APPLICATION OF TEM RESPONSE THEORY (IRT) MODEL IN ASSESSING PROGRAMMING PROFICIENCY

Gulnur S. Tleubaeva, PhD, Associate Professor
Nukus State Technical University
Email: gtleubaeva66@gmail.com

Abstract

This study outlines a methodology for applying Item Response Theory (IRT) to objectively evaluate the programming proficiency and latent abilities of higher education students. To overcome the limitations of Classical Test Theory (CTT), item difficulty (b) and discrimination (a) parameters were calibrated using the Two-Parameter Logistic (2PL) IRT model. Experimental findings demonstrate that an IRT-based adaptive diagnostic framework significantly improves the accuracy of assessing students' actual domain knowledge compared to traditional scoring methods.

Keywords: Item Response Theory, IRT, Computer Science Education, Latent Trait, Item Calibration, 2PL Model, Computerized Adaptive Testing.

1. Introduction

In higher education institutions, Classical Test Theory (CTT) is predominantly used for diagnosing student knowledge and skills in computer science and programming. Under CTT, a student's total raw score directly depends on overall test difficulty and random guessing probability. Consequently, this limits the objective evaluation of students' true programming competency and algorithmic thinking skills (latent traits).

In recent years, Item Response Theory (IRT) has been widely adopted in international educational measurement and psychometrics. The IRT framework models the mathematical relationship between a student's latent trait (θ) and the probability of correctly answering a specific item. The primary goal of this study is to

apply the IRT model to evaluate programming skills, calibrate test item banks, and enhance the measurement precision of student competence.

2. Methodology

Considering the specific nuances of programming assessments (syntax, logic, algorithmization, and code analysis), the **Two-Parameter Logistic (2PL IRT)** model was selected as the foundational mathematical framework:

Where:

θ- Student's latent trait (programming ability), typically normalized within the range [-3;+3];

$b_i -$ Difficulty parameter of the $i -$ th th test item;

a_i −Discrimination parameter of the i −th th test item:

$P_i(\theta)$- Probability that a student with ability level θ correctly answers the i −th item.

During the experimental phase, test bank items were categorized into 4 core competency domains:

Basic syntax and data structures;

Control structures and iterative algorithm logic;

Functional and object-oriented programming (OOP);

Code optimization and debugging.

The estimation of item parameters a_i and b_i was executed using the

Maximum Likelihood Estimation (MLE) method.

3. Results and Discussion

Item calibration was performed on the test bank constructed during empirical trials with students at Nukus State Technical University. The process established the parameter distribution across difficulty and discrimination metrics.

Table 1. IRT Calibration Parameters for Sample Programming Assessment Items

T/r	**Item Domain**	**Difficulty** (b_i)	**Discrimination** (a_i)	**Quality evaluation and interpretation**
1	Array Manipulation Syntax	-1.42	0.85	Easy item with low discrimination capacity
2	Recursive Function Execution Prediction	+0.68	2.14	Moderately challenging, high discrimination power
3	Algorithmic Complexity Analysis (Big-O)	+1.85	1.92	High difficulty, effectively discriminates high-performing students

Experimental findings revealed that items with, a_i > 1.5 provided clear differentiation between high-ability and lower-performing student groups.

Evaluating student latent ability θ using the IRT model demonstrated an **18% increase in diagnostic accuracy** compared to traditional cumulative scoring. This framework effectively prevents underestimating talented students who solve complex problems correctly but make minor careless errors on basic items.

4. Conclusion

Applying the IRT framework to evaluate student knowledge in programming overcomes the methodological limitations of classical testing approaches. The 2PL IRT model provides objective item calibration and impartial estimation of latent competencies.

These results serve as a theoretical and practical foundation for building future computerized adaptive testing (CAT) and intelligent diagnostic systems that dynamically tailor task difficulty to individual learning trajectories.

References

1. Lord, F. M. (2012). *Applications of item response theory to practical testing problems*. Routledge.
2. Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications.
3. Baker, F. B., & Kim, S. H. (2004). *Item response theory: Parameter estimation techniques*. CRC Press.