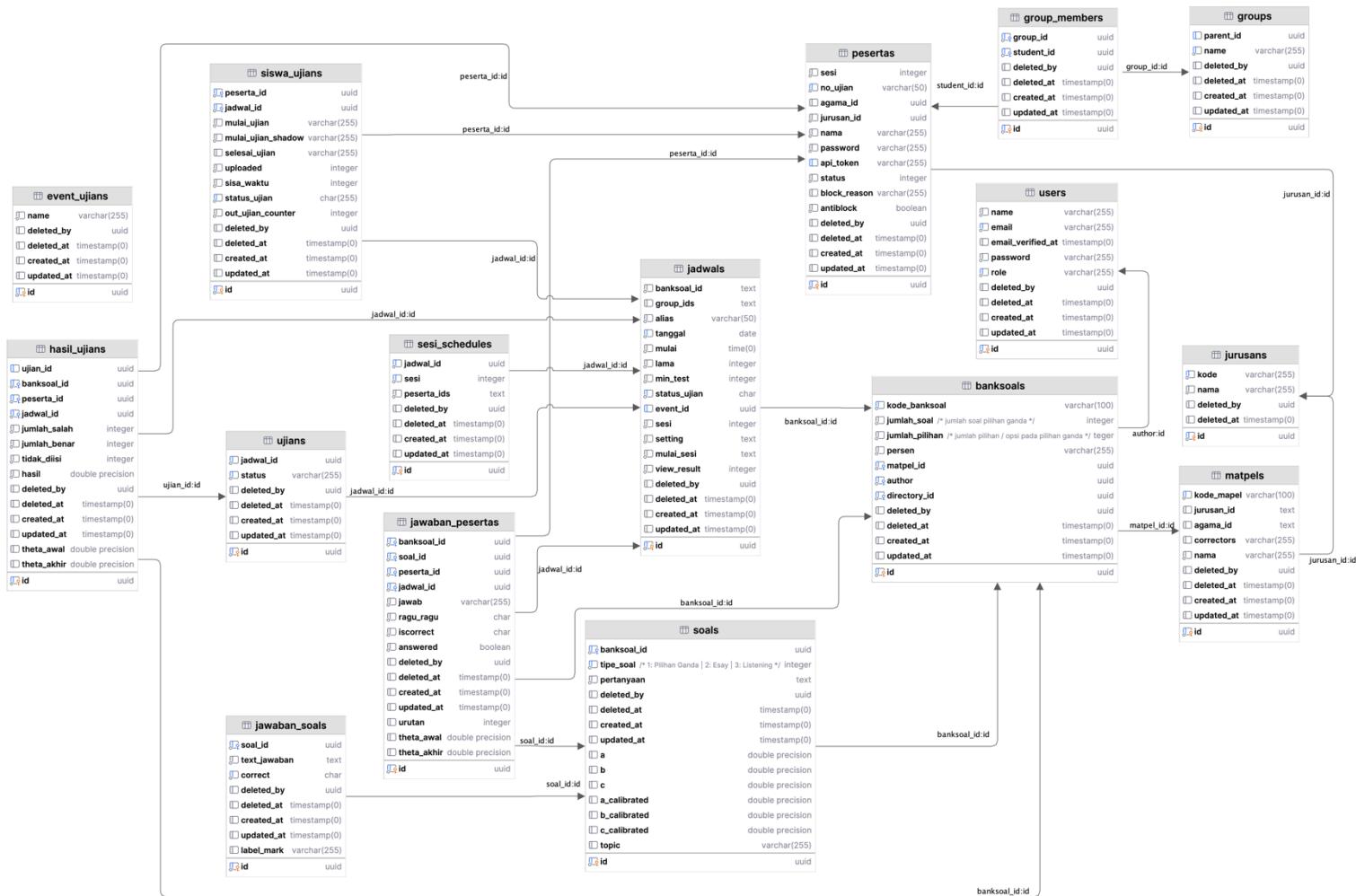
Setelah model IRT 3 parameter berhasil diestimasi, model tersebut dapat digunakan untuk mengembangkan tes adaptif, yang dikenal sebagai *Computer-based Adaptive Testing* (CAT). CAT menggunakan model IRT untuk memilih soal berikutnya yang akan diberikan kepada peserta berdasarkan kemampuannya yang diestimasi. Tujuannya adalah untuk memberikan soal yang sesuai dengan tingkat kesulitan peserta, sehingga tes dapat lebih efektif dalam mengukur kemampuan peserta. Jika kemampuan

peserta diestimasi tinggi, CAT akan memberikan soal yang lebih sulit. Sebaliknya, jika kemampuan peserta diestimasi rendah, CAT akan memberikan soal yang lebih mudah. Dengan cara ini, CAT dapat menyesuaikan kesulitan soal dengan kemampuan peserta secara *real-time*, sehingga tes menjadi lebih efisien dan akurat.

### 3.2 Perancangan *Database*

Pada tahap ini dijelaskan rancangan skema basis data untuk sistem *Computer-Based Adaptive Testing* (CAT). Skema ini terdiri dari beberapa tabel utama, termasuk tabel siswa\_ujian, event\_ujians, hasil\_ujians, ujian, sesi\_schedules, jawaban\_pesertas, jawaban\_soals, banksoals, peserta, groups, users, jurusans, dan matpels.

Hubungan antara tabel-tabel ini menggunakan kunci asing (foreign keys) untuk memastikan integritas dan konsistensi data. Misalnya, peserta\_id dalam tabel siswa\_ujian mengacu pada id dalam tabel peserta, dan jadwal\_id menghubungkan tabel ujian, sesi\_schedules, dan siswa\_ujian. Desain ini memungkinkan integrasi yang kuat dan penelusuran data yang efisien, memastikan bahwa semua aspek ujian dapat diakses dan dianalisis dengan mudah.

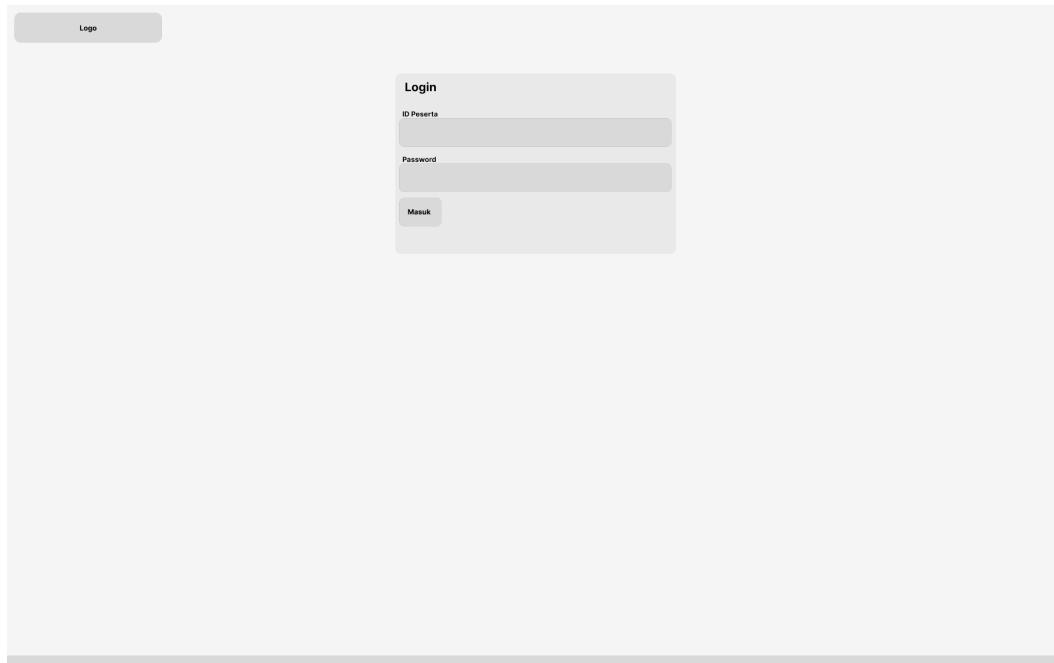


Gambar 3.3 Rancangan Database

### 3.3 Perancangan *Interface*

Perancangan interface atau antar muka merupakan salah satu bagian yang terpenting dalam membangun suatu sistem. Perancangan interface berfungsi untuk menghasilkan gambaran yang digunakan sebagai penghubung antara pengguna dengan sistem sehingga aplikasi dapat digunakan sesuai dengan fungsinya. Rancangan interface untuk sistem yang akan dibangun terdapat empat menu yaitu menu Home, Ujian, Bank Soal, Jadwal, dan Analisis.

#### 3.3.1 Peserta Ujian - Tampilan Perancangan *Home*

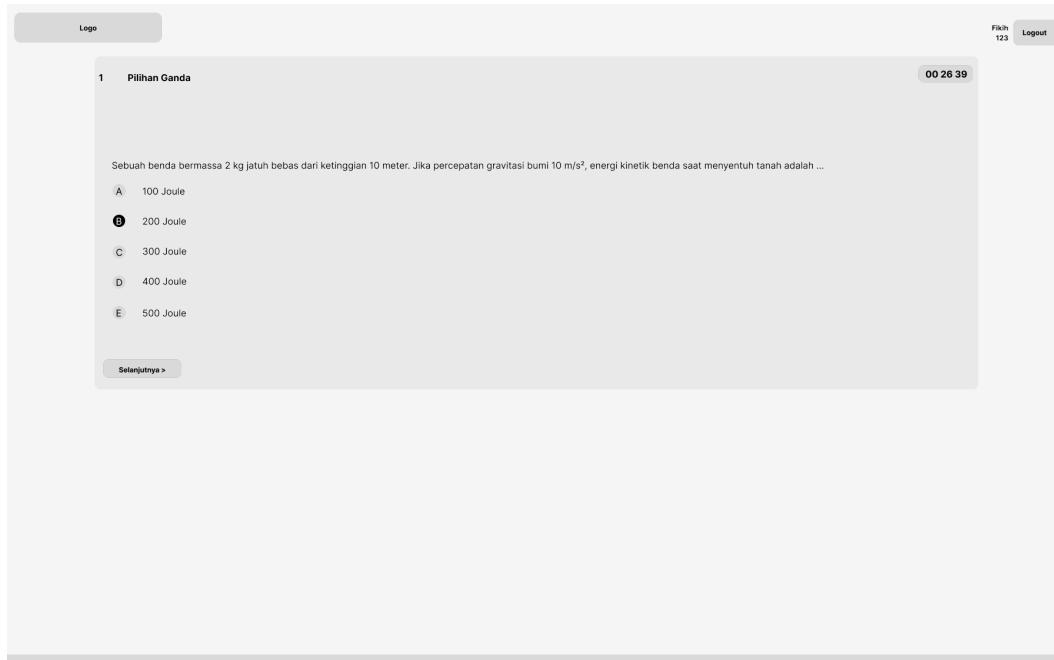


**Gambar 3.4:** Peserta Ujian - Tampilan Perancangan Home

Pada halaman utama aplikasi, terdapat header yang mencakup logo aplikasi, nama pengguna jika sudah login, dan tombol Logout untuk keluar dari akun. Di bawah header, informasi ujian yang sedang berlangsung akan ditampilkan, mencakup nama ujian dan waktu yang tersisa untuk menyelesaiannya. Jika tidak ada ujian yang sedang berlangsung, bagian ini akan kosong. Selanjutnya, terdapat bagian yang menampilkan daftar ujian yang akan datang, termasuk tanggal dan waktu pelaksanaannya, sehingga pengguna dapat mempersiapkan diri dengan baik. Di bagian bawah

halaman, terdapat tombol untuk melihat riwayat ujian, yang memungkinkan pengguna untuk meninjau hasil ujian sebelumnya dan mengevaluasi kinerja mereka. Desain halaman ini bertujuan untuk memberikan informasi yang jelas dan akses cepat ke fitur-fitur penting dalam aplikasi.

### 3.3.2 Peserta Ujian - Tampilan Perancangan Ujian



**Gambar 3.5:** Peserta Ujian - Tampilan Perancangan Ujian

Pada halaman ujian, header menampilkan logo aplikasi, nama ujian, dan waktu tersisa yang ditampilkan dengan countdown timer untuk memberikan informasi waktu yang terus berjalan. Di bawah header, konten utama terdiri dari soal yang sedang aktif, ditampilkan dengan jelas beserta pilihan jawabannya. Pengguna dapat menavigasi antar soal menggunakan tombol "Previous" dan "Next" yang terletak di bawah soal, memungkinkan mereka untuk bergerak maju atau mundur di antara soal-soal ujian.

Selain itu, terdapat tombol "Submit" yang memungkinkan pengguna untuk menyelesaikan dan mengirimkan ujian setelah menjawab semua soal. Indikator nomor soal juga ditampilkan, seperti "1/10," untuk menunjukkan posisi soal yang sedang dikerjakan dari total jumlah soal yang ada dalam

ujian. Desain ini memastikan pengguna dapat mengikuti ujian dengan mudah dan terorganisir, serta memantau waktu yang tersisa secara efektif.

### 3.3.3 Admin - Halaman Daftar Ujian

Ujian	Detail	Nama Ujian	Tanggal	Waktu Mulai	Durasi	Status	Action
Tambah Ujian	<a href="#">Detail</a>	Ujianan Harian 9	2024-05-22	13:00:00	30 Menit	Aktif	<a href="#">Edit</a> <a href="#">Delete</a>

**Gambar 3.6:** Admin - Halaman Daftar Ujian

Pada halaman pengelolaan ujian, sidebar menampilkan menu navigasi yang menyediakan akses cepat ke halaman lain, seperti Bank Soal, Analisis, dan lainnya. Sidebar ini memungkinkan pengguna untuk berpindah antar fitur aplikasi dengan mudah.

Konten utama halaman ini terdiri dari tabel yang menampilkan daftar ujian, dengan kolom yang mencakup nama ujian, tanggal pelaksanaan, status (seperti aktif atau selesai), dan tindakan (edit atau hapus). Tabel ini memberikan gambaran yang jelas dan terstruktur mengenai ujian-ujian yang ada, sehingga memudahkan pengelolaan dan pengawasan.

Di atas tabel, terdapat tombol "Buat Ujian Baru" yang memungkinkan pengguna untuk membuat ujian baru dengan cepat. Selain itu, fitur filter dan pencarian disediakan untuk memudahkan pengguna dalam menemukan ujian tertentu berdasarkan berbagai kriteria, seperti

nama atau tanggal pelaksanaan. Dengan demikian, halaman ini dirancang untuk meningkatkan efisiensi dan kemudahan dalam mengelola berbagai ujian dalam aplikasi.

### 3.3.4 Admin - Halaman Bank Soal

Ujian	Detail	Nama Ujian	Tanggal	Waktu Mulai	Durasi	Status	Action
Tambah Ujian	<a href="#">i</a>	Ulangan Harian 9	2024-05-22	13:00:00	30 Menit	Aktif	<a href="#">Edit</a> <a href="#">Delete</a>

**Gambar 3.7:** Admin - Halaman Bank Soal

Pada halaman pengelolaan soal, sidebar tetap menampilkan menu navigasi untuk akses cepat ke halaman lain seperti Bank Soal, Analisis, dan lainnya, memastikan navigasi yang mudah di dalam aplikasi.

Konten utama halaman ini terdiri dari tabel yang menampilkan daftar soal. Tabel ini memiliki kolom-kolom seperti teks soal, pilihan jawaban, parameter IRT (a, b, c), topik, dan tindakan (edit, hapus). Dengan demikian, pengguna dapat melihat dan mengelola detail setiap soal secara jelas dan terstruktur.

Di atas tabel, terdapat tombol "Tambah Soal Baru" yang memungkinkan pengguna untuk menambahkan soal baru ke bank soal dengan mudah. Fitur filter dan pencarian juga disediakan untuk membantu pengguna menemukan soal dengan cepat berdasarkan kriteria tertentu, seperti teks soal atau topik.

## **BAB 4**

### **IMPLEMENTASI DAN PENGUJIAN SISTEM**

#### **4.1 Implementasi Sistem**

Penelitian ini akan mencakup berbagai proses implementasi sistem Computer-Based Adaptive Testing (CAT) dengan menggunakan metode Item Response Theory (IRT) Three-Parameter Logistics (3PL). Tahapan implementasi meliputi pengumpulan dan analisis data responden, estimasi parameter soal seperti diskriminasi (a), kesulitan (b), dan tebakan (c), serta perbandingan skor klasik dengan penyetaraan IRT. Selain itu, penelitian ini juga akan mencakup analisis Item Characteristic Curve (ICC), evaluasi karakteristik butir soal, dan pengujian kesesuaian model.

##### **4.1.1 Sumber Data**

Data yang digunakan dalam penelitian ini berasal dari hasil simulasi ulangan harian berbasis komputer yang dilaksanakan oleh Madrasah Aliyah Negeri (MAN) 1 Medan pada mata pelajaran Fisika di kelas XI IPS 3 pada tanggal 28 Mei 2024. Pengumpulan data dilakukan secara online melalui platform TestMate, yang memungkinkan siswa untuk mengikuti tes dari rumah masing-masing (diluar jam pelajaran).

Prosedur pengumpulan data meliputi:

- 1) Alat Pengumpulan Data: Platform TestMate digunakan untuk administrasi ujian dan pengumpulan jawaban siswa.
- 2) Jumlah total data yang terkumpul adalah sebagai berikut:
  - Jumlah Responden : 20 siswa
  - Jumlah Butir Soal : 20 butir soal mata pelajaran Fisika

##### **4.1.2 Karakteristik Responden**

Responden dalam penelitian ini adalah siswa kelas XI IPS 3 MAN 1 Medan yang telah mengikuti simulasi ulangan harian berbasis komputer. Karakteristik responden adalah sebagai berikut.

- 1) Usia: Rentang usia responden antara 17 hingga 18 tahun.

- 2) Jenis Kelamin: 7 responden laki-laki (35%) dan 13 responden perempuan (65%).

#### **4.1.3 Distribusi Data**

Distribusi data yang dikumpulkan dalam penelitian ini adalah sebagai berikut:

**Tabel 4.1:** Distribusi Data Jawaban

User	Soa 11	Soa 12	Soa 13	Soa 14	Soa 15	Soa 16	Soa 17	Soa 18	Soa 19	Soal 10	Soal 11	Soal 12	Soal 13	Soal 14	Soal 15	Soal 16	Soal 17	Soal 18	Soal 19	Soal 20	Skor (* 5)
M. HANDY AULIA	1	0	1	1	0	1	1	1	0	1	0	1	1	0	1	1	0	1	1	0	65
VIA SURIYANTO	0	1	0	1	1	0	1	0	1	0	1	1	0	1	0	1	1	0	1	0	55
SHARLIZ AYLA SAMAIRO NST	1	1	1	1	0	0	1	1	1	1	0	0	1	1	1	1	1	0	0	1	70
ANISYAH TIARA PUTRI	0	0	0	1	1	1	1	0	0	1	1	1	1	1	0	1	0	0	1	1	60
NAZWA RAHMADANTIE RAMBE	1	1	0	0	1	1	0	0	1	1	1	0	0	1	1	1	1	1	0	1	65
DWI UTAMI RIZQINA	1	0	1	1	1	0	1	0	1	0	1	1	0	1	0	1	0	1	1	1	65
MAJDALUNA AINUN SYIFA	0	1	1	0	1	1	0	1	0	1	0	1	1	0	1	1	1	0	0	1	60
RESTU BINTANG PANGESTU	1	1	0	1	1	1	0	0	1	0	1	1	0	1	1	1	0	0	1	1	65
BUNGA SAHARA SIMBOLON	0	0	1	0	1	0	1	1	0	1	1	0	1	1	1	0	1	1	0	1	60
SAQILA MEINDI	1	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1	0	1	0	65
YUSUF KORY MARPAUNG	1	0	1	0	1	1	1	1	0	1	1	0	1	1	1	1	0	0	0	1	65
ATIQAH FARRAS ASRI	0	1	1	0	1	0	1	1	1	0	1	1	0	1	1	0	1	1	1	1	70
AZMI AZARIA	1	1	0	1	1	1	0	1	0	1	0	1	1	1	0	1	0	0	1	1	65
JIHAN MAHIRAH NASUTION	0	0	1	1	0	1	1	0	1	1	0	1	1	1	0	0	1	1	1	1	65
RAFA AZZIKRI BUTAR-BUTAR	1	0	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	0	70
RAHSYA HASBY SIREGAR	0	1	0	1	0	1	0	1	1	1	1	0	1	0	1	1	0	1	0	1	60
SYALFA KHAIRINA HAWANI HARAHAP	1	1	1	0	1	1	1	0	1	1	0	1	0	1	0	1	1	0	1	0	65
MAULANA IBRAHIM	0	1	0	1	0	1	1	0	1	0	1	1	1	0	1	0	1	1	1	1	65
FADHIL AHMAD GALY	1	0	1	0	1	1	1	1	0	1	1	0	1	0	1	1	0	0	1	1	65
NISRINA ANGGITA	0	1	1	0	1	0	1	0	1	1	1	0	1	1	1	1	0	1	1	1	70
<b>Total Penjawab Benar</b>	<b>11</b>	<b>12</b>	<b>12</b>	<b>11</b>	<b>15</b>	<b>14</b>	<b>14</b>	<b>10</b>	<b>12</b>	<b>15</b>	<b>13</b>	<b>13</b>	<b>14</b>	<b>14</b>	<b>13</b>	<b>16</b>	<b>11</b>	<b>10</b>	<b>13</b>	<b>15</b>	

## 1. Distribusi Skor:

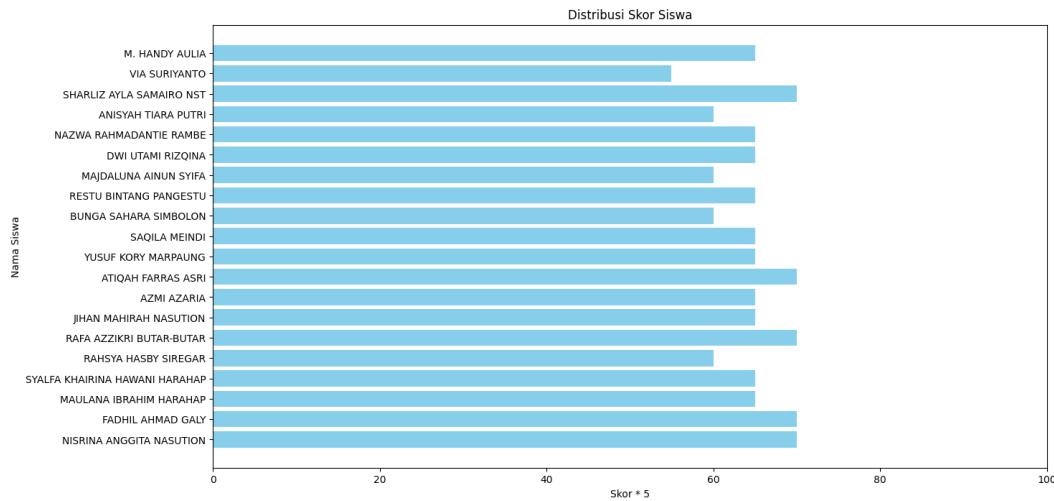
Skor dikalikan dengan 5 agar mencapai angka tertinggi 100 yang menjadi penilaian konvensional bagi para guru. Rata-rata skor ujian siswa 62.5 dengan standar deviasi 10. Median skor adalah 65, dan modus skor adalah 65. Distribusi frekuensi skor dapat dilihat pada Gambar 4.1.

**Tabel 4.2:** Distribusi Skor

Soal Nomor	a	b	c
1	1.0	0.5	0.2
2	1.0	1.0	0.2
3	1.0	0.5	0.2
4	1.0	1.5	0.2
5	1.0	1.5	0.2
6	1.0	1.5	0.2
7	1.0	0.5	0.2
8	1.0	1.0	0.2
9	1.0	0.5	0.2
10	1.0	0.5	0.2
11	1.0	0.5	0.2
12	1.0	0.5	0.2
13	1.0	0.5	0.2
14	1.0	1.0	0.2
15	1.0	1.0	0.2
16	1.0	1.0	0.2
17	1.0	1.0	0.2
18	1.0	1.5	0.2
19	1.0	1.5	0.2
20	1.0	0.5	0.2

## 2. Distribusi Butir Soal

Analisis tingkat kesulitan butir soal menunjukkan bahwa 45% soal tergolong mudah, 30% soal tergolong sedang, dan 25% soal tergolong sulit. Distribusi tingkat kesulitan soal dapat dilihat pada Gambar 4.2.



**Gambar 4.1:** Visualisasi Data Skor Siswa

### 3. Visualisasi Data

Untuk mempermudah interpretasi, distribusi data juga divisualisasikan dalam bentuk bar chart yang menggambarkan sebaran skor dan outliers.

#### 4.1.4 Pengujian Karakteristik Butir Soal

Bagian ini menjelaskan tentang pengujian karakteristik butir soal dengan menggunakan metode *Item Response Theory* (IRT) khususnya model *Three-Parameter Logistic* (3PL).

##### 4.1.4.1 Estimasi Parameter Soal

Estimasi parameter butir soal dilakukan untuk menentukan tiga parameter utama dalam model 3PL, yaitu diskriminasi (a), kesulitan (b), dan tebakan (c).

###### 4.1.4.1.1 Parameter Diskriminasi (a)

Rumus parameter diskriminasi (a) pada model Rasch yang sederhana adalah sebagai berikut:

$$P(X_i = 1 | \theta) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)}$$

di mana:

- 1)  $(P(X = 1|\theta = 1))$  adalah probabilitas jawaban benar untuk siswa dengan kemampuan tinggi ( $\theta = 1$ )
- 2)  $(P(X = 1|\theta = 0))$  adalah probabilitas jawaban benar untuk siswa dengan kemampuan rendah ( $\theta = 0$ )

Nilai parameter diskriminasi (a) awal adalah 1 yang dianggap selalu mampu mendiskriminasikan peserta ujian. Hasil dari perhitungan di atas adalah 0 yang menunjukkan bahwa butir soal tersebut tidak memiliki kemampuan diskriminasi yang baik antara siswa dengan kemampuan tinggi dan rendah. Semakin besar nilai  $a$ , semakin baik butir soal tersebut dalam membedakan antara siswa dengan kemampuan tinggi dan rendah.

Lalu kita implementasikan ke semua peserta penjawab soal 1:

- 1) Total penjawab benar adalah 11 dari 20 siswa.
- 2) Jumlah siswa dengan kemampuan tinggi ( $(\theta = 1)$ ) yang menjawab benar adalah 7 (M. Handy Aulia, Sharliz Ayla Samairo Nst, Nazwa Rahmadantie Rambe, Dwi Utami Rizqina, Restu Bintang Pangestu, Saqila Meindi, Yusuf Kory Marpaung).
- 3) Jumlah siswa dengan kemampuan rendah ( $(\theta = 0)$ ) yang menjawab benar adalah 4 (Via Suriyanto, Anisyah Tiara Putri, Bunga Sahara Simbolon, Atiqah Farras Asri, Azmi Azaria, Jihan Mahirah Nasution, Rafa Azzikri Butar-butar, Rahsya Hasby Siregar, Syalfa Khairina Hawani Harahap, Maulana Ibrahim Harahap, Fadhil Ahmad Galy, Nisrina Anggita Nasution).

Dengan menggunakan rumus yang telah disebutkan sebelumnya, kita dapat menghitung nilai parameter diskriminasi (a), misal pada soal nomor 1:

$$\begin{aligned}
 P(X = 1|\theta = 0) &= \frac{1}{1 + e^{1 \times (0 - 0.5)}} \approx \frac{1}{1 + e^{-0.5}} \approx \frac{1}{1 + 0.60653} \approx \frac{1}{1.60653} \\
 &\approx 0.6224593)(a = \ln \left( \frac{0.3775407 - 0}{1 - 0.3775407 - 0} \right) = \ln \left( \frac{0.3775407}{0.6224593} \right) \\
 &= \ln(0.6055536) \approx -0.4986228)
 \end{aligned}$$

$$(a = \ln \left( \frac{0.6224593 - 0}{1 - 0.6224593 - 0} \right) \approx \ln(0.6224593) \approx -0.4740769)$$

Misalnya, untuk soal nomor 1, nilai parameter diskriminasi (a) adalah 0.47. Ini berarti bahwa soal nomor 1 memiliki kemampuan diskriminasi sebesar 0.47, yang mengindikasikan bahwa soal tersebut cukup baik dalam membedakan antara siswa yang memiliki kemampuan tinggi dan rendah. Semakin tinggi nilai a, semakin baik soal tersebut dalam membedakan antara siswa yang memiliki kemampuan yang berbeda.

**Tabel 4.3:** Nilai (a) yang terkalibrasi

Soal	Nilai a (terkalibrasi)
1	0.47
2	1.31
3	0.47
4	0.80
5	0.80
6	0.80
7	0.47
8	1.31
9	0.47
10	0.47
11	0.47
12	0.47
13	0.47
14	1.31
15	1.31
16	1.31
17	1.31
18	0.80
19	0.80

Bisa dilihat telah terkalibrasi parameter a. Pada soal nomor 8, nilai parameter diskriminasi (a) adalah 1.31. Ini berarti bahwa soal nomor 8 memiliki kemampuan diskriminasi sebesar 1.31, yang mengindikasikan bahwa soal tersebut cukup baik dalam membedakan antara siswa yang

memiliki kemampuan tinggi dan rendah. Semakin tinggi nilai  $a$ , semakin baik soal tersebut dalam membedakan antara siswa yang memiliki kemampuan yang berbeda.

#### 4.1.4.1.2 Parameter Kesulitan (b)

Dalam konteks model Rasch, parameter kesulitan ( $b$ ) biasanya diinterpretasikan sebagai lokasi pada skala kemampuan di mana seorang peserta ujian memiliki peluang 50% untuk menjawab soal dengan benar. Rumus dasar yang sering digunakan dalam konteks ini adalah:

$$[b = \frac{\ln\left(\frac{p}{1-p}\right)}{a}]$$

Di mana:

- 1) ( $p$ ) adalah probabilitas seorang peserta menjawab soal dengan benar.
- 2) ( $a$ ) adalah parameter diskriminasi dari soal tersebut.

Mari kita ambil beberapa soal sebagai contoh untuk menghitung parameter kesulitan ( $b$ ) berdasarkan data yang diberikan sebelumnya. Misalkan kita ambil soal nomor 1, dengan nilai ( $a = 0.47$ ).

Untuk soal nomor 1, jumlah siswa yang menjawab benar adalah 11 dari 20 siswa, jadi probabilitas ( $p$ ) adalah 0.55 (atau 55%).

$$[p = \frac{11}{20} = 0.55]$$

$$[b = \frac{\ln\left(\frac{0.55}{1-0.55}\right)}{0.47} = \frac{\ln\left(\frac{0.55}{0.45}\right)}{0.47} = \frac{\ln(1.2222)}{0.47} = \frac{0.2007}{0.47} \approx 0.43]$$

Mari kita hitung parameter kesulitan ( $b$ ) untuk semua soal dengan cara yang sama dan buat summary-nya.

**Tabel 4.4:** Nilai (b) yang terkalibrasi

Soal	<i>p</i>	<i>a</i>	<i>b</i>
1	0.55	0.47	0.43
2	0.60	1.31	0.46
3	0.60	0.47	0.51
4	0.55	0.80	0.32
5	0.75	0.80	0.87
6	0.70	0.80	0.61
7	0.70	0.47	0.74
8	0.50	1.31	0.00
9	0.60	0.47	0.51
10	0.75	0.47	1.10
11	0.65	0.47	0.71
12	0.65	0.47	0.71
13	0.70	0.47	0.74
14	0.70	1.31	0.30
15	0.75	1.31	0.53
16	0.80	1.31	0.66
17	0.55	1.31	0.14
18	0.50	0.80	0.00
19	0.65	0.80	0.38
20	0.75	0.47	1.10

Soal nomor 1 memiliki probabilitas benar sebesar 0.55 atau 55%, yang menunjukkan bahwa lebih dari separuh peserta mampu menjawab soal ini dengan benar. Parameter diskriminasinya adalah 0.47, yang berarti soal ini memiliki kemampuan moderat dalam membedakan antara siswa yang memiliki kemampuan tinggi dan rendah. Nilai parameter kesulitan (b) untuk soal ini adalah 0.43, yang menunjukkan bahwa soal ini memiliki tingkat kesulitan sedang. Secara keseluruhan, soal nomor 1 tidak terlalu sulit dan cukup efektif dalam mengidentifikasi perbedaan kemampuan di antara siswa.

#### 4.1.4.1.3 Parameter Tebakan (c)

Dalam sistem ini, parameter tebakan tidak akan terkalibrasi karena hanya baku pilihan ganda.

$$tebakan = \frac{total\ pilihan\ jawaban\ yang\ benar}{total\ pilihan\ jawaban}$$

$$tebakan = \frac{1}{5}$$

$$tebakan = 0.2$$

#### 4.1.4.2 Estimasi Skor Laten (Theta)

Pada tahap ini dijelaskan proses estimasi skor laten (theta) yang merepresentasikan kemampuan siswa dalam menjawab soal menggunakan model Item Response Theory (IRT) dengan Three-Parameter Logistic (3PL). Estimasi theta dilakukan berdasarkan respon siswa terhadap soal dan parameter karakteristik soal (a, b, c).

Rumus yang digunakan untuk menghitung probabilitas (P) seorang siswa dengan kemampuan tertentu ( $\theta$ ) menjawab benar suatu soal adalah sebagai berikut:

$$P(X_i = 1 | \theta) = c_i + (1 - c_i) \frac{\exp(a_i(\theta - b_i))}{1 + \exp(a_i(\theta - b_i))}$$

Sebagai contoh, untuk soal nomor 1 dengan parameter a=0.47 b=0.43, dan c=0.2, serta respon siswa terhadap soal tersebut adalah [1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0], di mana '1' menunjukkan jawaban benar dan '0' menunjukkan jawaban salah.

Sehingga

$$P(\theta) = 0.2 + \frac{1 - 0.2}{1 + e^{-0.47(\theta - 0.43)}}$$

$$log - likelihood = 1 \cdot \log(P) + (1 - 1) \cdot \log(1 - P)$$

Lalu nilai P akan di ulang dan ditotalkan hingga menghitung semua response dengan bantuan fungsi Log-Likelihood. Dari anggapan  $\theta$  awal semua peserta adalah 0.5, Perulangan mendapatkan nilai  $0.46381465 \approx 0.46$ . Dimana fungsi Log-Likelihood telah mendapatkan batas kecocokan model, dengan  $\theta = 0.46$ . Sehingga dianggap kemampuan untuk bisa menjawab soal nomor 1 adalah 0.46 Dan seterusnya hingga semua soal dan semua respon peserta terkalkulasi.

**Tabel 4.5** Hasil perulangan theta menggunakan Log-Likelihood.

log	theta	parameter soal
[-0.49991963]	[0.5]	a: 0.47, b: 0.43, c: 0.2
[-1.44275355]	[0.5]	a: 1.31, b: 0.46, c: 0.2
[-1.95514706]	[0.5]	a: 0.47, b: 0.51, c: 0.2
[-2.41916808]	[0.5]	a: 0.8, b: 0.32, c: 0.2
[-3.19837107]	[0.5]	a: 0.8, b: 0.87, c: 0.2
[-3.73894936]	[0.5]	a: 0.8, b: 0.61, c: 0.2
[-4.04559011]	[0.5]	a: 0.47, b: 0.74, c: 0.5
[-4.36509445]	[0.5]	a: 1.31, b: 0.0, c: 0.2
[-5.27903794]	[0.5]	a: 0.47, b: 0.51, c: 0.2
...	...	...
[-0.50554209]	[0.46381465]	a: 0.47, b: 0.43, c: 0.2
[-1.42433454]	[0.46381465]	a: 1.31, b: 0.46, c: 0.2
[-1.94242189]	[0.46381465]	a: 0.47, b: 0.51, c: 0.2
[-2.41565473]	[0.46381465]	a: 0.8, b: 0.32, c: 0.2
[-3.18261258]	[0.46381465]	a: 0.8, b: 0.87, c: 0.2
[-3.73315497]	[0.46381465]	a: 0.8, b: 0.61, c: 0.2
[-4.04267802]	[0.46381465]	a: 0.47, b: 0.74, c: 0.5
[-4.37408352]	[0.46381465]	a: 1.31, b: 0.0, c: 0.2
[-5.27957959]	[0.46381465]	a: 0.47, b: 0.51, c: 0.2
...	...	...

#### 4.2.4.2 Perbandingan Skor Klasik dengan Penyetaraan IRT

**Tabel 4.6:** Perbandingan Skor Klasik dengan Penyetaraan

User	Soal 1	Soal 2	Soal 3	Soal 4	Soal 5	Soal 6	Soal 7	Soal 8	Soal 9	Soal 10	Soal 11	Soal 12	Soal 13	Soal 14	Soal 15	Soal 16	Soal 17	Soal 18	Soal 19	Soal 20	Skor Skala 100	Skor IRT Penyerataan 100	Selisih
M, HANDY AULIA	0,46	0	1,21	-0,02	0	0,17	0,67	0,74	0	0,43	0	1,14	0,76	0	1,15	0,5	0	0,47	0,31	0	65	67,91	2,91
VIA SURIYANTO	0	0,3	0	-0,02	0,96	0	0,67	0	0,4	0	0,49	1,14	0	0,23	0	0,5	0,61	0	0,31	0	55	47,51	-7,48
SHARLIZ AYLA SAMAIRO NST	0,46	0,3	1,21	-0,02	0	0	0,67	0,74	0,4	0,43	0	0	0,76	0,23	1,15	0,5	0,61	0	0	0,75	70	69,61	-0,38
ANISYAH TIARA PUTRI	0	0	0	-0,02	0,96	0,17	0,67	0	0	0,43	0,49	1,14	0,76	0,23	0	0,5	0	0	0,31	0,75	60	54,31	-5,68
NAZWA RAHMADANTIE RAMBE	0,46	0,3	0	0	0,96	0,17	0	0	0,4	0,43	0,49	0	0	0,23	1,15	0,5	0,61	0,47	0	0,75	65	58,82	-6,18
DWI UTAMI RIZQINA	0,46	0	1,21	-0,02	0,96	0	0,67	0	0,4	0	0,49	1,14	0	0,23	0	0,5	0	0,47	0,31	0,75	65	64,34	-0,65
MAJDALUNA AINUN SYIFA	0	0,3	1,21	0	0,96	0,17	0	0,74	0	0,43	0	1,14	0,76	0	1,15	0,5	0,61	0	0	0,75	60	74,12	14,12
RESTU BINTANG PANGESTU	0,46	0,3	0	-0,02	0,96	0,17	0	0	0,4	0	0,49	1,14	0	0,23	1,15	0,5	0	0	0,31	0,75	65	58,14	-6,86
BUNGA SAHARA SIMBOLON	0	0	1,21	0	0,96	0	0,67	0,74	0	0,43	0,49	0	0,76	0,23	1,15	0	0,61	0,47	0	0,75	60	71,99	11,99
SAQILA MEINDI	0,46	0,3	0	-0,02	0,96	0,17	0,67	0	0,4	0,43	0	1,14	0,76	0	0	0,5	0,61	0	0,31	0	65	56,86	-8,13
YUSUF KORY MARPAUNG	0,46	0	1,21	0	0,96	0,17	0,67	0,74	0	0,43	0,49	0	0,76	0,23	1,15	0,5	0	0	0	0,75	65	72,42	7,42
ATIQAH FARRAS ASRI	0	0,3	1,21	0	0,96	0	0,67	0,74	0,4	0	0,49	1,14	0	0,23	1,15	0	0,61	0,47	0,31	0,75	70	80,15	10,15
AZMI AZARIA	0,46	0,3	0	-0,02	0,96	0,17	0	0,74	0	0,43	0	1,14	0,76	0,23	0	0,5	0	0	0,31	0,75	65	57,20	-7,79
JIHAN MAHIRAH NASUTION	0	0	1,21	-0,02	0	0,17	0,67	0	0,4	0,43	0	1,14	0,76	0,23	0	0	0,61	0,47	0,31	0,75	65	60,60	-4,39
RAFA AZZIKRI BUTAR-BUTAR	0,46	0	1,21	0	0,96	0,17	0	0,74	0	0,43	0,49	1,14	0,76	0,23	1,15	0,5	0,61	0,47	0	0	70	79,22	9,22

RAHSYA HASBY SIREGAR	0	0,3	0	-0,02	0	0,17	0	0,74	0,4	0,43	0,49	0	0,76	0	1,15	0,5	0	0,47	0	0,75	60	52,19	-7,81
SYALFA KHAIRINA HAWANI HARAHAP	0,46	0,3	1,21	0	0,96	0,17	0,67	0	0,4	0,43	0	1,14	0	0,23	0	0,5	0,61	0	0,31	0	65	62,81	-2,18
MAULANA IBRAHIM HARAHAP	0	0,3	0	-0,02	0	0,17	0,67	0	0,4	0	0,49	1,14	0,76	0	1,15	0	0,61	0,47	0,31	0,75	65	61,2	-3,8
FADHIL AHMAD GALY	0,46	0	1,21	0	0,96	0,17	0,67	0,74	0	0,43	0,49	0	0,76	0	1,15	0,5	0	0	0,31	0,75	65	73,1	8,1
NISRINA ANGGITA NASUTION	0	0,3	1,21	0	0,96	0	0,67	0	0,4	0,43	0,49	0	0,76	0,23	1,15	0,5	0	0,47	0,31	0,75	70	73,35	3,35

Tabel di atas menunjukkan perbandingan skor antara *Classical Test* dan *Item Response Theory* (IRT) untuk sejumlah siswa dalam sebuah ujian. Skor *Classical Test* dihitung berdasarkan jumlah jawaban benar dari total soal, kemudian dikonversi ke skala 100. Sedangkan skor IRT merupakan estimasi kemampuan siswa (*theta*) berdasarkan karakteristik butir soal (*a*, *b*, *c*) menggunakan model IRT *Three-Parameter Logistic*.

Misalnya pada Siswa M. HANDY AULIA mendapatkan skor 65 dalam ujian Classical Test, namun setelah dihitung menggunakan model IRT, kemampuannya diperkirakan mencapai 67,91. Selisih antara kedua skor ini adalah sekitar 2,915. Ini menunjukkan bahwa model IRT memberikan estimasi yang sedikit lebih tinggi untuk kemampuan siswa ini dibandingkan dengan skor Classical Test. Selisih ini dapat diinterpretasikan sebagai perbedaan dalam sensitivitas pengukuran antara Classical Test dan IRT, di mana IRT mampu mengidentifikasi nuansa kemampuan siswa dengan lebih baik dengan skor dinamis berdasarkan karakteristik dari soal terjawab dan dipengaruhi oleh para siswa lain.

#### 4.2.4.2 Analisis *Item Characteristic Curve* (ICC)

*Item Characteristic Curve* (ICC) adalah grafik yang menunjukkan probabilitas seorang siswa menjawab benar suatu butir soal sebagai fungsi dari kemampuan siswa tersebut  $\theta$ . Kurva ini didasarkan pada model Teori Respons Butir (IRT), dan memvisualisasikan bagaimana butir soal tertentu membedakan siswa berdasarkan kemampuan mereka.

Dalam sistem ini, range  $\theta$  adalah -3 hingga 3, dan ini general pada setiap sistem yang mengadopsi CAT-IRT. Berikut adalah analisis ICC pada soal nomor 1.

Rumus IRT 3 parameter:

$$P(\theta) = c + \frac{1 - c}{1 + e^{-a(\theta-b)}}$$

Di mana:

- 1)  $P(\theta)$ = Probabilitas jawaban benar
- 2)  $a$  = Parameter diskriminasi

- 3)  $b$  = Parameter kesulitan
- 4)  $c$  = Parameter tebakan
- 5)  $\theta$  = Kemampuan siswa

Pada soal nomor 1, parameter ( $a = 1.0$ ), ( $b = 0.5$ ), ( $c = 0.25$ )

Untuk berbagai nilai ( $\theta$ ):

- 1) ( $\theta = -3$ )
- 2) ( $\theta = -2$ )
- 3) ( $\theta = -1$ )
- 4) ( $\theta = 0$ )
- 5) ( $\theta = 1$ )
- 6) ( $\theta = 2$ )
- 7) ( $\theta = 3$ )

Untuk ( $\theta = -3$ ).

$$P(-3) = 0.26 + \frac{1 - 0.26}{1 + e^{-1.14(-3-0.86)}}$$

$$P(-3) = 0.26 + \frac{0.74}{1 + e^{3.42}}$$

$$P(-3) = 0.26 + \frac{0.74}{1+30.6}$$

$$P(-3) = 0.26 + \frac{0.74}{31.6}$$

$$P(-3) = 0.26 + 0.0234$$

$$P(-3) = 0.2834$$

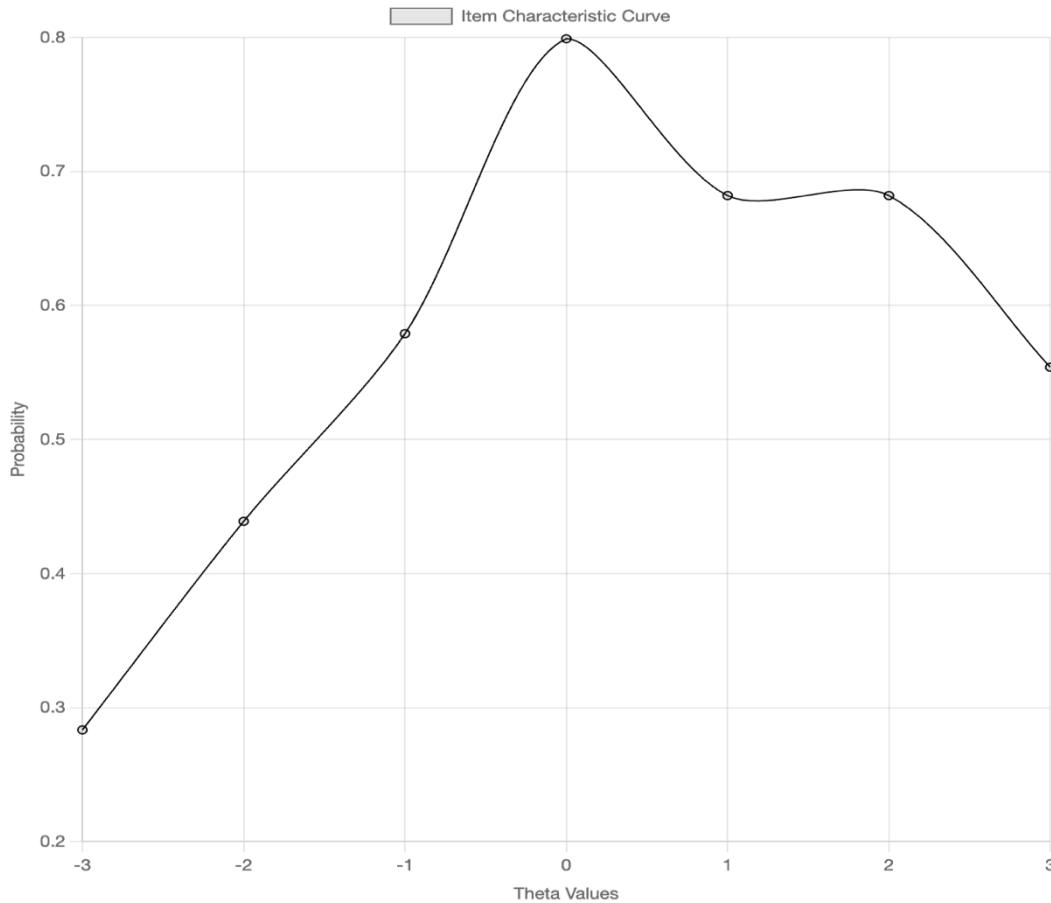
Dan seterusnya hingga untuk ( $\theta = 3$ ). Dengan menghitung probabilitas untuk setiap nilai ( $\theta$ ), kita dapat membuat plot ICC dengan sumbu x sebagai ( $\theta$ ) dan sumbu y sebagai probabilitas ( $P(\theta)$ ).

**Tabel 4.7:** Data plot ICC

$\theta$	$P(\theta)$
-3	0.2834
-2	0.439
-1	0.579
0	0.799
1	0.682

2	0.682
3	0.554

Hasil dari  $P(\theta)$  yang telah dihitung berdasarkan  $\theta$  adalah berupa plot ICC seperti berikut:



**Gambar 4.2:** Kurva Karakteristik Item (ICC)

Pada soal nomor 1 kurva ICC menunjukkan bahwa probabilitas peserta untuk menjawab benar meningkat seiring dengan peningkatan kemampuan  $\theta$ . Pada nilai  $\theta$  yang rendah (misalnya, -3), probabilitasnya masih rendah (sekitar 0.2834), tetapi kemudian meningkat secara signifikan seiring dengan peningkatan  $\theta$  hingga mencapai puncak probabilitas pada sekitar  $\theta=0$  (probabilitas sekitar 0.799). Setelah itu, probabilitasnya mulai menurun seiring dengan peningkatan  $\theta$  lebih lanjut. Ini menunjukkan bahwa soal memiliki tingkat kesulitan yang lebih tinggi di sekitar nilai  $\theta=0$ , di mana peserta dengan kemampuan sedang cenderung

menjawab dengan benar, sementara peserta dengan kemampuan ekstrem (sangat rendah atau sangat tinggi) cenderung menjawab dengan salah.

#### 4.2.4.3 Evaluasi Karakteristik Butir

Evaluasi karakteristik butir dilakukan untuk menilai apakah butir soal yang digunakan memenuhi kriteria yang diharapkan dalam model 3PL. Evaluasi ini melibatkan analisis parameter diskriminasi, kesulitan, dan tebakan. Kriteria evaluasinya adalah:

1) Diskriminasi (a)

Butir soal dianggap baik jika memiliki nilai  $a \geq 1$ .

2) Kesulitan (b)

Butir soal dianggap sesuai jika nilai  $b$  berada dalam rentang -2 hingga 2.

3) Tebakan (c)

Butir soal dianggap valid jika nilai  $c < 0.25$ .

Berdasarkan hasil estimasi parameter, sebagian besar butir soal memenuhi kriteria yang diharapkan. Namun, terdapat beberapa butir soal dengan nilai diskriminasi rendah atau nilai tebakan tinggi yang perlu ditinjau kembali.

**Tabel 4.8:** Hasil Evaluasi Karakteristik Butir

Soal	a	b	c	Diskriminasi (a)	Kesulitan (b)	Tebakan (c)
1	0.47	0.43	0.25	Tidak	Ya	Ya
2	1.31	0.46	0.25	Ya	Ya	Ya
3	0.47	0.51	0.25	Tidak	Ya	Ya
4	0.8	0.32	0.25	Tidak	Ya	Ya
5	0.8	0.87	0.25	Tidak	Ya	Ya
6	0.8	0.61	0.25	Tidak	Ya	Ya
7	0.47	0.74	0.25	Tidak	Ya	Ya
8	1.31	0	0.25	Ya	Ya	Ya
9	0.47	0.51	0.25	Tidak	Ya	Ya
10	0.47	1.1	0.25	Tidak	Ya	Ya
11	0.47	0.71	0.25	Tidak	Ya	Ya
12	0.47	0.71	0.25	Tidak	Ya	Ya
13	0.47	0.74	0.25	Tidak	Ya	Ya

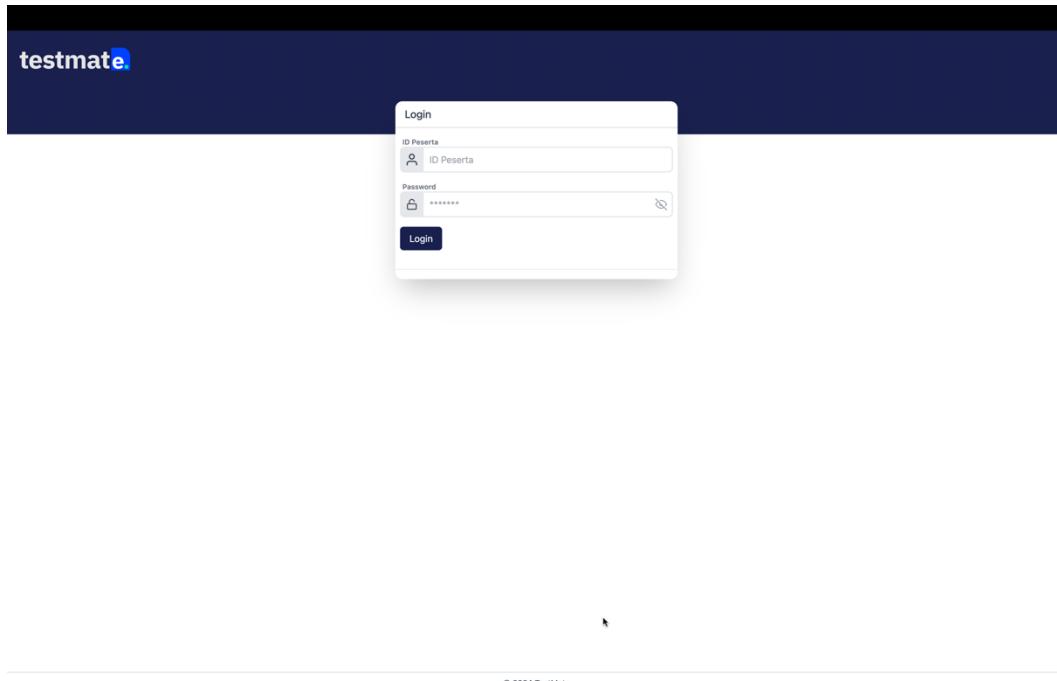
14	1.31	0.3	0.25	Ya	Ya	Ya
15	1.31	0.53	0.25	Ya	Ya	Ya
16	1.31	0.66	0.25	Ya	Ya	Ya
17	1.31	0.14	0.25	Ya	Ya	Ya
18	0.8	0	0.25	Tidak	Ya	Ya
19	0.8	0.38	0.25	Tidak	Ya	Ya
20	0.47	1.1	0.25	Tidak	Ya	Ya

Evaluasi menunjukkan bahwa mayoritas butir soal memiliki karakteristik yang baik, sehingga tes dapat dianggap valid dan reliabel untuk mengukur kemampuan siswa dalam bidang eksakta.

#### 4.2 Implementasi Interface

Setelah tahap analisis dan perancangan sistem *Computer-based Adaptive Test (CAT)* dengan metode *Item Response-Theory Three-Parameter Logistics* selesai, selanjutnya dilakukan implementasi dan pengujian sistem. Implementasi sistem dilakukan dengan menerjemahkan hasil analisis dan perancangan ke dalam kode program sehingga menghasilkan aplikasi yang berfungsi sesuai dengan tujuan, yaitu untuk mengevaluasi keandalan dan karakteristik soal, mengestimasi kemampuan siswa secara individual, dan menghasilkan laporan yang akurat.

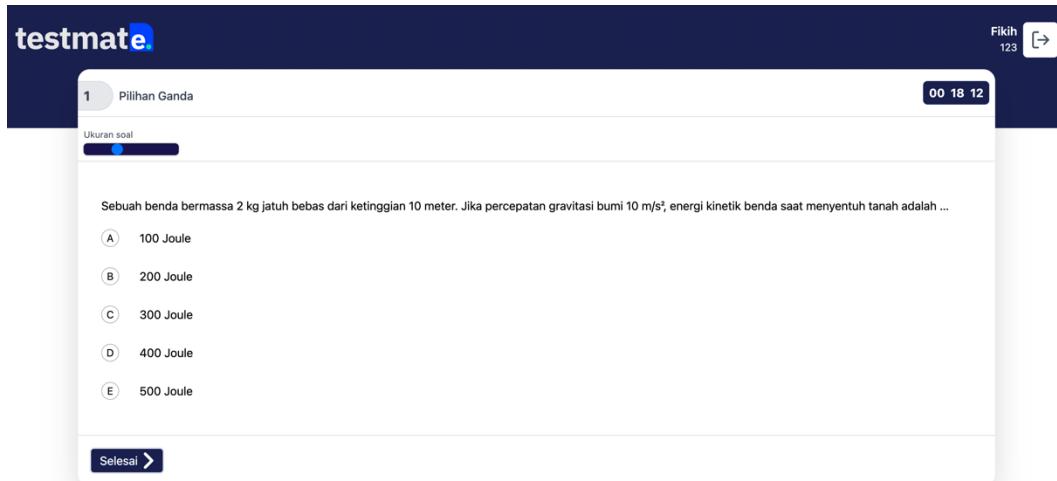
#### 4.2.1 Tampilan Halaman Home



**Gambar 4.3:** Peserta Ujian - Halaman Home

Halaman Home pada Sisi Siswa adalah halaman pertama yang muncul saat sistem dijalankan. Pada halaman ini, terlihat logo "testmate" di pojok kiri atas dengan latar belakang biru gelap. Di tengah halaman, terdapat form login yang terdiri dari dua kolom isian: ID Peserta dan Password. Form ini juga memiliki tombol "Login" berwarna biru gelap yang terletak di bawah kolom isian password. Terdapat ikon pengguna di kolom ID Peserta dan ikon kunci di kolom password, serta ikon mata di sebelah kanan kolom password untuk menampilkan atau menyembunyikan password yang dimasukkan. Di bagian paling bawah halaman, terdapat tulisan "© 2024 TestMate".

#### 4.2.2 Tampilan Halaman Ujian



**Gambar 4.4:** Tampilan Halaman Ujian

Halaman Ujian Siswa menampilkan soal ujian beserta instruksi, timer, navigasi soal, nomor soal, teks soal, pilihan jawaban, input jawaban, status soal, dan tombol submit. Halaman ini dirancang untuk membantu siswa fokus pada ujian dan menyelesaikannya dengan mudah.

#### 4.2.3 Tampilan Halaman *Login Admin*



**Gambar 4.5** Tampilan Login Admin

Halaman Login Admin memungkinkan administrator untuk masuk ke sistem dengan memasukkan nama pengguna dan kata sandi. Halaman ini memiliki elemen seperti formulir login, tombol login, pesan kesalahan, tautan lupa kata sandi, dan logo aplikasi. Fungsi utamanya adalah untuk memverifikasi identitas administrator dan menjaga keamanan sistem. Tampilannya bervariasi tergantung platform, namun umumnya memiliki desain sederhana dan mudah digunakan.

#### 4.2.4 Tampilan Halaman Daftar Peserta

Detail	ID Peserta	Nama Peserta	Status	Action
<input checked="" type="checkbox"/>	123	Fikih	Aktif	<button>Edit</button> <button>Hapus</button>
<input checked="" type="checkbox"/>	1234	Fikih	Aktif	<button>Edit</button> <button>Hapus</button>

2 peserta dari 2 total data peserta

Gambar 4.6: Tampilan Halaman Daftar Peserta

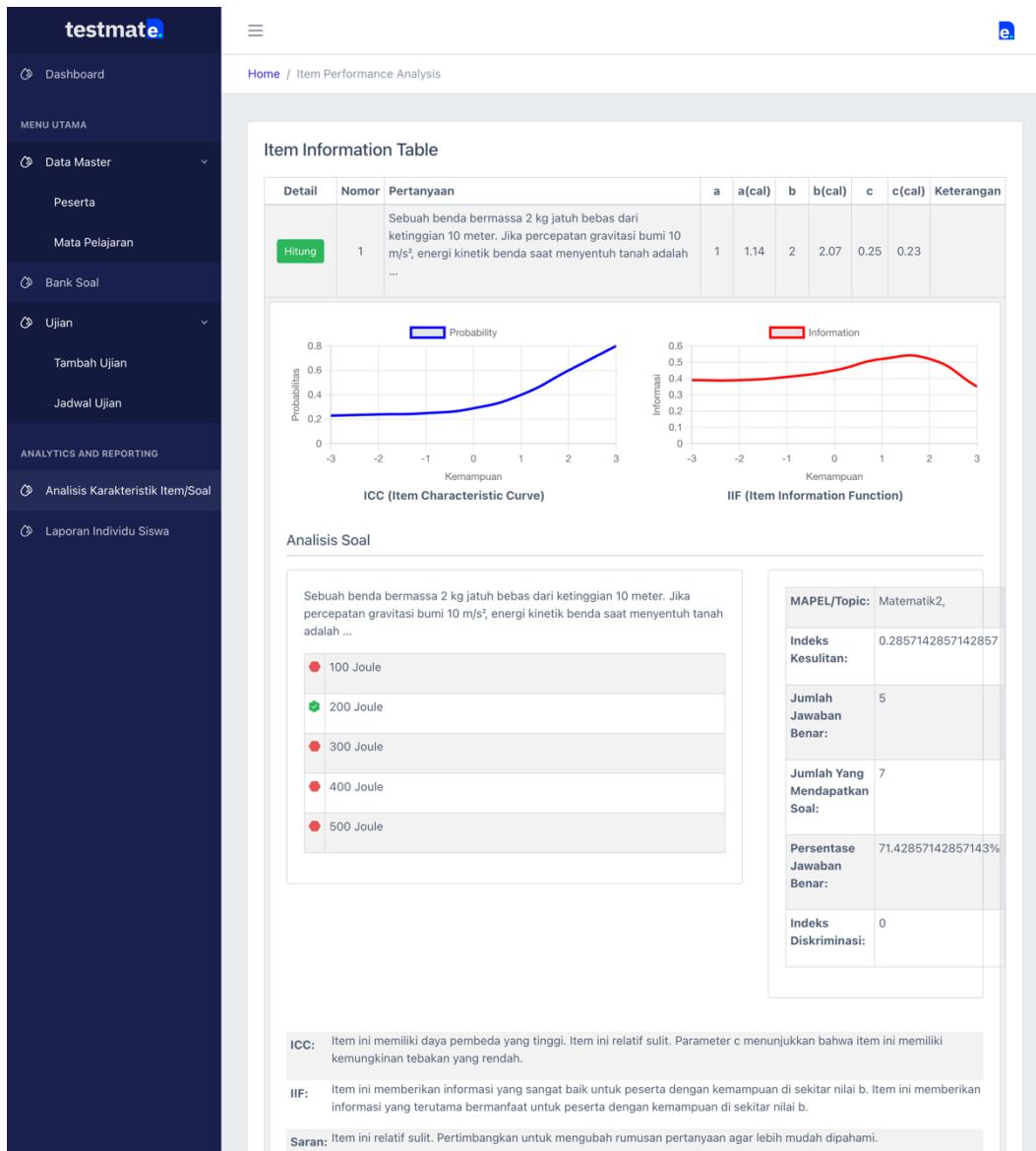
Halaman Daftar Peserta menampilkan daftar peserta yang terdaftar pada suatu acara atau kursus tertentu. Halaman ini memungkinkan administrator atau penyelenggara untuk mengelola dan melacak data peserta dengan mudah. Halaman ini memiliki elemen, seperti bar pencarian, informasi peserta, dan tombol aksi.

#### 4.2.5 Tampilan Halaman Daftar Bank Soal

Gambar 4.0.7: Tampilan Halaman Daftar Bank Soal

Halaman Daftar Soal Bank Soal menampilkan daftar soal yang tersimpan dalam bank soal. Halaman ini memungkinkan pengguna untuk mengelola soal dengan mudah, seperti mencari, melihat detail, mengedit, menghapus, dan menambahkan soal baru. Halaman ini memiliki elemen, seperti bar pencarian, data soal, dan tombol aksi.

#### 4.2.6 Tampilan Halaman Analisis Karakteristik Soal



**Gambar 4.8:** Tampilan Halaman Analisis Karakteristik Soal

Halaman Analisis Butir Soal menyediakan analisis mendalam tentang pertanyaan atau item individu dari suatu penilaian atau ujian. Halaman ini memungkinkan pendidik atau pengembang penilaian untuk mengevaluasi efektivitas dan kualitas setiap pertanyaan, membantu mereka meningkatkan penilaian dan hasil belajar siswa. Halaman ini menampilkan beberapa elemen kunci:

1) Informasi Pertanyaan

Menampilkan detail penting tentang pertanyaan, seperti nomor pertanyaan, batang pertanyaan, pilihan jawaban, dan jawaban yang benar.

2) Analisis Kinerja Pertanyaan

Memberikan wawasan tentang seberapa baik kinerja pertanyaan, seringkali menggunakan alat bantu visual seperti:

- Kurva Karakteristik Item (ICC)

Grafik yang menunjukkan probabilitas menjawab pertanyaan dengan benar pada tingkat kemampuan siswa yang berbeda.

- Fungsi Informasi Item (IIF)

Grafik yang menunjukkan jumlah informasi yang diberikan pertanyaan pada tingkat kemampuan siswa yang berbeda.

3) Interpretasi dan Rekomendasi Pertanyaan

Menawarkan saran untuk meningkatkan pertanyaan berdasarkan analisis, seperti merevisi batang pertanyaan, memodifikasi pilihan jawaban, atau menyesuaikan tingkat kesulitan.

### 4.3 Pengujian

Pada bagian ini, dilakukan pengujian fit atau kesesuaian model *Item Response Theory* (IRT) menggunakan metode *Chi-Square*.

#### 4.3.1 Metode Uji Kesesuaian Model

Model yang sesuai akan menghasilkan prediksi yang akurat dan dapat diandalkan. Pada penelitian ini, uji kesesuaian model dilakukan menggunakan uji fit *Chi-Square* ( $\chi^2$ ).

Uji fit dengan *Chi-Square* digunakan untuk membandingkan frekuensi yang diamati dengan frekuensi yang diharapkan berdasarkan model yang dihipotesiskan. Dalam konteks model pengukuran IRT, uji ini dapat digunakan untuk menentukan apakah model 3PL sesuai dengan data respons siswa.

**Tabel 4. 9:** Theta Akhir Peserta

Peserta	theta
M, HANDY AULIA	6,32
VIA SURIYANTO	5,29
SHARLIZ AYLA SAMAIRO NST	6,22
ANISYAH TIARA PUTRI	6,39
NAZWA RAHMADANTIE RAMBE	6,16
DWI UTAMI RIZQINA	5,9
MAJDALUNA AINUN SYIFA	7,21
RESTU BINTANG PANGESTU	6,08
BUNGA SAHARA SIMBOLON	7,26
SAQILA MEINDI	5,93
YUSUF KORY MARPAUNG	6,85
ATIQAH FARRAS ASRI	7,92
AZMI AZARIA	5,97
JIHAN MAHIRAH NASUTION	5,92
RAFA AZZIKRI BUTAR-BUTAR	7,65
RAHSYA HASBY SIREGAR	5,84
SYALFA KHAIRINA HAWANI HARAHAP	5,42
MAULANA IBRAHIM HARAHAP	6,9
FADHIL AHMAD GALY	6,93
NISRINA ANGGITA NASUTION	7,12

Tabel diatas adalah distribusi akhir theta atau kemampuan siswa dalam satu sesi ujian. Theta akhir ini akan menjadi dasar pengujian fit menggunakan *Chi-Square*.

Selanjutnya mentukan kelompok kemampuan dengan membagi data theta menjadi 4 kelompok berdasarkan kuartil cukup untuk mewakili kelompok:

Kelompok 1 (Terendah) :  $\theta < 5.915$

Kelompok 2 :  $5.915 \leq \theta < 6.19$

Kelompok 3 :  $6.19 \leq \theta < 7.015$

Kelompok 4 (Tertinggi) :  $7.015 \leq \theta$

Hasil dari pengelompokan berupa Frekuensi yang Diamati ( $O$ ) dalam kelompok seperti tabel dibawah.

**Tabel 4.10:** Frekuensi yang Diamati ( $O$ )

Kelompok	Frekuensi ( $O$ )
1	5
2	3
3	6
4	6

Setelah mendapatkan frekuensi yang diamati ( $O$ ), kita akan menghitung frekuensi yang diharapkan ( $E$ ).

- 1) Dengan asumsi distribusi normal untuk frekuensi yang diharapkan, setiap kelompok diharapkan memiliki jumlah siswa yang sama.
- 2)  $E = \text{Total Siswa} / \text{Jumlah Kelompok} = 20 / 4 = 5$ .

Setelah mengetahui frekuensi yang diharapkan, perlu menentukan Derajat Kebebasan (df).

$$df = \text{Jumlah Kelompok} - 1 = 4 - 1 = 3.$$

Selanjutnya menghitung Statistik Chi-Square ( $\chi^2$ ).

**Tabel 4.11** Hitung Statistik Chi-Square

Kelompok	$O$	$E$	$(O - E)^2$	$(O - E)^2 / E$
1	5	5	0	0
2	3	5	4	1
3	6	5	1	0
4	6	5	1	0
Total	20	20		2

Tentukan P-value melalui fungsi *Cumulative distribution function* pada python:

```
from scipy.stats import chi2
p_value = 1 - chi2.cdf(2, 3)
print(p_value)
```

Hasilnya adalah 0.57.

Akhirnya kita akan interpretasi dengan ( $\alpha = 0.05$  atau 5%). Dengan tingkat signifikansi  $\alpha = 0.05$ " berarti kita telah menetapkan ambang batas untuk menolak hipotesis nol sebesar 5%. Dengan kata lain, kita hanya akan menolak hipotesis nol jika ada kemungkinan kurang dari 5% bahwa hasil yang kita amati terjadi secara kebetulan.

Karena P-value (0.57)  $> \alpha$  (0.05), kita gagal menolak hipotesis nol.

### 4.3.2 Hasil Uji Kesesuaian Model

Uji fit *Chi-Square* digunakan untuk mengevaluasi kesesuaian antara distribusi kemampuan siswa yang dihasilkan dari model 3PL dengan distribusi yang diharapkan. Dalam contoh ini, data kemampuan 20 siswa dibagi menjadi empat kelompok berdasarkan kuartil, dengan asumsi distribusi normal.

Hasil uji *Chi-Square* menunjukkan nilai  $\chi^2 = 2$  dengan  $df = 3$  dan P-value = 0.57. Dengan  $\alpha = 0.05$ , karena P-value  $> \alpha$ , kita tidak dapat menolak hipotesis nol. Kesimpulannya, tidak ada cukup bukti untuk menyatakan perbedaan signifikan antara distribusi kemampuan siswa yang diamati dan distribusi yang diharapkan.

## BAB 5

### PENUTUP

#### **5.1 Kesimpulan**

Penelitian ini telah berhasil mengembangkan dan menguji sistem *Computer-Based Adaptive Testing* (CAT) yang dirancang untuk menyesuaikan tingkat kesulitan soal dengan kemampuan peserta ujian. Berikut adalah kesimpulan utama dari penelitian ini:

- 1) Penerapan model Item Response Theory (IRT) Three-Parameter Logistics (3PL) pada Computer-Based Adaptive Test (CAT) terbukti efektif dalam menganalisis karakteristik butir soal eksakta. Model ini mampu mengestimasi parameter diskriminasi (a), kesulitan (b), dan tebakan (c) dengan baik, memberikan informasi yang berharga tentang kualitas soal.
- 2) Perbandingan skor klasik dengan skor yang diperoleh melalui penyetaraan IRT menunjukkan bahwa model 3PL mampu membedakan kemampuan siswa dengan lebih baik dan akurat.
- 3) Hasil uji fit Chi-Square menunjukkan bahwa tidak terdapat perbedaan yang signifikan antara distribusi kemampuan siswa yang diamati dengan distribusi yang diharapkan. Hal ini mengindikasikan bahwa model 3PL cukup sesuai dengan data respons siswa.
- 4) Sistem CAT yang dikembangkan dalam penelitian ini berfungsi dengan baik, mampu mengadministrasikan tes, mengestimasi kemampuan siswa, dan menghasilkan laporan yang akurat.

#### **5.2 Saran**

Berdasarkan hasil penelitian ini, terdapat beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut dari sistem CAT dan kemungkinan penelitian lanjutan:

- 1) Untuk penelitian lanjutan, disarankan untuk menguji sistem CAT pada populasi peserta ujian yang lebih besar dan beragam. Selain itu, pengembangan dan pengujian algoritma adaptif lainnya yang dapat digunakan dalam sistem CAT untuk meningkatkan akurasi dan efisiensi penyesuaian soal juga perlu

dieksplorasi. Penggunaan model IRT yang lebih kompleks atau alternatif, seperti model multidimensional IRT, juga dapat menjadi fokus penelitian yang menarik untuk mengukur kemampuan yang lebih beragam.

- 2) Untuk pengembangan sistem CAT, disarankan untuk meningkatkan antarmuka pengguna agar lebih intuitif dan mudah digunakan oleh peserta ujian dari berbagai latar belakang. Selain itu, penambahan fitur analitik yang lebih canggih untuk analisis hasil ujian, termasuk laporan individual yang lebih terperinci, juga dapat menjadi langkah yang baik.
- 3) Dalam implementasi di berbagai konteks, disarankan untuk menerapkan sistem CAT dalam berbagai konteks pendidikan dan pelatihan, termasuk ujian sertifikasi profesional dan pelatihan keterampilan. Pengujian sistem dalam skenario ujian berbasis kompetensi juga dapat dilakukan untuk mengukur keterampilan praktis dan teoritis secara lebih komprehensif.

## DAFTAR PUSTAKA

- Adebule, S. (2013). Item Response Theory as a Basis for Measuring Latent Trait of Interest. *Greener Journal of Social Sciences*, 3, 378–382. <https://doi.org/10.15580/GJSS.2013.7.062513691>
- Anak Agung Purwa Antara. (2020). *PENYETARAAN VERTIKAL DENGAN PENDEKATAN KLASIK DAN ITEM RESPONSE THEORY (TEORI DAN APLIKASI)*.
- Artturi Jalli. (2022). *What is laravel? (Definition, uses, features) built in*. <https://builtin.com/software-engineering-perspectives/laravel>
- G. Jatobá, V. M., Farias, J. S., Freire, V., Ruela, A. S., & Delgado, K. V. (2020). ALICAT: a customized approach to item selection process in computerized adaptive testing. *Journal of the Brazilian Computer Society*, 26(1), 4. <https://doi.org/10.1186/s13173-020-00098-z>
- Gyamfi, A., & Acquaye, R. (2023). Parameters and Models of Item Response Theory (IRT): A Review of Literature. *Acta Educationis Generalis*, 13, 68–78. <https://doi.org/10.2478/atd-2023-0022>
- Hagander, M. (2024). PostgreSQL: About. In *What is PostgreSQL?* <https://www.postgresql.org/about/>
- Hakim, M. L., & Ramalis, T. R. (2019). Karakteristik Tes Hasil Belajar Ranah Kognitif Materi Elastisitas Menggunakan Analisis Item Response Theory. *Jurnal Penelitian Pembelajaran Fisika*, 10(1), 22–32. <http://journal.upgris.ac.id/index.php/JP2F>
- Hamidah, N., & Istiyono, E. (2022). The quality of test on National Examination of Natural science in the level of elementary school. *International Journal of Evaluation and Research in Education (IJERE)*. <https://api.semanticscholar.org/CorpusID:250227954>
- Herwin, H., Tenriawaru, A., & Fane, A. (2019). Math elementary school exam analysis based on the Rasch model. *Jurnal Prima Edukasia*, 7(2), 106–113. <https://doi.org/10.21831/jpe.v7i2.24450>
- Lang, J. W. B., & Tay, L. (2021). The Science and Practice of Item Response Theory in Organizations. *Annual Review of Organizational Psychology and*

- Organizational Behavior*, 8(Volume 8, 2021), 311–338.  
<https://doi.org/https://doi.org/10.1146/annurev-orgpsych-012420-061705>
- Nathan Thompson. (2023, May 21). *Computerized Adaptive Testing (CAT): An Introduction*. Assessment Systems Corporation.  
<https://assess.com/computerized-adaptive-testing>
- Reid, W. V., Ali, M. K., & Field, C. B. (2020). The future of bioenergy. *Global Change Biology*, 26(1). <https://doi.org/10.1111/gcb.14883>
- Ridho, A., Psikologi, F., & Malang, U. (2007). *KARAKTERISTIK PSIKOM ETRIK TES BERDASARKAN PENDEKATAN TEORI TES KLASIK DAN TEORI RESPON AITEM*.
- Sijtsma, K., & van der Ark, L. A. (2022). Advances in nonparametric item response theory for scale construction in quality-of-life research. *Quality of Life Research*, 31(1), 1–9. <https://doi.org/10.1007/s11136-021-03022-w>
- Sudaryono. (2011). Implementasi Teori Responsi Butir (Item Response Theory) Pada Penilaian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan Dan Kebudayaan*, 17(6), 719–732.
- Syamsuddin, S. (2023). Implementasi Classic Test dan Item Respon Theory Pada Penilaian Tes Pembelajaran Matematika. *EDUSCOPE: Jurnal Pendidikan, Pembelajaran, Dan Teknologi*, 8(2), 28–43.  
<https://doi.org/10.32764/eduscope.v8i2.3488>
- Yağcı, M. (2023). Evaluating knowledge levels of students with a Computerized Adaptive Test. *Journal of Teacher Education and Lifelong Learning*, 5(2), 921–932. <https://doi.org/10.51535/tell.1280384>