Studying the Educational Influence of Adaptive Learning Tools using Item Response Theory

by

Satanshu Mishra

B.Sc. Hons., The University of British Columbia, 2024

A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

B.S.C. COMPUTER SCIENCE HONOURS

in

Faculty of Science

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

April 28, 2024

© Satanshu Mishra, 2024

The following individual certifies that they have reviewed the report entitled:

STUDYING THE EDUCATIONAL INFLUENCE OF ADAPTIVE LEARNING TOOLS USING ITEM RESPONSE THEORY

submitted by SATANSHU MISHRA in partial fulfilment of the requirements of the degree B.Sc. Computer Science Honours.

Dr. Abdallah Mohammed, I. K. Barber Faculty of Science

**Supervisor**

Dr. Mostafa Mohammed, I. K. Barber Faculty of Science

**Supervisor**

**Abstract**

In today's technology-driven world, the demand for computing skills spans across disciplines, necessitating tailored learning experiences. Current e-learning systems often employ a "one size fits all" approach, which fails to accommodate for diverse student skill levels and learning styles. Adaptive learning, a technique that adjusts content based on individual performance, offers a solution to this challenge. This study introduces the uLearn Adaptive Learning tool, designed to personalize the learning journey for students in an introductory programming course. By leveraging Item Response Theory (IRT) and machine learning (ML) techniques, the system dynamically adjusts question difficulty based on overall student performance and individual abilities. Moreover, the uLearn tool employs a mastery tracking system to monitor student competency in various topics and categories, to provide targeted learning to students. The initial results of this study indicate promising improvements in student performance and comprehension, highlighting the potential of the uLearn adaptive learning tool to enhance learning outcomes in diverse educational settings. Additionally, students expressed increased engagement and a desire to continue using Adaptive Learning systems in the future, indicating a positive reception and potential for long-term adoption of the uLearn tool and Adaptive Learning tools in general.

Table of Contents

[Table of Contents iv](#_Toc165293349)

[List of Symbols and Abbreviations viii](#_Toc165293350)

[Acknowledgements ix](#_Toc165293351)

[1. Introduction 1](#_Toc165293352)

[2. Background 1](#_Toc165293353)

[2.1 Item Response Theory 1](#_Toc165293354)

[2.1.1 Classical Test Theory 1](#_Toc165293355)

[2.1.1.1 Classical Test Theory vs Item Response Theory 2](#_Toc165293356)

[2.1.2 IRT Assumptions 2](#_Toc165293357)

[2.1.2.1 Item Characteristic Curve 2](#_Toc165293358)

[2.1.3 Item Parameters 3](#_Toc165293359)

[2.1.4 IRT Model Types 3](#_Toc165293360)

[2.1.4.1 2-Parameter Logistic Model 3](#_Toc165293361)

[2.2 Literature Review 5](#_Toc165293362)

[3. Proposed System 5](#_Toc165293363)

[3.1 Design 5](#_Toc165293364)

[3.1.1 Data Collected 5](#_Toc165293365)

[3.1.2 Question Categories 6](#_Toc165293366)

[3.1.3 Mastery 6](#_Toc165293367)

[3.1.4 Modified (Question) Difficulty 6](#_Toc165293368)

[3.1.5 Difficulty Offset 7](#_Toc165293369)

[3.1.6 Bounding Ranges 7](#_Toc165293370)

[3.1.7 The Model 8](#_Toc165293371)

[3.2 Implementation 9](#_Toc165293372)

[4. Case Study 9](#_Toc165293373)

[4.1 Study Setting 9](#_Toc165293374)

[4.2 Results & Findings 10](#_Toc165293375)

[5. Discussion 12](#_Toc165293376)

[5.1 Significance of the System 12](#_Toc165293377)

[5.2 Analyzing Findings 12](#_Toc165293378)

[5.3 Limitations of the System 15](#_Toc165293379)

[6. Conclusion & Future Work 15](#_Toc165293380)

[7. References 17](#_Toc165293381)

[8. Appendices 18](#_Toc165293382)

[8.1 Individual Mastery Progression 18](#_Toc165293383)

[8.2 Individual Difficulty Offset Progression 20](#_Toc165293384)

**List of Tables**

[***Table 3.1 Structure of the studentKnowledge relation****. This relation is critical for the ML model. Only relevant columns are shown.* 6](#_Toc165266438)

[**Table 3.2:** **Structure of question relation.** This relation also holds critical information about each question used for the ML model. Only relevant columns are shown. 6](#_Toc165266439)

**List of Figures**

[**Figure 2.1: Graph representing the Item Response Function and the Item Characteristic Curve.** Plots the Respondents Ability vs the Probability of endorsing a correct response. 2](#_Toc165266450)

[**Figure 2.2:** **Graphs representing the transformations to ICC curve with changes to Item parameters.** Graph A represents changes with difficulty. Graph B represents changes with varying discrimination. 3](#_Toc165266451)

[**Figure 2.3:** **Item Characteristic Curves for 2-Parameter Logistical Model with Varying Item Parameters.** The Green curve acts as the base curve for this plot. The Blue curve shows the effect of increasing the Discrimination parameter. The Red Curve shows the effect of increasing the Difficulty parameter. 4](#_Toc165266452)

[**Figure 3.1: Model Overview visualization**. This figure shows the different scales and the default window for question selection (blue). 8](#_Toc165266453)

[**Figure 3.2: DFD-1 System Architecture Diagram.** This figure shows the data flow for the ML model. 8](#_Toc165266454)

[**Figure 4.1: Mastery Progression over Attempts (Overall Trend)**. This figure shows the over all trend for mastery all the categories combined into a single graph for each topic. 10](#_Toc165266455)

[**Figure 4.2: Difficulty Offset Progression Over Attempts (Overall Trend**). This figure shows the over all trend for difficulty offset all the categories combined into a single graph for each topic. 10](#_Toc165266456)

[**Figure 4.3: Cumulative Question Difficulty over Attempts**. This figure shows how the question difficulties across all topics as attempts increase focusing on the period after 14 attempts. 11](#_Toc165266457)

[**Figure 5.1: Mastery Progression (Attempts > 25).** This graph shows the mastery progression for the topic-category pairs with the most attempts to highlight the preliminary findings from the results. 12](#_Toc165266458)

[**Figure 5.2: System Effectiveness and Future Preference**. This figure shows the qualitative feedback received from students after they used the tool highlighting their opinion of the tool’s effectiveness and their future preferences with Adaptive Learning tools. 13](#_Toc165266459)

[**Figure 5.3: User Experience and Engagement**. This figure shows the qualitative feedback received from students after they used the tool highlighting their opinion using the tool and its impact. 14](#_Toc165266460)

[**Figure 8.1: Mastery Progression over Attempts for each Category in Error Handling.** 17](#_Toc165266461)

[**Figure 8.2: Mastery Progression over Attempts for each Category in I/O.** 17](#_Toc165266462)

[**Figure 8.3: Mastery Progression over Attempts for each Category in Recursion.** 18](#_Toc165266463)

[**Figure 8.4: Mastery Progression over Attempts for each Category in Lists, Stacks & Queues.** 18](#_Toc165266464)

[**Figure 8.5: Difficulty Offset Progression over Attempts for each Category in Error Handling.** 19](#_Toc165266465)

[**Figure 8.6: Difficulty Offset Progression over Attempts for each Category in Lists, Stacks & Queues.** 19](#_Toc165266466)

[**Figure 8.7: Difficulty Offset Progression over Attempts for each Category in Recursion.** 20](#_Toc165266467)

[**Figure 8.8: Difficulty Offset Progression over Attempts for each Category in I/O.** 20](#_Toc165266468)

# List of Symbols and Abbreviations

|  |  |
| --- | --- |
| 1-PL | The 1-Parameter Logistic Model |
| 2-PL | The 2-Parameter Logistic Model |
| ALS | Adaptive Learning System |
| ALT | Adaptive Learning Tool |
| CS | Computer Science |
| CS1 | First Year Computer Science Programming Course |
| CTT | Classical Test Theory |
| DFD | Data Flow Diagram |
| DSS | Decision Support System |
| ICC | Item Characteristic Curve |
| IRF | Item Response Function |
| IRT | Item Response Theory |
| ML | Machine Learning |
| OOP | Object-Oriented Programming |
| SD | Standard Deviation |
| SOM | Self-Organizing Map |
| FK | Foreign Key (Database Terminology) |
| V.S | Verses |

# Acknowledgements

I would like to express my sincere gratitude to Dr. Abdallah Mohammed for his invaluable guidance, unwavering support, and forbearance throughout this research endeavor. His expertise and mentorship have been instrumental in shaping the trajectory of this study. I am indebted to Dr. Mostafa Mohammed for his invaluable assistance in navigating the complexities of machine learning and for his insightful contributions to this research. I would like to extend my heartfelt thanks to Dr. Scott Fazackerley for graciously permitting the deployment and testing of my system in his first-year computer programming course. This opportunity provided invaluable practical experience and enriched my understanding of the subject matter. I am deeply appreciative of the encouragement and unwavering support of my parents and friends, whose belief in me has been a constant source of inspiration. This research would not have been possible without the contributions and support of these individuals, and for that, I will be forever grateful.

# Introduction

Technology surrounds us. It has found its way into every facet of our lives. From helping us get up in the morning, to organizing our day, or keeping us informed about the world that surrounds us. With this growing exposure to technology, there has also been an increasing demand for computing skills, especially from non-computer-science (CS) majors who recognize the mounting importance and relevance of computing in their fields of study [1]. Students in such classes often have different backgrounds, skills, interests, learning needs, styles, and educational levels. Furthermore, novice students at these levels often lack the necessary problem-solving skills critical for them to succeed in computer programming. This creates an increasing challenge in offering one-size-fits-all CS1 computer programming courses [2]. This is where adaptive learning comes in. Adaptive learning refers to the monitoring of student performance and using the data collected to modify the learning process [3][4]. This research designed and evaluated the uLearn Adaptive Learning Tool (ALT) to personalize the learning experience for students in a CS1 programming course.

For this study, we have formulated the following research questions:

1. What are the current trends of adaptive learning systems (ALS) in CS Education?
2. To what extent does the learning tool implementation enhance learners’ understanding and performance in a first-year CS course?
3. What future recommendations can be leveraged to advance CS1 education through adaptive learning systems?

# Background

The main objective of this thesis is to personalize students’ learning experiences using Item Response Theory. In this section, we introduce key concepts and provide background on Item Response Theory and related concepts.

## Item Response Theory

Item response theory (IRT) is a family of mathematical models designed to elucidate the relationship between latent traits, which are unobservable characteristics or attributes, and their manifestations, such as observed outcomes, responses, or performance. These models aim to establish a connection between the characteristics of items in an assessment tool, the individuals responding to those items, and the underlying trait being assessed. In IRT, it is assumed that the latent construct and the items of a measure are organized in an unobservable continuum. IRT focuses on establishing an individual’s position along this continuum [5], [6], [7], [8].

### Classical Test Theory

Classical Test Theory (CCT) preceded IRT but has the same objective; it aims to predict an individual’s latent trait based on an observed total score on an instrument. In CTT, the true score predicts the level of the latent variable and the observed score. The error is normally distributed with a mean of 0 and a standard deviation (SD) of 1. CTT is given by the following:

|  |  |
| --- | --- |
|  | (2. 1) |

Where:

* X = Total Score
* T = True Score
* e = Error

#### Classical Test Theory vs Item Response Theory

IRT offers many key advantages over CTT, offering item-level analysis, providing accurate insights into the item's latent traits without restrictions on the length and format of the test, and assuming item properties to be independent of the test takers. This nuanced approach enables the design of more personalized assessments and learning experiences.

### IRT Assumptions

IRT rests upon specific assumptions that shape its models and interpretations [6], [7], [8]. These include:

* **Monotonicity**: The probability of a correct response increases as the trait level (ability) increases.
* **One-Dimensionality**: A single dominant latent trait drives item responses.
* **Local Independence**: Item responses are independent and conditioned on a given ability level.
* **Invariance**: Item parameter estimates are independent of a specific sample of test-takers and their ability distribution.

#### Item Characteristic Curve

A graph with a blue line

Description automatically generated

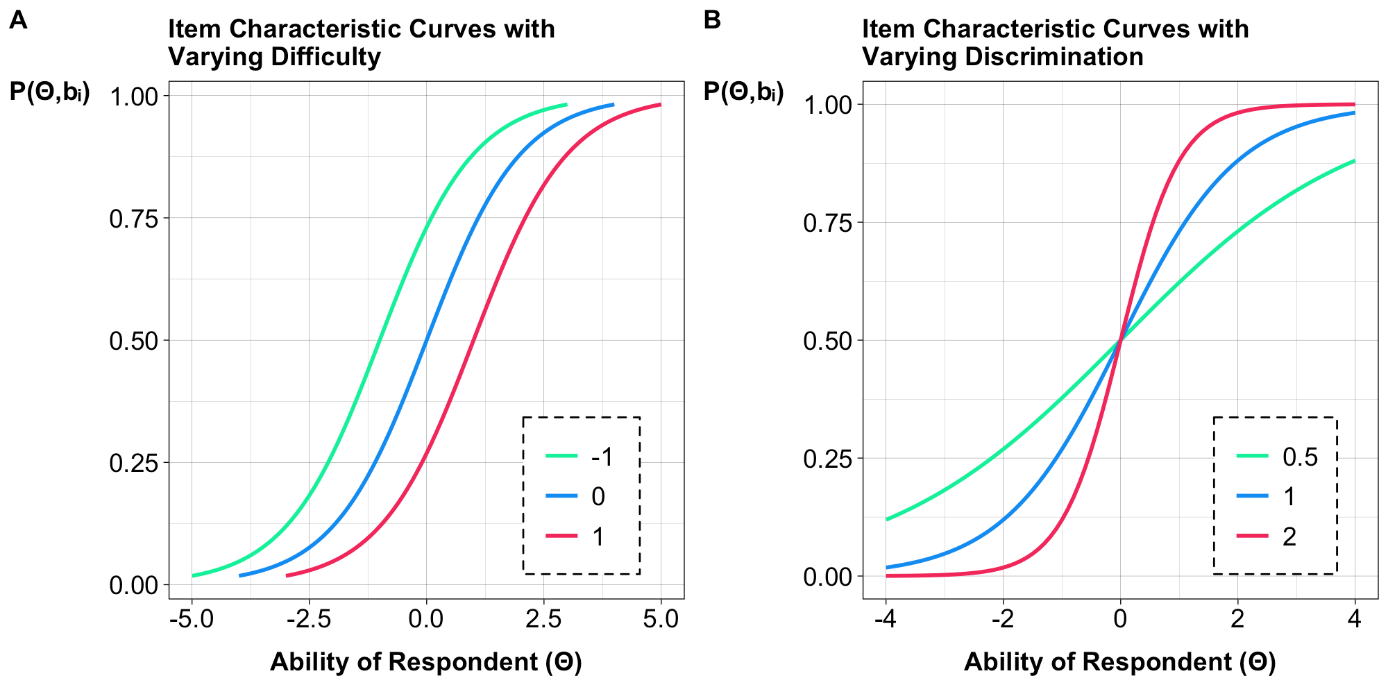
**Figure 2.1****: Graph representing the Item Response Function and the Item Characteristic Curve.** Plots the Respondents Ability vs the Probability of endorsing a correct response.

If the IRT assumptions hold, the differences in observing the correct responses between respondents will be due to variations in their latent trait. IRT models predict respondents’ answers based on their position on the latent trait continuum and item parameters. The Item Response Function (IRF) characterizes this relationship, implying that each response reveals information about the respondent’s ability level. A higher ability increases the probability of a correct response, visualized in the Item Characteristic Curve (ICC) shown in **Figure 2.1**.

### Item Parameters

People with different abilities will find themselves at different positions on the latent construct’s continuum, determined by the sample of respondents and item parameters. These parameters include [6], [7], [8]:

* **Item Difficulty**: Determines where on the ability scale an item functions. A more difficult item has its median probability of correct response at a higher ability level, which is shifted to the right on an ICC (**Figure 2.2.B**).
* **Item Discrimination**: Indicates how well an item distinguishes between individuals with different ability levels. Items with high discrimination accurately measure the latent trait in **Figure 2.2.B**.
* **Guessing**: Accounts for guessing on an item.



**Figure 2.2:** **Graphs representing the transformations to ICC curve with changes to Item parameters.** Graph A represents changes with difficulty. Graph B represents changes with varying discrimination.

### IRT Model Types

IRT encompasses a range of models designed to uncover deeper relationships between the latent trait and their observable outcomes. To review the simplicity and relevancy of IRT as a model, this thesis uses the 2-Parameter Logistic Model (2-PL).

#### 2-Parameter Logistic Model

The 2-Parameter Logistic Model (2-PL) model predicts the probability of a successful answer using two parameters, (Difficulty and Discrimination) [6], [7], [8]. The 2-PL model may be mathematically described as:

Where:

**Figure 2.3:** **Item Characteristic Curves for 2-Parameter Logistical Model with Varying Item Parameters.** The Green curve acts as the base curve for this plot. The Blue curve shows the effect of increasing the Discrimination parameter. The Red Curve shows the effect of increasing the Difficulty parameter.

* = Ability
* = Discrimination Parameter
* = Difficulty Parameter

A graph of different colored lines

Description automatically generatedThe discrimination parameter is allowed to vary between items. Thus, the ICC of different items can intersect and have different slopes. The steeper the slope, the higher the discrimination of the item. The Item Information function shows the amount of information each item provides [6], [7], [8]. For a 2-PL model, it is mathematically described as:

Where:

* = Ability
* = Difficulty Parameter
* = Correct Response
* = Incorrect Response

Information is derived from the but it’s weighted by the squared discrimination parameter. Consequently, items with higher discrimination provide more information, leading to potentially distinct ICC.

## Literature Review

The use of ALS has been investigated previously due to its high importance and growing popularity in supplementing classroom teaching. To build a good ALS for today’s learning needs, it is important to review the research already done to build on the successes of the past. In this section, we will review some past ALS implementations.

The paper by Pulido Vega, Y. L. et al. [9] used IRT to develop adaptive assessments for an introductory OOP course. Assessment tests were conducted to gather student responses for estimating item difficulty parameters. Proficiency levels were then identified from these responses. Through iterative calibration, the IRT model estimated student competence levels and guided item selection based on maximizing information function and considering item difficulty relative to students' competencies. This approach facilitated the creation of adaptive assessments tailored to individual student abilities, thereby enhancing learning effectiveness and efficiency.

In the paper by Yarandi, Jahankhani, and Tawil [10], the application of IRT within the adaptive e-learning Decision Support System (DSS) is demonstrated. IRT is utilized to analyze learners' responses to test items, allowing for the accurate assessment of their abilities. By leveraging IRT, the system dynamically generates personalized learning paths tailored to individual learners' proficiency levels, preferences, and learning histories. This approach facilitates improved learning effectiveness and learner satisfaction. Additionally, the paper highlights the importance of IRT in adaptive e-learning systems for providing accurate and personalized learning experiences.

A completely different approach to ALS can be seen in the study by Yuxin Huang et al. [11]. They propose leveraging smart techniques like knowledge graphs, which visually represent and organize knowledge, and Self-Organizing Map (SOM) clustering, an ML method for grouping similar data points, to better understand student needs and personalize learning experiences, accordingly, aiming to enhance learning outcomes for diverse learner profiles.

# Proposed System

## Design

This report proposes a new algorithm that combines IRT with commonly known mathematical approaches to build the ML model. In this section, we will be discussing the specific design and details of this model and its functionality.

### Data Collected

This system utilized a database to store and update values used within the ML model. The following tables highlight these important fields and tables.

***Table 3.1 Structure of the studentKnowledge relation****. This relation is critical for the ML model. Only relevant columns are shown.*

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **studentID** | Individual Student ID |
| **topicID** | Individual Question Topic ID |
| **categoryID** | Individual Question Category ID |
| **mastery** | Mastery Level for Topic– Category pair |
| **difficultyOffset** | Difficulty Offset calculated using IRT |

**Table 3.2:** **Structure of question relation.** This relation also holds critical information about each question used for the ML model. Only relevant columns are shown.

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **assignedDifficulty** | Pre-Assigned Question Difficulty |
| **modifiedDifficulty** | Personalized Question Difficulty based on overall performance |

### Question Categories

In our proposed model, we organized questions for each topic based on Bloom's Taxonomy levels: Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. This categorization system allows students to systematically develop their understanding by encompassing a range of cognitive skills, from simple recall to advanced critical thinking, thereby permitting deeper learning and analytical thinking.

### Mastery

The mastery level tracks a student's competency in each question topic–category pair. Initially, all students' mastery levels are set to 0.5, reflecting a basic understanding of the subject matter. After each question attempt, the mastery level for the relevant topic and category is adjusted. A correct answer increases the mastery level by 0.05, while an incorrect answer decreases it by 0.04. These values were determined through trial and error to ensure student mastery gradually increases or decreases as they attempt more questions. Additionally, to handle edge cases, such as a student getting all questions within a topic right or wrong, mastery levels are capped within the range of -1.7 and 1.7.

### Modified (Question) Difficulty

The Modified (Question) Difficulty is the effective difficulty assigned to each question. This is initially initialized to the “assignedDifficulty” of the question. This value is updated at each attempt based on the overall performance of students on the question. The algorithm utilizes a sigmoid function to modulate the adjustment of question difficulty, represented as

This function dynamically alters its output based on the number of attempts made on a question. As attempts increase, the sigmoid output approaches 1, favouring the cumulative student performance on the questions, while fewer attempts yield values closer to 0, favouring the pre-assigned difficulty value. The adjustment of question difficulty is governed by the equation:

=

This equation ensures a balanced adaptation by dividing the old difficulty level by the value, thereby increasing or decreasing difficulty based on attempt history. Additionally, it integrates the fraction of correct responses, allowing for consideration of student performance alongside attempt frequency. Through this mechanism, the algorithm dynamically tailors question difficulty to optimize learning experiences for students.

These equations can return values in the range of -∞ to ∞. For this report, this value was limited to the range of -10 to +10, inclusive. Constraining the difficulty adjustment values between -10 and +10 provides practical bounds within which the adaptive learning system can operate effectively. While the sigmoid function inherently outputs values between 0 and 1, constraining the difficulty adjustments ensures that questions remain appropriately challenging for students while maintaining system stability. By setting these boundaries, extreme adjustments are prevented, promoting smoother learning experiences and more predictable learning outcomes aligned with the system's objectives and content characteristics.

### Difficulty Offset

The backbone of this ML model lies in the 2-Parameter Logistic Item Response Theory (2-PL IRT) model. Traditionally, IRT has been applied to individual questions to gauge their difficulty levels. However, for a quiz platform aimed at first-year students, this approach can be exceedingly resource-intensive as both the question pool and user base expand. To mitigate this, we adopted an alternative strategy: rather than applying IRT to individual questions, we implemented it on specific question topics, generating "difficultyOffset" values for each category of questions within the topic. This ensured personalized difficulty adjustments for each student while minimizing resource consumption. Additionally, the 2-PL model accounted for discriminant scores, evaluating the quality of questions in determining their difficulty levels. These values were updated for each student after completing a full quiz (comprising 20 questions). Theoretically, the difficulty returned by the IRT model could range from -∞ to ∞. To maintain reasonable difficulty levels, we capped this range between -2.5 and +2.5.

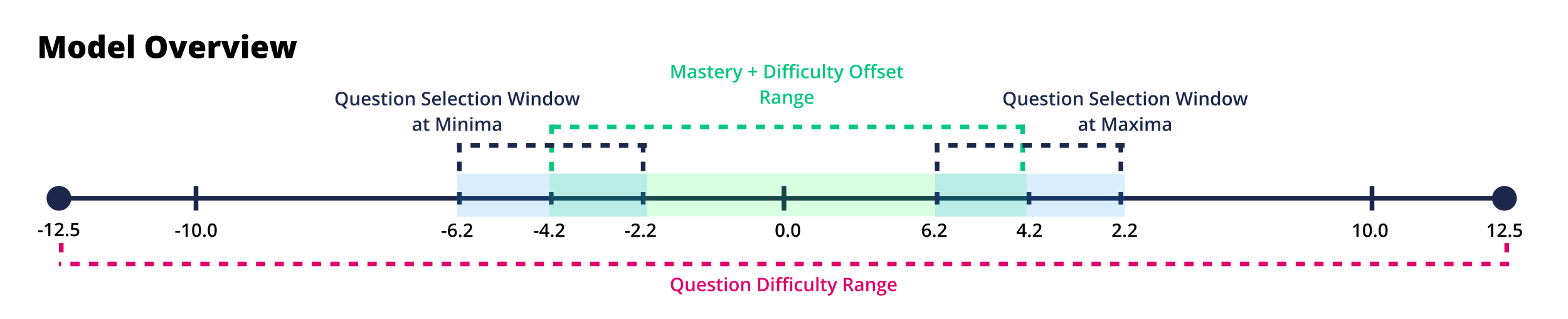
### Bounding Ranges

The decision to impose ranges on the variables of the ML model collectively ensures both mathematical stability and practicality within the adaptive learning system. By constraining the mastery level within -1.7 and 1.7, the modified question difficulty within -10 and +10, and the difficulty offset within -2.5 and +2.5, the model maintains a consistent framework where student competency, question difficulty adjustments, and topic-specific difficulty offsets operate within reasonable bounds. This collective constraint approach prevents extreme values that could disrupt the learning process while promoting smoother learning experiences aligned with the system's objectives.

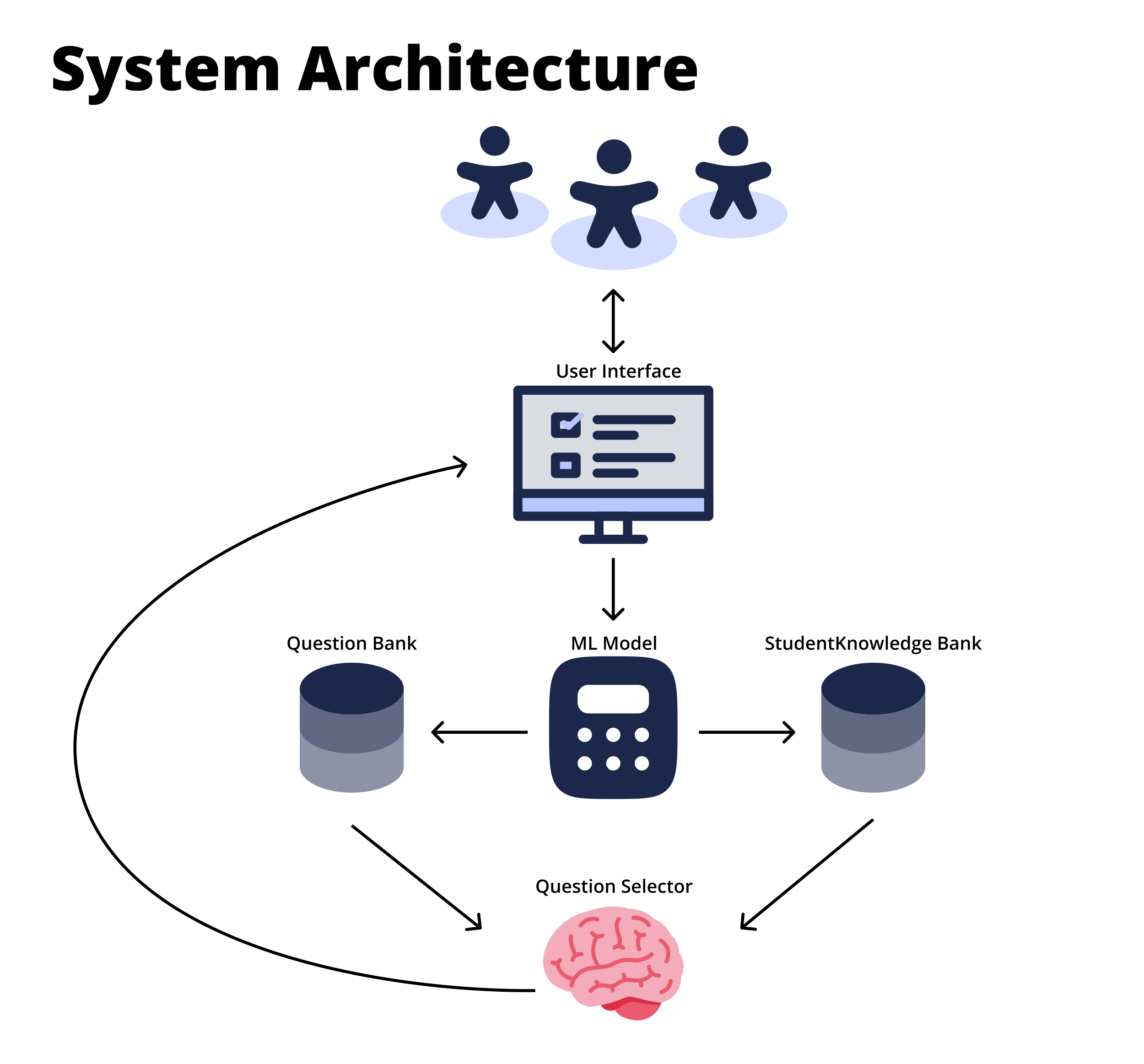
### The Model

Now that we have identified the individual aspects of the model, let us bring it all together to explain how the question selection process is personalized for each student. In conceptualizing the model's operation, envision a scale where each question, irrespective of its topic, possesses a potential difficulty range spanning from -10 to +10. However, to determine a question's final difficulty, we must incorporate the difficultyOffset specific to its topic and category. This difficultyOffset value, falling within the range of -2.5 to +2.5, effectively extends the maximum difficulty range for all questions to -12.5 and +12.5. Now, let's consider the selection of questions. Our objective is to tailor question selection to the individual student's proficiency level. To achieve this, we leverage the student's mastery level, constrained between -1.7 and +1.7. When combined with the difficultyOffset, which ranges from -2.5 to +2.5, the effective range for mastery becomes -4.2 and +4.2. Subsequently, we introduce a variable , initially set to 0.2, to create bounds around the mastery + difficultyOffset score. These bounds are calculated as the mastery + difficultyOffset - 0.2 for the lower bound and mastery + difficultyOffset + 0.2 for the upper bound. This bracketing mechanism ensures that students with higher mastery levels receive harder questions, while those with lower levels encounter easier ones.

To address the potential scarcity of questions within the desired difficulty bracket, the value of is iteratively incremented by 0.05 each time the required number of questions is not obtained. This adaptive adjustment guarantees that students receive an appropriate mix of questions, even if the ideal difficulty level is unavailable.



**Figure 3.1: Model Overview visualization**. This figure shows the different scales and the default window for question selection (blue).



**Figure 3.2: DFD-1 System Architecture Diagram.** This figure shows the data flow for the ML model.

## Implementation

This thesis used Next.js version 14.1.4 and MySQL Version 15.1 Distribution 10.3.39-MariaDB on a Linux (x86\_64) platform with Readline 5.1. Additionally, a virtual environment running Python 3.9 was set up to execute the IRT model using R through rpy2 version 3.5.13. The R analysis environment operated on version 4.3.3 ("Angel Food Cake"), running on a 64-bit x86\_64-redhat-linux-gnu platform.

The source code is hosted on GitHub and can be accessed at <https://github.com/SatanshuMishra/honours-2023-adaptive-learning-tool>.

# Case Study

## Study Setting

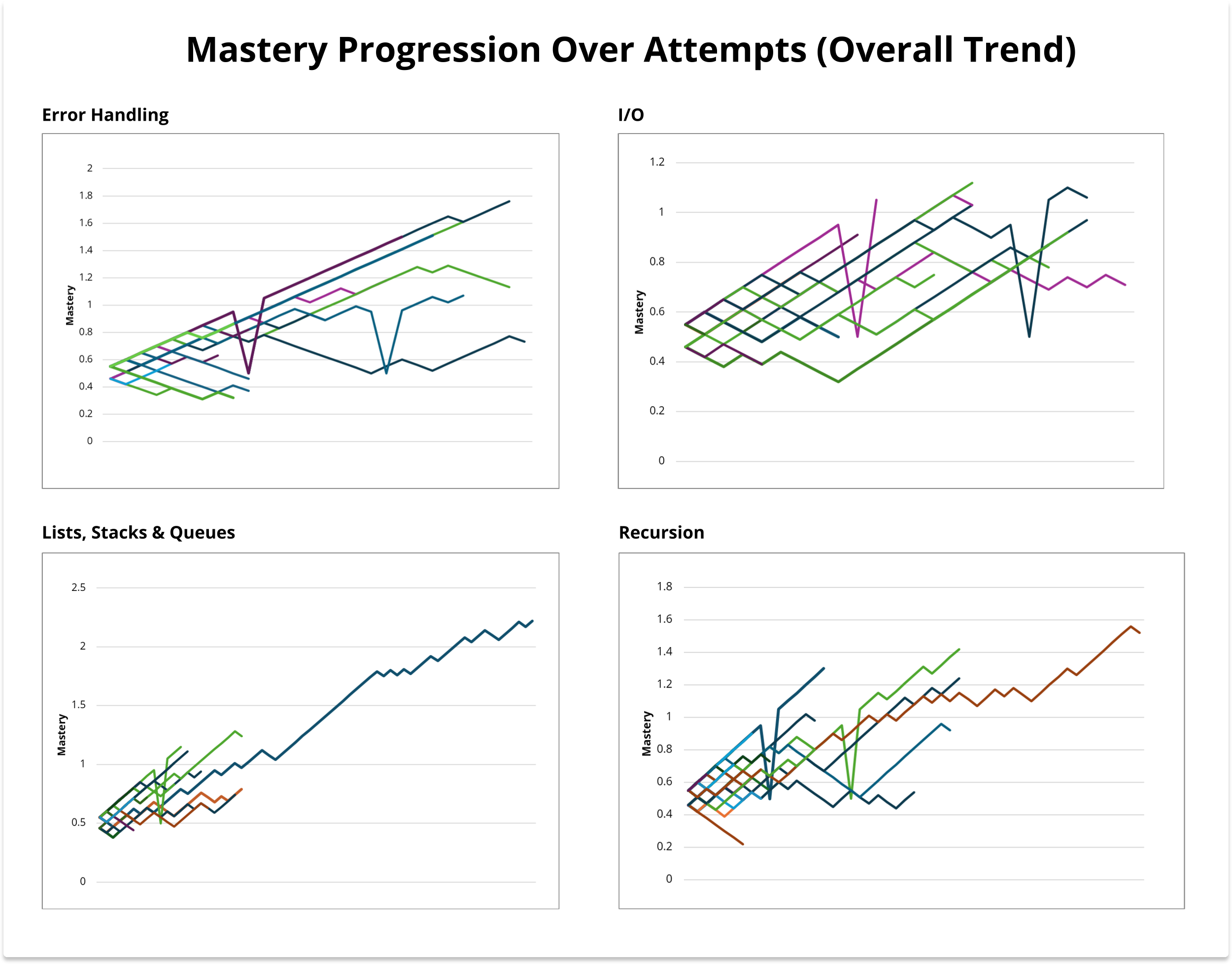
This system was implemented within a first-year introductory Computer Science course, focusing on fundamental programming concepts through the Java language. Four distinct question topics, including Error Handling, I/O, Recursion, and Lists, Queues & Stacks as one inclusive topic were developed for this system. Quizzes comprised of 20 questions each. Students were restricted to selecting questions from a single topic per quiz.

Students in the course were provided with a unique code enabling access to the adaptive learning tool. Upon logging in, students gained immediate access to all four topics, allowing them to initiate quizzes on any topic of their choice. Notably, there were no limitations on the number of quizzes students could undertake. A total of 137 students were enrolled in the course. Of these, 18 students provided consent to utilize their data from the adaptive learning system for this study. 63 students actively engaged with the system, completing at least one quiz.

Finally, to evaluate student satisfaction with the uLearn system, a questionnaire was designed to measure whether the proposed learning system improved student understanding and performance within the course and collect their opinions of ALS and their possible adaptation to future courses.

## Results & Findings

The data collected in this study provides initial insights into the performance of the adaptive learning system, although the findings are not yet sufficient to draw definitive conclusions. Nevertheless, the preliminary analysis reveals some interesting trends.



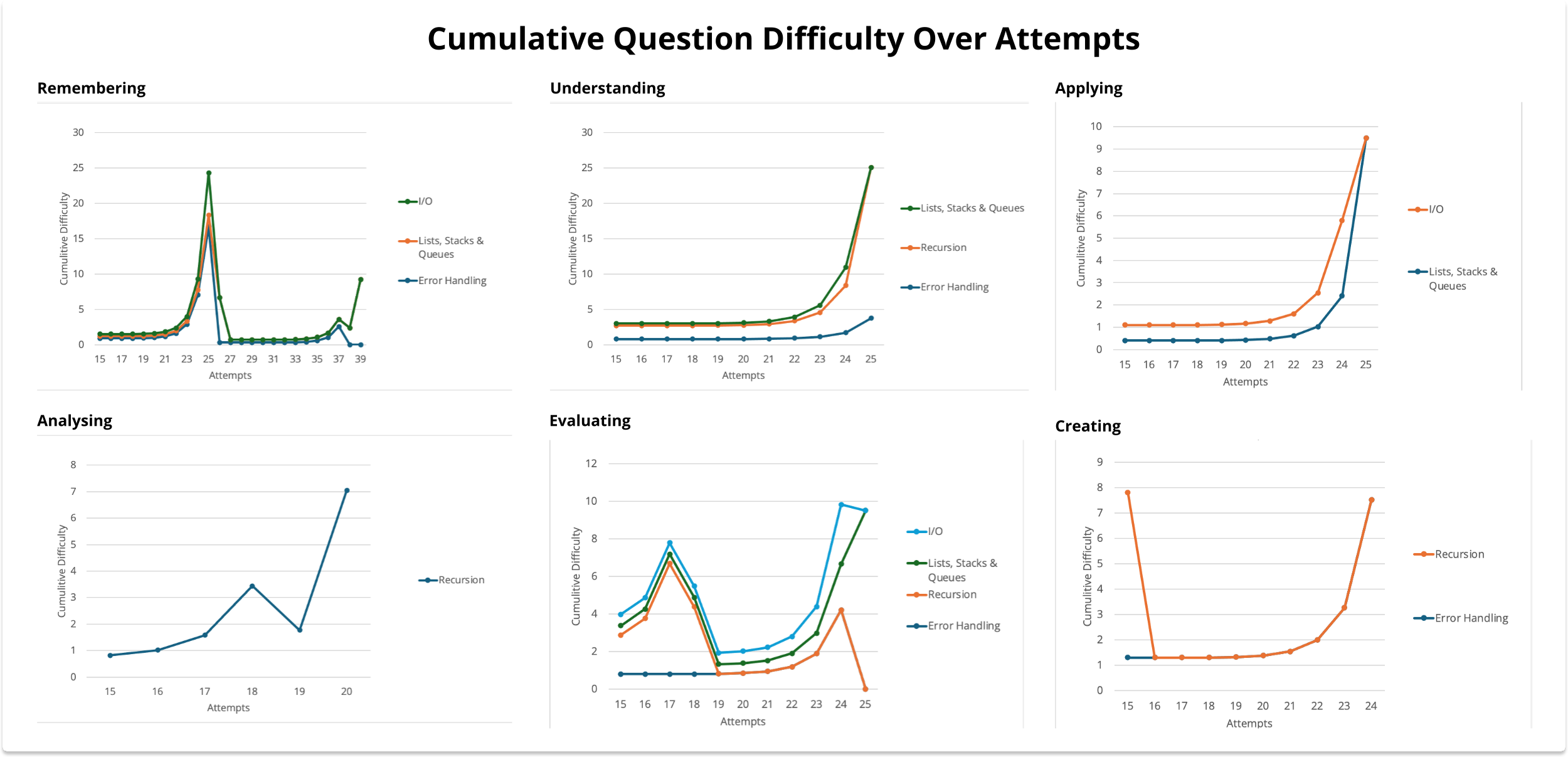
**Figure 4.1: Mastery Progression over Attempts (Overall Trend)**. This figure shows the over all trend for mastery all the categories combined into a single graph for each topic.

An upward trend was consistently observed when plotting mastery level against the number of attempts (**Figure 4.1**) across all categories and topics within the system. This suggests a general improvement in students' mastery as they used the learning tool over multiple attempts.



**Figure 4.2: Difficulty Offset Progression Over Attempts (Overall Trend**). This figure shows the over all trend for difficulty offset all the categories combined into a single graph for each topic.

Conversely, an overall downward trend was observed in the progression of difficulty offsets over the number of attempts for all categories and topics. This indicates a trend towards decreasing question difficulty as students continued to engage with the system. For a more detailed analysis of the trends observed in both mastery and difficulty offset, please refer to the individual graphs provided in the appendix 8.1 – 8.2.



**Figure 4.3: Cumulative Question Difficulty over Attempts**. This figure shows how the question difficulties across all topics as attempts increase focusing on the period after 14 attempts.

Observation of cumulative question difficulty after 14 attempts (**Figure 4.2**) revealed an interesting trend. At this point, the sigmoid function used to calculate new question difficulties begins to favour the cumulative performance of all students over the pre-assigned difficulty of the question. A sharp spike followed by a decline in question difficulty was observed.

# Discussion

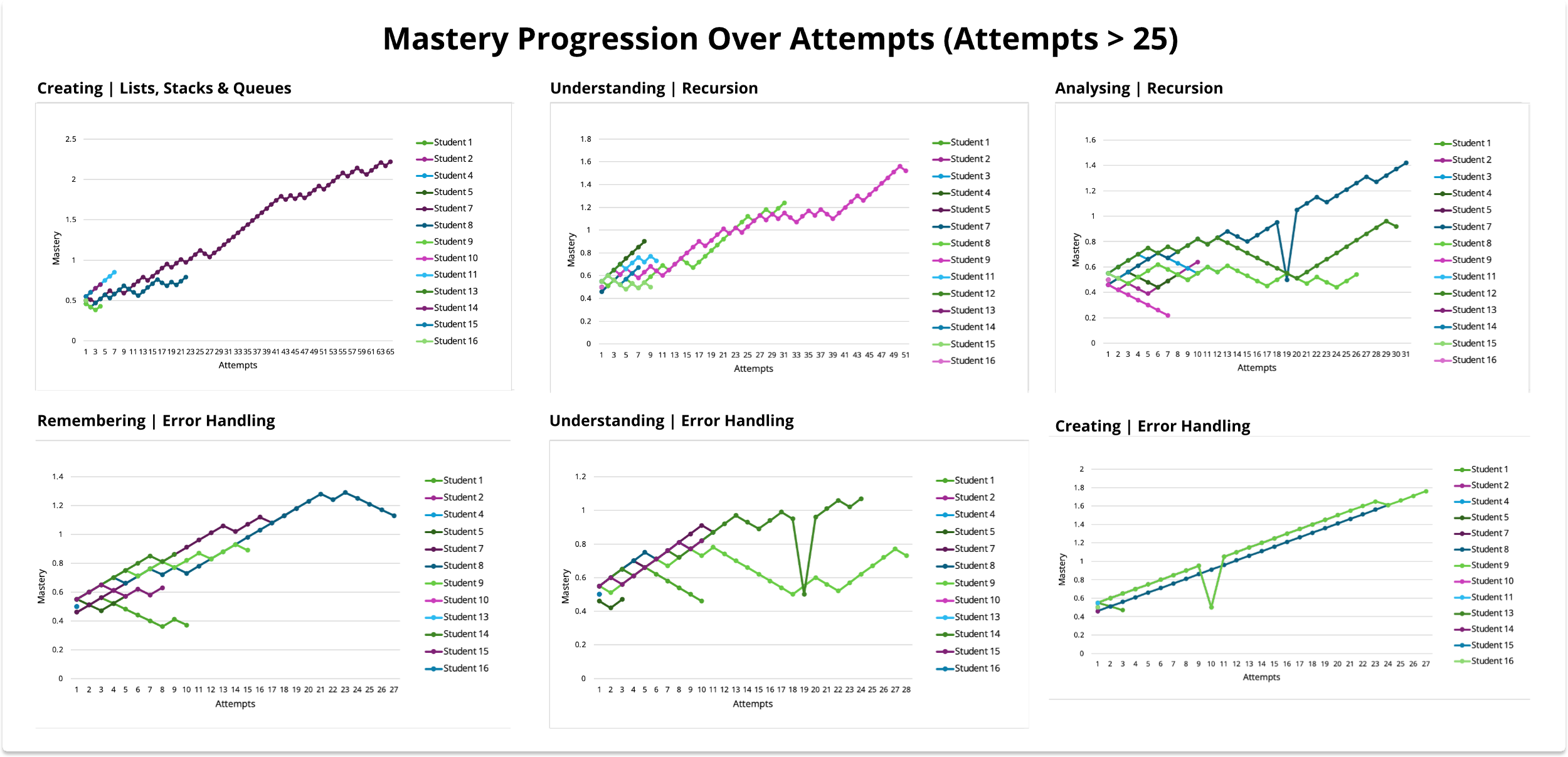
## Significance of the System

Our adaptive learning system has significance for several reasons. First and foremost, the system personalizes the questions students get to tailor to their individual proficiency levels, optimizing their learning experience. Additionally, our system is unique from traditional implementations of IRT. Our system utilizes IRT to cumulatively analyse questions across a specific topic rather than on individual questions as seen, for example, in the implementation by Pulido Vega, Y. L. et al. By aggregating IRT calculations at the topic level, we effectively personalize question difficulty assessment while optimizing the use of computational resources, making the system scalable.

Moreover, our system offers real-time feedback to students, providing immediate insights into the correctness of their responses. This feature allows students to grasp the rationale behind correct and incorrect answers, allowing a deeper understanding and improving their ability to tackle similar questions in the future.

Furthermore, as a proof of concept, our system represents just the beginning of what can be achieved. There is much room for refinement and enhancement, ongoing improvements such as fine-tuning parameters or using a modified IRT model to further elevate the system's performance and effectiveness. This potential for the improvement underscores the system's adaptability and its capacity to improve in response to changing educational needs and technological advancements the present day.

## Analyzing Findings



**Figure 5.1: Mastery Progression (Attempts > 25).** This graph shows the mastery progression for the topic-category pairs with the most attempts to highlight the preliminary findings from the results.

Analyzing topic-category pairings with more than 25 attempts (**Figure 5.1**) reveals a notable trend: as the number of attempts increases, students' overall mastery also tends to increase. This suggests that repeated use of the system leads to improved accuracy in answering questions within those topics, indicating enhanced understanding over time.

Similarly, a slight negative trend is observed in the progression of difficulty offsets (**Figure 4.2**). This indicates that questions became easier for students as they continued to attempt them within those topics, potentially reflecting a deeper grasp of the material with each subsequent attempt.

Examining cumulative question difficulty (**Figure 4.2**), we note a distinctive pattern: a spike followed by a subsequent drop. This pattern suggests an initial increase in question difficulty, followed by a decrease as students' accuracy in responses improves with repeated attempts. While these findings alone do not provide conclusive evidence of the system's effectiveness, they do indicate promising signs of improved student performance and understanding.

These quantitative findings are further supported by qualitative feedback collected anonymously from 55 students who utilized the uLearn system. As depicted in **Figure 5.2** a significant majority of respondents reported improvements in their understanding of concepts and increased confidence in their CS skills after using the learning tool. Additionally, **Figure 5.3** illustrates that most students found the tool effective in identifying and addressing their weaknesses in the course, leading to improvements in their overall performance.

In terms of student opinions on the learning tool (**Figure 5.3**), the majority expressed a desire to continue using adaptive learning tools in their future studies. Many also highlighted the tool's role in fostering self-paced and self-directed learning, indicating a preference for incorporating similar tools into future CS courses. Moreover, students generally viewed the learning tool as more effective than traditional studying methods and found it efficient in helping them achieve their learning goals.

**Figure 5.2: System Effectiveness and Future Preference**. This figure shows the qualitative feedback received from students after they used the tool highlighting their opinion of the tool’s effectiveness and their future preferences with Adaptive Learning tools.

**Figure 5.3: User Experience and Engagement**. This figure shows the qualitative feedback received from students after they used the tool highlighting their opinion using the tool and its impact.

## Limitations of the System

While our adaptive learning system has demonstrated numerous strengths and shown preliminary effectiveness in several areas, it is important to recognize that no system is without limitations. In this section, we explore some of the challenges and constraints encountered during the implementation and evaluation of our system.

The first limitation was that there weren’t enough questions developed for each topic. For a quiz system to be effective, there need to be many questions to ensure students aren’t getting the same questions over and over. Additionally, with a reduced question load, student is often able to remember the answers to specific questions this can give the false impression of improved understanding when realistically students are simply able to answer questions correctly since they know the correct answer.

Given that participation in the study was incentivized with bonus marks and did not mandate accurate quiz completion, there may have been instances where students rushed through quizzes, selecting answers indiscriminately. This guessing of responses may have led to inaccurate response data leading to false trends.

The question selection algorithm was a major part of the system. It is used to select the questions given to a student based on the question difficulty and the student’s personal mastery of the topic. This algorithm wasn’t fully fine-tuned to give the most optimized results. This resulted in students getting the same question repeatedly in multiple successive quizzes.

Finally, the data collected from this study is not sufficient to draw robust conclusions. More data will need to be collected to definitively evaluate the effectiveness of the system.

# Conclusion & Future Work

In this report, we developed the uLearn adaptive learning system to investigate the impact of adaptive learning tools in education, aiming to answer key research questions regarding current trends in adaptive learning systems in CS education, the effectiveness of the learning tool in enhancing learners' understanding and performance in a first-year CS course, and future recommendations for advancing CS1 education through adaptive learning systems. The system utilizes the 2-PL IRT model to personalize question selection based on question difficulty and student mastery levels, and it was tested in a first-year computer programming class across four distinct question topics: Error Handling, I/O, Recursion, and Lists, Queues & Stacks.

The findings provide promising results of the uLearn system in enhancing student learning outcomes and fostering deeper comprehension of complex programming concepts. Trends in mastery level progression, difficulty offset progression, and cumulative question difficulty highlight the positive impact of the system on student engagement, confidence, and overall learning experience.

This report serves as a proof-of-concept for the continued development of adaptive learning systems utilizing the modified IRT-based ML approach. While promising, further validation and confirmation of the findings are necessary through the collection of larger datasets and longer study periods. Additionally, fine-tuning of question selection parameters is essential to optimize question selection and avoid duplicate questions in successive quizzes. Another avenue of exploration could be the integration of additional ML algorithms with adaptive learning systems to enhance assessment accuracy and personalize learning experiences, as seen in the paper by T. Merembayev et al. [12]. Finally, incorporating secondary incentives such as achievements or rankings within the system could further encourage student involvement and utilization of the tool beyond the incentive of bonus marks, ultimately enhancing the learning experience.

# References

[1] L. J. Sax, K. J. Lehman, and C. Zavala, “Examining the Enrollment Growth: Non-CS Majors in CS1 Courses,” in *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, Seattle Washington USA: ACM, Mar. 2017, pp. 513–518. doi: 10.1145/3017680.3017781.

[2] A. Mohamed, “Designing a CS1 Programming Course for a Mixed-Ability Class,” in *Proceedings of the Western Canadian Conference on Computing Education*, Calgary AB Canada: ACM, May 2019, pp. 1–6. doi: 10.1145/3314994.3325084.

[3] *NMC horizon report... Higher Education Edition*. Austin (Texas): NMC, 2004.

[4] H. Peng, S. Ma, and J. M. Spector, “Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment,” *Smart Learn. Environ.*, vol. 6, no. 1, p. 9, Dec. 2019, doi: 10.1186/s40561-019-0089-y.

[5] S. E. Embretson and S. P. Reise, *Item Response Theory for Psychologists*. Hoboken: Taylor and Francis, 2013.

[6] W. J. van der Linden and R. K. Hambleton, *Handbook of modern item response theory*. New York: Springer, 1997.

[7] R. K. Hambleton and H. Swaminathan, *Item Response Theory: Principles and Applications*. Dordrecht: Springer Netherlands, 2013.

[8] X. Li, Z. Wang, X. Wu, Y. Li, and H. Dong, “The design of adaptive test paper composition algorithm based on the item response theory,” in *2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference*, Chongqing, China: IEEE, Aug. 2011, pp. 157–159. doi: 10.1109/ITAIC.2011.6030299.

[9] Y. L. P. Vega, G. M. F. Nieto, S. M. Baldiris, and J. C. Guevara Bolaños, “Application of item response theory (IRT) for the generation of adaptive assessments in an introductory course on object-oriented programming,” in *2012 Frontiers in Education Conference Proceedings*, Seattle, WA, USA: IEEE, Oct. 2012, pp. 1–4. doi: 10.1109/FIE.2012.6462377.

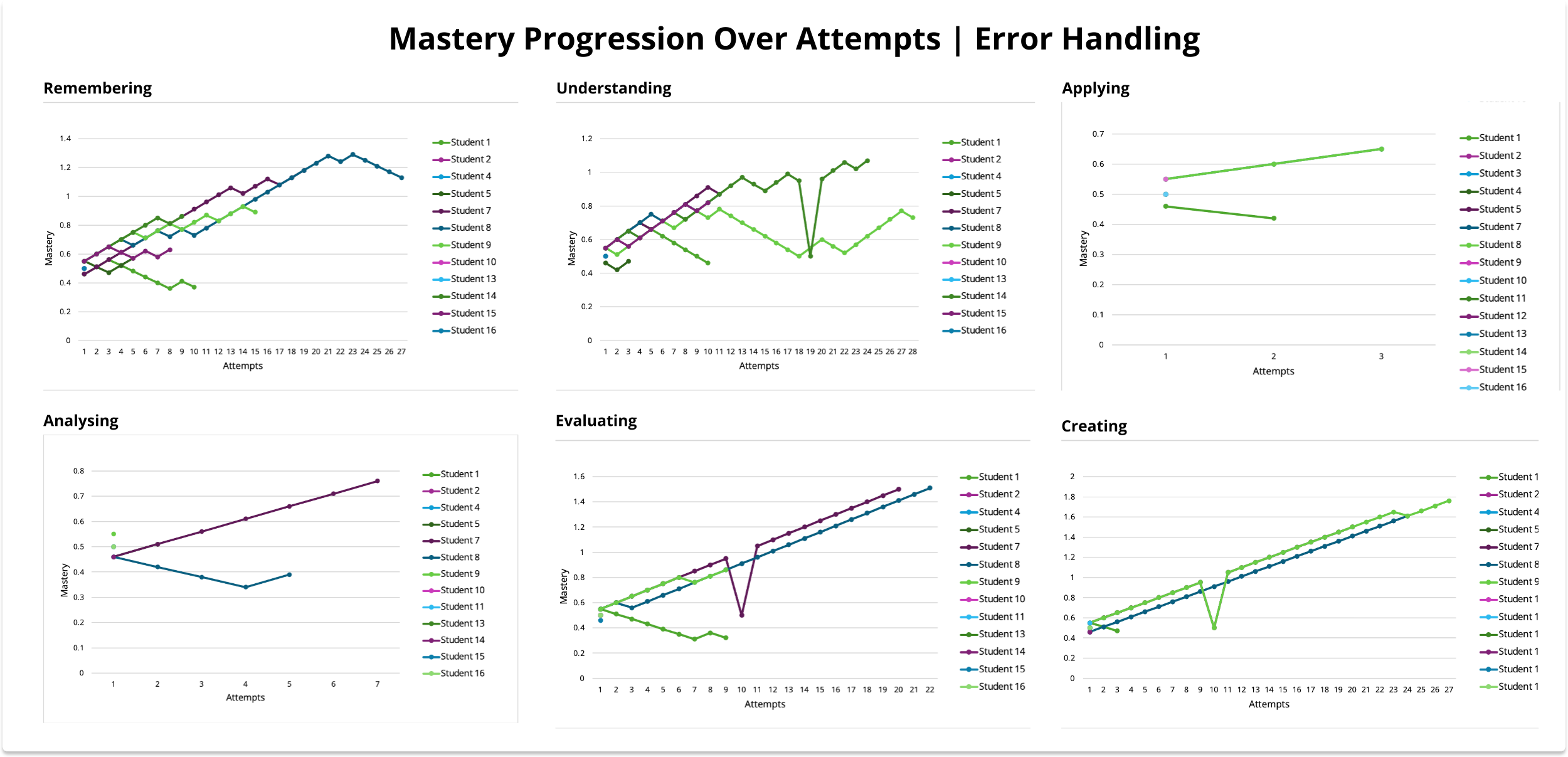
[10] M. Yarandi, H. Jahankhani, and A.-R. H. Tawil, “An adaptive e-learning Decision support system,” in *2012 15th International Conference on Interactive Collaborative Learning (ICL)*, Villach, Austria: IEEE, Sep. 2012, pp. 1–5. doi: 10.1109/ICL.2012.6402141.

[11] Y. Huang, Y. Fan, Z. Zhuang, and M. Tong, “The Research on Dynamical Model of Adaptive Learning System,” in *2023 5th International Conference on Computer Science and Technologies in Education (CSTE)*, Xi’an, China: IEEE, Apr. 2023, pp. 56–60. doi: 10.1109/CSTE59648.2023.00017.

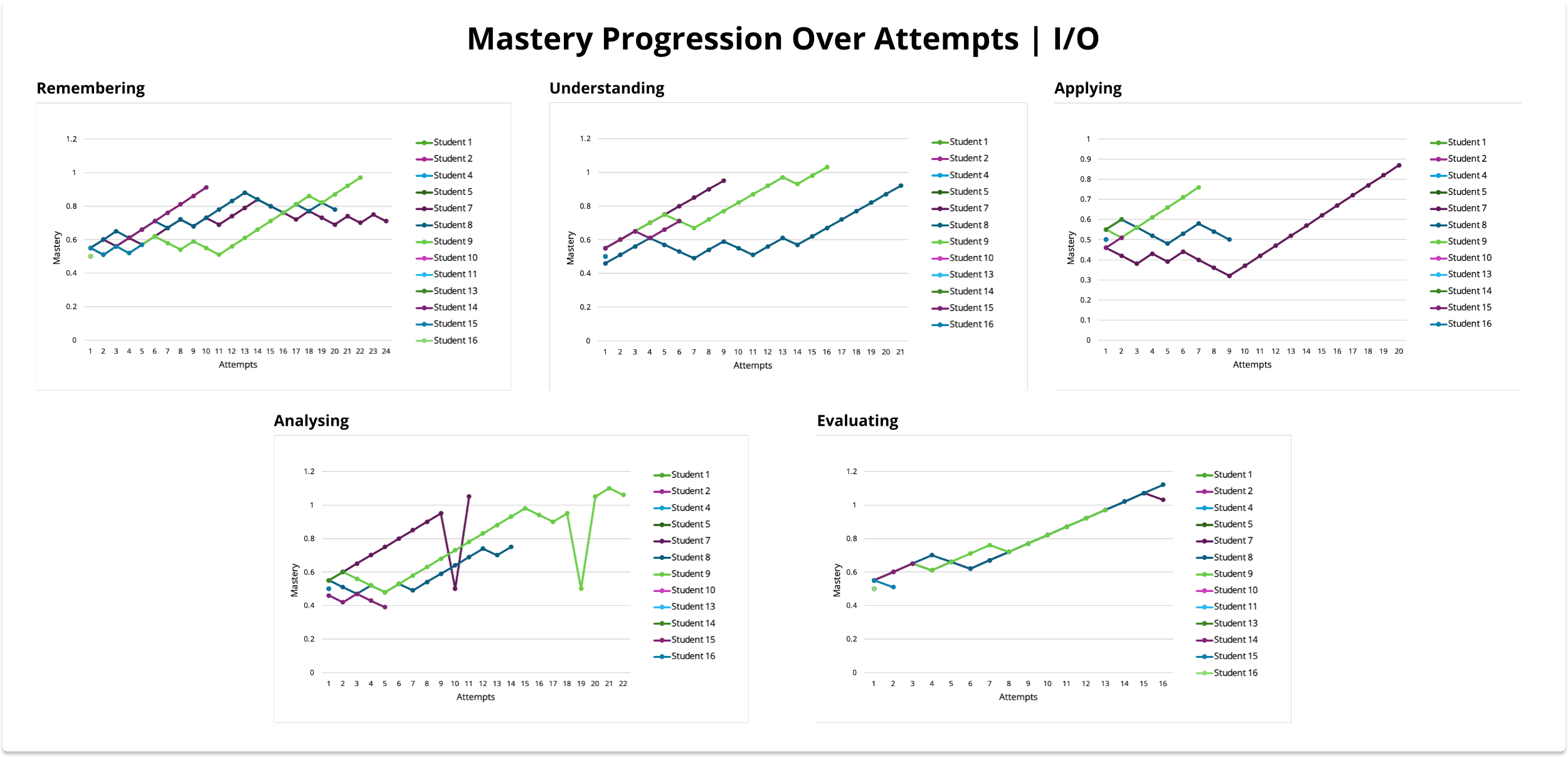
[12] M. T, A. S, and K. K, “Using item response theory in machine learning algorithms for student response data,” in *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Nur-Sultan, Kazakhstan: IEEE, Apr. 2021, pp. 1–5. doi: 10.1109/SIST50301.2021.9465896.

# Appendices

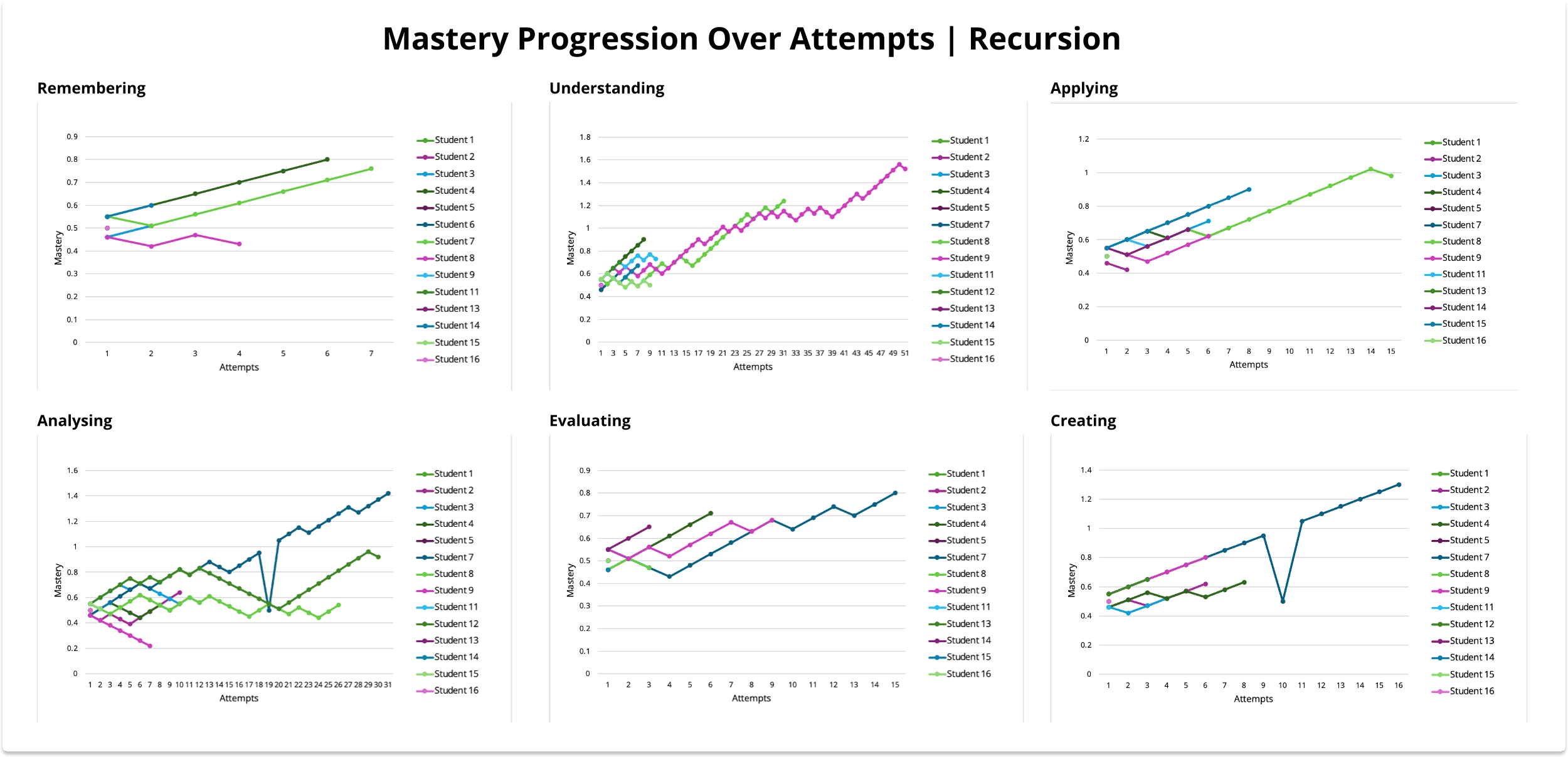
## Individual Mastery Progression



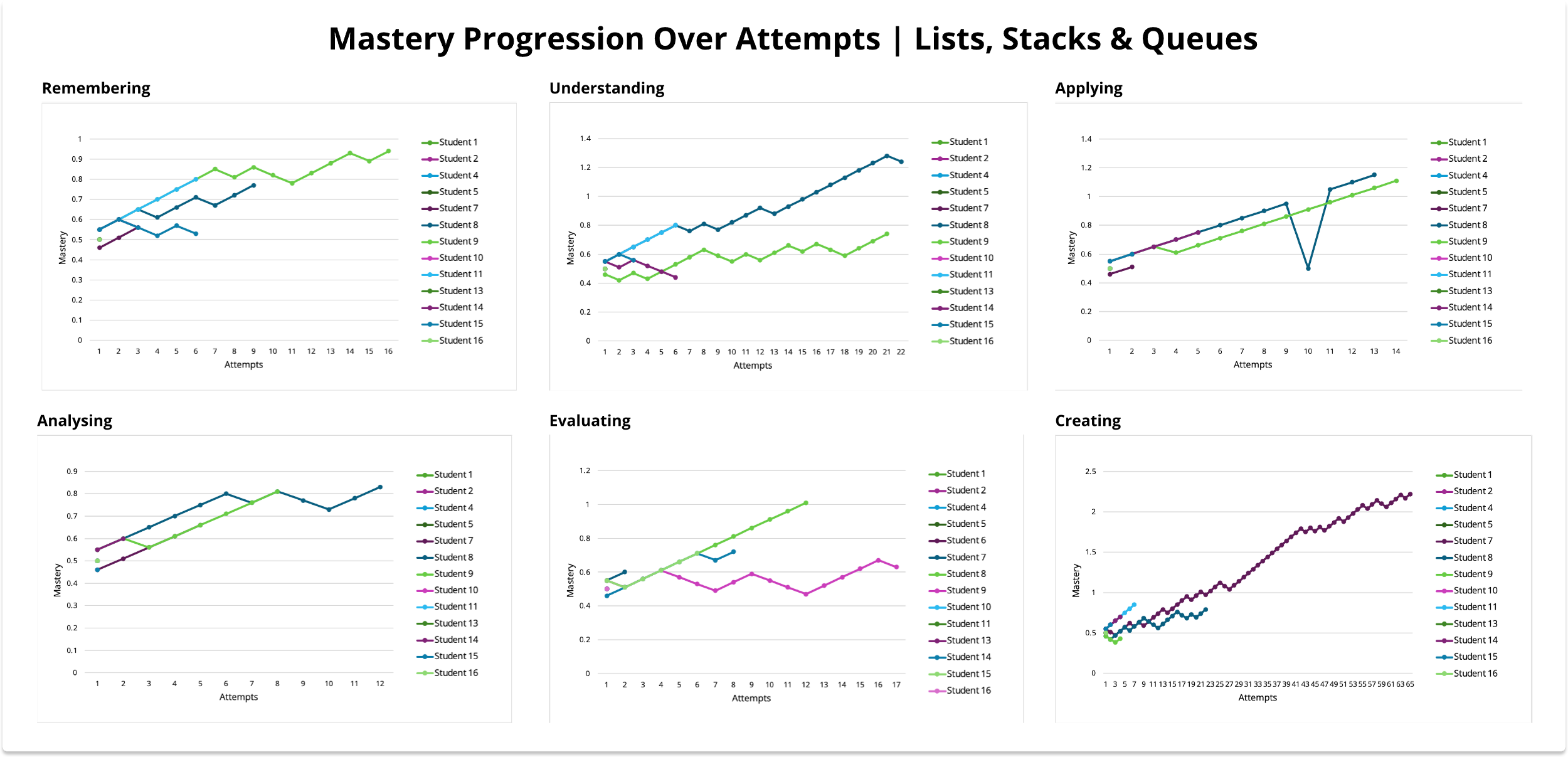
**Figure 8.1: Mastery Progression over Attempts for each Category in Error Handling.**



**Figure 8.2: Mastery Progression over Attempts for each Category in I/O.**



**Figure 8.3: Mastery Progression over Attempts for each Category in Recursion.**

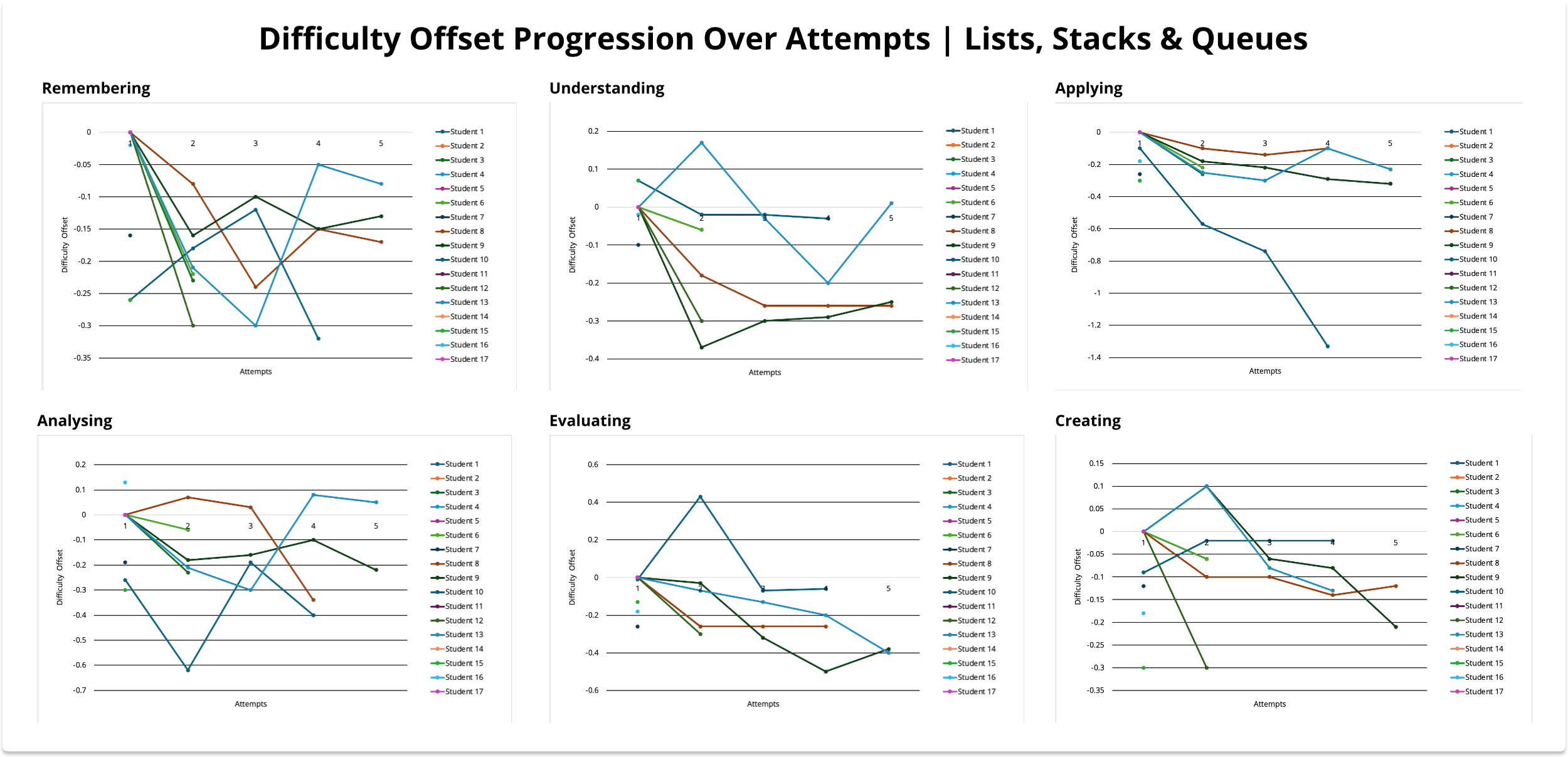


**Figure 8.4: Mastery Progression over Attempts for each Category in Lists, Stacks & Queues.**

## Individual Difficulty Offset Progression



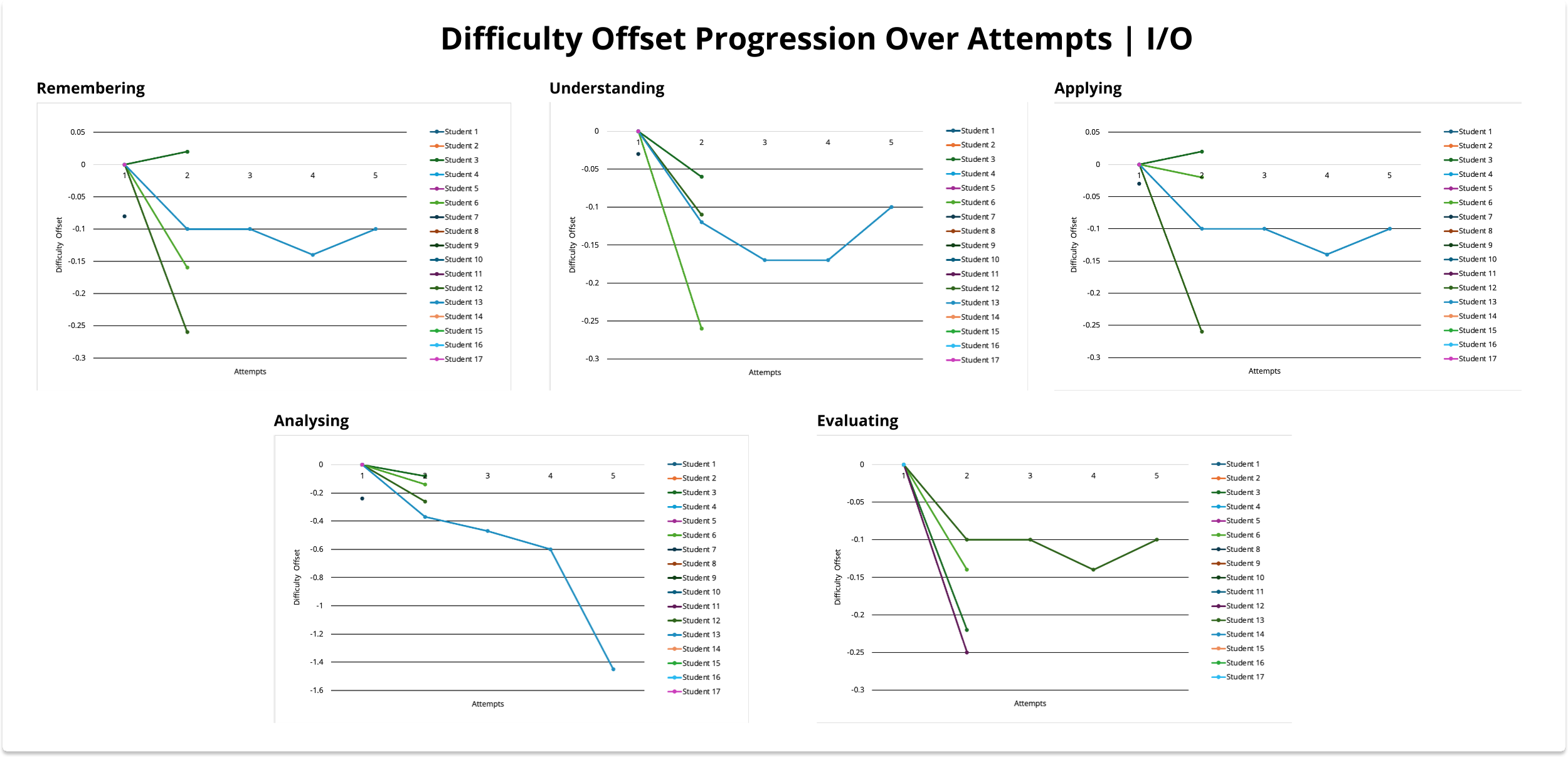
**Figure 8.5: Difficulty Offset Progression over Attempts for each Category in Error Handling.**



**Figure 8.6: Difficulty Offset Progression over Attempts for each Category in Lists, Stacks & Queues.**



**Figure 8.7: Difficulty Offset Progression over Attempts for each Category in Recursion.**



**Figure 8.8: Difficulty Offset Progression over Attempts for each Category in I/O.**