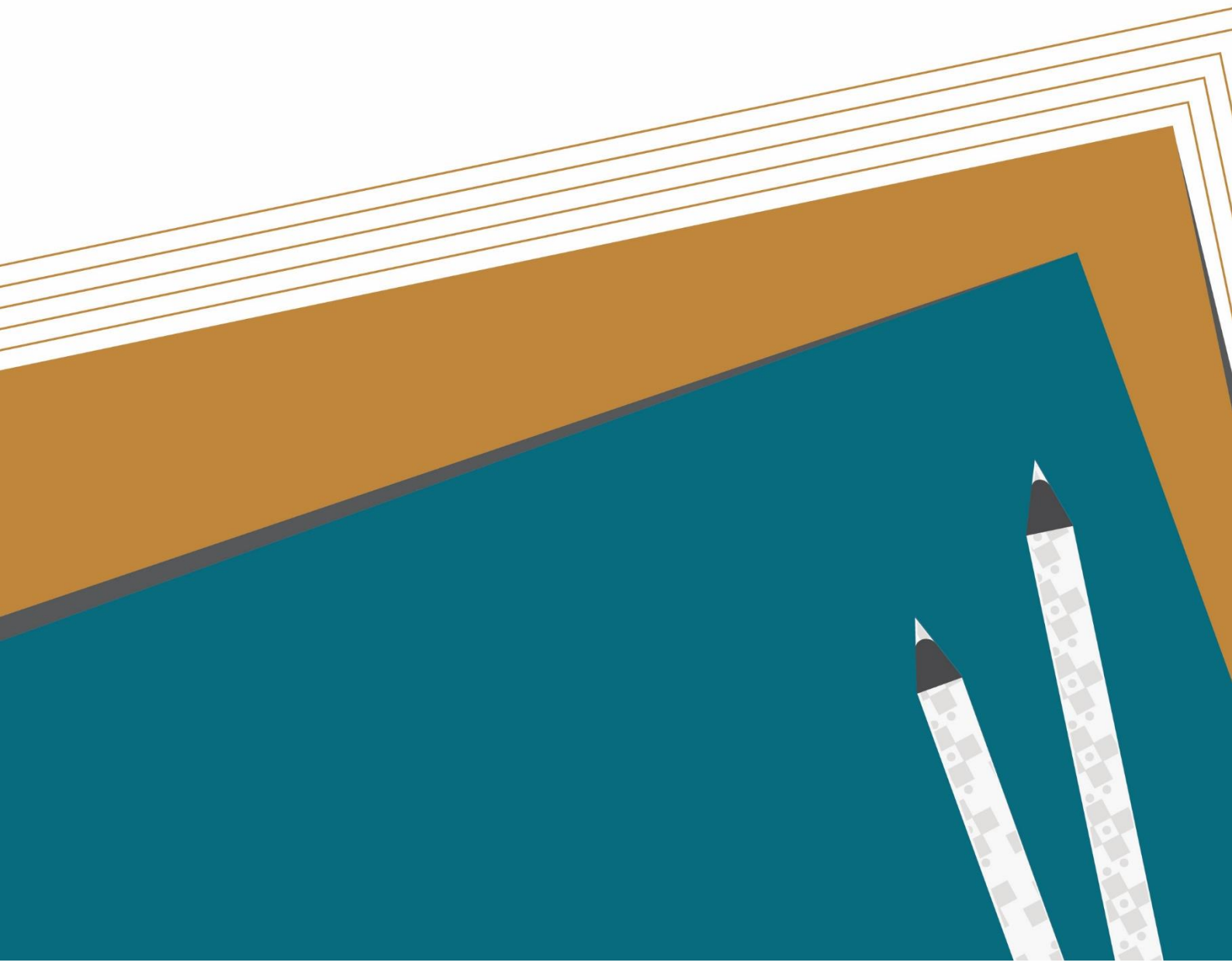


An Introduction to Item Response Theory



An Introduction to Item Response Theoryⁱ

Introduction

In an educational set-up, the most common approach to assess examinees' ability is to design a question paper with some assessment items and use it with the examinees to allot them a numerical score. This score reflects an examinee's ability level. The question paper (test) can have a combination of multiple choice and constructed response questions. Item Response Theory (IRT) is a statistical theory about examinee's performance in item and test and how performance relates to the examinee abilities that are measured by the items in the test.

Examinee 'ability' is used within IRT to refer to unobservable or latent traits such as reading ability, arithmetic ability, and intelligence. While these attributes can be described, they cannot be measured the way physical variables such as height or weight can be measured. In IRT, to measure examinee's "ability", a scale having a midpoint zero that ranges from positive to negative infinity is proposed. Any examinee can be placed on this scale as per his/ her ability. This scale also allows to compare many examinees based on their ability.

Item response theory rest upon the individual items of a test hence an individual's performance in each assessment item is considered rather than the raw test score. Since each examinee responding to a test item possesses some amount of the underlying ability which can be quantified, they can be placed somewhere on the ability scale.

This document describes the basic concepts for IRT model which will help the readers to infer the test results better.

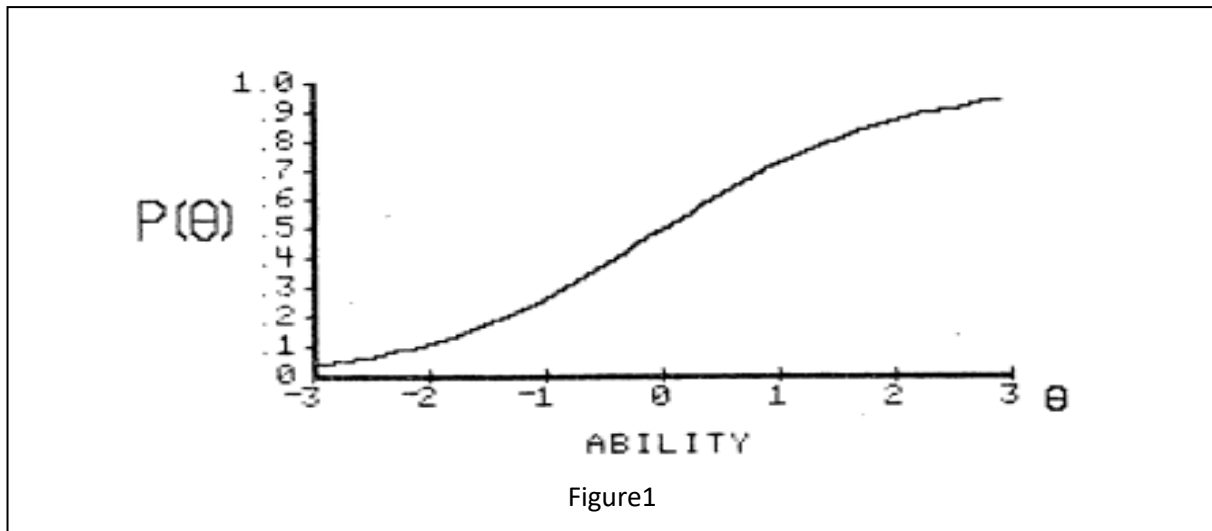
Item Characteristics Curve

Definition

Ability score is represented by θ (Greek letter theta). $P(\theta)$ is the probability that an examinee of ability (θ) will respond correctly to an assessment item. This probability will be small for examinees of low ability and large for examinees of high ability. Item Characteristics Curve (ICC) represents the relationship between probability of correct response to an item ($P(\theta)$) and the ability (θ).

Interpreting ICC

A smooth S shaped curve results on plotting $P(\theta)$ as a function of ability (θ). Each item in a test will have its own item characteristic curve. In Figure 1, ability is assumed to range between -3 to +3. The probability of correct response is near zero at the lower levels of ability (less than 0). It increases until at the highest levels of ability (in the range above 0), the probability of correct response approaches.



Difficulty level

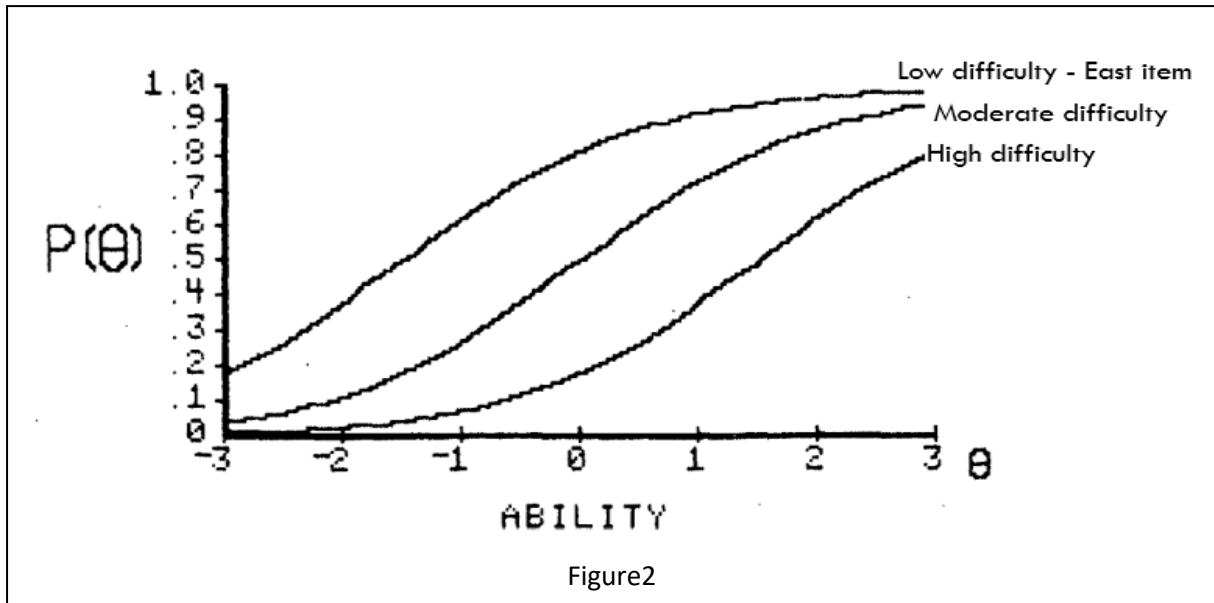
Definition

The difficulty of an item describes where the item functions along the ability scale. It is assumed that there is 50% probability of receiving correct response on every item. Therefore, an easy item functions among the low-ability examinees and a hard item function among the high-ability examinees. Thus, difficulty is a location index. Difficulty will have the following levels: very easy, easy, medium, hard, and very hard.

Interpreting difficult level in ICC

In Figure2, three ICCs are presented on the same graph for three different items. All items have the same level of discrimination but differ with respect to difficulty.

- The left-hand curve represents an easy item because the probability of correct response is high for low-ability examinees and approaches 1 for high-ability examinees.
- The center curve represents an item of moderate difficulty because the probability of correct response is low at the lowest ability levels, around 0.5 in the middle of the ability scale and near 1 at the highest ability levels.
- The righthand curve represents an item of hard difficulty. The probability of correct response is low for most of the ability scale and increases only when the higher ability levels are reached. Even at the highest ability level shown (+3), the probability of correct response is only .8 for the most difficult item.



Locating difficulty level in ICC

Identify the axis with the Probability distribution. Draw a horizontal line at probability=0.5. Let this line cut the ICCs. Drop a vertical line on the ability axis where the horizontal line cuts the ICCs.

- If ability = 0 – Item is of moderate difficult level
- If ability < 0 – Item is of low difficulty level
- If ability > 0 – Item is of high difficulty level

Discrimination Index

Definition

It describes how well an item can differentiate between examinees having abilities below the item location and those having abilities above the item location. This property essentially reflects the steepness of the item characteristic curve in its middle section. Discrimination has the following levels: none, low, moderate, high and perfect. The steeper the curve, the better the item can discriminate. The flatter the curve, the less the item can discriminate since the probability of correct response at low ability levels is nearly the same as it is at high ability levels.

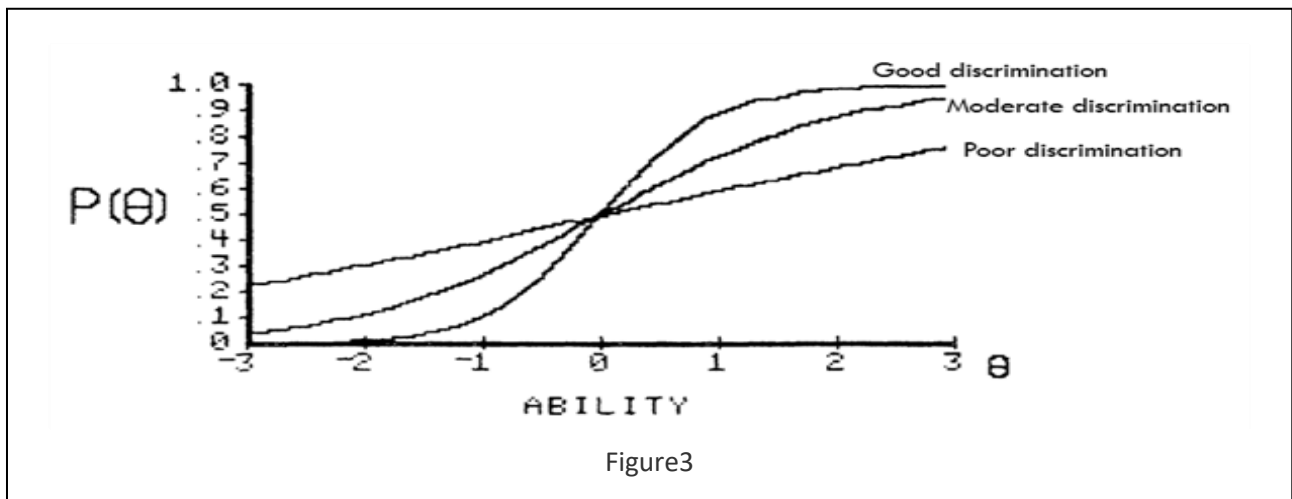
Interpreting discrimination index in ICC

In Figure3, there are three ICCs for three different items having the same difficulty level (at 0.5 probability, all the ICCs have the ability 0) but differing with respect to discrimination.

- The upper curve has a high level of discrimination since the curve is quite steep in the middle where the probability of correct response changes very rapidly as ability increases. Just a short distance to the left of the middle of the

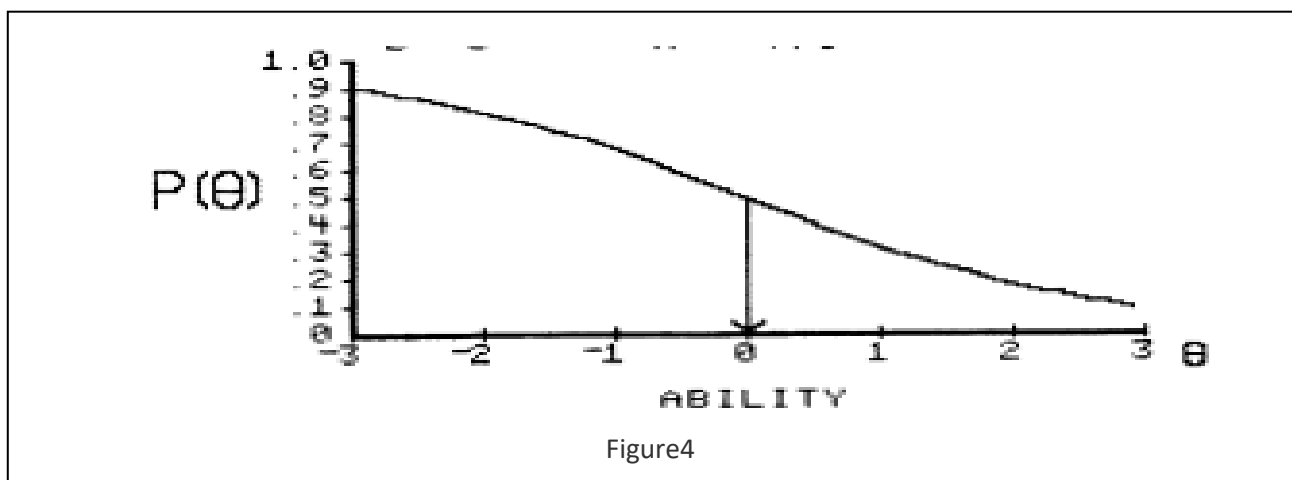
curve, the probability of correct response is much less than 0.5, and a short distance to the right the probability is much greater than 0.5.

- The middle curve represents an item with a moderate level of discrimination. The slope of this curve (at the ability level corresponding to the difficulty level) is much less than the previous curve and the probability of correct response changes less dramatically than the previous curve as the ability level increases. However, the probability of correct response is near zero for the lowest-ability examinees and near 1 for the highest-ability examinees.
- The third curve represents an item with low discrimination. The curve has a very small slope and the probability of correct response changes slowly over the full range of abilities shown. Even at low ability levels, the probability of correct response is reasonably large, and it increases only slightly when high ability levels are reached.



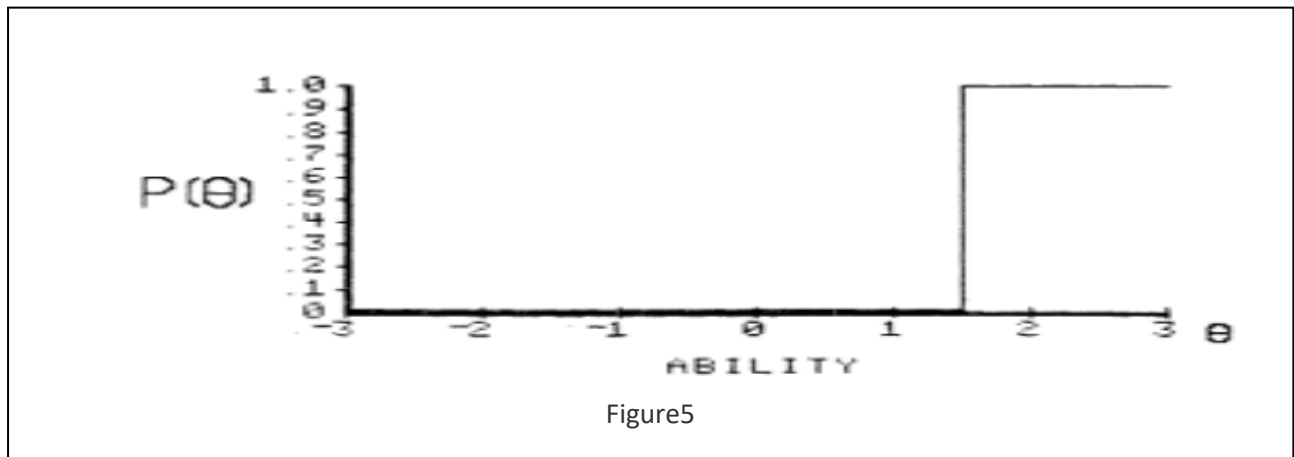
Negative Discrimination

While most test items discriminate in a positive manner (i.e., the probability of correct response increases as the ability level increases), some items have negative discrimination. In such items, the probability of correct response decreases as the ability level increases from low to high. Figure4 is an example of such item. This tells that something is wrong with the item: Either it is poorly written or there is some misinformation prevalent among the high-ability examinees.



Perfect Discrimination Index

One special case is of interest--namely, that of an item with perfect discrimination. The item characteristic curve of such an item is a vertical line at some point along the ability scale. In Figure5, to the left of the vertical line at $\theta = 1.5$, the probability of correct response is zero; to the right of the line, the probability of correct response is 1. Thus, the item discriminates perfectly between examinees whose abilities are above and below an ability score of 1.5. Such items would be ideal for distinguishing between examinees with abilities just above and below 1.5.



Locating discrimination index in ICC

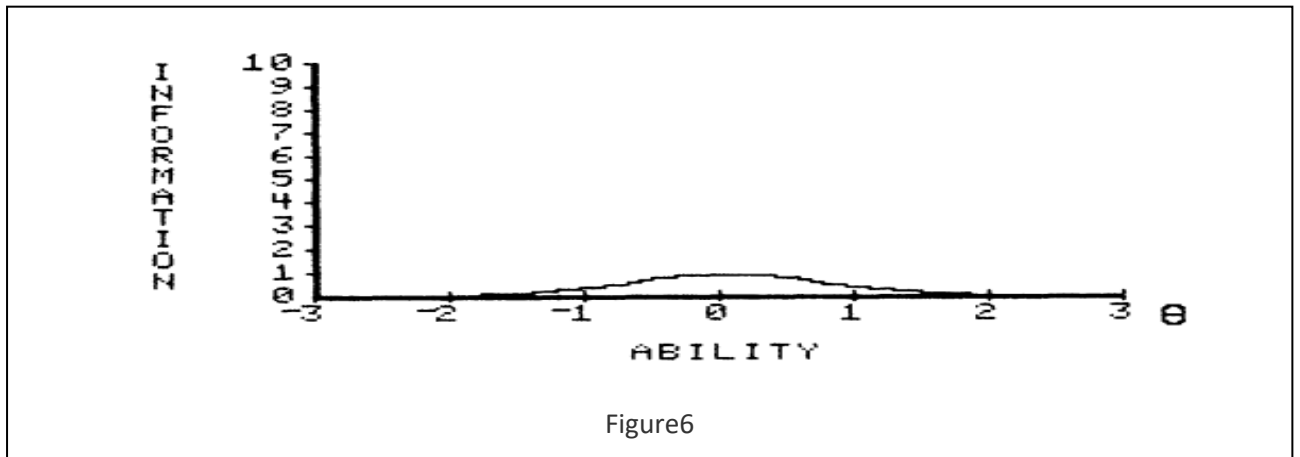
Identify the axis where ability is plotted. Identify ability = 0, plot a vertical line at ability=0. This line meets the ICCs. Identify the axis where probability is plotted. From the point where the vertical line meets the ICCs, draw a horizontal line to the axis with probability. Identify the probability for each ICC

- A low probability signifies high discrimination
- A high probability signifies low discrimination

Item Information Function

The item information function tells us how well each ability level is being estimated using an item. The typical information function is shown in Figure6. Different ability levels are estimated with differing degrees of precision. The item precisely estimates the ability of examinees around 0 (at the peak and its neighbourhood the information estimated is in the range of 1 to 1.5 approx.). The amount of item information decreases as the ability level departs from the item difficulty and approaches zero at the extremes of the ability scale. This item precisely estimates the ability of examinees with moderate ability. This becomes of considerable importance to both the test constructor and the test consumer

since it means that the precision with which an examinee's ability is estimated depends upon where the examinee's ability is located on the ability scale.

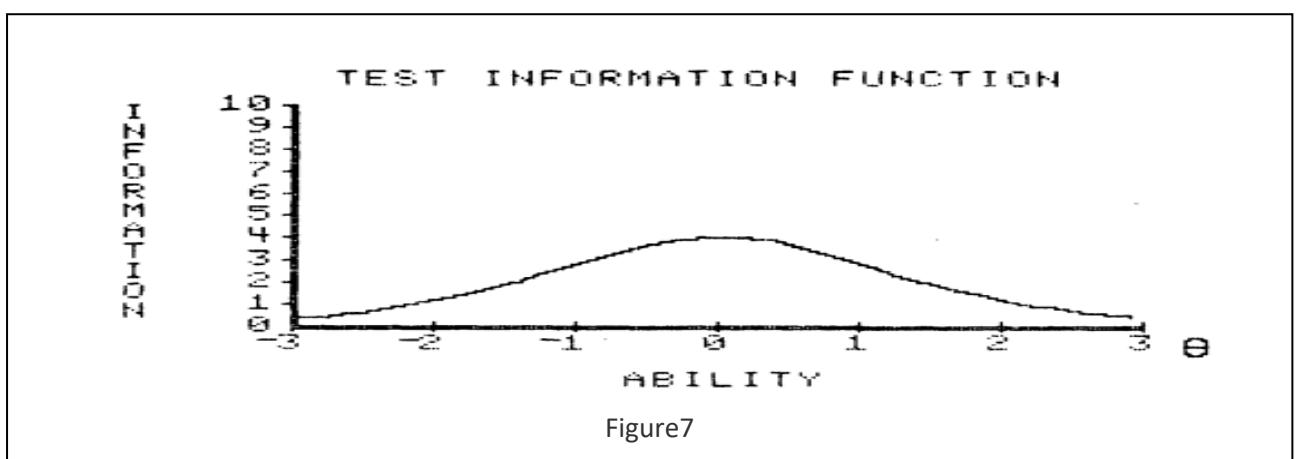


Test Information Function

Definition

Test information function (TIF) helps to estimate the overall performance of the question paper. A question paper is designed using a set of items of varied difficulty and discrimination levels. Therefore, the TIF is the total of the item information functions of all the items in the question paper.

The general level of the test information function will be much higher than that for a single item information function. Thus, a test measures ability more precisely than does a single item. The more items in the test, the greater the amount of information. Thus, longer tests will measure an examinee's ability with greater precision than a shorter test.



Interpreting TIF

Plotting the amount of test information against ability yields a graph of the test information function as shown in Figure 7. The maximum value of the test information function is modest. The amount of information decreases rather steadily as the ability level differs from that corresponding to the maximum. Thus, ability is estimated with some precision near the center of the ability scale. However, as the ability level approaches the extremes of the scale, the amount of test information decreases significantly.

The test information function is an extremely useful feature of item response theory. It helps in identifying how well the test is doing in estimating ability over the whole range of ability scores. An ideal test information function often may be a horizontal line. In practical situations, TIF peaked around 0 ability (moderate difficulty level) and moderate discrimination index is desirable

Summary

Item parameters	Interpretation
Difficulty level	Low – Less than -1; Moderate – -1 to +1; High – More than 1
Discrimination index	Poor – Less than 0.65; Moderate- 0.65 - 1.34, Good – More than 1.34
Item characteristics curve	Ideal item - Should start from zero and should be steeper in the middle
Item information function	Ideal item – Horizontal line
Test information function	Ideal item – Horizontal line Practical situation- Peak of the curve should be at zero ability level (indicates moderate difficulty level) and should have a tall peak (indicates good discrimination)

Reference(s)

1. Baker, F. B. (2001). *The basics of item response theory*.

ⁱⁱ This article is authored by Bijay Das. He works as Lecturer in the School of Continuing Education at Azim Premji University. He can be reached at bijay.das@azimpremijfoundation.org.

This article can be cited as-

An introduction to item response theory, Assessment resources, 2021, Azim Premji University

An Introduction to Item Response Theory

