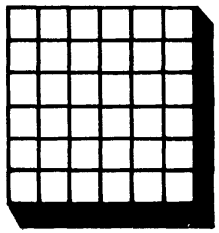

CHAPTER 1
The Item Characteristic Curve

CHAPTER 1

The Item Characteristic Curve

In many educational and psychological measurement situations, there is an underlying variable of interest. This variable is often something that is intuitively understood, such as "intelligence." When people are described as being bright or average, the listener has some idea as to what the speaker is conveying about the object of the discussion. Similarly, one can talk about scholastic ability and its attributes, such as getting good grades, learning new material easily, relating various sources of information, and using study time effectively. In academic areas, one can use descriptive terms such as reading ability and arithmetic ability. Each of these is what psychometricians refer to as an unobservable, or latent, trait. Although such a variable is easily described, and knowledgeable persons can list its attributes, it cannot be measured directly as can height or weight, for example, since the variable is a concept rather than a physical dimension. A primary goal of educational and psychological measurement is the determination of how much of such a latent trait a person possesses. Since most of the research has dealt with variables such as scholastic, reading, mathematical, and arithmetic abilities, the generic term "ability" is used within item response theory to refer to such latent traits.

If one is going to measure how much of a latent trait a person has, it is necessary to have a scale of measurement, i.e., a ruler having a given metric. For a number of technical reasons, defining the scale of measurement, the numbers on the scale, and the amount of the trait that the numbers represent is a very difficult task. For the purposes of the first six chapters, this problem shall be solved by simply defining an arbitrary underlying ability scale. It will be assumed that, whatever the ability, it can be measured on a scale having a midpoint of zero, a unit of measurement of one, and a range from negative infinity to positive infinity. Since there is a unit of measurement and an arbitrary zero point, such a scale is referred to as existing at an interval level of measurement. The underlying idea here is that if one could physically ascertain the ability of a person, this ruler would be used to tell how much ability a given person has, and the ability of several persons could be compared. While

the theoretical range of ability is from negative infinity to positive infinity, practical considerations usually limit the range of values from, say, -3 to +3. Consequently, the discussions in the text and the computer sessions will deal only with ability values within this range. However, you should be aware that values beyond this range are possible.

The usual approach taken to measure an ability is to develop a test consisting of a number of items (questions). Each of these items measures some facet of the particular ability of interest. From a purely technical point of view, such items should be free-response items for which the examinee can write any response that seems appropriate. The person scoring the test must then decide whether the response is correct or not. When the item response is determined to be correct, the examinee receives a score of one; an incorrect answer receives a score of zero, i.e., the item is dichotomously scored. Under classical test theory, the examinee's raw test score would be the sum of the scores received on the items in the test. Under item response theory, the primary interest is in whether an examinee got each individual item correct or not, rather than in the raw test score. This is because the basic concepts of item response theory rest upon the individual items of a test rather than upon some aggregate of the item responses such as a test score.

From a practical point of view, free-response items are difficult to use in a test. In particular, they are difficult to score in a reliable manner. As a result, most tests used under item response theory consist of multiple-choice items. These are scored dichotomously: the correct answer receives a score of one, and each of the distractors yields a score of zero. Items scored dichotomously are often referred to as binary items.

A reasonable assumption is that each examinee responding to a test item possesses some amount of the underlying ability. Thus, one can consider each examinee to have a numerical value, a score, that places him or her somewhere on the ability scale. This ability score will be denoted by the Greek letter theta, θ. At each ability level, there will be a certain probability that an examinee with that ability will give a correct answer to the item. This probability will be denoted by $P(\theta)$. In the case of a typical test item, this probability will be small for examinees of low ability and large for examinees of high ability. If one plotted $P(\theta)$ as a function of ability, the result would be a smooth S-shaped

curve such as shown in Figure 1-1. The probability of correct response is near zero at the lowest levels of ability. It increases until at the highest levels of ability, the probability of correct response approaches 1. This S-shaped curve describes the relationship between the probability of correct response to an item and the ability scale. In item response theory, it is known as the item characteristic curve. Each item in a test will have its own item characteristic curve.

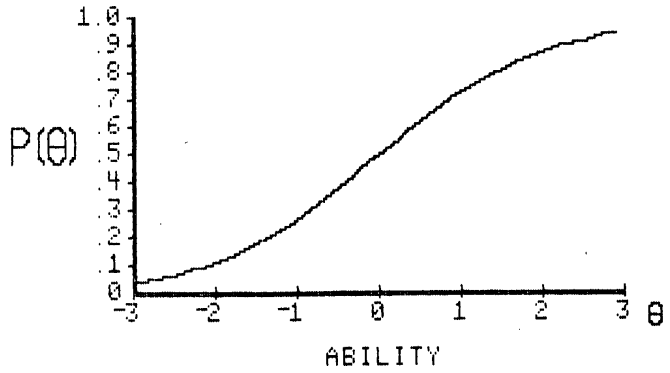

FIGURE 1-1. A typical item characteristic curve

The item characteristic curve is the basic building block of item response theory; all the other constructs of the theory depend upon this curve. Therefore, considerable attention will be devoted to this curve and its role within the theory. There are two technical properties of an item characteristic curve that are used to describe it. The first is the difficulty of the item. Under item response theory, the difficulty of an item describes where the item functions along the ability scale. For example, an easy item functions among the low-ability examinees and a hard item functions among the high-ability examinees; thus, difficulty is a location index. The second technical property is discrimination, which 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. The steeper the curve, the better the item can discriminate. The flatter the curve, the less the item is able to discriminate since the probability of correct response at low ability levels is

nearly the same as it is at high ability levels. Using these two descriptors, one can describe the general form of the item characteristic curve. These descriptors are also used to discuss the technical properties of an item. It should be noted that these two properties say nothing about whether the item really measures some facet of the underlying ability or not; that is a question of validity. These two properties simply describe the form of the item characteristic curve.

The idea of item difficulty as a location index will be examined first. In Figure 1-2, three item characteristic curves are presented on the same graph. All 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 medium difficulty because the probability of correct response is low at the lowest ability levels, around .5 in the middle of the ability scale and near 1 at the highest ability levels. The right-hand curve represents a hard item. 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.

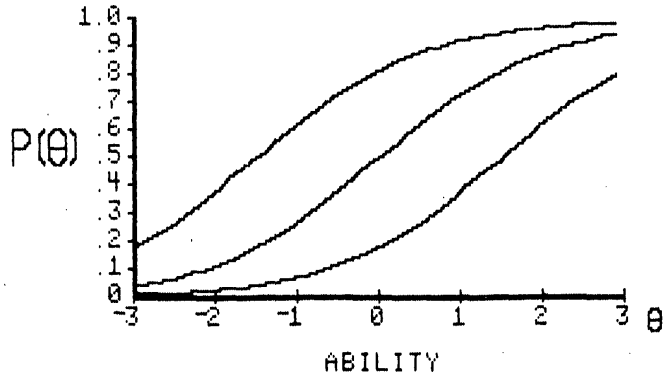

FIGURE 1-2. Three item characteristic curves with the same discrimination but different levels of difficulty

The concept of discrimination is illustrated in Figure 1-3. This figure contains three item characteristic curves having the same difficulty level 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 .5, and a short distance to the right the probability is much greater than .5. The middle curve represents an item with a moderate level of discrimination. The slope of this curve 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. The reader should be warned that although the figures only show a range of ability from -3 to +3, the theoretical range of ability is from negative infinity to positive infinity. Thus, all item characteristic curves of the type used here actually become asymptotic to a probability of zero at one tail and to 1.0 at the other tail. The restricted range employed in the figures is necessary only to fit the curves on the computer screen reasonably.

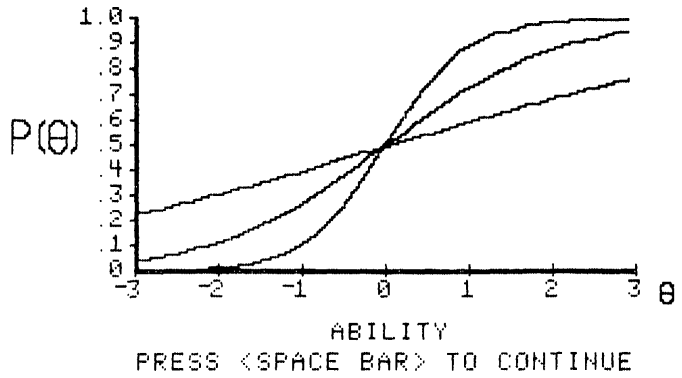

FIGURE 1-3. Three item characteristic curves with the same difficulty but with different levels of discrimination

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. Figure 1-4 shows such an item. 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. However, such an item makes no distinction among those examinees with abilities above 1.5 nor among those examinees with abilities below 1.5.

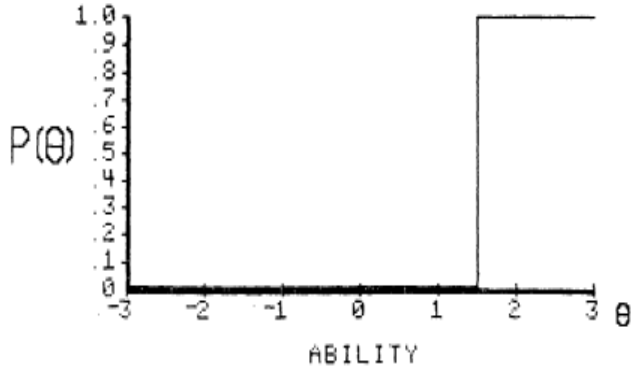

FIGURE 1-4. An item that discriminates perfectly at $\theta = 1.5$

At the present point in the presentation of item response theory, the goal is to allow you to develop an intuitive understanding of the item characteristic curve and its properties. In keeping with this goal, the difficulty and discrimination of an item will be defined in verbal terms.

Difficulty will have the following levels:

very easy
easy
medium
hard
very hard

Discrimination will have the following levels:

none
low
moderate
high
perfect

These terms will be used in the computer session to specify item characteristic curves.

Chapter 1: *The Item Characteristic Curve* 2

Computer Session for Chapter 1

The purpose of this session is to enable you to develop a sense of how the shape of the item characteristic curve is related to item difficulty and discrimination. To accomplish this, you will be able to select verbal terms describing the difficulty and discrimination of an item. The computer will then calculate and display the corresponding item characteristic curve on the screen. You should do the exercises, then try various combinations of levels of difficulty and discrimination and relate these to the resulting curves. After a bit of such exploratory practice, you should be able to predict what the item characteristic curve will look like for a given combination of difficulty and discrimination.

Procedures for an Example Case

a. Follow the start-up procedures described in the Introduction.

b. Use the mouse to highlight the ITEM CHARACTERISTIC CURVE session and click on [SELECT] to activate the session.

c. Read the explanatory screen and move to the next screen by clicking on [CONTINUE]. The SELECT CHARACTERISTICS screen will appear.

d. Use the left mouse button to click on medium difficulty and then click on moderate discrimination.

e. Click on [CONTINUE] to display the plot of the item characteristic curve.

f. The computer will display an item characteristic curve for an item with medium difficulty and moderate discrimination, shown in Figure 1-5.

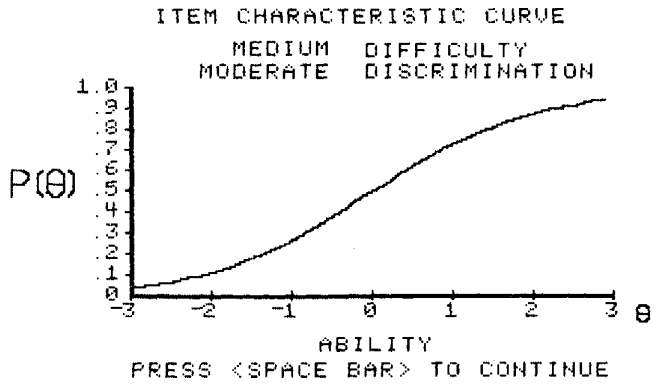

FIGURE 1-5. Example item characteristic curve

g. After you have studied the curve, click on [CONTINUE].

h. Respond to the message DO ANOTHER ITEM? by clicking on the YES button.

i. Respond to the message PLOT ON THE SAME GRAPH? by clicking on the YES button.

j. Now select easy difficulty and low discrimination and click on [CONTINUE] to see the new graph.

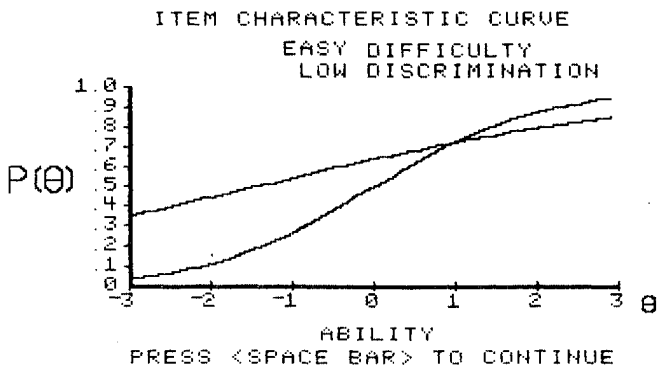

FIGURE 1-6. Two item characteristic curves

k. This second curve was overlaid on the same graph as the previous curve for comparison purposes. The new curve is rather flat and has higher probabilities of correct response in the lower range of abilities than did the previous item. This is because it was an easier item and low-ability examinees should do well on it. The low discrimination shows up in the curve having only a slight bend over the range of ability scores employed. At the high ability levels, the probability of correct response was somewhat lower than that of the previous item. This is a reflection of the lower discrimination of the new item.

l. Click on [CONTINUE] and respond to the DO ANOTHER ITEM message by clicking on the YES button.

m. To clear the graph for the next problem, respond to the message PLOT ON THE SAME GRAPH? by clicking on the NO button.

Exercises

Exercise 1

a. Use the menu to select an item with easy difficulty and high discrimination.

b. From the graph it can be seen that the probability of correct response will be rather high over most of the ability scale. The item characteristic curve will be steep in the lower part of the ability scale.

c. After you have studied the curve, respond to the message DO ANOTHER ITEM? by clicking on the YES button.

d. Respond to the message PLOT ON THE SAME GRAPH? by clicking on the NO button.

Exercise 2

a. Use the menu to select an item with hard difficulty and low discrimination.

b. From the graph it can be seen that the probability of correct response will have a low general level over most of the ability scale. The item characteristic curve will not be very steep.

c. After you have studied the curve, respond to the message DO ANOTHER ITEM? by clicking on the YES button.

d. Respond to the message PLOT ON THE SAME GRAPH? by clicking on the NO button.

Exercise 3

a. Use the menu to select an item with medium difficulty and low discrimination.

b. From the graph it can be seen that the probability of correct

Chapter 1: *The Item Characteristic Curve* 16

response will be between .2 and .8 over the range of ability shown. The item characteristic curve will be nearly linear over the range of ability employed.

c. After you have studied the curve, respond to the message DO ANOTHER ITEM? by clicking on the YES button.

d. Respond to the message PLOT ON THE SAME GRAPH? by clicking on the NO button.

Exercise 4

In this exercise, all the items will have the same difficulty but different levels of discrimination. The intent is to relate the steepness of the curves to the level of discrimination.

e. Use the menu to select an item with medium difficulty and moderate discrimination.

f. From the graph it can be seen that the probability of correct response will be small at low ability levels and large at high ability levels. The item characteristic curve will be moderately steep in the middle part of the ability scale.

g. After you have studied the curve, respond to the message DO ANOTHER ITEM? by clicking on the YES button.

h. Respond to the message PLOT ON SAME GRAPH? by clicking on the YES button.

i. Now repeat steps a through d several times using medium difficulty for each item and discrimination values of your choosing.

j. After the last item characteristic curve has been shown, clear the graph for the next problem by responding to the message PLOT ON THE SAME GRAPH? by clicking on the NO button.

Exercise 5

In this exercise, all the items will have the same level of discrimination but different difficulty levels. The intent is to relate the location of the item on the ability scale to its difficulty level.

a. Use the menu to select an item with very easy difficulty and moderate discrimination.

b. From the graph it can be seen that the probability of correct response will be reasonably large over most of the ability scale. The item characteristic curve will be moderately steep in the lower part of the ability scale.

c. After you have studied the curve, respond to the message DO ANOTHER ITEM? by clicking on the YES button.

d. Respond to the message PLOT ON SAME GRAPH? by clicking on the YES button.

e. Now repeat steps a through d several times using items with moderate discrimination and difficulty levels of your choosing.

f. After the last item characteristic curve has been shown, clear the graph for the next problem by responding to the message PLOT ON THE SAME GRAPH? by clicking on the NO button.

Exercise 6

Experiment with various combinations of difficulty of your own choosing until you are confident that you can predict the shape of the item

Chapter 1: *The Item Characteristic Curve* 13

characteristic curve corresponding to the values chosen. You may find it useful to make a rough sketch of what you think the curve will look like before you have the computer display it on the screen.

Things To Notice

1. When the item discrimination is less than moderate, the item characteristic curve is nearly linear and appears rather flat.

2. When discrimination is greater than moderate, the item characteristic curve is S-shaped and rather steep in its middle section.

3. When the item difficulty is less than medium, most of the item characteristic curve has a probability of correct response that is greater than .5.

4. When the item difficulty is greater than medium, most of the item characteristic curve has a probability of correct response less than .5.

5. Regardless of the level of discrimination, item difficulty locates the item along the ability scale. Therefore item difficulty and discrimination are independent of each other.

6. When an item has no discrimination, all choices of difficulty yield the same horizontal line at a value of $P(\theta) = .5$. This is because the value of the item difficulty for an item with no discrimination is undefined.

7. If you have been very observant, you may have noticed the point at which $P(\theta) = .5$ corresponds to the item difficulty. When an item is easy, this value occurs at a low ability level. When an item is hard, this value corresponds to a high ability level.