

---

## ***Item Response Theory (IRT) Vs Classical Test Theory (CTT): A Practical Comparison***

***Neha Bisht<sup>1</sup>, Rakesh Kumar Yadav<sup>2</sup>***

*Assistant Professor<sup>1</sup>, Research Scholar<sup>2</sup>*

*Department of Education*

*Govt. Degree College, Bageshwar, Uttarakhand, India*

*Email ID: neha\_bishtedu@gmail.com<sup>1</sup>, rkyadav.research@yahoo.co.in<sup>2</sup>*

### **ABSTRACT**

*Measurement and evaluation play a central role in educational research, assessment design, and psychological testing. Among the most widely used frameworks for test analysis are Classical Test Theory (CTT) and Item Response Theory (IRT). While CTT has been historically dominant due to its simplicity and ease of application, IRT has emerged as a more sophisticated and flexible approach, offering deeper insights into item-level performance and test-taker ability. This paper presents a comprehensive and practical comparison between CTT and IRT, focusing on their theoretical foundations, assumptions, advantages, limitations, and real-world applicability. The discussion highlights how CTT relies on total test scores and assumes equal measurement precision across all ability levels, whereas IRT models the interaction between item characteristics and individual ability. A comparative analysis is supported through tables and a conceptual figure to demonstrate key differences. The paper further explores scenarios where each theory is more appropriate, particularly in modern large-scale assessments and adaptive testing environments. The findings suggest that while CTT remains relevant for small-scale assessments, IRT provides superior accuracy and flexibility in contemporary educational measurement.*

**KEYWORDS:** *Item Response Theory, Classical Test Theory, Educational Measurement, Test Reliability, Psychometrics, Assessment Models*

## **INTRODUCTION**

Educational assessment is fundamental for evaluating student performance, guiding instructional practices, and shaping policy decisions. Over time, different measurement theories have been developed to ensure that assessments are valid, reliable, and fair. Two of the most prominent frameworks are Classical Test Theory (CTT) and Item Response Theory (IRT).

CTT, developed in the early 20th century, focuses on observed scores and treats measurement error as a random component. It has been widely used due to its straightforward computation and interpretation. However, its limitations, particularly in handling item-level data and varying ability levels, have led to the development of IRT.

IRT, a modern psychometric approach, addresses many of CTT's limitations by analyzing individual item characteristics and modeling the probability of a correct response based on a test-taker's ability. It has gained prominence in large-scale testing programs such as standardized exams and computer adaptive testing systems.

This paper aims to provide a practical comparison between CTT and IRT, highlighting their theoretical differences and applications in educational settings.

## **THEORETICAL FOUNDATIONS (ELABORATED)**

### **Classical Test Theory (Ctt)**

Classical Test Theory (CTT) forms the traditional backbone of educational and psychological measurement. Developed during the early 20th century, CTT provides a simple yet powerful framework for understanding test scores and measurement error. The theory is grounded in the idea that every observed score is composed of two components: a true score and random error.

### **Fundamental Model of CTT**

$$X = T + EX = T + E$$

Where:

- **X** = Observed score
- **T** = True score (actual ability)
- **E** = Error component

This equation implies that the observed score obtained by a test-taker is not perfectly accurate but is influenced by random fluctuations such as fatigue, guessing, environmental distractions, and test conditions.

## **CORE ASSUMPTIONS OF CTT**

CTT operates on several key assumptions:

### **1. Randomness of Error**

The error component is assumed to be random and normally distributed, meaning that over multiple test administrations, positive and negative errors cancel each other out.

### **2. Constancy of True Score**

A test-taker's true score remains stable across repeated testing under similar conditions.

### **3. Independence of Error and True Score**

The error component does not systematically relate to the true score.

### **4. Parallel Test Assumption**

Different forms of a test (parallel forms) are assumed to measure the same construct with equal difficulty and variance.

## **KEY CONCEPTS IN CTT**

### **1. Reliability**

Reliability refers to the consistency of test scores. In CTT, reliability is defined as the ratio of true score variance to observed score variance. Common reliability measures include:

- Cronbach's Alpha
- Test-retest reliability
- Split-half reliability

Higher reliability indicates lower measurement error.

### **2. Item Difficulty Index (p-value)**

This indicates how easy or difficult an item is, calculated as the proportion of students who answered correctly.

- High value → Easy item
- Low value → Difficult item

### 3. Item Discrimination Index

This measures how well an item differentiates between high-performing and low-performing students. Items with higher discrimination are more effective in assessment.

#### LIMITATIONS IN THEORETICAL TERMS

Despite its usefulness, CTT has inherent theoretical limitations:

- Item parameters are **sample-dependent**
- Test reliability is **population-specific**
- Measurement error is assumed **constant across all ability levels**, which is unrealistic

These limitations restrict its applicability in high-stakes and adaptive testing environments.

#### ITEM RESPONSE THEORY (IRT)

Item Response Theory (IRT) represents a paradigm shift in measurement theory by focusing on the interaction between test-taker ability and item characteristics. Unlike CTT, which emphasizes total scores, IRT models the probability of a correct response at the item level.

#### Fundamental Concept of IRT

IRT assumes that the probability of a correct response to a test item is a mathematical function of:

- The individual's latent trait (ability)
- Specific characteristics of the item

#### General IRT Model (Logistic Function)

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

Where:

- **P(θ)** = Probability of a correct response
- **θ (theta)** = Ability level of the individual
- **a** = Discrimination parameter
- **b** = Difficulty parameter

In more advanced models (3PL), an additional parameter is included:

- **c** = Guessing parameter

## **Key Assumptions of IRT**

### **1. Unidimensionality**

The test measures a single underlying trait (e.g., mathematical ability).

### **2. Local Independence**

Responses to individual items are independent once ability is controlled.

### **3. Monotonicity**

The probability of a correct response increases as ability increases.

## **Key Parameters in IRT**

### **1. Difficulty Parameter (b)**

Represents the ability level at which a test-taker has a 50% chance of answering correctly. Items are placed along an ability scale.

### **2. Discrimination Parameter (a)**

Indicates how sharply an item differentiates between individuals of different ability levels. Higher values imply better discrimination.

### **3. Guessing Parameter (c)**

Accounts for the probability of low-ability individuals answering correctly by chance, particularly relevant in multiple-choice tests.

## **Strengths of IRT From A Theoretical Perspective**

- Provides **item-level precision**
- Allows **ability estimation independent of specific test forms**
- Supports **adaptive testing models**
- Measurement error varies with ability, making it more realistic

## **Theoretical Limitations of IRT**

- Requires **large sample sizes** for stable parameter estimation
- Assumptions like unidimensionality may not always hold
- Complex mathematical modeling makes interpretation difficult

## **SYNTHESIS OF THEORETICAL DIFFERENCES**

From a theoretical standpoint, the fundamental distinction between CTT and IRT lies in their conceptualization of measurement:

- **CTT** views measurement as a combination of true score and random error, focusing on overall test performance.
- **IRT** views measurement as a probabilistic interaction between ability and item properties, focusing on individual item responses.

Thus, while CTT provides a **macro-level understanding**, IRT offers a **micro-level analytical framework**, making it more suitable for modern, data-driven educational assessment systems.

**ITEM RESPONSE THEORY (IRT)**

IRT models the relationship between an individual's ability and the probability of answering an item correctly. It is based on logistic models such as:

- One-Parameter Model (Rasch Model)
- Two-Parameter Model (2PL)
- Three-Parameter Model (3PL)

A general IRT equation is:

Probability of correct response = function (ability, item parameters)

Key parameters:

- Difficulty (b)
- Discrimination (a)
- Guessing (c)

**COMPARATIVE ANALYSIS OF CTT AND IRT**

**Table 1: Key Differences Between CTT and IRT**

<b>Feature</b>	<b>Classical Test Theory (CTT)</b>	<b>Item Response Theory (IRT)</b>
Focus	Total test score	Individual item response
Error Measurement	Same for all	Varies with ability
Item Parameters	Sample dependent	Sample independent
Ability Estimation	Based on raw score	Based on probability model

Feature	Classical Test Theory (CTT)	Item Response Theory (IRT)
Complexity	Simple	Complex
Application	Small-scale tests	Large-scale & adaptive tests

Dimension	CTT	IRT
Parameter Dependency	Depends on the specific sample	Sample-invariant
Error Estimation	Global SEM	Individualized information function
Item Analysis	Based on the total score (item-total correlation)	Based on latent traits ( $\theta$ and item parameter interaction)
Model Complexity	Simple linear model	Nonlinear probability model
Application Scenarios	Small samples, conventional tests	Large samples, high-precision measurement

*Figure 1: Difference between CTT and IRT*

### ADVANTAGES OF CLASSICAL TEST THEORY

1. Easy to understand and implement
2. Requires smaller sample sizes
3. Widely used in classroom assessments
4. Minimal computational requirements

### LIMITATIONS OF CLASSICAL TEST THEORY

1. Item parameters depend on the sample
2. Assumes equal measurement error
3. Cannot provide item-level precision
4. Not suitable for adaptive testing

### ADVANTAGES OF ITEM RESPONSE THEORY

1. Provides detailed item-level analysis
2. Independent of sample characteristics
3. Supports computer adaptive testing
4. Measures ability more precisely

**LIMITATIONS OF ITEM RESPONSE THEORY**

1. Requires large sample sizes
2. Computationally complex
3. Needs specialized software
4. Difficult to interpret for beginners

**PRACTICAL APPLICATIONS**

**In Classroom Assessment**

CTT is more practical for:

- Teacher-made tests
- Small student groups
- Quick evaluation

**In Large-Scale Testing**

IRT is preferred for:

- Competitive exams
- Standardized testing (e.g., GRE, TOEFL)
- Online adaptive assessments

*Table 2: Practical Usage Scenarios*

Scenario	Preferred Theory
Small classroom test	CTT
National-level exam	IRT
Adaptive testing system	IRT
Quick performance analysis	CTT

**DISCUSSION**

The comparison between CTT and IRT reveals that both frameworks serve important roles in educational measurement. While CTT offers simplicity and practicality, it lacks the precision required in modern testing environments. IRT, on the other hand, provides a robust analytical framework that captures the interaction between item characteristics and learner ability.

One of the most significant advantages of IRT is its ability to produce item characteristic curves, which visually represent how different items perform across ability levels. This allows educators to design more balanced and fair assessments.

However, the adoption of IRT is often limited by resource constraints, particularly in developing educational systems where access to large datasets and computational tools is restricted.

## CONCLUSION

Both Classical Test Theory and Item Response Theory have their unique strengths and limitations. CTT remains a valuable tool for small-scale assessments due to its simplicity and ease of use. However, in the context of modern educational assessment, where precision, adaptability, and fairness are critical, IRT offers a more advanced and reliable framework.

The choice between CTT and IRT should be guided by the purpose of the assessment, available resources, and the level of accuracy required. Future research should focus on integrating both approaches to create hybrid models that combine simplicity with precision.

## REFERENCES

1. Kline, T. J. B. (2005). *Psychological Testing: A Practical Approach to Design and Evaluation*. Sage Publications.
2. Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Springer.
3. Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Routledge.
4. Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum.
5. Crocker, L., & Algina, J. (2006). *Introduction to Classical and Modern Test Theory*. Cengage Learning.
6. Singh, A. (2018). Educational Measurement and Evaluation. *Indian Journal of Educational Research and Assessment Studies*, 7(2), 45–60.
7. Sharma, P., & Verma, R. (2019). Comparative Analysis of Psychometric Models. *Journal of Advanced Educational Psychology and Measurement Systems*, 8(1), 23–34.

8. Gupta, S. (2020). Application of IRT in Indian Competitive Exams. *International Journal of Educational Measurement and Policy Research*, 6(3), 55–67.
9. Thorndike, R. M. (2005). *Measurement and Evaluation in Psychology and Education*. Pearson.
10. Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse.
11. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. McGraw-Hill.
12. Misra, K. (2021). Modern Trends in Educational Assessment. *Journal of Indian Educational Innovations and Testing Frameworks*, 5(4), 78–89.
13. De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. Guilford Press.